Comprehensive Vulnerability Evaluation of Face Recognition Systems to Template Inversion Attacks via 3D Face Reconstruction

Hatef Otroshi Shahreza and Sébastien Marcel

***Abstract*—In this article, we comprehensively evaluate the vulnerability of state-of-the-art face recognition systems to template inversion attacks using 3D face reconstruction. We propose a new method (called GaFaR) to reconstruct 3D faces from facial templates using a pretrained geometry-aware face generation network, and train a mapping from facial templates to the intermediate latent space of the face generator network. We train our mapping with a semi-supervised approach using real and synthetic face images. For real face images, we use a generative adversarial network (GAN)-based framework to learn the distribution of generator intermediate latent space. For synthetic face images, we directly learn the mapping from facial templates to the generator intermediate latent code. Furthermore, to improve the success attack rate, we use two optimization methods on the camera parameters of the GNeRF model. We propose our method in the whitebox and blackbox attacks against face recognition systems and compare the transferability of our attack with state-of-the-art methods across other face recognition systems on the MOBIO and LFW datasets. We also perform practical presentation attacks on face recognition systems using the digital screen replay and printed photographs, and evaluate the vulnerability of face recognition systems to different template inversion attacks.**

***Index Terms*—Face recognition, face reconstruction, facial template, generative adversarial network (GAN), geometry-aware, neural radiance fields (NeRF), preseantation attck, semi-supervised learning, template inversion (TI), transferability, vulnerability evaluation.**

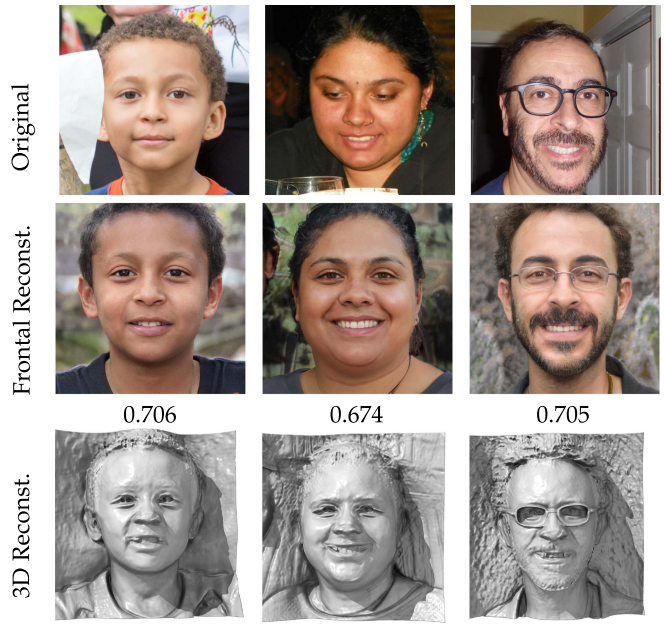

Fig. 1. Sample face images from the FFHQ dataset (first row) and frontal 2D image (second row) from our 3D reconstruction (third row) in the *whitebox* template inversion attack against ArcFace. The values below each image of the second row show the cosine similarity between the templates of the original and frontal reconstruction face images. The decision threshold for FMR $= 10^{-3}$ is 0.24 on the LFW dataset.

I. Introduction

FACE recognition (FR) is one of the most well-known biometric authentication tools, and its applications tend toward ubiquity, including smart phone unlock,[1] e-banking[2] national identity system,[3] border control,[4] etc. In addition to the security applications, FR is also being used in entertainment[5] applications. Generally in FR systems, some features (also known as templates or embeddings) are extracted from each face image. The extracted templates are stored in the system's database during the enrollment stage, and are later used for recognition.

Among different types of attacks against FR systems that are studied in the literature [1], [2], [3], [4], [5], template inversion (TI) attack can considerably jeopardize both security and privacy of users. In a TI attack, the adversary gains access to the templates stored in the system's database and tries to invert facial templates to reconstruct the underlying face image. Then, the

Manuscript received 28 June 2023; revised 31 August 2023; accepted 31 August 2023. Date of publication 5 September 2023; date of current version 3 November 2023. This work was supported by the H2020 TReSPAsS-ETN Marie Skłodowska-Curie Early Training Network under Grant 860813. Recommended for acceptance by W. Scheirer. *(Corresponding author: Hatef Otroshi Shahreza.)*

Hatef Otroshi Shahreza is with the Biometrics Security and Privacy Group, Idiap Research Institute, 1920 Martigny, Switzerland, and also with the École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland (e-mail: hatef.otroshi@epfl.ch).

Sébastien Marcel is with the Biometrics Security and Privacy Group, Idiap Research Institute, 1920 Martigny, Switzerland, and also with the Université de Lausanne (UNIL), 1015 Lausanne, Switzerland (e-mail: marcel@idiap.ch).

The project page is available at https://www.idiap.ch/paper/gafar.

Digital Object Identifier 10.1109/TPAMI.2023.3312123

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

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

[3]https://bbc.in/3QeIsO2

[4]https://nyti.ms/3XEIbaW

[5]https://apple.co/3ZOxW5S

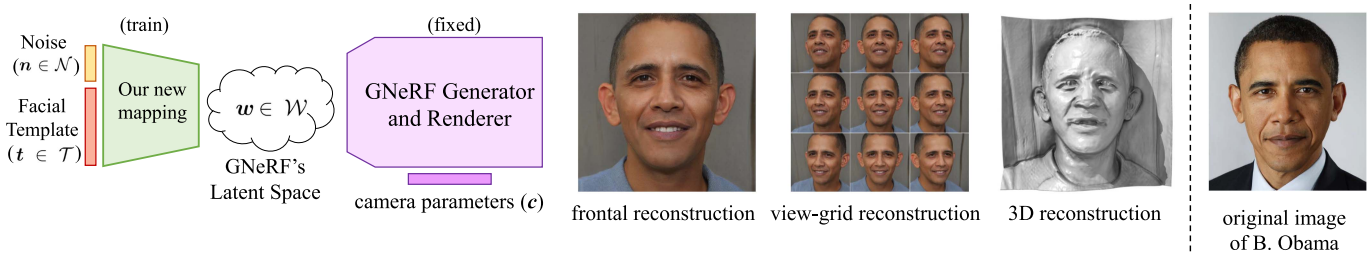

Fig. 2. General block diagram of the proposed method: we train a mapping network from facial templates (input) to the intermediate latent space $\mathcal{W}$ of GNeRF model. The mapped latent codes along with camera parameters are fed to the GNeRF generator and renderer network (fixed) to generate face image from desired view. Sample outputs of our model (frontal image, view-grid, and 3D face reconstruction) for face reconstruction from B. Obama's facial template are depicted.

adversary can use the reconstructed face image to impersonate and enter the system (security threat). In addition, the reconstructed face image may reveal privacy-sensitive information of the enrolled user, such as age, gender, ethnicity, etc. (privacy threat). In this paper, we focus on TI attacks in FR systems and present a comprehensive vulnerability evaluation of FR systems to TI attacks using 3D face reconstruction. We propose a new method (called geometry-aware face reconstruction, shortly *GaFaR*) to 3D reconstruct faces from facial templates using a geometry-aware face generator network. To our knowledge, this is the first work to reconstruct 3D faces from facial templates. Fig. 1 illustrates sample face images from the FFHQ [6] dataset and their corresponding 3D reconstruction from ArcFace [7] templates using our proposed method.

In recent years, the neural radiance fields (NeRF) [8] has attracted attentions in the computer vision community because of its impressive results in the novel-view generation problem. Generative NeRF (GNeRF) methods such as [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21] combine conditional NeRF with generative models, such as a generative adversarial network (GAN), for geometry-aware image generation tasks. In GNeRF methods, a generative model is used to embed the appearance and shape of an object into a latent space. Then, the camera parameters along with the latent code of the generative model are fed into a NeRF model for the rendering process. Among GNeRF methods, several works proposed geometry-aware 3D face generation models that can generate face images from different views [13], [14], [15], [16], [17], [18], [19], [20].

In our proposed 3D face reconstruction method, we use a geometry-aware face generator network based on GNeRF, and learn a mapping from facial templates to the *intermediate* latent space of the GNeRF model. We train our model with a *semi-supervised* approach using real and synthetic face images. For real training face images, where we do not have the corresponding GNeRF latent codes, we train our mapping within a GAN-based framework to learn the distribution of GNeRF *intermediate* latent space (*unsupervised* learning). However, for the synthetic training face images, we have the corresponding GNeRF latent codes, and directly learn the mapping from facial templates to the GNeRF *intermediate* latent space (*supervised* learning). At the inference stage, we have the 3D reconstructed face and can generate a face image from any arbitrary pose. Thus, we apply optimization on the camera parameters to generate face images with a pose that can increase the success attack rate against the FR system. Fig. 2 illustrates the general block diagram of our proposed template inversion attack.

We introduce our face reconstruction method for *whitebox* and *blackbox* TI attacks against FR systems. In the *whitebox* scenario, the adversary knows the internal functioning and parameters of the feature extraction model. However, in the *blackbox* scenario, the adversary does not have any knowledge about the internal functioning of the feature extraction model and can only use it to extract features from an arbitrary image. We consider the scenario where the adversary uses another FR model, with known internal functioning and parameters (i.e., *whitebox* knowledge), and uses this FR model for training the face reconstruction network. We present a comprehensive vulnerability evaluation of state-of-the-art (SOTA) FR systems to our TI attacks in *whitebox* and *blackbox* scenarios. We evaluate the *transferability* of the reconstructed face images by considering the situation where the adversary tries to reconstruct face images of the templates leaked from a FR system and use the reconstructed face images to impersonate the same users in another FR system (with a different feature extraction model) that the users are enrolled. Indeed, the transferability of TI attacks reveals a critical threat to FR systems, since the reconstructed face images can be used to enter other FR systems that the victim is enrolled in. Considering the *whitebox/blackbox* scenario and the adversary's knowledge of the target FR system, we define five different TI attacks, and comprehensively evaluate the vulnerability of SOTA FR systems to TI attacks. Furthermore, we perform practical evaluations based on presentation attacks using the digital replay and printed photographs of the reconstructed face images, and evaluate the vulnerability of SOTA FR systems.

To elaborate on the contributions of our paper, we summarize them hereunder:

- We present a *comprehensive* vulnerability evaluation of SOTA FR system to TI attacks using 3D face reconstruction from facial templates. Considering the *whitebox/blackbox* scenarios and the adversary's knowledge of the target FR system, we define five different TI attacks and evaluate the vulnerability of SOTA FR systems to different TI attacks as well as *transferability* of reconstructed face images in TI attacks. We also perform a practical evaluation based on presentation attacks using the digital replay and printed

photograph of the reconstructed face images in TI attacks against SOTA FR systems.

- We propose a new method to reconstruct 3D faces from facial templates using a geometry-aware face generator network based on GNeRF. We use the proposed 3D face reconstruction method to introduce *whitebox* and *blackbox* TI attacks against FR systems. To our knowledge, this is the first work to reconstruct 3D faces from facial templates. To use 3D reconstructed face in TI attack against 2D FR systems during the inference stage, we apply optimization on the camera parameters in the input of the GNeRF model and find a pose that improves the success attack rate.
- We learn a mapping from facial templates to the *intermediate* latent space of GNeRF. We train our mapping network with a *semi-supervised* approach, using real and synthetic face images. For the real training face images, we train our mapping within a GAN-based framework to learn the distribution of *intermediate* latent space of GNeRF. For the synthetic training face images, we directly learn the mapping from facial templates to the GNeRF *intermediate* latent codes.

The remainder of this paper is structured as follows. First, we review the related works in Section II. Then, we describe the threat model, our five different defined attacks, and our proposed method in Section III. Next, in Section IV, we present our experiments and discuss our results. Finally, the paper is concluded in Section V.

II. Related Works

Methods in the literature for face reconstruction in TI attacks against FR systems can be generally categorized from different aspects, including the basis of the method (optimization/learning-based), the type of attack (*whitebox*/*blackbox* attack), and the resolution of reconstructed face images (high/low resolution). However, all previous methods generate 2D images in TI attacks against FR systems.

Several methods have been proposed for reconstructing low-resolution 2D face images from facial templates [22], [23], [24], [25], [26], [27]. In [22], authors proposed two *whitebox* methods to reconstruct 2D low-resolution face images from facial templates. In the first method (optimization-based), they used a gradient-descent-based approach on a guiding image or random (noise) image to find an image that minimizes the distance between the template of the reconstructed face image and the target template. In addition, they used several regularization terms to generate a smooth image, including the total variation and Laplacian pyramid gradient normalization [33] of the reconstructed face image. In their learning-based method, they trained a deconvolutional neural network with the same loss function as in their optimization-based method, to generate reconstructed face images. For the evaluation of their method, they only discussed the visual reconstruction quality and did not provide any security evaluation on a FR system.

In [23], authors trained a multi-layer perceptron (MLP), to find the facial landmark coordinates, and a convolutional neural network (CNN), to generate face texture from the given facial template. Next, they used a differentiable warping to combine the estimated landmarks (from MLP) with the generated textures (from CNN) and reconstruct low-resolution 2D face images. They used their method for *whitebox* and *blackbox* attacks. In the *whitebox* attack, they trained their MLP and CNN by minimizing the distance between templates of the original and reconstructed face images. However, for their *blackbox* attack, they trained MLP and CNN separately, and used the warping in the inference only. For the security evaluation, they only reported the histogram of scores between the templates extracted from the original and reconstructed face images and compared it with the histogram of genuine scores.

In [24], authors proposed a learning-based method to generate low-resolution 2D face images in the *blackbox* attacks against FR systems. They proposed two new deconvolutional networks, called NbBlock-A and NbBlock-B, and trained them with either pixel loss (ℓ_1 norm of pixel-level reconstruction error) or perceptual loss (distance of middle layers of VGG-19 [34] when given the original and reconstructed face images). For the security evaluation, they considered two types of attacks and evaluated vulnerability of FR systems. In their first type of attack, they compared the templates extracted from the original and reconstructed face images, and in their second type of attack, they compared the templates extracted from reconstructed images with templates of a different face image of the same user.

In [25] and [26], a same method based on bijection learning is used to train GAN networks with PO-GAN [35] and TransGAN [36] structures, respectively. In the *whitebox* attack, authors minimized the distance between target templates and templates extracted from the reconstructed face images using the FR model. To extend their method to the *blackbox* attack, they proposed to use the distillation of knowledge to train a student network that mimics the target FR model. However, they did not report any detail about the training of the student network (e.g., network structure, etc.) nor published their source code. For the security evaluation, they reported the matching accuracy between the reconstructed image and another original image in each positive pair in their TI attacks. However, they did not evaluate the vulnerability of FR systems at different threshold configurations.

In [27], authors proposed a 3-step method to reconstruct low-resolution 2D face images in the *blacbox* attack. In the first step, they trained a general face generator network based on GAN. In the second step, they trained a MLP to map the templates to the templates of a known (i.e., *whitebox* knowledge) FR model. In the third step, they used an optimization on the latent space of their face generator to find a latent code that can generate a face image that maximizes two terms; the cosine similarity between the templates (mapped templates and the templates extracted by the known FR model) and the discriminator score (for being a real face image). For their security evaluation, they reported the adversary's success attack rate (SAR), but they did not specify the system's operation configuration, such as the system's recognition false match rate (FMR).

In contrast to the most works in the literature that generate low-resolution 2D face images, recently few methods are

TABLE I
COMPARISON WITH RELATED WORKS

Reference	Method Basis	2D/3D	Resolution	Whitebox/ Blackbox	Transferability Evaluation	Practical Presentation Attack Evaluation	Available source code
[22]	1) optimization 2) learning	2D	low	whitebox	✗	✗	✗
[23]	learning	2D	low	both*	✗	✗	✗
[24]	learning	2D	low	blackbox	✗	✗	✓
[25]	learning	2D	low	both†	✗	✗	✗
[26]	learning	2D	low	both†	✗	✗	✗
[27]	learning + optimization	2D	low	blackbox	✗	✗	✗
[28]	learning	2D	high	blackbox	✗	✗	✓
[29]	1) learning 2) optimization	2D	high	blackbox	✗	✗	✗
[30]	learning	2D	low + high§	blackbox	✗	✗	✗
[31]	optimization	2D	high	blackbox	✗	✗	✓
[32]	optimization	2D	high	blackbox	✗	✗	✗
[Ours]	**learning**	**3D**	**high**	**both‡**	✓	✓	✓

*The method is based on the *whitebox* attack, and is also applied in the *blackbox* scenario by removing a loss term that required the FR model.
†The method is based on the *whitebox* attack, and is extended to the *blackbox* with knowledge distillation of the FR model.
‡The method is based on the *whitebox* attack, and is extended to *blackbox* using a different FR model.
§They first reconstruct low-resolution face images and then apply a super-resolution model to generate high resolution face images.

proposed for high-resolution 2D face reconstruction. In [28], authors proposed a learning-based method to reconstruct high-resolution 2D face images in the *blackbox* attack. They used a pretrained StyleGAN2 [37] to generate some face images and extracted the templates using the FR model. Then, they trained a MLP to map facial templates to the input latent codes of StyleGAN2 [37]. For the security analysis, they considered two types of attacks as defined in [24] and evaluated the vulnerability of FR systems. They also evaluated their reconstructed face images with a commercial-off-the-shelf (COTS) presentation attack detection (PAD) system, also known as face liveness detection in their paper. However, the authors did not perform a *practical* presentation attack scenario, in which the images should have been recaptured by camera prior to be fed to the COTS PAD. Similarly, in [29], authors proposed a learning-based method for high-resolution 2D face reconstruction in the *blackbox* attack. They learned three mapping networks from the facial templates to three separate parts in the intermediate latent space of StyleGAN. Each of these mapping networks is composed of a MLP and is used to reconstruct coarse to fine information of face image. They also proposed to find this mapping with optimization instead of learning the mapping networks. For the security analysis, they did not report success attack rate (percentage) for any configuration. They only reported the histogram of the distance between templates of reconstructed and original face images and compared it with the histogram of templates for random pair of images (i.e., zero-effort impostor).

In [30], authors used a learning-based method based on a conditional denoising diffusion probabilistic model to reconstruct 2D face images in *blackbox* attack. They used the conditional diffusion model in [38] and iteratively denoise an input Gaussian noise conditioned with facial templates to generate low resolution (i.e., 64×64) face images from facial templates. Then, they used a super-resolution network to generate face images with a higher resolution (i.e., 256×256). Compared to other learning-based methods, their proposed method is relatively very slow,[6] because of iterative reconstruction in the inference stage. In addition, compared to other methods, that directly generate high-resolution face images, the method in [30] first reconstructs low-resolution face images and then uses a super-resolution to generate high-resolution face images. For security analysis, similar to [25], [26], they reported the matching accuracy between the reconstructed and a different original image in each positive pair, and did not evaluate the vulnerability of FR systems at different threshold configurations.

In [31], authors proposed a optimization on the latent vector (i.e., input noise) of StyleGAN2 [37] to find latent codes which generates face images with templates similar to the target templates. They solved this optimization with a grid-search and simulated annealing [39] approach for the *blackbox* scenario. However, since their method is computationally expensive,[7] they evaluated their method on only 20 face images and reported distance between the original templates and templates of the reconstructed face images. Along the same lines, in [32] authors considered a similar optimization to [31] on the latent vector of StyleGAN2 [37], but instead of grid-search, they solved the optimization using the standard genetic algorithm [40] for the *blackbox* attack. For the security analysis, they also considered two types of attacks as defined in [24] and evaluated the vulnerability of FR systems. Moreover, they evaluated their reconstructed face images using three COTS PAD systems (called liveness detection in their paper). However, similar to [28], they did not perform a *practical* presentation attack scenario by recapturing the reconstructed face images.

Table I compares our paper with the previous works in the literature. To our knowledge, our proposed method is the first method on 3D face reconstruction from facial templates (which

[6]They reported four minutes to reconstruct 64×64 face images and the super-resolution to 256×256 on a NVIDIA RTX 3090 GPU.
[7]They reported 5 minutes execution time to reconstruct each single image on a system equipped with graphic card.

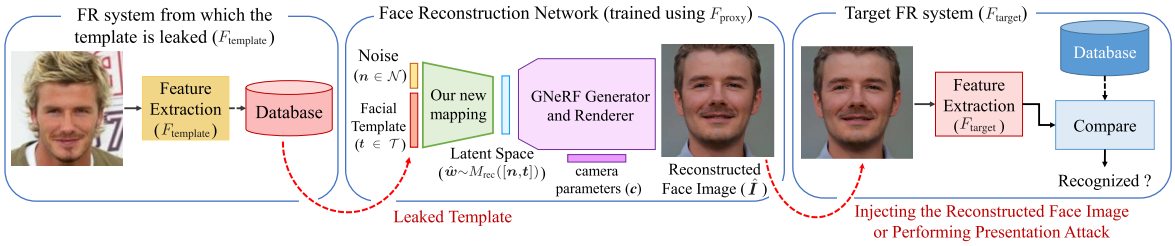

Fig. 3. Block diagram of our threat model.

are extracted from 2D face recognition models). Moreover, in contrast to most works in the literature, our method generates high-resolution face images. We also propose our method for both *whitebox* and *blackbox* attacks against FR systems and evaluate the *transferability* of our reconstructed face images (which has not been reported before for TI attacks). Furthermore, we perform practical presentation attacks against FR systems using the reconstructed face images. Last but not least, the source code of all the experiments in this paper is publicly available to facilitate the reproducibility of our work.

III. Proposed Method

We describe our threat model and define different TI attacks against FR systems in Section III-A (as depicted in Fig. 3). Then, we describe our proposed method to reconstruct 3D faces from facial templates in Section III-B. In the inference stage, we optimization on the camera parameters to generate a face image that can improve the success attack rate, as described in Section III-C. Fig. 4 illustrates the block diagram of the proposed TI attack, including our 3D face reconstruction method and our optimization on camera parameters during the inference stage.

A. Threat Model

We consider the situation where the adversary gains access to the database of a FR system (F_{template}), and aims to invert its templates. The adversary is also assumed to have access[8] to a feature extractor model F_{proxy} (which can be the same or different than F_{template}). The adversary trains a face reconstruction model to reconstruct face images from templates extracted by F_{template}, and uses the reconstructed face images to impersonate into the same or a different FR system (F_{target}). Therefore, we consider the following properties for the adversary:

- *Adversary's goal:* The adversary aims to reconstruct face images from templates stored in the database of a FR system (F_{template}), and use the reconstructed face images to enter the same or a different FR system (we call it the target FR system, F_{target}).
- *Adversary's knowledge:* The adversary has the following information:
  - The leaked face templates $\boldsymbol{t}_{\text{leaked}}$ of users, which are enrolled in the database of F_{template}.
  - The adversary also has the *whitebox* knowledge of a feature extractor model (F_{proxy}). It is worth mentioning that F_{proxy} can be similar to or different from F_{template} and F_{target}.
- *Adversary's capability:* We consider two scenarios for the adversary's capability:
  - The adversary can perform a presentation attack using the reconstructed face images to impersonate and enter the target FR system (e.g., using digital replay attacks or printed photographs).
  - The adversary can inject the reconstructed face image as a query to the target FR system.
- *Adversary's strategy:* The adversary trains a face reconstruction model to invert the leaked facial templates $\boldsymbol{t}_{\text{leaked}}$. Then, based on the adversary's capability, the adversary can use the reconstructed face images to either perform a presentation attack or inject the reconstructed face image as a query to the target FR system.

In our threat model, we consider three different feature extraction models, including $F_{\text{template}}(.)$, $F_{\text{proxy}}(.)$, and $F_{\text{target}}(.)$. Fig. 3 illustrates the block diagram of our threat model. Based on the target FR system and the adversary's knowledge, we can define five different attacks:

- *Attack 1:* The adversary has the *whitebox* knowledge of the feature extractor of the FR system from which the template is leaked and aims to impersonate to the same FR system (i.e., $F_{\text{template}} = F_{\text{proxy}} = F_{\text{target}}$).
- *Attack 2:* The adversary has the *whitebox* knowledge of the feature extractor of the FR system from which the template is leaked, but aims to impersonate to a different FR system (i.e., $F_{\text{template}} = F_{\text{proxy}} \neq F_{\text{target}}$).
- *Attack 3:* The adversary aims to impersonate to the same FR system from which the template is leaked, but has only the *blackbox* access to the feature extractor of the FR system. Instead, the adversary has the *whitebox* knowledge of another FR model to use for training the face reconstruction model (i.e., $F_{\text{template}} = F_{\text{target}} \neq F_{\text{proxy}}$).
- *Attack 4:* The adversary aims to impersonate to a different FR system than the one which from the template is leaked. In addition, the adversary has the *whitebox* knowledge of the feature extractor of the target FR system (i.e., $F_{\text{template}} \neq F_{\text{proxy}} = F_{\text{target}}$).
- *Attack 5:* The adversary aims to impersonate to a different FR system from which the template is leaked, and has only

[8] The adversary can use F_{proxy} for training the face reconstruction network.

TABLE II
DIFFERENT TI ATTACKS AGAINST FR SYSTEMS IN OUR THREAT MODEL

Attack Type	$\mathbf{F}_{\mathbf{template}}$ [1]	$\mathbf{F}_{\mathbf{proxy}}$ [2]	$\mathbf{F}_{\mathbf{target}}$ [3]
Attack 1	□	□	□
Attack 2	□	□	♦
Attack 3	■	△	■
Attack 4	■	△	△
Attack 5	■	△	♦

[1] $\mathbf{F}_{\mathbf{template}}$: the FR system from which the template is leaked.
[2] $\mathbf{F}_{\mathbf{proxy}}$: the FR model which adversary has access to and use it for training the TI model (i.e., always whitebox).
[3] $\mathbf{F}_{\mathbf{target}}$: the target FR system that the adversary aims to enter using the reconstructed face image from the TI attack.

FR models are indicated with symbols, where having the same (different) symbol means the same (different) FR models are used. Each symbol is also either filled with white or black, indicating whitebox or blackbox knowledge to the corresponding model, respectively.

the *blackbox* knowledge of the both the FR systems. However, the adversary instead has the *whitebox* knowledge of another FR model to use for training the face reconstruction model (i.e., $F_{\text{template}} \neq F_{\text{proxy}} \neq F_{\text{target}}$).

Table II summarizes different TI attack types in our threat model as well as the adversary's knowledge of different FR models in each type of attack. In all types of attacks, the leaked facial templates to be reconstructed are from F_{template} and the reconstructed face image is used to attack target FR system F_{target}. In attack 1 and attack 3, the target FR system is the same as the FR system from which the template is leaked (i.e., $F_{\text{template}} = F_{\text{target}}$). However, in attacks 2, 4, and 5, the target FR system is different from the FR system from which the template is leaked (i.e., $F_{\text{template}} \neq F_{\text{target}}$), and therefore in attack 2, 4, and 5, the *transferability* of reconstructed face images in attacks against different FR systems is evaluated. Comparing different types of attacks, in attack 1 the adversary has knowledge of the FR system from which the template is leaked and aims to enter the same FR system, therefore it is expected that attack 1 may be the easiest attack. In contrast, in attack 5 the adversary does not have the whitebox knowledge of the FR system from which the template is leaked or the target FR system, and thus attack 5 may be the hardest attack for the adversary.

B. Proposed 3D Face Reconstruction

To reconstruct 3D faces from facial templates, we use a pretrained EG3D [18] model as a geometry-aware face generator network based on GNeRF. This model consists of two networks, a mapping network and a generator and renderer network. The mapping network M_{GNeRF} takes a random noise $\boldsymbol{z} \in \mathcal{Z}$ in the input and generates an *intermediate* latent code $\boldsymbol{w} = M_{\text{GNeRF}}(\boldsymbol{z}) \in \mathcal{W}$. The *intermediate* latent code $\boldsymbol{w}$ provides more control over the generated face images than input random noise $\boldsymbol{z}$. The generator and renderer network $G(\cdot,\cdot)$ takes the *intermediate* latent code $\boldsymbol{w}$ and camera parameters $\boldsymbol{c}$, to generate a face image $\boldsymbol{I} = G(\boldsymbol{w}, \boldsymbol{c})$ from an arbitrary view. To reconstruct 3D faces from facial templates, we learn a new mapping $M_{\text{rec}} : \mathcal{T} \rightarrow \mathcal{W}$ from the facial templates $\boldsymbol{t} \in \mathcal{T}$ to the *intermediate* latent space $\mathcal{W}$ of the GNeRF model. Then, we feed the mapped *intermediate* latent vector $\hat{\boldsymbol{w}}$ along with camera parameters $\boldsymbol{c}$ into the GNeRF model $G(\cdot,\cdot)$ to generate a face image $\hat{\boldsymbol{I}} = G(\hat{\boldsymbol{w}}, \boldsymbol{c})$ from an arbitrary view corresponds to the camera parameters $\boldsymbol{c}$. We train our mapping network M_{rec} simultaneously using real and synthetic training data with a *semi-supervised* approach as follows:

1) Unsupervised Learning Using Real Training Data: To train our mapping network $M_{\text{rec}}(.)$ with the real training data, we use a set of real face images $\{\boldsymbol{I}_{\text{real},i}\}_{i=0}^{N}$ and extract the facial template $\boldsymbol{t}_{\text{real},i} = F_{\text{template}}(\boldsymbol{I}_{\text{real},i})$ from each face image $\boldsymbol{I}_{\text{real},i}$ using the FR model $F_{\text{template}}(.)$. We assume that the adversary does not have any information about the training dataset of $F_{\text{template}}(.)$and $F_{\text{target}}(.)$, and thus use another dataset for training the face reconstruction model. Since we do not have the true value of the *intermediate* latent space $\mathcal{W}$ of the GNeRF model for the real face images in $\{\boldsymbol{I}_{\text{real},i}\}_{i=0}^{N}$, we consider training our mapping network using the real training data as *unsupervised* learning. For the real training data, we train our mapping $M_{\text{rec}}(.)$ within a GAN-based framework based on Wasserstein GAN (WGAN) [41] algorithm to learn the distribution of *intermediate* latent space $\mathcal{W}$ of the GNeRF model. In this framework, our mapping network M_{rec} acts as the generator of our WGAN training and generates a latent code $\hat{\boldsymbol{w}} = M_{\text{rec}}([\boldsymbol{n}, \boldsymbol{t}])$ from a random vector $\boldsymbol{n} \in \mathcal{N}$ and the facial template $\boldsymbol{t}$. In our WGAN framework, we can also generate the real latent code $\boldsymbol{w} = M_{\text{GNeRF}}(\boldsymbol{z}) \in \mathcal{W}$ using the GNeRF mapping function M_{GNeRF} and a random vector $\boldsymbol{z} \in \mathcal{Z}$. Then, we can use a critic network $C(.)$ to score the latent codes generated by GNeRF mapping (as real) and our mapping (as fake). Hence, we can train our mapping M_{rec} along with the the critic network $C(.)$ in the WGAN framework using the following loss functions:

$$\mathcal{L}_{C}^{\text{WGAN}} = \mathbb{E}_{\boldsymbol{w} \sim M_{\text{GNeRF}}(\boldsymbol{z})}[C(\boldsymbol{w})] - \mathbb{E}_{\hat{\boldsymbol{w}} \sim M_{\text{rec}}([\boldsymbol{n},\boldsymbol{t}])}[C(\hat{\boldsymbol{w}})] \quad (1)$$

$$\mathcal{L}_{M_{\text{rec}}}^{\text{WGAN}} = \mathbb{E}_{\hat{\boldsymbol{w}} \sim M_{\text{rec}}([\boldsymbol{n},\boldsymbol{t}])}[C(\hat{\boldsymbol{w}})] \quad (2)$$

In addition to the WGAN training, we feed the generated latent code $\hat{\boldsymbol{w}} = M_{\text{rec}}([\boldsymbol{n}, \boldsymbol{t}])$ to the GNeRF model to generate the face image $\hat{\boldsymbol{I}} = G(\hat{\boldsymbol{w}}, \boldsymbol{c})$, and then use the generated face image $\hat{\boldsymbol{I}}$ to optimize our mapping network $M_{\text{rec}}(.)$ using the following multi-term loss function:

$$\mathcal{L}_{\text{real}}^{\text{rec}} = \mathcal{L}^{\text{Pixel}} + \mathcal{L}^{\text{ID}}, \quad (3)$$

where $\mathcal{L}^{\text{Pixel}}$ and $\mathcal{L}^{\text{ID}}$ are pixel loss and ID loss, respectively, and are defined as:

$$\mathcal{L}^{\text{Pixel}} = \mathbb{E}_{\hat{\boldsymbol{w}} \sim M_{\text{rec}}([\boldsymbol{n},\boldsymbol{t}])}[\|\boldsymbol{I} - G(\hat{\boldsymbol{w}}, \boldsymbol{c})\|_2^2] \quad (4)$$

$$\mathcal{L}^{\text{ID}} = \mathbb{E}_{\hat{\boldsymbol{w}} \sim M_{\text{rec}}([\boldsymbol{n},\boldsymbol{t}])}[\|F_{\text{proxy}}(\boldsymbol{I}) - F_{\text{proxy}}(G(\hat{\boldsymbol{w}}, \boldsymbol{c}))\|_2^2] \quad (5)$$

The pixel loss $\mathcal{L}^{\text{Pixel}}$ minimizes the pixel-level reconstruction error and the ID loss $\mathcal{L}^{\text{ID}}$ optimizes the model to generate face images that have similar facial templates (extracted by F_{proxy}) to the templates of the original image $\boldsymbol{I}$.

2) Supervised Learning Using Synthetic Training Data: To train our mapping network $M_{\text{rec}}(.)$ with the synthetic training face images, we use the pretrained GNeRF model to generate a set of random face images $\{\boldsymbol{I}_{\text{syn},i}\}_{i=0}^{K}$. Therefore, as opposed to real training data, we have the true value of *intermediate* latent space $\boldsymbol{w} \in \mathcal{W}$ to generate the same synthetic face image, and therefore can directly learn the GNeRF *intermediate* latent code

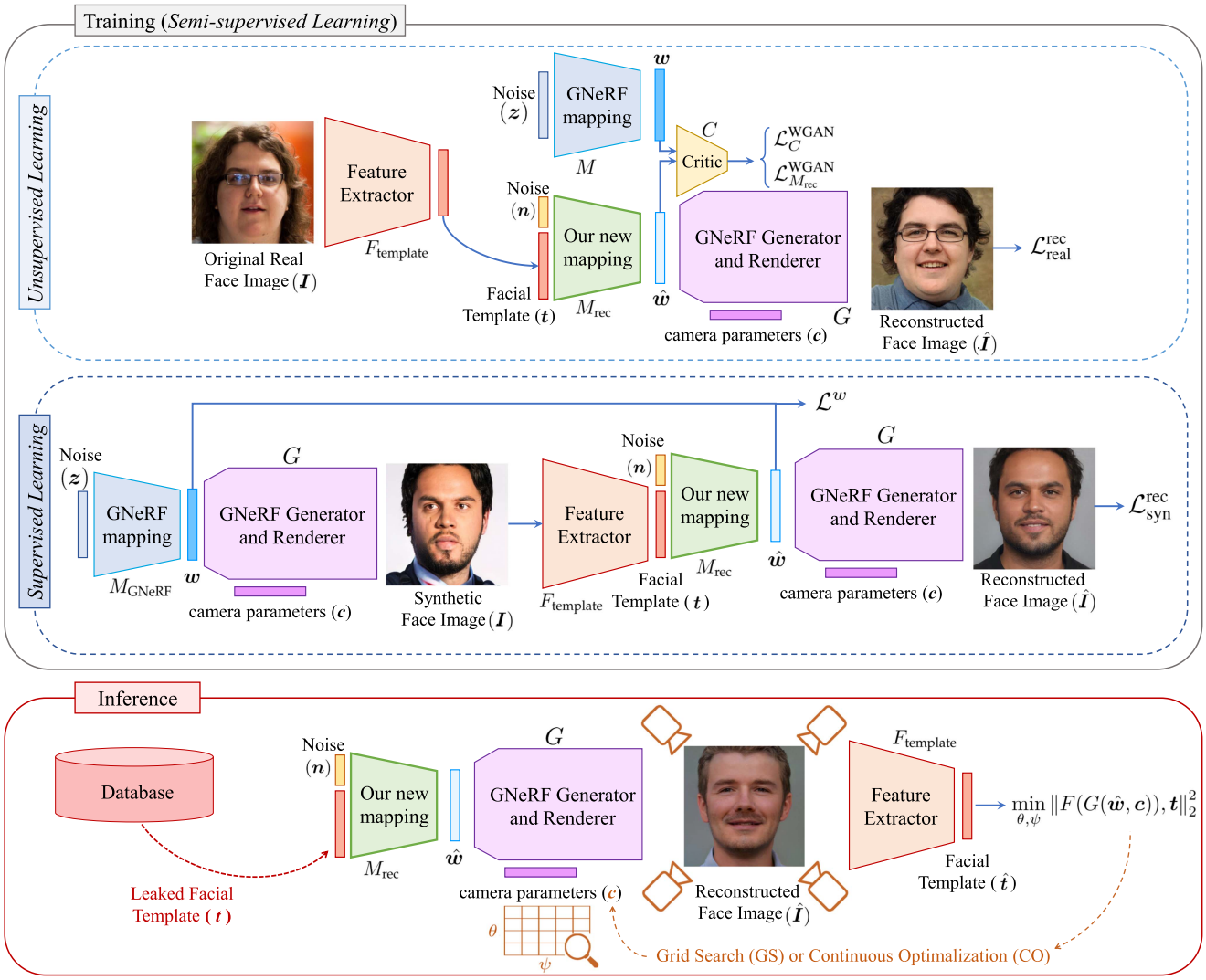

Fig. 4. Block diagram of our proposed TI attack: during the training process, a *semi-supervised* approach is used to learn our mapping M_{rec} (illustrated as a green block) from the facial templates to the *intermediate* latent space of the GNeRF model. We use *real* training data (where we don't have the corresponding latent code) and *synthetic* training data (where we have the corresponding latent code $\boldsymbol{w}$), simultaneously, for *unsupervised* and *supervised* learning in our method. In the inference stage, the leaked template $\boldsymbol{t}$ is fed into our mapping network to find corresponding vector $\hat{\boldsymbol{w}} = M_{\text{rec}}([\boldsymbol{n}, \boldsymbol{t}])$ in the *intermediate* latent space of the GNeRF. Then, camera parameters $\boldsymbol{c}$ along with $\hat{\boldsymbol{w}}$ are given to the generator and renderer of GNeRF G to generate a reconstructed face image $\hat{\boldsymbol{I}} = G(\hat{\boldsymbol{w}}, \boldsymbol{c})$. To enhance the attack, we propose an optimization (grid search or continuous optimization) on two of the camera parameters, θ and ψ, from $\boldsymbol{c}$, to find the best pose, which minimizes the distance between the template of reconstructed face image and the leaked template $\boldsymbol{t}$.

$\boldsymbol{w} = M_{\text{GNeRF}}(\boldsymbol{z})$ from template $\boldsymbol{t}_{\text{syn},i} = F_{\text{template}}(\boldsymbol{I}_{\text{syn},i})$. Hence, we consider training our mapping network using the synthetic data as *supervised* learning. In addition to directly learning the *intermediate* latent code $\boldsymbol{w}$, we use the generated face image to optimize our mapping network by minimizing the following multi-term loss function:

$$\mathcal{L}_{\text{syn}}^{\text{rec}} = \mathcal{L}^{w} + \mathcal{L}^{\text{Pixel}} + \mathcal{L}^{\text{ID}}, \tag{6}$$

where $\mathcal{L}^{\text{Pixel}}$ and $\mathcal{L}^{\text{ID}}$ are the pixel loss (4) and ID loss (5), respectively. Moreover, $\mathcal{L}^{w}$ is w-loss to directly learn the latent space of GNeRF by minimizing the mean squared error between $\boldsymbol{w}$ and $\hat{\boldsymbol{w}} = M_{\text{rec}}([\boldsymbol{n}, \boldsymbol{t}])$ as follows:

$$\mathcal{L}^{w} = \mathbb{E}_{\boldsymbol{w} \sim M_{\text{GNeRF}}(\boldsymbol{z})}[\|\boldsymbol{w} - M_{\text{rec}}([\boldsymbol{n}, \boldsymbol{t}])\|_2^2] \tag{7}$$

To train our networks, we use Adam [42] optimizer and optimize the parameters of our new mapping network $M_{\text{rec}}(.)$ for $\mathcal{L}_{\text{real}}^{\text{rec}}$ (i.e., (3)) and $\mathcal{L}_{\text{syn}}^{\text{rec}}$ (i.e., (6)) losses in every iteration of our training process (also shown in Fig. 4). However, in the WGAN framework, we update weights of our new mapping network $M_{\text{rec}}(.)$ and critic network $C(.)$ every n_M^{WGAN} (for minimizing $\mathcal{L}_{M_{\text{rec}}}^{\text{WGAN}}$ in (2)) and every n_C^{WGAN} (for minimizing $\mathcal{L}_C^{\text{WGAN}}$ in (1)) iterations, respectively. Algorithm 1 represents our training process. We should note that our mapping network M_{rec} has 2 fully-connected layers with Leaky ReLU activation function.

C. Camera Parameters Optimization

After generating a 3D reconstruction of face from the facial template using our proposed method described in Section III-B, the adversary needs to select a pose to generate a 2D reconstructed face image to inject into the system or perform a presentation attack. To this end, during the inference stage we

Algorithm 1: Training Process of Our New Mapping Network.

Require: θ_M, parameters of $M_{\text{rec}}(.)$ network. θ_C, parameters of network $C(.)$.
Require: n_{epoch}, no. epochs. $n_{\text{iteration}}$, no. iterations in each epoch. n_M^{WGAN}, no. training iterations after which to optimize θ_M in WGAN. n_C^{WGAN}, no. training iterations after which to optimize θ_C in WGAN. δ, the WGAN clipping parameter.
Require: α_M^{real}, learning rate for optimizing θ_M based on $\mathcal{L}_{\text{real}}^{\text{rec}}$. α_M^{syn}, learning rate for optimizing θ_M based on $\mathcal{L}_{\text{syn}}^{\text{rec}}$. α_M^{WGAN}, learning rate for optimizing θ_M in WGAN. α_C^{WGAN}, learning rate for optimizing θ_C in WGAN.
Require: $\mathcal{D}_{\text{real}}$, a dataset of real face images and corresponding facial templates extracted using F_{template}.

procedure Training
  Initialize θ_C and θ_M
  for epoch $= 1, \ldots, n_{\text{epoch}}$ **do**
    for itr $= 1, \ldots, n_{\text{iteration}}$ **do**
      Sample a batch from $\mathcal{Z}$ and calculate:
        $g_{\theta_M}^{\text{syn}} \leftarrow \nabla_{\theta_M} \mathcal{L}_{\text{syn}}^{\text{rec}}$
        $\theta_M \leftarrow \theta_M - \alpha_M^{\text{syn}} \cdot \text{Adam}(\theta_M, g_{\theta_M}^{\text{syn}})$
      Sample a batch from $\mathcal{D}_{\text{real}}$ and calculate:
        $g_{\theta_M}^{\text{real}} \leftarrow \nabla_{\theta_M} \mathcal{L}_{\text{real}}^{\text{rec}}$
        $\theta_M \leftarrow \theta_M - \alpha_M^{\text{real}} \cdot \text{Adam}(\theta_M, g_{\theta_M}^{\text{real}})$
      if itr $\mod n_M^{\text{WGAN}} = 0$ **then**
        $g_{\theta_M}^{\text{WGAN}} \leftarrow \nabla_{\theta_M} \mathcal{L}_M^{\text{WGAN}}$
        $\theta_M \leftarrow \theta_M - \alpha_M^{\text{WGAN}} \cdot \text{Adam}(\theta_M, g_{\theta_M}^{\text{WGAN}})$
      end if
      if itr $\mod n_C^{\text{WGAN}} = 0$ **then**
        Sample a batch $\boldsymbol{w} \sim \mathcal{W}$ and calculate:
          $g_{\theta_C}^{\text{WGAN}} \leftarrow \nabla_{\theta_C} \mathcal{L}_C^{\text{WGAN}}$
          $\theta_C \leftarrow \theta_C - \alpha_C^{\text{WGAN}} \cdot \text{Adam}(\theta_C, g_{\theta_C}^{\text{WGAN}})$
          $\theta_C \leftarrow \text{clip}(\theta_C, -\delta, \delta)$
      end if
    end for
  end for
end procedure

can optimize the camera parameters to find a pose that increases the success attack rate (SAR). In other words, having the 3D reconstruction of a face, we would like to find the camera parameters so that the 2D generated face image has a facial template that is more similar to the leaked templates than the templates of any other pose. Among different camera parameters $\boldsymbol{c}$, we consider the parameters that corresponds to the camera rotations and therefore can change the pose of the generated face image. It is noteworthy that by changing the camera rotations, we want to vary the pitch and yaw rotations of the reconstructed face and do not want to modify the roll rotation. As a matter of fact, the effect of any roll rotation will be eliminated in the FR system through the face alignment in the pre-processing step of the feature extraction. We consider two different approaches to optimize camera parameters as follows:

1) Grid Search (GS): In our grid search approach, we consider pre-defined steps to change the camera pitch $\theta \in \Theta$ and yaw $\psi \in \Psi$ and generate corresponding camera parameters $\boldsymbol{c}$. We generate the 2D face images for all values of camera rotation steps (θ_{step} and ψ_{step}) and find the facial templates for each generated image. Finally, we select the face image $\hat{\boldsymbol{I}} = G(M_{\text{rec}}([\boldsymbol{n}, \boldsymbol{t}]), \boldsymbol{c})$ which has a template $\hat{\boldsymbol{t}} = F_{\text{template}}(\hat{\boldsymbol{I}})$ that minimizes the mean squared error with the leaked template $\boldsymbol{t}$:

$$\min_{\theta,\psi} \left\| \hat{\boldsymbol{t}} - \boldsymbol{t} \right\|_2^2, \tag{8}$$

Note that the grid search can be applied in both *whitebox* and *blackbox* scenarios (i.e., all attacks defined in Section III-A) using the FR model F_{template}.

2) Continuous Optimization (CO): For continuous optimization, we start from the frontal camera parameters and use the Adam [42] optimizer to solve the following minimization using the mapped latent code $\hat{\boldsymbol{w}} = M_{\text{rec}}([\boldsymbol{n}, \boldsymbol{t}])$:

$$\min_{\theta,\psi} \left\| F_{\text{template}}(G(\hat{\boldsymbol{w}}, \boldsymbol{c})) - \boldsymbol{t} \right\|_2^2, \tag{9}$$

By solving this optimization, we can find the θ and ψ rotations and the corresponding camera parameters $\boldsymbol{c}$ that lead to a face image with the template close to the leaked template $\boldsymbol{t}$. In contrast to the grid search, the continuous optimization approach can be applied only when the adversary has the *whitebox* knowledge of F_{template} (i.e., attack 1 and attack 2).

IV. Experiments

In this section, we evaluate the vulnerability of SOTA FR systems to our TI attacks defined in Section III. First, in Section IV-A we describe our experimental setup. In Section IV-B, we consider the case where the adversary can inject the reconstructed face image as a query to the system to impersonate, and present our experimental results. In Section IV-C, we consider the situation where the adversary uses the reconstructed face images to perform presentation attacks and evaluate the vulnerability of SOTA FR systems. Finally, we discuss our findings in Section IV-D.

A. Experimental Setup

1) Face Recognition Models: In our experiments, we evaluate the vulnerability of different SOTA FR models to our TI attacks. We consider two SOTA models, including ArcFace [7], ElasticFace [43], as the models from which templates are leaked (i.e., F_{template}) and use our proposed method to reconstruct face images. Then, to evaluate the transferability of reconstructed face images, we also use four different FR models with SOTA backbones from FaceX-Zoo [44] for the target FR system (i.e., F_{target}), including AttentionNet [45], HRNet [46], RepVGG [47], and Swin [48]. The recognition performances of these models are reported in Table III.

2) Datasets: All the FR models used in our experiments are trained on the MS-Celeb1M dataset [49]. However, we assume that the adversary does not have knowledge about the training data of the FR network (either F_{template} or F_{target}), and uses another dataset for training the face reconstruction model. We

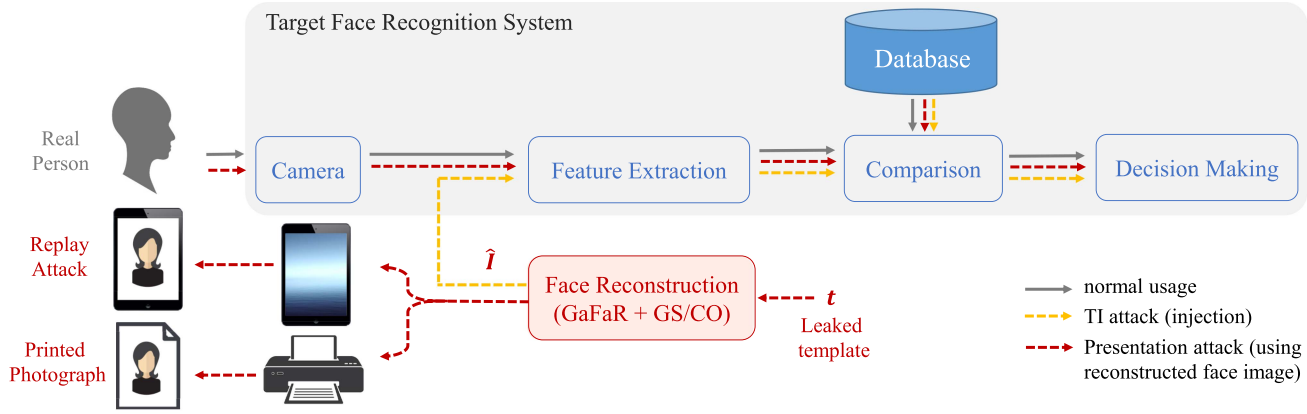

Fig. 5. Block diagram of a FR system and data flows in normal usage (gray solid arrows), TI attack by injecting the reconstructed face image (orange dashed arrows), and performing presentation attack using the reconstructed face image (red dashed arrows).

TABLE III
RECOGNITION PERFORMANCE OF FACE RECOGNITION MODELS USED IN OUR EXPERIMENTS IN TERMS OF TRUE MATCH RATE (TMR) AT THE THRESHOLDS CORRESPOND TO FALSE MATCH RATES (FMRS) OF 10^{-2} AND 10^{-3} EVALUATED ON THE MOBIO AND LFW DATASETS

model	MOBIO		LFW	
	FMR=10^{-2}	FMR=10^{-3}	FMR=10^{-2}	FMR=10^{-3}
ArcFace	100.00	99.98	97.60	96.40
ElasticFace	100.00	100.00	96.87	94.70
AttentionNet	99.71	97.73	84.27	72.77
HRNet	98.98	98.23	89.30	78.43
RepVGG	98.75	95.80	77.20	58.07
Swin	99.75	98.98	91.70	87.83

The values are in percentage.

use the Flickr-Faces-HQ (FFHQ) dataset [6], which consists of 70,000 high-resolution (i.e., 1024×1024) face images crawled from the internet (without identity labels), for training our 3D face reconstruction model. We randomly split the FFHQ dataset to train (90%) and validation (10%) subsets.

To evaluate the vulnerability of FR systems to TI attacks, we consider two other different face image datasets with identity labels, including the MOBIO [50] and Labeled Faces in the Wild (LFW) [51] datasets. The MOBIO dataset includes face images captured using mobile devices from 150 people in 12 sessions (6-11 samples in each session). The LFW dataset includes 13,233 face images of 5,749 people collected from the internet, where 1,680 people have two or more images.

3) Evaluation Protocol: To implement each of the attacks described in Section III-A, we build one or two separate FR systems using the same or two different SOTA feature extractor models (based on the attack type). If the target FR system is the *same* as the system from which the template is leaked (i.e., $F_{\text{template}} = F_{\text{target}}$, as in attack 1 and attack 3), we have only one FR system. Otherwise, if the target system is *different* than the system from which the template is leaked (i.e., $F_{\text{template}} \neq F_{\text{target}}$, as in attack 2, attack 4, and attack 5), we have two FR systems with *two* different feature extractors. We should note that in the transferability evaluations, we need that the subjects whose templates are leaked to be enrolled in the target system too. Therefore, to implement any of the attacks which require two FR systems (i.e., attack 2, attack 4, and attack 5), we use one of our evaluation datasets to build both FR systems (i.e., F_{template} and F_{target}).

To evaluate the vulnerability to all our TI attacks, we assume that the target FR system is configured at the threshold corresponding to a false match rate (FMR) of 10^{-2} or 10^{-3}, and we evaluate the adversary's success attack rate (SAR) in entering that system. In our experiments, we consider two situations, where the adversary can inject the reconstructed face image as a query to the FR system (Section IV-B), or use the reconstructed face image to perform a presentation attack (Section IV-C). Fig. 5 depicts and compares two scenarios of injecting the reconstructed face image or performing a presentation attack. In our evaluation of TI attacks by injecting the reconstructed face image (Section IV-B), we directly inject the reconstructed face images into the feature extractor of the FR system and evaluate the TI attack in terms of SAR. However, in our evaluation of the presentation attack using the reconstructed face image (Section IV-C), we present the reconstructed face image (using either a digital screen or a printed photograph) in front of the camera and evaluate the attack in terms of SAR.

4) Implementation Details and Source Code: To build the FR pipeline and evaluate the TI attacks against FR systems, we use the Bob[9] [52] toolbox. We use the PyTorch package and trained all the networks on a system equipped with an NVIDIA GeForce RTX™ 3090. For the GNeRF model, we use the pretrained model of EG3D[10] with StyleGAN [37] backbone to generate 3D faces with 512×512 high-resolution images from any arbitrary view. For the FR models, we use the pretrained models[11] form Bob and FaceX-Zoo [44] toolboxes.

To train our 3D face reconstruction networks, we consider $n_{\text{epoch}} = 15$, $n_C^{\text{WGAN}} = 4$ and $n_M^{\text{WGAN}} = 2$ in Algorithm 1. Furthermore, the input noise vectors to the mapping network of

[9] Available at https://www.idiap.ch/software/bob/
[10] Available at https://github.com/NVlabs/eg3d
[11] Available at https://gitlab.idiap.ch/bob/bob.bio.face

GNeRF's pretrained network (i.e., $\boldsymbol{z} \in \mathcal{Z}$) and to our mapping network M_{rec} (i.e., $\boldsymbol{n} \in \mathcal{N}$) are both from the standard normal distribution and with 512 and 16 dimensions, respectively. The *intermediate* latent space of GNeRF model has 14×512 dimensions, i.e., $\mathcal{W} \subset \mathbb{R}^{14\times512}$. The templates extracted by the FR models in Table III have 512 dimensions. For simplicity in training our mapping network, we assume that our training face images from the FFHQ dataset (i.e., real data) are frontal.

In our experiments, we use the continuous optimization (in *whitebox* attacks only) and grid search optimization (in both *whitebox* and *blackbox* attacks) in the inference stage, as described in Section III-C, to optimize camera parameters. In the grid search approach, we consider $\psi \in [-45°, +45°]$ and $\theta \in [-30°, +30°]$ for a 11×11 grid with step sizes of $\psi_{\text{step}} = 9°$ and $\theta_{\text{step}} = 6°$. For the continuous optimization, we use Adam optimizer [42] with the learning rate of 10^{-2} and 121 iterations. An ablation study on the effect of these hyperparameters and the corresponding execution times are reported in Section IV-D.

We should note that the source code and the captured images for our presentation attack evaluation are publicly available to help reproduce our results.[12]

B. TI Attack by Injecting Reconstructed Face Images

In this section, we consider the situation where the adversary can inject the reconstructed face image to the feature extractor of the target FR system. We consider SOTA FR models and evaluate the vulnerability of these systems to different TI attacks described in Section III-A in the whitebox (attacks 1-2) and blackbox (attacks 3-5) scenarios.

1) Whitebox Scenario: In attacks 1-2, we assume that the adversary has the *whitebox* knowledge of the FR system from which the template is leaked (i.e., F_{template}) and uses the same feature extraction model for training (i.e., F_{proxy}) the face reconstruction network. We considered ArcFace and ElasticFace models for the system from which the template is leaked (i.e., F_{template}) and evaluate the vulnerability of SOTA FR systems as the target FR systems against attacks 1-2. Table IV compares the vulnerability of different target systems to attacks 1-2 using our method[13] in terms of adversary's SAR at the system's FMR of 10^{-3}. As this table shows, our proposed face reconstruction method achieves considerable SAR values against ArcFace and ElasticFace target FR systems in attack 1. Comparing the SAR values between attack 1 and attack 2, the SAR values degrade for different target FR models in attack 2. However, the reconstructed face images are transferable and can still be used to enter a target system with a different feature extractor. It is also noteworthy that considering the recognition performances in Table III, we can conclude that the target FR system with a higher recognition accuracy is generally more vulnerable to attack 2. For example, when ArcFace is used for F_{template} in Table IV, attacks against ElasticFace and Swin as target FR systems result in the highest SAR, and there is the same order for their recognition performance in Table III. Comparing the frontal reconstructed face images by our proposed method (iGaFaR) with our camera parameter optimizations methods (GaFaR+GS and GaFaR+CO), the results show that camera parameter optimization methods improve SAR in both attack 1 and attack 2. Therefore, camera parameter optimization methods not only enhance the attack against the same system (i.e., attack 1), but are also transferable to other FR systems (i.e., attack 2). Comparing the grid search and continuous optimization methods for camera parameter optimization, the results show that the continuous optimization method achieves higher SAR values, and therefore further enhances our TI attack. Fig. 6 illustrates sample face images and their corresponding frontal face reconstruction as well as a sub-grid of reconstructed face images with different poses from ArcFace templates in the *whitebox* TI attacks (i.e., attacks 1-2). We should note that the reconstructed face images in attack 1 and attack 2 are the same, however, they are used to enter different target FR systems.

TABLE IV
EVALUATION OF WHITEBOX ATTACKS (I.E., ATTACKS 1-2) AGAINST SOTA FR MODELS IN TERMS OF ADVERSARY'S SUCCESS ATTACK RATE (SAR) WHEN INJECTING RECONSTRUCTED FACE IMAGE GENERATED USING OUR FACE RECONSTRUCTION METHODS EVALUATED ON THE MOBIO AND LFW DATASETS

$\mathbf{F}_{\text{database}}$	$\mathbf{F}_{\text{target}}$	MOBIO			LFW		
		M1	M2	M3	M1	M2	M3
ArcFace	ArcFace	84.29	86.67	**89.52**	79.74	82.38	**84.43**
	ElasticFace	78.10	78.10	**80.00**	65.19	67.81	**70.37**
	AttentionNet	65.24	67.14	**69.05**	30.20	33.43	**35.36**
	HRNet	62.86	61.43	**67.14**	30.41	33.26	**35.42**
	RepVGG	45.24	49.05	**55.24**	18.38	20.32	**21.39**
	Swin	70.95	71.90	**77.14**	51.18	53.91	**55.91**
ElasticFace	ArcFace	51.43	59.52	**61.43**	53.07	58.89	**61.08**
	ElasticFace	78.10	83.33	**84.29**	63.06	68.94	**71.78**
	AttentionNet	48.10	51.43	**56.67**	21.69	25.35	**26.92**
	HRNet	51.43	50.95	**54.29**	22.39	26.45	**28.23**
	RepVGG	37.14	40.95	**48.10**	12.80	14.72	**15.97**
	Swin	54.29	54.29	**60.00**	40.47	43.29	**45.59**

All the values are in percentage and SAR values correspond to the threshold where the target system has FMR = 10^{-3}. M1: GaFaR [ours], M2: GaFaR+GS [ours], and M3: GaFaR+CO [ours]. Cells are color-coded according to the type of attack as defined in Section 3 for attack 1 (dark green) and attack 2 (light green).

2) Blackbox Scenario: In attacks 3-5, we assume that the adversary has the *blackbox* knowledge of the feature extractor of the FR system from which the template is leaked (i.e., F_{template}) and uses another feature extraction model for training (i.e., F_{proxy}). Similar to Section IV-B1, we consider ArcFace and ElasticFace models for F_{template} and evaluate the vulnerability of SOTA FR systems in the target FR systems against attacks 3-5. In each case, we also use the other model for F_{proxy} (i.e., ArcFace as F_{template} and ElasticFace as F_{proxy} or ElasticFace as F_{template} and ArcFace as F_{proxy}). Table V compares the performance of our method with *blackbox* methods in the literature [24], [28], [31] for attacks 3-5 in terms of adversary's SAR at system's FMR of 10^{-3}. As the results in this table show, the frontal face reconstruction by our method (i.e, GaFaR) achieves superior performance than previous methods in the literature. Moreover, when we apply camera parameter optimization (i.e., GaFaR+GS) the

[12]Project page: https://www.idiap.ch/paper/gafar

[13]Note that as reported in Table I, none of the *whitebox* face reconstruction methods in the literature has an available source code, and we neither could reproduce their results.

TABLE V
EVALUATION OF BLACKBOX ATTACKS (I.E., ATTACKS 3-5) AGAINST SOTA FR MODELS IN TERMS OF ADVERSARY'S SUCCESS ATTACK RATE (SAR) WHEN INJECTING RECONSTRUCTED FACE IMAGE GENERATED USING DIFFERENT FACE RECONSTRUCTION METHODS EVALUATED ON THE MOBIO AND LFW DATASETS

$F_{database}$	F_{loss}	F_{target}	MOBIO								LFW							
			M1	M2	M3	M4	M5	M6	M7	M8	M1	M2	M3	M4	M5	M6	M7	M8
ArcFace	ElasticFace	ArcFace	5.24	15.24	0.0	1.90	3.33	26.19	50.95	**62.86**	16.83	40.25	4.32	10.97	13.24	57.44	52.72	**61.71**
		ElasticFace	4.29	10.95	0.0	1.43	3.81	17.14	52.38	**55.24**	13.09	34.41	3.32	8.56	6.25	29.06	43.59	**47.57**
		AttentionNet	3.81	6.67	0.0	2.86	2.86	5.71	29.05	**36.19**	1.89	7.21	0.51	1.22	2.12	9.79	14.58	**17.13**
		HRNet	4.29	6.19	0.0	1.43	3.33	10.48	31.90	**41.90**	2.03	7.77	0.41	1.44	1.70	9.51	14.77	**17.35**
		RepVGG	1.90	2.38	0.0	1.90	2.38	3.81	28.10	**32.86**	0.86	4.06	0.23	0.76	1.39	4.48	8.42	**10.11**
		Swin	4.29	13.33	0.0	0.95	4.29	13.81	43.33	**50.00**	8.44	23.82	1.60	4.79	6.22	20.75	29.69	**33.16**
ElasticFace	ArcFace	ArcFace	5.71	18.57	0.0	2.38	3.81	11.43	74.29	**77.62**	20.38	48.66	7.50	15.33	12.23	36.80	71.48	**74.30**
		ElasticFace	16.19	43.81	2.38	3.33	8.10	38.10	84.76	**88.10**	26.96	58.15	10.88	21.45	12.69	53.06	74.77	**78.18**
		AttentionNet	1.43	18.1	0.0	1.90	4.29	7.14	62.38	**65.24**	3.85	16.37	1.53	2.89	3.16	11.16	34.28	**37.34**
		HRNet	6.19	20.0	0.0	0.48	4.76	11.43	61.90	**65.71**	4.10	18.36	1.74	3.45	2.47	11.81	35.87	**39.19**
		RepVGG	7.62	13.81	0.0	0.0	4.29	5.71	47.62	**51.43**	1.75	9.14	0.66	1.54	2.01	6.04	21.12	**22.84**
		Swin	16.19	26.19	0.0	0.95	6.67	17.62	67.14	**70.95**	15.72	38.76	4.13	9.18	8.51	24.22	55.51	**58.12**

All the values are in percentage and SAR values correspond to the threshold where the target system has FMR $= 10^{-3}$. **M1**: NbNetB-M [24], **M2**: NbNetB-P [24], **M3**: NbNetA-M [24], **M4**: NbNetA-P [24], **M5**: Dong *et al.* [28], **M6**: Vendrow and Vendrow [31], **M7: GaFaR [ours]**. and **M8: GaFaR+GS [ours]**. Cells are color-coded according to the type of attack as defined in Section 3 for attack 3 (yellow), attack 4 (orange), and attack 5 (red).

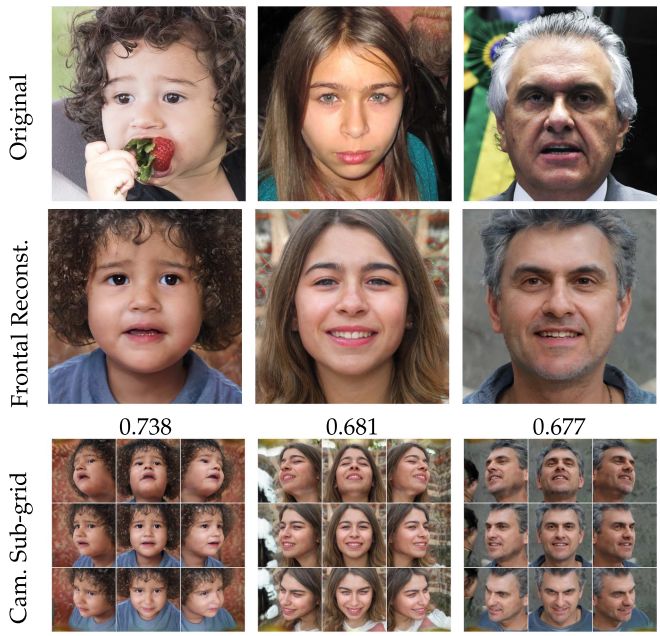

Fig. 6. Sample face images from the FFHQ dataset (first row) and their corresponding frontal face reconstruction (second row) as well as reconstructed face images within the camera parameters sub-grid (third row) using our method in the *whitebox* TI attacks (i.e., attacks 1-2) against ArcFace. The values below each image show the cosine similarity between templates of original and frontal reconstructed face images.

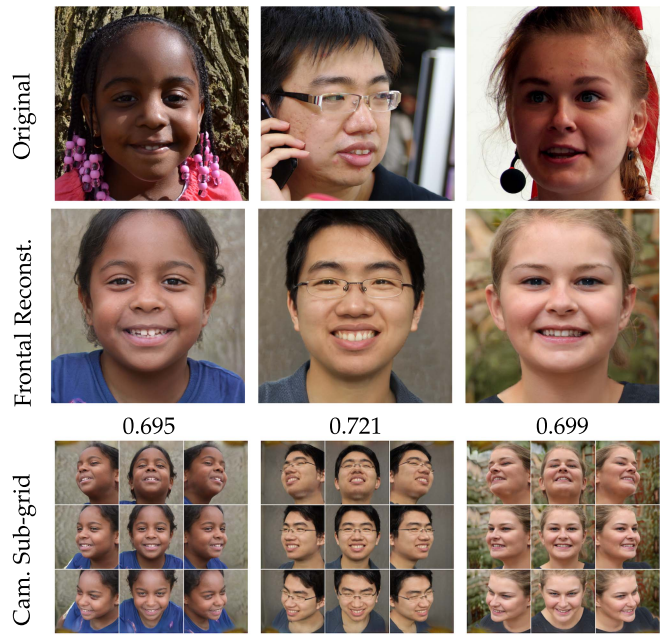

Fig. 7. Sample face images from the FFHQ dataset (first row) and their corresponding frontal (second row) reconstructed face images using our method in the *blackbox* attack against ElasticFace using ArcFace as F_{proxy}. The values below each image show the cosine similarity between templates of original and frontal reconstructed face images.

performance of our attack improves up to 11.91%, 3.98%, and 10.00% compared to our frontal face reconstruction (i.e, GaFaR) in attack 3, attack 4, and attack 5, respectively. Comparing the use of ArcFace and ElasticFace as F_{proxy}, the results show that the SAR values in attacks with the ArcFace model are higher. This can be due to the fact that according to Table III, ArcFace has a better recognition performance than ElasticFace.

Table V also shows that SOTA FR systems are vulnerable to our TI attacks in the *blackbox* scenario. In particular, in attack 5 which is the hardest TI attack, where F_{target}, F_{template}, and F_{proxy} are different, the results show that SOTA FR models (as the target FR system) are still vulnerable to our TI attack. The results of attack 5 for our proposed method also show the transferability of our attack to different FR systems. In addition, similar to the *whitebox* scenario, we can also observe that for TI attacks in the *blackbox* scenario, the FR model with a higher recognition performance is generally more vulnerable to our TI attacks. Comparing the results in Tables IV and V and as expected, attack 1 is the easiest attack with the highest SAR, where F_{template}, F_{proxy}, and F_{target} are the same, and attack 5 is the most difficult attack, where F_{template}, F_{proxy}, and F_{target} are different. Fig. 7 shows sample face images and their corresponding frontal face reconstruction as well as their sub-grids of reconstructed face images with different poses from ElasticFace templates in the *blackbox* TI attack (i.e., attacks 3-5) using ArcFace as

(a) replay attack (b) printed photograph

Fig. 8. Our evaluation setup for performing different types of presentation and capturing presentation using mobile devices: (a) replay attack using Apple iPad Pro, and (b) presentation attack using printed photograph.

F_{proxy}. Similar to attacks 1-2, the reconstructed face images in attacks 3-5 are the same, however, they are used to enter different target FR system.

C. Practical Presentation Attack Using Reconstructed Face Images

In this section, we consider the situation where the adversary uses the reconstructed face image to perform a presentation attack to enter the target FR system. We consider reconstructed face images from ArcFace templates using our proposed face reconstruction method and camera parameter optimizations (i.e., GaFaR, GaFaR+GS, and GaFaR+CO) in both *whitebox* and *blackbox* scenarios, and use the reconstructed face images in each case to perform presentation attacks. We perform our presentation attacks against different SOTA FR systems based on the various TI attacks described in Section III-A. Therefore, we similarly have five different presentation attacks according to the adversary's knowledge of the FR system from which the template is leaked (i.e., F_{template}) and the target FR system (i.e., F_{target}). We also assume that the adversary can use the reconstructed face images to perform two types of attacks as follows:

- *Presentation attack via digital replay (replay attack):* In this type of presentation attack, the adversary presents the reconstructed face image using a digital display in front of the camera. To perform this attack, we use a tablet (Apple iPad Pro) showing the reconstructed face image and put it in front of the camera of the target FR system.
- *Presentation attack via printed photograph:* In this type of presentation attack, the adversary prints the reconstructed face image and presents the printed photograph. To perform this attack, we print the reconstructed face images with a colorful laser printer (Develop Ineo+C364e) on typical papers and present the printed photograph in front of the camera of the target FR system.

To perform the presentation attacks (with either digital replay or printed photograph), the reconstructed image should be presented in front of the camera of the target FR system. For each of these cases, we considered three different mobile devices, including Apple iPhone 12, Xiaomi Redmi 9 A, and Samsung Galaxy S9, as the camera of the target FR system and capture images from the presentations. Fig. 8 shows our evaluation setup for capturing presentation attacks from tablet and printed photographs using different mobile cameras. It is

Fig. 9. Sample image from the MOBIO dataset, its corresponding reconstructed face images using our face reconstruction methods (i.e., GaFaR, GaFaR+GS, and GaFaR+CO) in the *whitebox* and *blackbox* scenarios, the corresponding digital replay attacks and presentation attacks using printed photographs captured with different mobile devices.

TABLE VI
VULNERABILITY EVALUATION OF THE SIMULATION (I.E., INJECTION) AND PRACTICAL *WHITEBOX* AND *BLACKBOX* TI ATTACKS USING ARCFACE TEMPLATES AGAINST DIFFERENT FR SYSTEMS AS TARGET IN TERMS OF SAR/IAPMR FOR FR SYSTEMS WITH FMR OF 10^{-3} EVALUATED ON THE MOBIO DATASET

Scenario	Attack type	Device	Reconstruction Method	F_{tagret} (SAR/IAPMR)					
				ArcFace	ElasticFace	AttentionNet	HRNet	RepVGG	Swin
whitebox	injection	N/A	GaFaR	84.29	78.10	65.24	62.86	45.24	70.95
			GaFaR+GS	86.67	78.10	67.14	61.43	49.05	71.9
			GaFaR+CO	**89.52**	**80.00**	**69.05**	**67.14**	**55.24**	**77.14**
	Replay Attack	iPhone 12	GaFaR	80.48	75.71	61.90	59.52	47.14	68.57
			GaFaR+GS	**85.71**	**79.05**	66.19	**61.43**	**50.95**	71.90
			GaFaR+CO	83.81	76.67	**68.10**	60.95	50.00	**72.86**
		Redmi 9A	GaFaR	80.00	76.67	62.86	61.43	47.14	69.52
			GaFaR+GS	**86.19**	**79.05**	67.62	**65.71**	50.00	**74.29**
			GaFaR+CO	**86.19**	78.10	**70.48**	65.24	**51.90**	75.24
		Galaxy S9	GaFaR	75.71	72.38	58.10	49.52	40.95	60.95
			GaFaR+GS	80.95	73.81	62.86	**55.24**	42.38	63.81
			GaFaR+CO	**81.90**	**75.24**	**64.76**	**55.24**	**43.33**	**64.29**
	Print Photograph	iPhone 12	GaFaR	65.24	56.19	49.52	49.05	37.62	53.33
			GaFaR+GS	82.86	71.43	**66.67**	61.43	46.67	68.10
			GaFaR+CO	**83.81**	**73.81**	64.76	**62.38**	**50.00**	**71.43**
		Redmi 9A	GaFaR	74.76	66.19	57.14	54.76	44.29	64.29
			GaFaR+GS	**85.24**	73.33	65.71	**63.33**	47.14	68.10
			GaFaR+CO	83.81	**74.76**	**67.62**	62.86	**51.90**	**69.05**
		Galaxy S9	GaFaR	71.90	64.29	58.57	54.76	42.86	64.76
			GaFaR+GS	83.33	70.48	**65.71**	60.48	48.10	68.57
			GaFaR+CO	**83.33**	**72.86**	64.76	**61.90**	**51.43**	**69.05**
blackbox	injection	N/A	GaFaR	50.95	52.38	29.05	31.90	28.10	43.33
			GaFaR+GS	**62.86**	**55.24**	**36.19**	**41.90**	**32.86**	**50.00**
	Replay Attack	iPhone 12	GaFaR	47.14	**51.43**	30.95	32.38	26.19	42.38
			GaFaR+GS	**54.76**	50.95	**38.10**	**39.05**	**32.86**	**47.14**
		Redmi 9A	GaFaR	48.10	50.48	28.57	33.33	26.67	43.81
			GaFaR+GS	**58.57**	**52.86**	**36.19**	**39.05**	**31.43**	**47.62**
		Galaxy S9	GaFaR	42.86	**47.62**	27.62	28.57	23.81	41.9
			GaFaR+GS	**50.95**	46.67	**34.29**	**36.19**	**27.14**	**42.86**
	Print Photograph	iPhone 12	GaFaR	42.86	**46.19**	30.95	32.86	25.24	43.81
			GaFaR+GS	**51.90**	**46.19**	**38.57**	**35.71**	**34.29**	**49.52**
		Redmi 9A	GaFaR	41.90	47.62	29.05	28.10	28.10	40.95
			GaFaR+GS	**54.29**	**49.05**	**38.10**	**36.67**	**33.33**	**47.14**
		Galaxy S9	GaFaR	44.76	**48.10**	28.10	30.95	**32.86**	44.76
			GaFaR+GS	**54.29**	47.14	**36.19**	**36.19**	**32.38**	**50.00**

The values are in percentage and the best values of SAR for different reconstruction methods are embolden in each attack. Cells are color-coded according to the type of attack as defined in Section 3 for attack 1 (dark green), attack 2 (light green), attack 3 (yellow), attack 4 (orange), and attack 5 (red).

noteworthy that we used the default display scale on the digital screen (i.e., iPad), in which the reconstructed face images with 512×512 resolution do not cover all the screen. However, the face area in the captured images is still larger than the required resolution to feed to be used in the target FR systems.

Fig. 9 illustrates a sample face image from the MOBIO dataset, its reconstructed face images from ArcFace templates using our different methods (GaFaR, GaFaR+GS, and GaFaR+CO) in the *whitebox* and *blackbox* (using ElasticFace as F_{proxy}) scenarios, and captured images from the reconstructed face images using different mobile devices in replay attacks and presentation attacks using printed photographs. As this figure shows, the captured images from replay attacks are more similar to the reconstructed face images, while the ones from printed photographs suffer from quality degradation. In addition, different mobile devices introduce different sensor qualities, and therefore different image qualities for the captured images in our experiment. We use the captured images[14] by each mobile device from presentation attacks as inputs to different SOTA FR systems as target FR systems, and evaluate the vulnerability of these FR systems to the presentation attack using the reconstructed face images.

Table VI reports the result of the vulnerability evaluation against SOTA FR systems to TI attacks (by injecting the reconstructed face images in our simulation), and different presentation attacks (digital replay attack and printed photograph) in the whitebox and blackbox scenarios in terms of SAR.[15] It is noteworthy that based on the presentation type, we have two

[14]The reconstructed face images and all captured images for our presentation attack evaluation are publicly available.

[15]According to the ISO/IEC 30107-3 standard [53], the adversary's success attack rate in the evaluation of presentation attack is reported in terms of the Impostor Attack Presentation Match Rate (IAPMR). However, for consistency with our experiments in Section IV-B, we use "SAR" to report the success attack rate in the evaluation of our presentation attacks using reconstructed face images too.

TABLE VII
COMPARISON OF OUR PROPOSED METHOD WITH PREVIOUS *BLACKBOX* TI METHODS IN PRACTICAL PRESENTATION ATTACKS (REPLAY ATTACKS CAPTURED BY IPHONE 12) USING ARCFACE TEMPLATES AGAINST DIFFERENT FR SYSTEM (I.E., ATTACKS 3-5) IN TERMS OF SAR/IAPMR AT FMR OF 10^{-3} ON THE MOBIO DATASET

Reconstruction Method	F_{tagret} (SAR/IAPMR)					
	ArcFace	ElasticFace	AttentionNet	HRNet	RepVGG	Swin
NBNetB-P [24]	9.05	2.38	3.81	3.81	0.95	6.19
Vendrow & Vendrow [31]	25.24	10.48	7.14	10.95	7.62	15.71
GaFaR [ours]	47.14	**51.43**	30.95	32.38	26.19	42.38
GaFaR+GS [ours]	**54.76**	50.95	**38.10**	**39.05**	**32.86**	**47.14**

The values are in percentage and the best values are embolden in attack against each FR system. Cells are color-coded according to the type of attack as defined in Section 3 for attack 3 (yellow), attack 4 (orange), and attack 5 (red).

TABLE VIII
ABLATION STUDY ON THE PROPOSED *SEMI-SUPERVISED* LEARNING APPROACH AND EVALUATION OF THE EFFECT OF LOSS TERMS IN ATTACK 1 AGAINST ARCFACE MODEL IN TERMS OF SUCCESS ATTACK RATE (SAR) ON THE MOBIO AND LFW DATASETS

approach	Loss Functions	MOBIO		LFW	
		FMR=10^{-2}	FMR=10^{-3}	FMR=10^{-2}	FMR=10^{-3}
supervised	$\mathcal{L}_{syn}^{rec} = \mathcal{L}^{w} + \mathcal{L}^{Pixel} + \mathcal{L}^{ID}$	90.96	82.38	83.80	69.467
	$\mathcal{L}_{syn}^{rec} = \mathcal{L}^{w} + \mathcal{L}^{Pixel}$	43.81	8.57	31.75	13.92
	$\mathcal{L}_{syn}^{rec} = \mathcal{L}^{w} + \mathcal{L}^{ID}$	0	0	0.86	0.30
	$\mathcal{L}_{syn}^{rec} = \mathcal{L}^{w}$	32.38	9.52	33.69	15.43
unsupervised	$\mathcal{L}_{real}^{rec} = \mathcal{L}_{Pixel} + \mathcal{L}_{ID}$ [without WGAN]	0	0	0.44	0.15
	$\mathcal{L}_{real}^{rec} = \mathcal{L}_{Pixel} + \mathcal{L}_{ID}$ [with WGAN]	70.48	31.90	67.72	45.76
	$\mathcal{L}_{real}^{rec} = \mathcal{L}_{ID}$ [with WGAN]	52.86	19.52	54.51	30.83
	$\mathcal{L}_{real}^{rec} = \mathcal{L}_{Pixel}$ [with WGAN]	0	0	2.21	0.40
semi-supervised	**Eqs. 1,2,3,6**	**95.71**	**82.86**	**89.27**	**79.84**

The SAR values are in percentage and for an attack without any camera parameter optimization (i.e., GS/CO).

types of presentation attacks (replay attack and printed photograph), and based on the adversary's knowledge of the FR system from which the template is leaked (i.e., F_{template}) and the target FR system (i.e., F_{target}), we have five different TI attacks (as described in Section III-A) and thus five different corresponding presentation attacks. The results in Table VI show that SOTA FR models as target systems are vulnerable to our attacks. In general, and as also seen in Section IV-B, attack 1 is the easiest attack, and as the adversary's knowledge becomes more limited, the attack gets more difficult in attack 2, attack 3, attack 4, and attack 5, respectively. Comparing our different reconstruction methods (i.e., GaFaR, GaFaR+GS, and GaFaR+CO), we can observe that camera parameter optimizations improve SAR values. The results also show that replay attacks achieve higher SAR values compared to presentation attacks using printed photographs. Comparing the results in Table VI for different mobile devices, the SAR values are comparable across different methods and in different attack types.

We also compare the performance of our method with two best *blackbox* methods in the literature from Table V (i.e., NBNetB-P [24] and Vebdrow and Vendrow [31]) in presentation attacks based on TI attacks 3-5 against SOTA FR models. Table VII reports this evaluation for digital replay presentation attack (captured by Apple iPhone 12) based on TI attacks using ArcFace templates against SOTA FR models in terms of adversary's SAR at the system's FMR of 10^{-3} on the MOBIO dataset. The results in this table show that our method still achieves superior performance than previous methods in the literature. Comparing this table with Table V, we can see there are in average -4.7%, 0%, -0.87%, and -2.69% changes in the SAR values in presentation attacks than the injection of reconstructed face images (Table V) for NBNetB-P [24], Vebdrow and Vendrow [31], GaFaR, GaFaR+GS, respectively.

D. Discussion

Our experiments in Section IV-B show that our proposed method outperforms previous methods in the literature in TI attacks against FR systems. To evaluate the effect of each part in our proposed method, we perform an ablation study and train different models. To this end, we evaluate the effect of *semi-supervised* learning approach in our method compared to fully *supervised* learning (i.e, using only synthetic data where we have the corresponding latent code for each template) and fully *unsupervised* learning approach (i.e., using only real data where we do not have the corresponding latent code for each template). In each of fully *supervised* learning and fully *unsupervised* learning approaches, we also evaluate the effect of each loss function. In the case of the fully *unsupervised* learning approach, we also evaluate the effect of adversarial learning in our method. Table VIII reports our ablation study on the effect of each part in our proposed method in attack 1 (injection) against ArcFace model on the MOBIO and LFW datasets in terms of SAR at system's FMR of 10^{-2} and 10^{-3}. As the results of our ablation study show, the proposed *semi-supervised* approach has a better reconstruction performance (in terms of SAR) than fully *supervised* learning and fully *unsupervised* learning approaches. Moreover, our ablation study on the effect of loss terms shows that each of the loss terms has an important impact on the performance of our face reconstruction network. In particular, using WGAN for our *unsupervised* learning (i.e., using real training data where we don't have the true value of *intermediate* latent codes for each training data) helps our mapping network M_{rec} to learn the distribution of GNeRF *intermediate* latent space $\mathcal{W}$. However, if we do not use WGAN in training with real data, our mapping network M_{rec} cannot learn the distribution of GNeRF *intermediate* latent space $\mathcal{W}$, and therefore the generated latent codes by our mapping network M_{rec} will be out of distribution $\mathcal{W}$. This will cause the generator part of GNeRF to generate non-face-like images. In addition to WGAN training, the results in Table VIII show that each of the pixel loss and ID loss terms enhances the reconstruction performance of our method in training with either synthetic (*supervised* learning) or real (*unsupervised* learning) data.

As another ablation study, we evaluate the effect of hyperparameters in the camera parameter optimization for our proposed grid search (GS) and continuous optimization (CO) approaches. For the grid search optimization approach, in our experiments in Sections IV-B and IV-C, we considered $\psi \in [-45°, +45°]$ and $\theta \in [-30°, +30°]$ for a 11×11 grid with step sizes of $\psi_{\text{step}} = 9°$ and $\theta_{\text{step}} = 6°$. Fig. 10 illustrates a sample face image from the

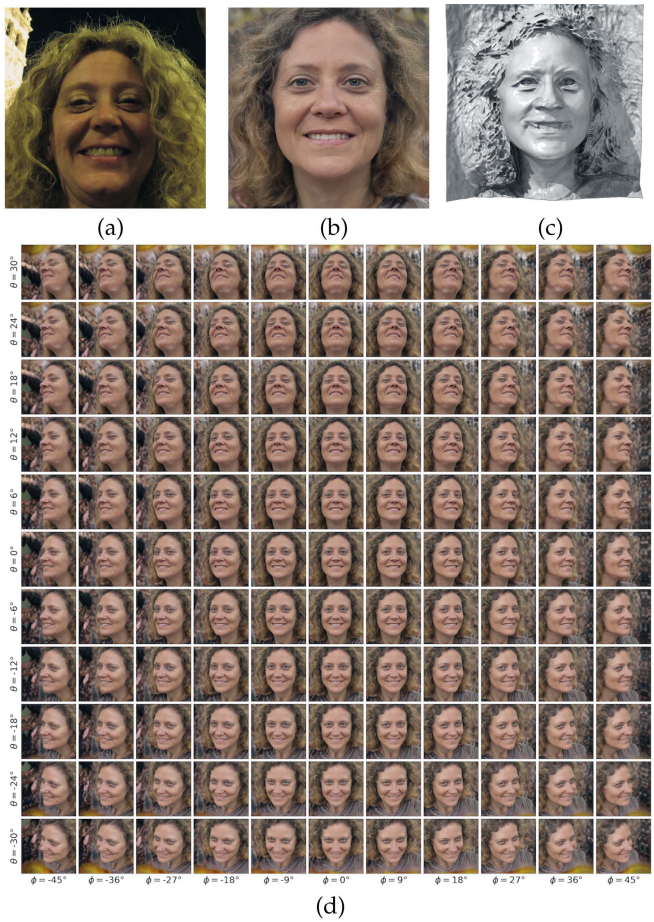

Fig. 10. (a) Sample face image from the FFHQ dataset, (b) its frontal reconstructed face image, (c) its 3D face reconstruction, and (d) the corresponding reconstructed face images with camera parameters grid using our method in the *whitebox* attack against ArcFace. The cosine similarity between templates of original (a) and frontal (b) reconstructed face images is 0.679.

FFHQ dataset and its frontal and 3D reconstruction as well as the grid of reconstruction with the size of 11×11 and camera parameters $\psi \in [-45°, +45°]$ and $\theta \in [-30°, +30°]$. For our ablation study, we use the same hyperparameters and only change one of these hyperparameters (i.e., grid size, interval of Φ, and interval of Θ) to evaluate its effect on the performance of our method in terms of SAR and average execution time. Fig. 11 reports our ablation study in the attack 1 (injection) against the ArcFace FR system configured at FMR $= 10^{-3}$ on the MOBIO dataset. The results in this figure show that the intervals of Φ and Θ are not required to be very large. Moreover, by increasing the size of our search grid (i.e., the number of steps) we can achieve a better SAR with the cost of a higher execution time. For the continuous optimization approach, in our experiments in Sections IV-B and IV-C, we considered $\psi \in [-45°, +45°]$ and $\theta \in [-30°, +30°]$ and used Adam optimizer [42] with 121 iterations and the learning rate of 10^{-2}. Similarly, for the ablation study, we use the same hyperparameters and only change one of these hyperparameters (i.e., learning rate, number of iterations, interval of Φ, and interval of Θ) to evaluate its effect on the performance of our method in terms of SAR and average execution time. Fig. 12 reports our ablation study in the attack 1 (injection) against the ArcFace FR system configured at FMR $= 10^{-3}$ on the MOBIO dataset. According to these results, similar to the ablation study for the grid search optimization, the intervals of Φ and Θ should not be necessarily very large. In addition, similar to the effect of the grid size in the grid search optimization, by increasing the number of iterations we can achieve a better SAR with the cost of a higher execution time.

According to the results in Tables IV, V, and VI, our camera parameter optimization methods improve the performance of our face reconstruction network. In particular, we observe that GaFaR+GS and GaFaR+CO also improve the SAR in attacks against different target FR systems (i.e., transferability evaluation in attacks 2, 4, and 5) too. This shows that our camera parameter optimization methods improve the attacks in the way that the reconstructed face images have more similar templates to templates of the original face images, even if extracted by a different FR model. Achieving such improvements in attacks against different target FR systems shows the transferability of our pose-optimized reconstructed face images.

We further investigate the effect of our camera parameter optimization methods on our attacks. In attack 1 against ArcFace, our grid search method increases the similarity between templates of original and reconstructed face images for 89.52% and 88.70% of cases on the MOBIO and LFW datasets, respectively. Moreover, our continuous optimization method increases the similarity between templates for 99.04% and 98.66% of reconstructed face images on the MOBIO and LFW datasets, respectively.[16] We also use the pose estimation model in [54] to find the histograms of the pose of original and reconstructed face images in attack 1 against[17] ArcFace on the MOBIO and LFW datasets. As the histograms in this figure show, most of the pose-optimized reconstructed face images have a small variation around the frontal pose. This observation is also consistent with our ablation study in Figs. 11 and 12, where we see that the intervals of Φ and Θ are not required to be very large. In addition, Fig. 13 also shows that the pose of reconstructed face images does not have the same distribution as that of the original face images. This demonstrates that our camera parameter optimization methods (CO or GS) do not try to find the same pose as the original images, but rather try to find a pose that has a template with higher similarity to the leaked template. Our transferability evaluations in Tables IV, V, and VI (i.e., attacks 2, 4, and 5) also confirm that the pose-optimized reconstructed face images also achieve better performance in attacks (either inject or even presentation attack) against different FR systems. Therefore, 3D reconstruction is essentially more useful than 2D reconstruction to generate better 2D reconstructed face images in our attacks. Fig. 14 shows sample reconstructed face images from the MOBIO dataset in *whitebox* and *blackbox* (using ElasticFace) TI attacks using our different reconstruction methods.

[16] These results can also explain the superiority of GaFaR+CO compared to GaFaR+GS in Tables IV and VI.

[17] We should note that since we use the same reconstructed face images for injection and presentation attacks, the histograms in Fig. 13 are valid for both injection and presentation attacks.

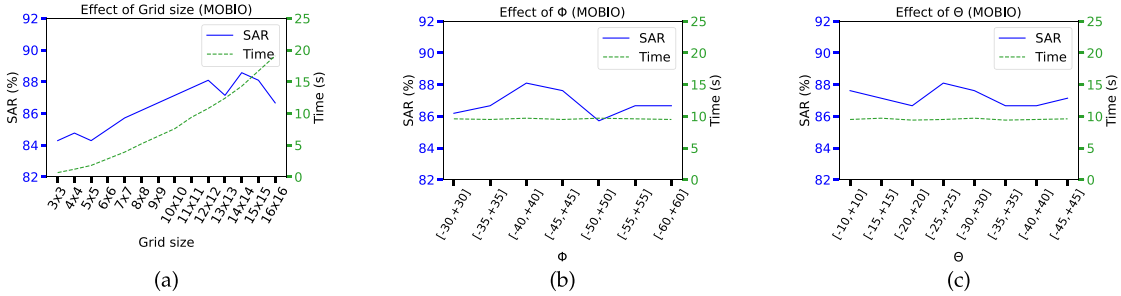

Fig. 11. Ablation study on the effect of different hyperparameters in grid search for camera parameters optimization in terms of success attack rate (SAR) and average execution time for each image reconstruction for whitebox attack (i.e., attack 1) against a FR system based on ArcFace configured at FMR=10^{-3} on the MOBIO dataset: a) grid size, b) interval of Φ, and c) interval of Θ.

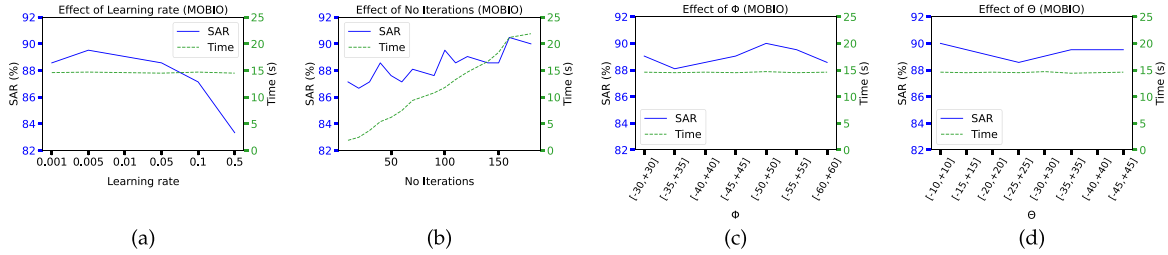

Fig. 12. Ablation study on the effect of different hyperparameters in continuous optimization for camera parameters in terms of success attack rate (SAR) and average execution time for each image reconstruction for whitebox attack (i.e., attack 1) against a FR system based on ArcFace configured at FMR = 10^{-3} on the MOBIO dataset: a) learning rate, b) number of iterations, c) interval of Φ, and d) interval of Θ.

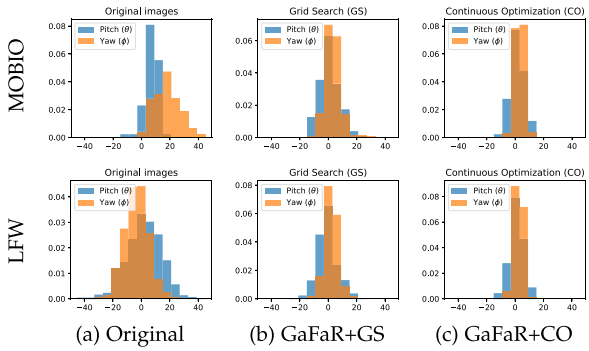

Fig. 13. Histogram of pitch and yaw in (a) original, (b) GaFaR+GS, and (c) GaFaR+CO for attack 1 against ArcFace on the MOBIO (first row) and LFW (second row) datasets. Note that for GaFaR without any camera parameter optimization, the reconstructed face images are frontal (i.e., pitch and yaw values are zero), and thus the histogram for GaFaR is not depicted in this figure.

Fig. 14. Reconstruction of sample images from the MOBIO dataset in *whitebox* and *blackbox* (using ElasticFace) TI attacks against ArcFace templates using our methods.

We can observe that our camera paramter optimization leads to different poses to increase SAR.

Comparing our result in *whitebox* (Table IV) and *blackbox* (Table V) attacks in Section IV-B, we observe that our proposed face reconstruction network, GaFaR, achieves better performance in *whitebox* attacks (attacks 1-2) than *blackbox* attacks (attacks 1-2) when inverting ArcFace templates (i.e., ArcFace as F_{template}). However, in inverting ElasticFace templates, the results show that GaFaR achieves better performance in *blackbox* attacks (attacks 3-5) than *whitebox* attacks (attacks 1-2). As a matter of fact, the difference in *whitebox* and *blackbox* attacks in our method is the FR model used as F_{proxy} for training our network. In *blackbox* attacks against ElasticFace templates, the ArcFace model is used as F_{proxy} while in *whitebox* attacks, the ElasticFace model is used as F_{proxy}. Similarly, Table III also shows that ArcFace has a superior recognition performance than ElasticFace, and thus it can more help the training of the face reconstruction network. To further investigate the effect of F_{proxy} for difference attacks, as another experiment we compare the performance of our method in *whitebox* attacks (attack 1) and *blackbox* attacks (attack 3 using ArcFace as F_{proxy}) against different FR systems on the MOBIO and LFW datasets. As the results in Table IX show, in all cases except attacks against Swin, *blackbox* attacks with ArcFace as F_{proxy} achieve superior performance than *whitebox* attacks for templates of different FR

TABLE IX
WHITEBOX (ATTACK 1) AND BLACKBOX (ATTACK 3) TI ATTACKS WITH OUR METHOD, GAFAR, AGAINST DIFFERENT TARGET FR SYSTEMS IN TERMS OF SAR AT FMR OF 10^{-3} ON THE MOBIO AND LFW DATASETS

	MOBIO					LFW				
	Elas.Face	Att.Net	HRNet	RepVGG	Swin	Elas.Face	Att.Net	HRNet	RepVGG	Swin
whitebox	78.10	64.29	71.43	53.81	**94.76**	63.06	27.00	31.87	17.33	**74.08**
blackbox	**84.76**	**72.38**	**76.67**	**72.86**	89.05	**74.77**	**33.59**	**37.80**	**25.40**	67.11

In whitebox attacks the same model, and in blackbox attacks the ArcFace model is used as F_{proxy}.

models. In contrast to other FR models in our experiments which are CNN-based, Swin is a transformer-based FR model, which can be the reason why in *blackbox* attacks with Swin templates using ArcFace (which is a CNN-based FR model) as F_{proxy} could not lead to superior performance.

In drawing our discussion to a close, our experiments in Section IV-B show the vulnerability of SOTA FR systems to TI attacks using our face reconstruction methods (GaFaR, GaFaR+GS, and GaFaR+CO). Similarly, our experiments in Section IV-C show that the reconstructed face images by our proposed methods can be used for presentation attacks against the same FR system or different FR systems that the corresponding user is enrolled (i.e., transferability of the reconstructed face images). In fact, our experiments show potential threats that can seriously jeopardize the security and privacy of users if the facial templates are leaked. In addition to the experiments in Sections IV-B and IV-C, we should note that our proposed method can generate 3D face from facial templates (as shown in Figs. 1 and 10). Such 3D reconstruction can be used for more sophisticated presentation attacks (e.g., 3D face mask, etc.) against FR systems, which require further studies in future works.

V. CONCLUSION

In this article, we presented a comprehensive vulnerability evaluation of SOTA FR systems to TI attacks using 3D face reconstruction from facial templates. We proposed a new method (called GaFaR) to reconstruct 3D faces from facial templates using a geometry-aware face generation network based on GNeRF. We learned a mapping from facial templates to the *intermediate* latent space of the GNeRF model with a *semi-supervised* learning approach using real and synthetic training data. For the real data, where we do not have correct *intermediate* latent code, we used a GAN-based training to learn the distribution of *intermediate* latent space of the GNeRF model (*unsupervised* learning). For the synthetic data, we have the corresponding *intermediate* latent code and directly learn the mapping (*supervised* learning). In addition, we proposed two optimization methods on the camera parameters in GNeRF to find a pose that improves the TI attack: grid search and continuous optimization. In the grid search method, we considered a grid for pitch and yaw rotations of the reconstructed face, and in continuous optimization, we used a gradient-based optimizer to optimize camera parameters.

We proposed our method in the *whitebox* and *blackbox* attacks against face recognition systems and comprehensively evaluated the vulnerability of SOTA FR systems to our method. Considering *whitebox* and *blackbox* blackbox scenarios and adversary's knowledge of target FR system, we defined five types of TI attacks and evaluated the *transferability* of our reconstructed face images across other FR systems on the MOBIO and LFW datasets. We evaluated the TI attacks by injecting reconstructed face images as queries to the target FR systems. In addition, we performed practical presentation attacks against SOTA FR systems using digital screen replay and printed photographs of reconstructed frontal and pose-optimized face images. Our experiments showed the vulnerability of SOTA FR models to our TI attacks and also presentation attacks using our reconstructed face images.

Last but not least, our proposed method can generate 3D faces from facial images, and we used the 3D reconstruction to find a pose that improves the adversary's success attack rate. However, 3D reconstruction of users' faces paves the way for new types of attacks (e.g., 3D face masks, etc.), which need to be investigated in the future.

ACKNOWLEDGMENTS

The authors would like to thank Karine Vaucher (Idiap Research Institute, Switzerland) for her help in conducting data collection in the presentation attack experiments.

REFERENCES

[1] J. Galbally, C. McCool, J. Fierrez, S. Marcel, and J. Ortega-Garcia, "On the vulnerability of face verification systems to hill-climbing attacks," *Pattern Recognit.*, vol. 43, no. 3, pp. 1027–1038, 2010.
[2] B. Biggio, G. Fumera, P. Russu, L. Didaci, and F. Roli, "Adversarial biometric recognition: A review on biometric system security from the adversarial machine-learning perspective," *IEEE Signal Process. Mag.*, vol. 32, no. 5, pp. 31–41, Sep. 2015.
[3] A. Hadid, N. Evans, S. Marcel, and J. Fierrez, "Biometrics systems under spoofing attack: An evaluation methodology and lessons learned," *IEEE Signal Process. Mag.*, vol. 32, no. 5, pp. 20–30, Sep. 2015.
[4] J. Galbally, S. Marcel, and J. Fierrez, "Biometric antispoofing methods: A survey in face recognition," *IEEE Access*, vol. 2, pp. 1530–1552, 2014.
[5] S. Marcel, J. Fierrez, and N. Evans, *Handbook of Biometric Anti-Spoofing: Presentation Attack Detection and Vulnerability Assessment*. Berlin, Germany: Springer, 2023.
[6] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 4401–4410.
[7] J. Deng, J. Guo, X. Niannan, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 4685–4694.
[8] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing scenes as neural radiance fields for view synthesis," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 405–421.
[9] K. Schwarz, Y. Liao, M. Niemeyer, and A. Geiger, "GRAF: Generative radiance fields for 3D-aware image synthesis," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2020, pp. 20 154–20 166.
[10] Q. Meng et al., "GNeRF: GAN-based neural radiance field without posed camera," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 6351–6361.
[11] J. Zhang et al., "3D-aware semantic-guided generative model for human synthesis," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 339–356.
[12] S. Cai, A. Obukhov, D. Dai, and L. Van Gool, "Pix2NeRF: Unsupervised conditional π-GAN for single image to neural radiance fields translation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 3981–3990.
[13] E. R. Chan, M. Monteiro, P. Kellnhofer, J. Wu, and G. Wetzstein, "pi-GAN: Periodic implicit generative adversarial networks for 3D-aware image synthesis," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 5799–5809.
[14] M. Niemeyer and A. Geiger, "GIRAFFE: Representing scenes as compositional generative neural feature fields," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 11 453–11 464.

[15] R. Or-El, X. Luo, M. Shan, E. Shechtman, J. J. Park, and I. Kemelmacher-Shlizerman, "StyleSDF: High-resolution 3D-consistent image and geometry generation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 13 503–13 513.
[16] J. Gu, L. Liu, P. Wang, and C. Theobalt, "StyleNeRF: A style-based 3D aware generator for high-resolution image synthesis," in *Proc. 10th Int. Conf. Learn. Representations*, 2022, pp. 1–25.
[17] Y. Xu, S. Peng, C. Yang, Y. Shen, and B. Zhou, "3D-aware image synthesis via learning structural and textural representations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 18 430–18 439.
[18] E. R. Chan et al., "Efficient geometry-aware 3D generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 16 123–16 133.
[19] J. Sun, X. Wang, Y. Shi, L. Wang, J. Wang, and Y. Liu, "IDE-3D: Interactive disentangled editing for high-resolution 3D-aware portrait synthesis," *ACM Trans. Graph. (ToG)*, vol. 41, no. 6, pp. 1–10, 2022.
[20] S. Galanakis, B. Gecer, A. Lattas, and S. Zafeiriou, "3DMM-RF: Convolutional radiance fields for 3D face modeling," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2023, pp. 3536–3547.
[21] D. Rebain, M. Matthews, K. M. Yi, D. Lagun, and A. Tagliasacchi, "LOLNeRF: Learn from one look," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 1558–1567.
[22] A. Zhmoginov and M. Sandler, "Inverting face embeddings with convolutional neural networks," 2016, *arXiv:1606.04189*.
[23] F. Cole, D. Belanger, D. Krishnan, A. Sarna, I. Mosseri, and W. T. Freeman, "Synthesizing normalized faces from facial identity features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 3703–3712.
[24] G. Mai, K. Cao, P. C. Yuen, and A. K. Jain, "On the reconstruction of face images from deep face templates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 5, pp. 1188–1202, May 2019.
[25] 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 *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 6132–6141.
[26] T.-D. Truong, C. N. Duong, N. Le, M. Savvides, and K. Luu, "Vec2Face-V2: Unveil human faces from their blackbox features via attention-based network in face recognition," 2022. [Online]. Available: https://arxiv.org/abs/2209.04920
[27] M. Akasaka, S. Maeda, Y. Sato, M. Nishigaki, and T. Ohki, "Model-free template reconstruction attack with feature converter," in *Proc. Int. Conf. Biometrics Special Int. Group*, 2022, pp. 1–5.
[28] X. Dong, Z. Jin, Z. Guo, and A. B. J. Teoh, "Towards generating high definition face images from deep templates," in *Proc. Int. Conf. Biometrics Special Int. Group*, 2021, pp. 1–11.
[29] M.-H. Le and N. Carlsson, "IdDecoder: A face embedding inversion tool and its privacy and security implications on facial recognition systems," in *Proc. 13th ACM Conf. Data Appl. Secur. Privacy*, 2023, pp. 15–26.
[30] M. Kansy et al., "Controllable inversion of black-box face-recognition models via diffusion," 2023, *arXiv:2303.13006*.
[31] E. Vendrow and J. Vendrow, "Realistic face reconstruction from deep embeddings," in *Proc. NeurIPS Workshop Privacy Mach. Learn.*, 2021, pp. 1–6. [Online]. Available: https://openreview.net/forum?id=-WsBmzWwPee
[32] X. Dong et al., "Reconstruct face from features using GAN generator as a distribution constraint," 2022, *arXiv:2206.04295*.
[33] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," in *Readings in Computer Vision*. Amsterdam, The Netherlands: Elsevier, 1987, pp. 671–679.
[34] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Representations*, San Diego, CA, USA, 2015, pp. 1–14.
[35] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," in *Proc. Int. Conf. Learn. Representations*, 2018.
[36] Y. Jiang, S. Chang, and Z. Wang, "TransGAN: Two pure transformers can make one strong GAN, and that can scale up," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2021, pp. 14 745–14 758.
[37] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of StyleGAN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 8110–8119.
[38] P. Dhariwal and A. Nichol, "Diffusion models beat GANs on image synthesis," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2021, pp. 8780–8794.
[39] P. J. Van Laarhoven and E. H. Aarts, "Simulated annealing," in *Simulated Annealing: Theory and Applications*. Berlin, Germany: Springer, 1987, pp. 7–15.
[40] M. Srinivas and L. M. Patnaik, "Genetic algorithms: A survey," *Computer*, vol. 27, no. 6, pp. 17–26, 1994.
[41] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 214–223.
[42] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[43] F. Boutros, N. Damer, F. Kirchbuchner, and A. Kuijper, "ElasticFace: Elastic margin loss for deep face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 1578–1587.
[44] J. Wang, Y. Liu, Y. Hu, H. Shi, and T. Mei, "FaceX-Zoo: A PyTorch toolbox for face recognition," in *Proc. 29th ACM Int. Conf. Multimedia*, 2021, pp. 3779–3782.
[45] F. Wang et al., "Residual attention network for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 3156–3164.
[46] J. Wang et al., "Deep high-resolution representation learning for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3349–3364, Oct. 2021.
[47] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding, and J. Sun, "RepVGG: Making VGG-style ConvNets great again," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 13 733–13 742.
[48] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 10012–10022.
[49] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, "MS-Celeb-1M: A dataset and benchmark for large-scale face recognition," in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2016, pp. 87–102.
[50] C. McCool, R. Wallace, M. McLaren, L. El Shafey, and S. Marcel, "Session variability modelling for face authentication," *IET Biometrics*, vol. 2, no. 3, pp. 117–129, Sep. 2013.
[51] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Univ. Massachusetts, Amherst, MA, USA, Tech. Rep. 07–49, Oct. 2007.
[52] A. Anjos, M. Günther, T. de Freitas Pereira, P. Korshunov, A. Mohammadi, and S. Marcel, "Continuously reproducing toolchains in pattern recognition and machine learning experiments," in *Proc. ICML Reproducibility Mach. Learn. Workshop*, 2017, pp. 1–8. [Online]. Available: https://openreview.net/forum?id=BJDDItGX-
[53] Information Technology – Biometric Presentation Attack Detection – Part 3: Testing and Reporting, ISO/IEC 30107–3:2017(E), International Organization for Standardization International Standard, Jun. 2017.
[54] N. Ruiz, E. Chong, and J. M. Rehg, "Fine-grained head pose estimation without keypoints," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018, pp. 2155–215509.

Hatef Otroshi Shahreza received the BSc (hons) degree in electrical engineering from the University of Kashan, Iran, in 2016, and the MSc degree in electrical engineering from the Sharif University of Technology, Iran, in 2018. He is currently working toward the PhD degree with the École Polytechnique Fédérale de Lausanne (EPFL) and is a research assistant with the Biometrics Security and Privacy Group, Idiap Research Institute, Switzerland, where he received H2020 Marie Skłodowska-Curie Fellowship (TReSPAsS-ETN) for his doctoral program. During his PhD, he also experienced six months as a visiting scholar with the Biometrics and Internet Security Research Group, Hochschule Darmstadt, Germany. He is also the winner of the European Association for Biometrics (EAB) Research Award 2023. His research interests include deep learning, computer vision, biometrics, and biometric template protection.

Sébastien Marcel received the PhD degree in signal processing from the Université de Rennes I in France, in 2000 at CNET, Research Center of France Telecom (now Orange Labs). He heads the Biometrics Security and Privacy Group, Idiap Research Institute, Switzerland and conducts research on face recognition, speaker recognition, vein recognition, attack detection (presentation attacks, morphing attacks, and deepfakes), and template protection. He is professor with the University of Lausanne (School of Criminal Justice), and a lecturer with 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.