SlerpFace: Face Template Protection via Spherical Linear Interpolation

Zhizhou Zhong[1*], Yuxi Mi[1*], Yuge Huang[2†], Jianqing Xu[2], Guodong Mu[2], Shouhong Ding[2], Jingyun Zhang[3], Rizen Guo[3], Yunsheng Wu[2], Shuigeng Zhou[1†]

[1] Fudan University
[2] Youtu Lab, Tencent
[3] WeChat Pay Lab33, Tencent
zzzhong22@m.fudan.edu.cn, {yxmi20, sgzhou}@fudan.edu.cn
{yugehuang, joejqxu, gordonmu, ericshding, simonwu}@tencent.com
{naskyzhang, rizenguo}@tencent.com

Abstract

Contemporary face recognition systems use feature templates extracted from face images to identify persons. To enhance privacy, face template protection techniques are widely employed to conceal sensitive identity and appearance information stored in the template. This paper identifies an emerging privacy attack form utilizing diffusion models that could nullify prior protection. The attack can synthesize high-quality, identity-preserving face images from templates, revealing persons' appearance. Based on studies of the diffusion model's generative capability, this paper proposes a defense by rotating templates to a noise-like distribution. This is achieved efficiently by spherically and linearly interpolating templates on their located hypersphere. This paper further proposes to group-wisely divide and drop out templates' feature dimensions, to enhance the irreversibility of rotated templates. The proposed techniques are concretized as a novel face template protection technique, SlerpFace. Extensive experiments show that SlerpFace provides satisfactory recognition accuracy and comprehensive protection against inversion and other attack forms, superior to prior arts.

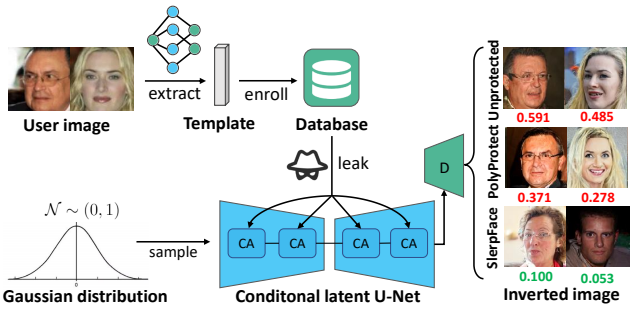


Figure 1: Paradigm of DM inversion attacks. It receives templates as conditional contexts and synthesizes identity-descriptive images. While unprotected templates and prior FTP arts are experimentally found vulnerable to inversion attacks, this paper presents a novel SlerpFace method as an effective defense. It deteriorates DM to let it generate images with obfuscated facial semantics and lower similarity scores, hence preserving privacy.

1 Introduction

Face recognition (FR) is a biometric way to identify persons by facial appearance. Contemporarily, face recognition is enabled by comparing identity-discriminative feature vectors, or *face templates*, extracted from face images via deep neural networks (DNN).

Face templates are commonly considered sensitive data, as they carry identity and appearance information inferable of a specific person. To meet growing regulatory demands, face template protection (FTP) methods are proposed to conceal original templates, and securely represent them with an irreversible and revocable reference form (ISO/IEC 2022), known as *protective templates*. These methods can be broadly divided into three categories: Crypto-based methods (Boddeti 2018; Jindal et al. 2020; Engelsma, Jain, and Boddeti 2022) process templates with encryption or security protocols in high latency and computation costs. Hash-based methods (Mohan et al. 2019; Dang et al. 2020; Kim et al. 2021; Rathgeb et al. 2022; Dusmanu et al. 2021) turn templates into randomized codewords. Yet, they are less tolerant of minor facial attribute variations and hence could downgrade recognition. Recently, transform-based methods (Phillips et al. 2019; Abdellatef et al. 2020; Hahn and Marcel 2022; Shahreza, Hahn, and Marcel 2023) have gained increasing attention as they are appealing in both accuracy and cost. They obtain protective templates via carefully designed transformations that obfuscate their binding with the original ones.

Transform-based methods, however, could bear two privacy bottlenecks. First, they usually must pre-negotiate some secure parameters or *secrets* for the transformation, which once exposed would compromise privacy. Second, they could also be susceptible to privacy attacks, where reconstruction (Mai et al. 2018; Shahreza, Hahn, and Marcel 2022; Shahreza and Marcel 2023a) models using generative adversarial networks (GAN) or autoencoders (AE) and score-based techniques (Razzhigaev et al. 2021; Dong et al. 2023; Lai et al. 2021; Wang et al. 2021) may manage to

[*] Authors contributed equally to this paper.
[†] Corresponding authors.

recover partial facial appearance from protective templates, rendering protection less effective.

This paper further investigates an emerging type of attack, based on recent grave advances in diffusion models (DM) (Ho, Jain, and Abbeel 2020). DM synthesizes high-quality images from randomly sampled noise through a learned denoising process, where the image's content can be designated by an optional *context* condition (Lu et al. 2024; Ren et al. 2024). This paper identifies a new privacy threat from DM's capability, referred to as *inversion attacks*: Taking templates as the context, DM may invert face images that preserve the templates' identity, hence revealing the persons' face, as depicted in Fig. 1. Such DMs have been made possible by recent image synthesis arts (Kansy et al. 2023; Boutros et al. 2023). This paper identifies inversion attacks as *more dreadful than previous attack forms*, warranting special attention in future research: For the first time, it enables recovering both *high-quality* and *identity-preserving* face images from templates, imposing exacerbated threats. Prior transform-based arts are also proven vulnerable to inversion attacks, later demonstrated in Sec. 4.3.

To address privacy issues, this paper proposes a novel transform-based FTP method, SlerpFace. It effectively improves prior arts' inadequate protection against secret exposure and different attacks, thus improving privacy.

Considering inversion attacks as the primary threat, this paper finds that DM's performance can deteriorate by altering the context's distribution. Specifically, when replacing authentic templates with randomly sampled Gaussian noise as context, DM is obfuscated from producing face images with a consistent identity, which is desirable for privacy. Drawing insights, this paper proposes to transform the original templates toward being alike sample-wise noises while maintaining discriminative identities. This is efficiently achieved by spherically and linearly interpolating (slerp) templates on their located hypersphere. The noises serve as secrets for transformation. This paper further addresses the noises' exposure by grouping and randomly dropping out protective templates' feature dimensions, where the division of groups is learnable to optimize recognizability. To the authors' knowledge, SlerpFace is the first FTP method to study resistance to inversion attacks. Extensive experiments suggest that it effectively prevents inversion and other attack forms as the protective templates are securely obfuscated. It also outperforms prior arts in accuracy and cost.

This paper's contributions are three-fold: (1) It identifies the inversion attack as a severe privacy threat to transform-based template protection and analyzes the attack model's generative capability. (2) It suggests spherical linear interpolation as an effective defense, by rotating templates towards sample-wise noises. It further proposes feature grouping and dropout to enhance templates' privacy under secret exposure, and learnable feature grouping to improve recognizability. (3) It presents a novel FTP method, SlerpFace. Experiments demonstrate its superior privacy protection, better accuracy, and lower cost than prior arts.

2 Related Work

2.1 Face Recognition

Modern FR systems recognize persons by comparing their face templates. Templates are feature vectors extracted from face images using DNNs, where angular margin losses (Deng et al. 2019; Huang et al. 2020; Kim, Jain, and Liu 2022) are most employed during training to earn templates with identity discrepancy that facilitates recognition. They produce normalized templates that can be considered as unit vectors onto a hypersphere. During inference, cosine similarities are calculated to find the closest match.

2.2 Face Template Protection

This paper divides FTP methods into three branches: Crypto-based methods (Boddeti 2018; Jindal et al. 2020; Engelsma, Jain, and Boddeti 2022) use homomorphic encryption to turn templates into ciphertexts and perform calculations thereon. Their shortages involve high computation costs and reliance on the secrecy of encryption keys.

Hash-based methods use one-way schemes such as fuzzy commitment (Mohan et al. 2019; Gilkalaye, Rattani, and Derakhshani 2019), fuzzy vault (Dong et al. 2021; Rathgeb et al. 2022), locality-sensitive hashing (Dang et al. 2020), discretization (Xu et al. 2020), and trainable models (Chen et al. 2019) to map templates into irreversible protective codewords or hash values. Unfortunately, they often fail to achieve satisfactory recognition accuracy, as their means are less tolerant of the intra-identity variability inherent in facial attributes, resulting in false negatives. Recently, IronMask (Kim et al. 2021) and ASE (Dusmanu et al. 2021) achieve protection with improved accuracy, by rotating templates to randomly chosen codewords and affining them into random subspace, respectively.

Transform-based methods apply task-specific transformations like feature reduction (Pillai et al. 2011; Hahn and Marcel 2022), mixing (Phillips et al. 2019; Abdellatef et al. 2020) and rotation (Shahreza, Hahn, and Marcel 2023) to convert templates into protective forms, obscuring their bindings with the original ones. They differ from hash-based methods in must keep confidential some pre-negotiated secrets that designate the transformation. They could be susceptible to privacy attacks, as later experiments reveal. This paper proposes a novel transform-based method, SlerpFace, that can address the above inadequacies.

A research direction parallels to us is image protection (Mi et al. 2023, 2024, 2022; Zhang et al. 2024b,a; Liu et al. 2024; Yuan et al. 2022, 2024). They focus on the protection of the image transmitted to service providers, whereas SlerpFace is dedicated to ensuring the security of the stored face feature.

2.3 Privacy Attacks

This paper studies attacks that attempt to recover facial appearances from templates and compromise privacy. They can be divided into three folds by attack means.

Reconstruction attacks query the FR system with attacker-owned face images to obtain corresponding templates. They then use generative adversarial networks (Truong et al.

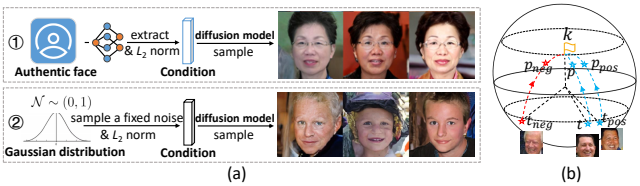


Figure 2: (a) DM's generative capability: Consistent identity faces from authentic templates, but deteriorates with noise templates, causing semantic variation. (b) Slerp rotation at d=3: Query, positive, and negative templates rotate in the same direction, maintaining margin differences.

2022; Shahreza and Marcel 2023a,b) or autoencoders (Cole et al. 2017; Mai et al. 2018) trained on image-template pairs to learn an inverse fit that generates synthetic images.

Score-based attacks (Razzhigaev et al. 2021; Vendrow and Vendrow 2021; Dong et al. 2023; Lai et al. 2021; Wang et al. 2021; Shahreza and Marcel 2023b) instigate through the knowledge of similarity scores. It recursively optimizes model-generated images by querying and maximizing their FR similarity scores with images in the database.

This paper further introduces inversion attacks, where diffusion models (Boutros et al. 2023; Kansy et al. 2023) receive templates as their contexts to generate images preserving the same identity, thus revealing persons' faces. Inversion attacks impose more dreadful threats as they can generate high-quality, identity-preserving images.

3 Methodology

3.1 Motivation

In practical applications, FR typically occurs between a server and its clients, with the server acting as an FR service provider that pre-trains the recognition model and enrolls identity templates to create a database. The database and model are shared with local devices or clients, who use them to recognize query faces locally. To protect templates' sensitive information, FTP's goal is to design a transformation that turns database templates $\vec{t}$ into protective forms $\vec{p}$ via secret parameters $\vec{k}$, making them safer to share.

We consider inversion attacks as the primary threat to shared templates. A DM is a generative model concretized as $g : (\vec{\epsilon}, \vec{t}) \rightarrow X$. Taking random Gaussian noise $\vec{\epsilon}$ as input and template $\vec{t}$ as context, it synthesizes a face image X descriptive of $\vec{t}$'s identity, thus nullifying privacy. Testifying the attack's generative capability, we first train a DM using a pipeline from IDiff-Face (Boutros et al. 2023). Then, we infer it multiple times with a *fixed* template extracted from an authentic face image via a pre-trained FR model. The details of DM and its training are aligned with IDiff-Face. As shown in Fig. 2(a1), the inverted images exhibit consistent facial appearances, *i.e.*,, the same elderly woman wearing glasses. This suggests that DM can generate identity-preserving face images of high quality.

Prior image synthesis arts (Lugmayr et al. 2022; Boutros et al. 2023) also use randomly sampled contexts as a practice to let DM generate unseen concepts, *e.g.*, new identities. In Fig. 2(a2), we further choose a *fixed* noise template that each of its feature dimensions is randomly sampled from Gaussian distribution. Taking it as DM's context, we find the inversion deteriorates: Though high-quality face images are still being generated, they no longer preserve a "hypothesized" identity but vary in semantics such as gender and age. We attribute the downgrade to the distributional discrepancy between noise and authentic templates. Studies (Li et al. 2021; Shen et al. 2020; Chen et al. 2022) prove that templates follow *a priori* distributions that help preserve semantics. Randomly drawn noise most likely falls in a different distribution that is close to semantics' decision bounds, rendering uncertainties in images' facial appearance.

We leverage the observation as a means to prevent inversion attacks. Given that randomly sampled noise templates deteriorate the attack, an intuitive idea to protect templates is to "move" them toward the noise distribution, hence diminishing their discriminative semantics.

3.2 Rotation via Spherical Linear Interpolation

Let d be the feature dimension of templates. Recall in Sec. 2.1 that templates can be considered as unit vectors located on a d-dimensional hypersphere for most modern FR. To move template $\vec{t}$ toward noise distribution thus equals rotating $\vec{t}$ on the hypersphere to the direction of a noise template $\vec{k}$. We refer to $\vec{k}$ as *key template* as it represents a secret that designates the rotation. Prior methods implement rotation of $\vec{t}$ by multiplying it with an orthogonal matrix $\vec{M}^{d\times d}$. However, to derive a random $\vec{M}$ is rather time-consuming, especially for large d (Schreiber and Van Loan 1989; Chen et al. 2023). Instead, we adopt an efficient way in light of a computer graphic study (Shoemake 1985). It suggests that a 3D object can be smoothly rotated by interpolating its coordinates on a sphere. We refer to the technique as spherical linear interpolation, or *slerp*. We generalize slerp to d-dimension and rotate $\vec{t}$ as:

$$\vec{p} = \frac{\sin((1-\alpha)\theta)}{\sin\theta}\vec{t} + \frac{\sin(\alpha\theta)}{\sin\theta}\vec{k}, \tag{1}$$

where $\vec{p}$ denotes the rotated protective template and $\vec{k}$ is *sample-wisely* chosen for each $\vec{t}$. $\theta = \arccos\left(\frac{\vec{t}^{\top}\vec{k}}{\|\vec{t}\|\|\vec{k}\|}\right)$ denotes the included angle between $\vec{t}$ and $\vec{k}$, and α is a hyperparameter that controls the degree of rotation.

Based on previous discussions, we expect rotating $\vec{t}$ toward $\vec{k}$ to deteriorate inversion attacks by obfuscating DMs from generating face images aligned with $\vec{t}$'s identity. Section 4.3 later testified to its effectiveness of protection.

Rotation also maintains templates' recognizability. Figure 2(b) exemplifies the effect of slerp at d=3. Let $\vec{t}, \vec{t}_{pos}, \vec{t}_{neg}$ denote a query template and two templates with the same or different identities in the database, respectively. Initially, the angular margin between $\vec{t}, \vec{t}_{pos}$ is smaller than that between $\vec{t}, \vec{t}_{neg}$ as FR encourages templates to have small intra-class and large inter-class margins. Slerp rotates them toward key $\vec{k}$ by the same degree to obtain corresponding protective $\vec{p}, \vec{p}_{pos}, \vec{p}_{neg}$. We highlight that the relative

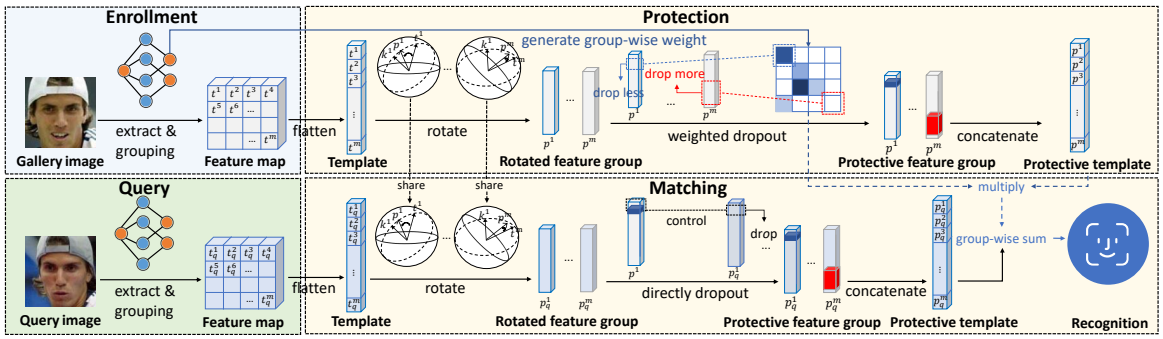


Figure 3: Pipeline of SlerpFace: (1) Train an FR model to extract and group feature maps as templates. (2) Protect templates by independently rotating feature groups toward a key template and applying random dropout based on learnable weights. (3) During inference, extract a query template. (4) Match it with enrolled templates using the same rotation and dropout.

difference among their margins is maintained as $\vec{p}_{pos}$ remains a closer match of $\vec{p}$. Hence, $\vec{p}$ will not be falsely recognized. To apply slerp in practice, denote $\vec{P} = \{\vec{p}_1, \ldots, \vec{p}_n\}$ as a set of n database templates, protected by the server using corresponding keys $\vec{K} = \{\vec{k}_1, \ldots, \vec{k}_n\}$. The server shares $\{\vec{P}, \vec{K}\}$ with its clients. Given a query template $\vec{t}_q$, a client try rotating it with each $\vec{k}_{i\in[n]}$ to obtain $\vec{p}_{q_i}$ and match with $\vec{p}_i$. The relative marginal difference before and after rotation is maintained, hence the client can compare similarities to find the closest match.

3.3 Feature Grouping and Dropout

Rotating templates effectively gains resiliency against inversion attacks. Its protection is yet not intact, as we find the original $\vec{t}$ can still be recovered by any malicious client knowing of $\{\vec{p}, \vec{k}\}$. Specifically, let $\{t_i, p_i, k_i\}_{i\in[d]}$ be the respective feature dimensions of $\{\vec{t}, \vec{p}, \vec{k}\}$. Equation 1 can be rewritten as a full-rank linear system of d equations with d unknowns $\{t_1, \ldots, t_d\}$:

$$p_i = \frac{\sin((1-\alpha)\theta)}{\sin\theta} t_i + \frac{\sin(\alpha\theta)}{\sin\theta} k_i, \quad i \in [d]. \tag{2}$$

The client can thus employ numerical calculation techniques, *e.g.*, the Newton-Raphson method (Lindstrom and Bates 1988), to approximately solve Eq. 2 and obtain an estimated $\tilde{\vec{t}} \approx \vec{t}$. This will break the irreversibility of FTP and nullify its protection.

This section introduces a two-fold solution to enhance privacy. First, we intuitively observe that reducing the effective number of equations in Eq. 2 will make it an underdetermined system, leading to imprecise approximations of $\tilde{\vec{t}}$. We achieve the reduction by *feature dropout*, *i.e.*,, randomly resetting a specific ratio β of $\vec{p}$'s feature dimensions to 0.

To evaluate the effectiveness of feature dropout, fixing a $\vec{k}$, $\tilde{\vec{t}}$'s precision can be quantified as its angle $\tilde{\theta}$ between $\vec{k}$, compared to θ between $\{\vec{t}, \vec{k}\}$, *i.e.*,, $\Delta_\theta = |\tilde{\theta} - \theta|$. A larger Δ_θ indicates $\tilde{\vec{t}}$ is further away from the authentic $\vec{t}$, which benefits privacy. Setting β=0.5, we experimentally draw $\tilde{\vec{t}}$ 10000 times for templates with different dimensions d, and depict their range of Δ_θ in Fig. 4(a). We find that feature dropout is less satisfactory for a large d (*e.g.*, 512), as its $\Delta_\theta < 5°$. On the other hand, $\Delta_\theta < 28°$ is salient when choosing a smaller d (*e.g.*, 16), suggesting that the client's estimation of $\tilde{\vec{t}}$ is more uncertain.

Though a small d would favor privacy, we cannot simply reduce $\vec{t}$'s dimension since such feature compression compromises recognizability heavily. Instead, we propose *feature grouping* to leverage the findings, by dividing $\vec{t}$ into smaller groups and protecting each group separately. Specifically, let $\{\vec{t}^1, \ldots, \vec{t}^m\}$ be m equal-dimension groups divided from a d-dimension template $\vec{t}$ and $\{\vec{k}^1, \ldots, \vec{k}^m\}$ be the division of its key $\vec{k}$. We normalize each group to a (d/m)-dimension hypersphere, and rotate it via slerp:

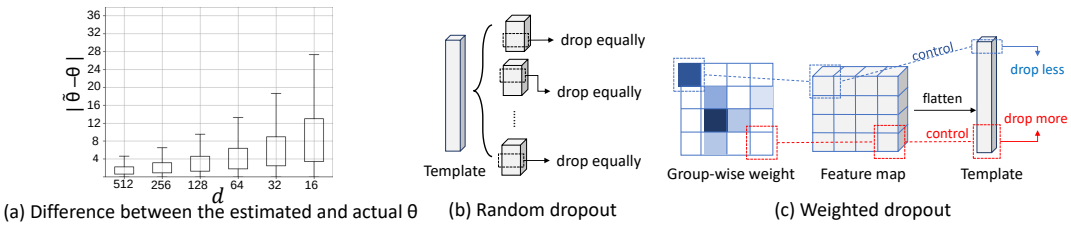
$$\vec{p}^i = \frac{\sin((1-\alpha)\theta^i)}{\sin\theta^i} \vec{t}^i + \frac{\sin(\alpha\theta^i)}{\sin\theta^i} \vec{k}^i, \quad i \in [m], \tag{3}$$

under angles $\vec{\theta} = \{\theta^i\}_{i\in[m]}$. We perform feature dropout on each of $\{\vec{p}^1, \ldots, \vec{p}^m\}$, as shown in Fig. 4(b), and concatenate them to form the protective $\vec{p}$. During inference, the similarity score is derived as the *group-wise sum* similarity between a query $\vec{t}_q$ and database templates.

3.4 Learnable Feature Grouping

We experimentally find feature grouping is at the cost of salient FR performance downgrade, later in Sec. 4.4. To address this issue, we propose *learnable feature grouping* inspired by a model interpretability study (Lin et al. 2021). It suggests that facial recognizability can be regarded as the collective effort of different feature groups. Drawing insight, we aim to incorporate feature dimensions into learnable groups that achieve the same recognizability as the entire face together. Figure 5 describes our approach.

Specifically, let $\{X_a, X_b\}$ be a pair of images with either the same or different identities. We train an FR model to extract their templates $\{\vec{t}_a, \vec{t}_b\}$ and denote their similarity as $\mathrm{sim}(\vec{t}_a, \vec{t}_b)$. To facilitate the model in producing appropriate groupings of features, we branch out before its final fully connection layer with a 1×1 convolution layer



(a) Difference between the estimated and actual θ (b) Random dropout (c) Weighted dropout

Figure 4: (a) Ranges of Δ_θ for templates with different feature dimensions d. Δ_θ gradually increases as d decreases. (b) Random dropout. Each feature group randomly discards equal dimensions of features. (c) Weighted dropout. Feature groups with larger weights discard fewer feature dimensions to better preserve crucial features.

to simultaneously obtain a pair of learnable feature maps $\{\vec{t}_{a'}, \vec{t}_{b'}\}$, each with the shape of (c, h, w). We regard each c-dimension spatial feature slice as a feature group, with a total of $m = hw$ groups. We calculate the similarity between each pair of feature groups $\{\vec{t}^{\,i}_{a'}, \vec{t}^{\,i}_{b'}\}$, denoted as $\mathrm{sim}(\vec{t}^{\,i}_{a'}, \vec{t}^{\,i}_{b'})_{i\in[m]}$. We then calculate their weighted sum by group-wise weights $\vec{w} = \{w^1, \ldots, w^m\}$, which are derived from the self-attention of feature maps. To encourage feature groups $\{\vec{t}^{\,i}_{a'}, \vec{t}^{\,i}_{b'}\}_{i\in[m]}$ together to be as recognizable as templates $\{\vec{t}_a, \vec{t}_b\}$, let n be the batch size, we establish loss:

$$\mathcal{L}_g = \frac{1}{n}\sum_{i=1}^{n}\left\|\mathrm{sim}(\vec{t}_{a_i}, \vec{t}_{b_i}) - \sum_{j=1}^{m} w^j\,\mathrm{sim}(\vec{t}^{\,j}_{a'_i}, \vec{t}^{\,j}_{b'_i})\right\|_1, \quad (4)$$

to bridge and align the weighted similarity sum to the template pair's similarity. The general FR loss (*e.g.*, ArcFace) is $\mathcal{L} = \mathcal{L}_{fr} + \gamma\mathcal{L}_g$, where γ is a hyper-parameter.

During inference, given a query image, we extract its learned feature map $\vec{t}'$, divide it into feature groups, and then concatenate all groups to form a flattened template, still denoted as $\vec{t}'$. $\vec{t}'$ is protected by group-wise rotation, as discussed in Sec. 3.3. We further advocate an alternative *weighted dropout*: As weights $\vec{w}$ reflect each group's contribution in calculating similarity, we propose to drop less or more feature dimensions for groups of $\vec{t}'$ with higher or lower weight, respectively, as shown in Fig. 4(c). It helps preserve more crucial features compared to random dropout.

4 Experiments

4.1 Experimental Setup

We employ an IR-50 model, trained on the MS1Mv2 (Guo et al. 2016) dataset on 8 GPUs in parallel with ArcFace loss as $\mathcal{L}_{fr}$, as the FR backbone. We train the model for 24 epochs using a stochastic gradient descent (SGD) optimizer, choosing the total batch size, initial learning rate, momentum, and weight decay as $256, 0.01, 0.9, 0.0005$, respectively. We set parameters $(\alpha, \beta, \gamma, c, m)$ as $(0.9, 0.5, 1, 16, 49)$. Evaluation is done on 5 regular-size datasets, LFW (Learned-Miller 2014), CFP-FP (Sengupta et al. 2016), AgeDB (Moschoglou et al. 2017), CPLFW (Zheng and Deng 2018), and CALFW (Zheng, Deng, and Hu 2017), and 2 large-scale datasets, IJB-B (Whitelam et al. 2017) and IJB-C (Maze et al. 2018).

4.2 Recognition Performance

Compared methods. We compare SlerpFace with an unprotected baseline and five FTP methods: **ArcFace** as the baseline; **Boddeti** (Boddeti 2018) using Fully Homomorphic Encryption; **IronMask** (Kim et al. 2021) and **ASE** (Dusmanu et al. 2021), both hash-based—IronMask employs orthogonal matrices for template rotation into random codewords, while ASE maps templates into random affine subspaces; and transform-based **MLP-Hash** (Shahreza, Hahn, and Marcel 2023) and **PolyProtect** (Hahn and Marcel 2022), with MLP-Hash rotating templates via pre-negotiated orthogonal matrices and PolyProtect translating templates into polynomials with specific exponents and coefficients.

Recognizability and time cost. We perform face recognition by verifying if two templates refer to the same person. We report *accuracy* for LFW, CFP-FP, AgeDB, CPLFW, and CALFW, and *TPR@FPR(1e-4)* for IJB-B and IJB-C. Results are summarized in Tab. 1. Here, results for IronMask and MLP-Hash on IJB-B/C are marked as "N/A", as IronMask's approach to precisely match two codewords is viable only for accuracy results, and recognition for MLP-Hash on IJB-B/C takes prohibitive time.

Among the compared methods, FHE offers Bodddti the highest recognizability but is vulnerable if its key is compromised. Methods below the horizontal line can still provide protection when the key and template leakage. IronMask and ASE demonstrate subpar performance due to their hash-based methods' low tolerance for intra-identity facial attribute variations, impacting accuracy. SlerpFace surpasses previous models on most datasets, enhancing recognizability; however, it slightly trails PolyProtect on CFP-FP and CPLFW, which focus on facial pose variations. The accuracy gap is attributed to a reduction in descriptive capability from feature grouping, despite the groups being learnable. Nonetheless, note that the gap is marginal and serves as an efficient trade-off, as SlerpFace significantly surpasses PolyProtect in time cost and privacy.

The last two columns in Tab. 1 show the average enrollment (to register into a database) and matching (to match once with the database) time (ms) for a single template on a personal laptop, highlighting SlerpFace's advantage.

4.3 Protection Against Privacy Attacks

Prior transform-based arts can be susceptible to privacy attacks, where attack models may reveal facial appearances

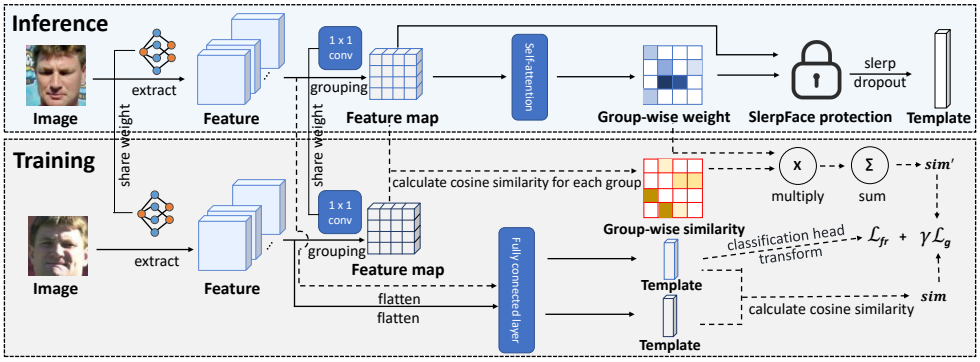


Figure 5: Paradigm of learnable feature grouping: Modify the FR model's output layer to generate feature maps, which are split into groups. Self-attention layers provide group-wise weights. For face image pairs, $\mathcal{L}_g$ aligns group-wise weighted similarities with original template similarities, trained alongside $\mathcal{L}_{fr}$. During inference, feature maps are reorganized as templates.

from templates. We compare SlerpFace with transform-based MLP-Hash and PolyProtect using 3 SOTA attacks.

Attack setup. We set up one representative attack for each of the 3 attack forms discussed in Sec. 2.3. (1) For inversion attacks, we designate IDiff-Face (Boutros et al. 2023). It is a dataset synthesis work that uses latent DMs conditioned by templates to generate high-quality, identity-preserving images. We turn it into attacks using its DM's generative capability. (2) For score-based attacks, we consider EG3D (Chan et al. 2022), which surpasses the classic score-based attack in recovery quality. It combines score-based techniques with reconstruction, optimizing the similarity scores of GAN-generated images. (3) For reconstruction attacks, we refer to NbNet (Mai et al. 2018). It deconvolutes templates into face image form via a CNN structure.

Using the LFW dataset as the source of ground truth images, we extract templates with a pre-trained FR model and secure them using MLP-Hash, PolyProtect, and SlerpFace. To conduct attacks, we train IDiff-Face, EG3D, and NbNet models from scratch for the unprotected baseline and three protection strategies, creating a total of 12 models.

Attack evaluation. Qualitatively, Fig. 6(a) exemplifies 2 sample ground truth images and the recovery from their unprotected and 3 method-corresponding protective templates. In corners, we mark the cosine similarity score between the templates of ground truth and recovered images. Dissimilar images and lower scores indicate better protection. Here, we highlight: (1) Inversion attacks impose more severe threats than other attack forms, as they generate both high-quality and identity-preserving images for unprotected baseline. (2) Prior arts fail to resist inversion as their recovered images display both high visual similarity and scores. SlerpFace provides better protection as its recoveries have low scores and show changes in semantics (*e.g.*, gender). (3) Though score-based and reconstruction attacks themselves exhibit lower capacities than inversion attacks, we can observe that SlerpFace outperforms prior arts with far lower scores. Notably, score-based attacks recover similar images for both samples protected by SlerpFace. We suspect this occurs as the protective templates do not provide effective semantics that influence the recovery, implying enhanced privacy.

Quantitatively, Tab. 2 compares the dataset-wise similarity scores among different settings. We further quantify recovered images' quality by similarity (Sim) and the success rate of replay attacks (SRRA) (Shahreza and Marcel 2023a). It replays templates of recovered images as queries to determine if they can be recognized under a FAR threshold of 1e-3. Lower Sim and SRRA suggest better protection.

4.4 Ablation Study

Component's contribution. We proposed learnable feature grouping and weighted dropout to improve the recognizability of protective templates. In Tab. 3, "w/o LW" and "w/o W" denote SlerpFace without both learnable feature grouping and weighted dropout (equals Sec. 3.3 setting), and without weighted dropout alone, respectively. We observe that SlerpFace initially suffers salient accuracy downgrades due to feature grouping. Its accuracy is promoted for 5% by learnable feature grouping and 2% further by weighted dropout.

4.5 Security Analysis

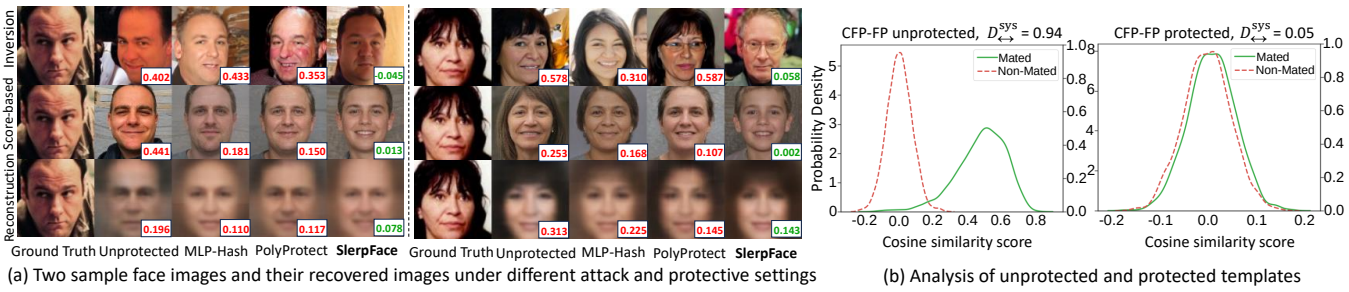
This section further discusses that SlerpFace satisfies important identity protection criteria (ISO/IEC 2022), known as *irreversibility*, *revocability*, and *unlinkability*.

Irreversibility. It requires that revealing the original templates $\vec{t}$ from the protective $\vec{p}$ should be computationally infeasible. A client with knowledge of $\vec{p}$ and key $\vec{k}$ might approximate $\vec{\tilde{t}} \approx \vec{t}$ using numerical methods like Newton-Raphson (NR) (Lindstrom and Bates 1988), which approximately solve Eq. 2. We find that feature dropout hinders approximation by reducing the equations in Eq. 2. Then, we design an NR-based attack to show SlerpFace's irreversibility and the effectiveness of feature dropout. We consider two protective settings with (d, m)=$(16, 49)$. NR can be considered as a root-finding algorithm that produces successively better estimates $\vec{\tilde{t}}$ for $\vec{t}$. It begins with a random initial guess, succeeding with a close estimate or failing at maximum iterations, after which NR reinitializes and retries.

Taking the templates from LFW, CFP-FP, and AgeDB, we measure the cost by NR's average count of reruns r. To obtain results within finite time, we measure r for each $d = 16$

Method	LFW	CFP-FP	AgeDB	CALFW	CPLFW	IJB-B	IJB-C	Enrollment	Matching
ArcFace	99.73	98.00	97.87	95.92	92.50	93.93	95.52	-	-
Boddeti	99.73	97.86	97.81	95.84	92.41	93.88	95.47	1.99	22.4
IronMask	84.42	52.70	53.22	50.00	50.00	N/A	N/A	832.22	0.25
ASE	98.77	86.80	88.48	84.12	83.42	0.00	0.00	0.79	0.19
MLP-Hash	98.82	91.40	93.67	93.07	87.12	N/A	N/A	134.31	131.76
PolyProtect	99.30	**94.00**	95.28	94.77	**89.22**	87.37	89.96	1.86	1.88
SlerpFace	**99.42**	92.79	**95.70**	**94.82**	88.90	**89.96**	**92.25**	**0.35**	**0.17**

Table 1: Comparison of recognition accuracy and time cost among SlerpFace, baseline, and SOTAs.



(a) Two sample face images and their recovered images under different attack and protective settings

(b) Analysis of unprotected and protected templates

Figure 6: (a) Two sample face images and their recovered images under different attack and protective settings. Values in the corner mark the similarity score. Dissimilar images and lower scores indicate better protection. (b) SSWL score between mated and non-mated template pairs for unprotected baseline and SlerpFace.

	Inversion		Score-based		Recon	
	Sim	SRRA	Sim	SRRA	Sim	SRRA
Unprotected	0.46	98.83%	0.16	17.15%	0.12	6.50%
MLP-Hash	0.19	27.55%	0.06	0.12%	0.07	0.50%
PolyProtect	0.32	73.83%	0.06	0.07%	0.07	0.52%
SlerpFace	**0.05**	**0.10%**	**0.05**	**0.06%**	**0.06**	**0.43%**

Table 2: Quantitative privacy analysis by Sim and SRRA.

Method	LFW	CFP-FP	AGEDB
ArcFace	99.73	98.00	97.87
w/o LW	99.37	87.13	90.73
w/o W	99.00	92.56	93.67
SlerpFace	**99.42**	**92.79**	**95.70**

Table 3: Components' role to SlerpFace's recognizability.

feature group. Hence, r^m would be the ideal cost to recover full $\vec{t}$, as the attacker must succeed within each group *simultaneously*. Without dropout, it takes NR $1.015^{49} \approx 2$ reruns to find a $\vec{t}$. However, with dropout, it takes around 3.6^{49} reruns, which is computationally infeasible. This suggests that dropout provides strong irreversibility.

Using the framework from Mai et al. (2020), we also *theoretically* analyzed the irreversibility of SlerpFace, showing it has better entropy (69.58 compared to 59.41) and matching accuracy (86.20% compared to 85.36%) on CFP-FP dataset.

Revocability. It mandates that any compromised protective templates be revocable and replaceable. This can be easily achieved by re-enrolling $\vec{t}$ with a distinct key $\vec{k}'$.

Unlinkability. It requires that when generating different protective templates for the same person's identity, the generated templates cannot be associated with each other. We measure SlerpFace's unlinkability via system score-wise linkability or SSWL score $D_{\leftrightarrow}^{sys}$, an evaluation metric proposed by Gomez-Barrero et al. (2017) and used in prior arts (Mai et al. 2020). In essence, it is a $[0, 1]$ value that measures the distributional discrepancy between template pairs describing the same or different identities, referred to as *mated* or *non-mated* pairs. To achieve unlinkability, the distributions of mated and non-mated pairs should be close and with small $D_{\leftrightarrow}^{sys}$. Figure 6(b) exhibits the pair-wise distribution for unprotected baseline and SlerpFace on the CFP-FP dataset. While baseline having $D_{\leftrightarrow}^{sys}$=0.94 indicates distinguishable distributions and undermines privacy, SlerpFace achieves $D_{\leftrightarrow}^{sys}$=0.05 and close distributions that are satisfactory for unlinkability.

5 Conclusion

This paper studies face template protection in face recognition. It first identifies diffusion-based inversion attacks as an exacerbated privacy threat. It then proposes a novel FTP method, SlerpFace, that effectively prevents inversion. SlerpFace rotates templates to a noise-like distribution that deteriorates attack models' capability, efficiently via spherical linear interpolation. It further proposes feature grouping and dropout, optimizable via a learnable approach, to enhance irreversibility. Extensive experiments demonstrate that SlerpFace outperforms SOTAs in both privacy protection and recognition performance.

References

Abdellatef, E.; Ismail, N. A.; Abd Elrahman, S. E. S.; Ismail, K. N.; Rihan, M.; and Abd El-Samie, F. E. 2020. Cancelable multi-biometric recognition system based on deep learning. *The Visual Computer*, 36: 1097–1109.

Boddeti, V. N. 2018. Secure face matching using fully homomorphic encryption. In *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 1–10. IEEE.

Boutros, F.; Grebe, J. H.; Kuijper, A.; and Damer, N. 2023. IDiff-Face: Synthetic-based Face Recognition through Fizzy Identity-conditioned Diffusion Models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.

Chan, E. R.; Lin, C. Z.; Chan, M. A.; Nagano, K.; Pan, B.; Mello, S. D.; Gallo, O.; Guibas, L.; Tremblay, J.; Khamis, S.; Karras, T.; and Wetzstein, G. 2022. Efficient Geometry-aware 3D Generative Adversarial Networks. In *CVPR*.

Chen, Y.; Wo, Y.; Xie, R.; Wu, C.; and Han, G. 2019. Deep secure quantization: On secure biometric hashing against similarity-based attacks. *Signal Processing*, 154: 314–323.

Chen, Z.; Zhang, J.; Lai, Z.; Chen, J.; Liu, Z.; and Li, J. 2022. Geometry-aware guided loss for deep crack recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4703–4712.

Chen, Z.; Zhang, J.; Lai, Z.; Zhu, G.; Liu, Z.; Chen, J.; and Li, J. 2023. The devil is in the crack orientation: A new perspective for crack detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 6653–6663.

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

Dang, T. M.; Tran, L.; Nguyen, T. D.; and Choi, D. 2020. FEHash: Full entropy hash for face template protection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 810–811.

Deng, J.; Guo, J.; Xue, N.; and Zafeiriou, S. 2019. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 4690–4699.

Dong, X.; Kim, S.; Jin, Z.; Hwang, J. Y.; Cho, S.; and Teoh, A. B. J. 2021. Secure chaff-less fuzzy vault for face identification systems. *ACM Transactions on Multimidia Computing Communications and Applications*, 17(3): 1–22.

Dong, X.; Miao, Z.; Ma, L.; Shen, J.; Jin, Z.; Guo, Z.; and Teoh, A. B. J. 2023. Reconstruct face from features based on genetic algorithm using GAN generator as a distribution constraint. *Computers & Security*, 125: 103026.

Dusmanu, M.; Schonberger, J. L.; Sinha, S. N.; and Pollefeys, M. 2021. Privacy-preserving image features via adversarial affine subspace embeddings. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14267–14277.

Engelsma, J. J.; Jain, A. K.; and Boddeti, V. N. 2022. HERS: Homomorphically encrypted representation search. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 4(3): 349–360.

Gilkalaye, B. P.; Rattani, A.; and Derakhshani, R. 2019. Euclidean-distance based fuzzy commitment scheme for biometric template security. In *2019 7th International Workshop on Biometrics and Forensics (IWBF)*, 1–6. IEEE.

Gomez-Barrero, M.; Galbally, J.; Rathgeb, C.; and Busch, C. 2017. General framework to evaluate unlinkability in biometric template protection systems. *IEEE Transactions on Information Forensics and Security*, 13(6): 1406–1420.

Guo, Y.; Zhang, L.; Hu, Y.; He, X.; and Gao, J. 2016. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In *European Conference on Computer Vision*, 87–102. Springer.

Hahn, V. K.; and Marcel, S. 2022. Towards protecting face embeddings in mobile face verification scenarios. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 4(1): 117–134.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33: 6840–6851.

Huang, Y.; Wang, Y.; Tai, Y.; Liu, X.; Shen, P.; Li, S.; Li, J.; and Huang, F. 2020. Curricularface: adaptive curriculum learning loss for deep face recognition. In *proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5901–5910.

ISO/IEC. 2022. Information security, cybersecurity and privacy protection — Biometric information protection. Document ISO/IEC 24745:2022.

Jindal, A. K.; Shaik, I.; Vasudha, V.; Chalamala, S. R.; Ma, R.; and Lodha, S. 2020. Secure and privacy preserving method for biometric template protection using fully homomorphic encryption. In *2020 IEEE 19th international conference on trust, security and privacy in computing and communications (TrustCom)*, 1127–1134.

Kansy, M.; Raël, A.; Mignone, G.; Naruniec, J.; Schroers, C.; Gross, M.; and Weber, R. M. 2023. Controllable Inversion of Black-Box Face Recognition Models via Diffusion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 3167–3177.

Kim, M.; Jain, A. K.; and Liu, X. 2022. Adaface: Quality adaptive margin for face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 18750–18759.

Kim, S.; Jeong, Y.; Kim, J.; Kim, J.; Lee, H. T.; and Seo, J. H. 2021. IronMask: Modular architecture for protecting deep face template. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16125–16134.

Lai, Y.; Jin, Z.; Wong, K.; and Tistarelli, M. 2021. Efficient known-sample attack for distance-preserving hashing biometric template protection schemes. *IEEE Transactions on Information Forensics and Security*, 16: 3170–3185.

Learned-Miller, G. B. H. E. 2014. Labeled Faces in the Wild: Updates and New Reporting Procedures. Technical Report UM-CS-2014-003, University of Massachusetts, Amherst.

Li, S.; Xu, J.; Xu, X.; Shen, P.; Li, S.; and Hooi, B. 2021. Spherical confidence learning for face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 15629–15637.

Lin, Y.-S.; Liu, Z.-Y.; Chen, Y.-A.; Wang, Y.-S.; Chang, Y.-L.; and Hsu, W. H. 2021. xCos: An explainable cosine metric for face verification task. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 17(3s): 1–16.

Lindstrom, M. J.; and Bates, D. M. 1988. Newton—Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association*, 83(404): 1014–1022.

Liu, Y.; Jia, C.; Xiao, R.; Jia, X.; Wei, H.; Jiang, K.; and Wang, Z. 2024. Local Features Meet Stochastic Anonymization: Revolutionizing Privacy-Preserving Face Recognition for Black-Box Models. arXiv:2412.08276.

Lu, Y.; Zhang, M.; Ma, A. J.; Xie, X.; and Lai, J. 2024. Coarse-to-fine latent diffusion for pose-guided person image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6420–6429.

Lugmayr, A.; Danelljan, M.; Romero, A.; Yu, F.; Timofte, R.; and Van Gool, L. 2022. Repaint: Inpainting using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11461–11471.

Mai, G.; Cao, K.; Lan, X.; and Yuen, P. C. 2020. Secureface: Face template protection. *IEEE Transactions on Information Forensics and security*, 16: 262–277.

Mai, G.; Cao, K.; Yuen, P. C.; and Jain, A. K. 2018. On the reconstruction of face images from deep face templates. *IEEE transactions on pattern analysis and machine intelligence*, 41(5): 1188–1202.

Maze, B.; Adams, J.; Duncan, J. A.; Kalka, N.; Miller, T.; Otto, C.; Jain, A. K.; Niggel, W. T.; Anderson, J.; Cheney, J.; et al. 2018. Iarpa janus benchmark-c: Face dataset and protocol. In *2018 international conference on biometrics (ICB)*, 158–165. IEEE.

Mi, Y.; Huang, Y.; Ji, J.; Liu, H.; Xu, X.; Ding, S.; and Zhou, S. 2022. Duetface: Collaborative privacy-preserving face recognition via channel splitting in the frequency domain. In *Proceedings of the 30th ACM International Conference on Multimedia*, 6755–6764.

Mi, Y.; Huang, Y.; Ji, J.; Zhao, M.; Wu, J.; Xu, X.; Ding, S.; and Zhou, S. 2023. Privacy-preserving face recognition using random frequency components. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 19673–19684.

Mi, Y.; Zhong, Z.; Huang, Y.; Ji, J.; Xu, J.; Wang, J.; Wang, S.; Ding, S.; and Zhou, S. 2024. Privacy-preserving face recognition using trainable feature subtraction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 297–307.

Mohan, D. D.; Sankaran, N.; Tulyakov, S.; Setlur, S.; and Govindaraju, V. 2019. Significant Feature Based Representation for Template Protection. In *CVPR Workshops*, 2389–2396.

Moschoglou, S.; Papaioannou, A.; Sagonas, C.; Deng, J.; Kotsia, I.; and Zafeiriou, S. 2017. Agedb: the first manually collected, in-the-wild age database. In *proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 51–59.

Phillips, T.; Zou, X.; Li, F.; and Li, N. 2019. Enhancing biometric-capsule-based authentication and facial recognition via deep learning. In *Proceedings of the 24th ACM Symposium on Access Control Models and Technologies*, 141–146.

Pillai, J. K.; Patel, V. M.; Chellappa, R.; and Ratha, N. K. 2011. Secure and robust iris recognition using random projections and sparse representations. *IEEE transactions on pattern analysis and machine intelligence*, 33(9): 1877–1893.

Rathgeb, C.; Merkle, J.; Scholz, J.; Tams, B.; and Nesterowicz, V. 2022. Deep face fuzzy vault: Implementation and performance. *Computers & Security*, 113: 102539.

Razzhigaev, A.; Kireev, K.; Udovichenko, I.; and Petiushko, A. 2021. Darker than black-box: Face reconstruction from similarity queries. *arXiv:2106.14290*.

Ren, Y.; Xia, X.; Lu, Y.; Zhang, J.; Wu, J.; Xie, P.; Wang, X.; and Xiao, X. 2024. Hyper-sd: Trajectory segmented consistency model for efficient image synthesis. *arXiv:2404.13686*.

Schreiber, R.; and Van Loan, C. 1989. A storage-efficient WY representation for products of Householder transformations. *SIAM Journal on Scientific and Statistical Computing*, 10(1): 53–57.

Sengupta, S.; Chen, J.-C.; Castillo, C.; Patel, V. M.; Chellappa, R.; and Jacobs, D. W. 2016. Frontal to profile face verification in the wild. In *Winter conference on applications of computer vision*, 1–9.

Shahreza, H. O.; Hahn, V. K.; and Marcel, S. 2022. Face reconstruction from deep facial embeddings using a convolutional neural network. In *2022 IEEE International Conference on Image Processing (ICIP)*, 1211–1215. IEEE.

Shahreza, H. O.; Hahn, V. K.; and Marcel, S. 2023. Mlp-hash: Protecting face templates via hashing of randomized multi-layer perceptron. In *2023 31st European Signal Processing Conference (EUSIPCO)*, 605–609. IEEE.

Shahreza, H. O.; and Marcel, S. 2023a. Comprehensive Vulnerability Evaluation of Face Recognition Systems to Template Inversion Attacks Via 3D Face Reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Shahreza, H. O.; and Marcel, S. 2023b. Face reconstruction from facial templates by learning latent space of a generator network. In *37th Conference on Neural Information Processing Systems*.

Shen, Y.; Yang, C.; Tang, X.; and Zhou, B. 2020. Interfacegan: Interpreting the disentangled face representation learned by gans. *IEEE transactions on pattern analysis and machine intelligence*, 44(4): 2004–2018.

Shoemake, K. 1985. Animating rotation with quaternion curves. In *Proceedings of the 12th annual conference on Computer graphics and interactive techniques*, 245–254.

Truong, T.-D.; Duong, C. N.; Le, N.; Savvides, M.; and Luu, K. 2022. Vec2Face-v2: Unveil Human Faces from their Blackbox Features via Attention-based Network in Face Recognition. *arXiv:2209.04920*.

Vendrow, E.; and Vendrow, J. 2021. Realistic face reconstruction from deep embeddings. In *NeurIPS 2021 Workshop Privacy in Machine Learning*.

Wang, H.; Dong, X.; Jin, Z.; Teoh, A. B. J.; and Tistarelli, M. 2021. Interpretable security analysis of cancellable biometrics using constrained-optimized similarity-based attack. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 70–77.

Whitelam, C.; Taborsky, E.; Blanton, A.; Maze, B.; Adams, J.; Miller, T.; Kalka, N.; Jain, A. K.; Duncan, J. A.; Allen, K.; et al. 2017. Iarpa janus benchmark-b face dataset. In *proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 90–98.

Xu, Z.; Hao, D.; Wu, Y.; and Peng, J. 2020. A random binarization scheme for deep face feature protection. In *Proceedings of the 4th International Conference on Computer Science and Application Engineering*, 1–5.

Yuan, L.; Chen, W.; Pu, X.; Zhang, Y.; Li, H.; Zhang, Y.; Gao, X.; and Ebrahimi, T. 2024. PRO-Face C: Privacy-preserving Recognition of Obfuscated Face via Feature Compensation. *IEEE Transactions on Information Forensics and Security*.

Yuan, L.; Liu, L.; Pu, X.; Li, Z.; Li, H.; and Gao, X. 2022. PROface: A generic framework for privacy-preserving recognizable obfuscation of face images. In *Proceedings of the 30th ACM international conference on multimedia*, 1661–1669.

Zhang, D.; Peng, Y.; Wu, A.; and Zheng, W.-s. 2024a. Privacy-Preserving Face Recognition with Adaptive Generative Perturbations. In *International Conference on Pattern Recognition*, 210–227. Springer.

Zhang, D.; Peng, Y.-X.; Wu, X.-M.; Wu, A.; and Zheng, W.-S. 2024b. PixelFade: Privacy-preserving Person Re-identification with Noise-guided Progressive Replacement. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 6326–6334.

Zheng, T.; and Deng, W. 2018. Cross-pose lfw: A database for studying cross-pose face recognition in unconstrained environments. *BUPT, Tech. Rep*, 5(7).

Zheng, T.; Deng, W.; and Hu, J. 2017. Cross-age lfw: A database for studying cross-age face recognition in unconstrained environments. *arXiv:1708.08197*.