Privacy-Preserving Face Recognition Using Trainable Feature Subtraction

Yuxi Mi[1]* Zhizhou Zhong[1]* Yuge Huang[2]† Jiazhen Ji[2] Jianqing Xu[2]
Jun Wang[3] Shaoming Wang[3] Shouhong Ding[2] Shuigeng Zhou[1]†
[1] Fudan University [2] Youtu Lab, Tencent [3] WeChat Pay Lab33, Tencent
yxmi20@fudan.edu.cn, zzzhong22@m.fudan.edu.cn, sgzhou@fudan.edu.cn
{yugehuang, royji, joejqxu, ericshding}@tencent.com
{earljwang, mangosmwang}@tencent.com

Abstract

The widespread adoption of face recognition has led to increasing privacy concerns, as unauthorized access to face images can expose sensitive personal information. This paper explores face image protection against viewing and recovery attacks. Inspired by image compression, we propose creating a visually uninformative face image through feature subtraction between an original face and its model-produced regeneration. Recognizable identity features within the image are encouraged by co-training a recognition model on its high-dimensional feature representation. To enhance privacy, the high-dimensional representation is crafted through random channel shuffling, resulting in randomized recognizable images devoid of attacker-leverageable texture details. We distill our methodologies into a novel privacy-preserving face recognition method, MinusFace. Experiments demonstrate its high recognition accuracy and effective privacy protection. Its code is available at https://github.com/Tencent/TFace.

1. Introduction

Face recognition (FR) is a biometric way to identify persons through their face images. It has seen prevalent methodological and application advancements in recent years. Currently, considerable parts of FR are implemented as online services to overcome local resource limitations: Local clients, such as cell phones, outsource captured face images to an online service provider. Using its model, the provider extracts identity-representative templates from the face images and matches them with its database.

It has been common sense that face images are sensitive biometric data and should be protected. Increasing regulatory demands [51] call for privacy-preserving face recognition (PPFR), to avoid leakage of face images during the outsourcing. They attempt to ensure that the faces' appearances are both *visually concealed* from inadvertent view by third parties and *difficult to recover* by deliberate attackers.

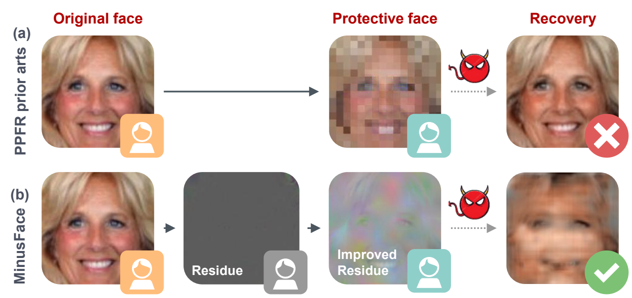
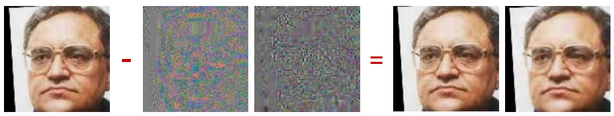


Figure 1. Comparison between SOTAs and MinusFace. (a) SOTAs gradually remove the most visually informative features. Inadequacy of removal can result in successful recovery, which undermines privacy. (b) MinusFace first obtains a *fully* visually uninformative residue representation, then improves its recognizability. It exhibits better privacy protection than all SOTAs.

State-of-the-art (SOTA) PPFR primarily employs two approaches: Cryptographic methods protect face images with encryption or security protocols. Recently, transform-based methods have gained popularity due to their low latency and budget-saving computational costs. They convert images into protective representations by minimizing visual details, rendering them safe to share.

Transform-based methods yet face a persistent challenge in balancing accuracy and privacy. In face images, the recognizable *identity features* and appearance-revealing *visual features* are closely intertwined. To achieve privacy while preserving optimal recognizability, prior arts invest significant efforts to locate and minimize the most visually informative feature components while retaining the rest. They commonly employ either a heuristic or adversarial training approach: For instance, some [24, 35, 36, 56] turn face images into frequency domain and heuristically prune the most human-perceivable frequency channels. Others exploit deep steganography [64] or cyclically add adversarial noise [57].

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



(a) Original image (b) Residual representations (c) Compressed images

Figure 2. Examples of image compression. Subtle details like texture are removed from (a) the original image to obtain (c) the compressed ones. The removed (b) residual representations are visually uninformative, yet carry descriptive features of the origin.

While these methods succeed in concealing faces from human inspection, they can be *largely susceptible to recovery attacks* [9, 15, 30]. Their challenge lies in ensuring an adequate removal of visual features, as subtle features may remain, providing attackers with potential leverage.

This paper advocates a novel approach to more effectively minimize visual features, drawing inspiration from image compression. Image compression reduces image size while preserving fidelity by discarding subtle features such as texture details and color variations. The paper observes that the discarded features, *i.e.*, the *residue* between original and compressed images, exhibit properties closely aligned with the desired protective face representation: They are both visually uninformative and preserve descriptive features of the original image, as shown in Fig. 2.

Emulating the production of discarded features, this paper introduces a trainable *feature subtraction* strategy to craft protective representations. In this approach, a generative model is first trained to faithfully produce a regeneration of the original face, where the regenerated face simulates a compressed image. The residue between the original and regenerated faces is expected to be devoid of visual features if the model is well-optimized. It is later exploited to produce a protective representation. To retain recognizability within the residue, a recognition model is co-trained, taking the residue as input to learn identity features.

Two techniques are subsequently proposed to enhance both the recognizability and privacy of the residue. To address specific training constraints of the FR model (detailed in Sec. 3.3), the residue is generated as *high-dimensional representations* instead of spatial images, enabling better preservation of identity features. Privacy is heightened through *random channel shuffling*, which obscures facial texture signals and increases randomness to hinder recovery attacks. The shuffled high-dimensional residue is ultimately mapped back as a spatial image, serving as the protective representation. The methodology is concretized into a novel PPFR framework, MinusFace. Figure 1 compare it with SOTA prior arts in paradigm. Experimental results show that MinusFace achieves high recognition accuracy and better privacy protection than SOTAs.

This paper presents three-fold contributions:

- It introduces *feature subtraction*, a new methodology to generate protective face representation, by capturing residue between an original image and its regeneration.
- It proposes two specific techniques, high-dimensional mapping and random channel shuffling, to ensure recognizability and accuracy for the residue.
- It presents a novel PPFR method, MinusFace. Experimental results demonstrate its high recognition accuracy and superior privacy protection to SOTAs.

2. Related work

2.1. Face recognition

Current FR systems identify persons by comparing their face templates, *i.e.*, one-dimensional feature embeddings. The service provider trains a convolutional neural network (CNN) to extract templates from face images. With angular-margin-based losses [7, 8, 21, 28, 53], the templates are encouraged to have large inter-identity and small intra-identity discrepancies that facilitate recognition.

2.2. Privacy-preserving face recognition

Many approaches have been proposed to protect face privacy [33, 34, 55]. We divide them into two categories.

Cryptographic methods perform recognition on encrypted face images. To allow necessary computations in the cipher space, many prior arts employ homomorphic encryption [11, 18, 20, 23, 43, 62] or secure multiparty computation [29, 41, 60, 63] to extract encrypted features and calculate their pair-wise distances. Others leverage different crypto-primitives including matrix encryption [26], one-time-pad [10], functional encryption [1], and locality-sensitive hashing [12, 66]. These methods, however, mostly bear high latency and heavy computational overheads.

Transform-based methods convert face images into protective representations that cannot be directly viewed. Pioneering arts obfuscate the image by adding crafted noise [4, 27, 31, 58, 65], performing clustering [16], or extracting coarse representations [5, 25, 37, 47]. Some regenerate the faces' features to obtain different visual appearances using autoencoders [38, 48], adversarial generative networks [2, 27, 39], and diffusion models [3, 22]. However, these methods suffer from compromised recognition accuracy as the obfuscation and regeneration often indiscriminately degrade the faces' visual and identity features. Recent methods locate and modify the images' most visually informative components. [24, 35, 36, 56] transform images to the frequency domain, where human-perceivable low-frequency channels are pruned. [64] uses deep steganography to conceal the face under distinct carrier images and aligns identity features via contrastive loss. [57] generates protective features by cyclically adding adversarial noise to sensitive signals. These methods visually conceal face appearance quite successfully and maintain decent recognition accuracy. However, we later experimentally show that they can be vulnerable to recovery attacks.

3. Methodology

This section describes our proposed MinusFace. The name comes from the key methodology to produce the protective representation, by subtracting between the original face and its regeneration, *i.e.*, the "*minus*". We begin by learning a visually uninformative representation in Sec. 3.2 via feature subtraction. In Sec. 3.3, we improve the representation in high-dimension to let it preserve identity features. In Sec. 3.4, we further address its privacy and to produce the final protective representation.

3.1. Motivation

The general goal of transform-based PPFR is to design a privacy-preserving transformation F that converts any original face image X to a *protective representation* $X_p = F(X)$. In prior arts, X_p can be concretized as a spatial image [4, 20, 50, 57] or high-dimensional feature channels [24, 35, 36, 56, 57]. Either way, we expect X_p to *preserve identity features* and *minimize visual features*.

Our approach is enlightened by image compression, a technique that reduces image size while preserving fidelity. Specifically, lossy image compression methods [46, 49, 52] exploit human perceptual insensitivity to discard subtle features like texture details or color variations. Interestingly, the discarded features possess the desired properties for our X_p: Since they are considered insignificant to image fidelity, they should be visually uninformative; otherwise, the compression would be too lossy. Meanwhile, viewing these features as the *residual representation*, denoted as R, between the original and compressed image, they still contain the image's clues. If the residue R can be utilized for recognition, the compression factually provides us with a natural X_p that is both visually indiscernible and recognizable.

Figure 2(b) demonstrates the residual representations of an example face image under JPEG [52] and JPEG 2000 [46] compression standards. We can observe feature clues from both residual representations. Regretfully, we find directly using them for face recognition ends up quite ineffective because their features are not specifically manufactured to keep the identity. In fact, they behave more like random noise from the perception of FR models.

3.2. Feature subtraction: minimize visual features

Imitating image compression, we can produce a residual representation R that is recognizable through a trainable *feature subtraction* strategy: To minimize visual features, we train a model that regenerates a face image X' taking the original face X as input. It simulates the compression process. We produce R as the subtraction between X and X', *i.e.* their *minus*, which should be visually uninformative

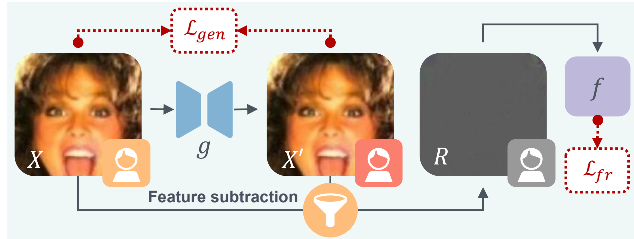


Figure 3. The core idea of MinusFace. Imitating image compression, a visually uninformative residue R is generated from *feature subtraction*: the original face *minus* its regeneration. R is also optimized with an FR model to preserve identity features.

if the regeneration is successful. Unlike image compression, crucially, we meanwhile train an FR model that tries to recognize R. By balancing the training of two models, R should also preserve identity features once the FR model is optimized. Such an R hence may serve as our protective representation X_p. Figure 3 demonstrates our idea.

We first concretize the minimization of visual features. Specifically, let g be a generative model (*wlog.*, a CNN autoencoder). We regenerate a face image from the original face as $X' = g(X)$. To make X' visually close to X, we employ l_1-norm as the model's objective:

$$\mathcal{L}_{gen} = \left\| X, X' \right\|_1 . \tag{1}$$

Prior studies [35, 36] suggest optimizing Eq. (1) is trivial provided the original face is not further obfuscated, which is our case. Therefore, we can confidently obtain a regeneration with high fidelity, $X' \approx X$. We produce the residue as their subtraction, $R = X - X'$.

As earlier discussed, prior arts invest huge efforts in removing the most visually informative features from X_p. However, their removals of features are often inadequate, resulting in unsatisfactory privacy. Leveraging feature subtraction, we efficiently transform the feature-minimizing objective of X_p to the feature-maximizing goal of X', which is easier to quantify: Instead of explicitly removing R's visual features, since Eq. (1) can be rewritten as

$$\mathcal{L}_{gen} = \left\| X, X' \right\|_1 = \left\| X - X', 0 \right\|_1 = \left\| R \right\|_1 , \tag{2}$$

we can expect R to be visually uninformative simply by producing high-quality X'.

3.3. Preserve identity features in high-dimension

Next, we aim to obtain the identity features for R to make it recognizable. As illustrated in Fig. 3, the most intuitive strategy is to incorporate an FR model f that takes R as input. Let f be end-to-end trained with the generative model g, aiming to predict the face's identity y. Thus, R should acquire identity features as long as f is also optimized.

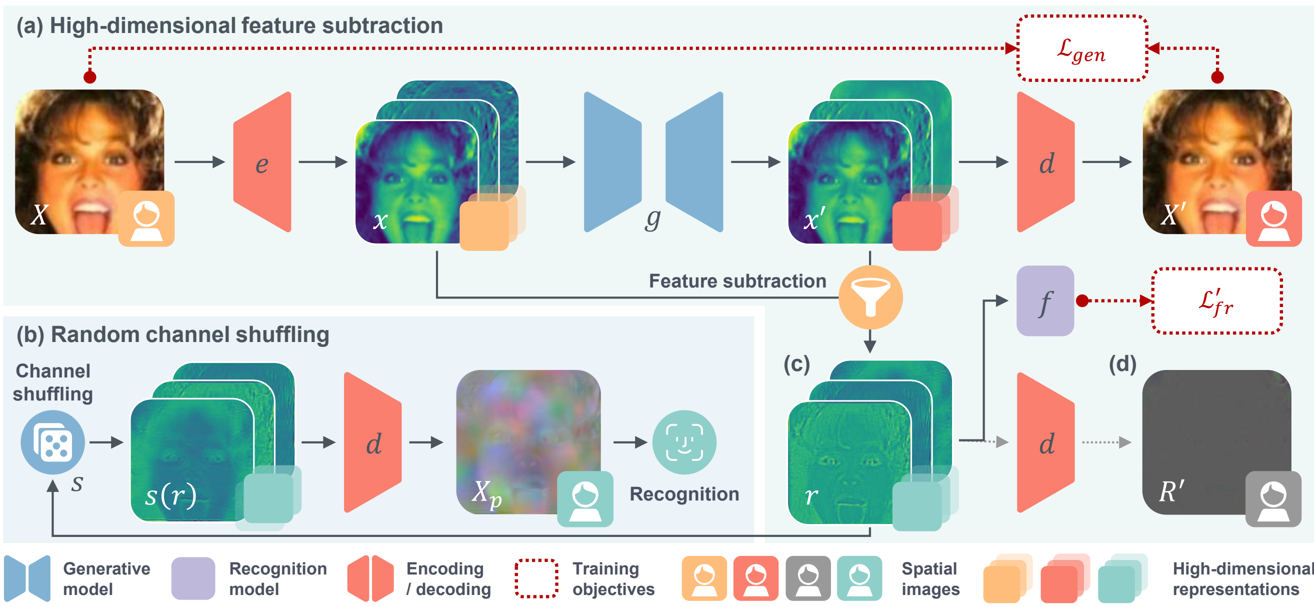


Figure 4. The MinusFace pipeline. (a) It centers around the idea of feature subtraction, where the protective representation X_p is derived from the residue between the original face X and its regeneration X'. Both regeneration and feature subtraction occur in high-dimension to preserve identity features within the trained residue r. (b) The residue r further undergoes random channel shuffling and decoding to produce the protective representation X_p. (c-d) All face figures are experimentally obtained and illustrate their representations faithfully.

However, we find training f can be challenging as it often ends up in poor convergence. We owe it to a slight drawback of feature subtraction: By optimizing Eq. (1), it in fact indiscriminately removes both visual and identity features, encouraging R to be blank. In other words, feature subtraction is trading off recognizability for privacy.

We propose a strategy that circumvents the trade-off, inspired by the property of high-dimensional spaces. Specifically, high-dimensional spaces often contain *significant redundancy of features*. If we map a spatial image $X \in \mathbf{M}$ to a high-dimensional representation $x \in \mathbf{N}$, we can expect X's visual appearance to be described by very few of x's components, *i.e.*, the *principal components*. The remaining features of x can then be reorganized without changing X. Let x, x' be the high-dimensional representations of X, X', respectively. While feature subtraction enforces $X' \to X$, likely making the principal components of x, x' identical, we can make a difference in their less visually descriptive and abundant remaining features. This allows us to produce non-blank high-dimensional residue $r = x - x' \neq 0$, which can carry identity features.

We establish a pair of differentiable, deterministic encoding $e : \mathbf{M} \to \mathbf{N}$ and decoding $d : \mathbf{N} \to \mathbf{M}$ mappings to handle the conversion between $\mathbf{M}, \mathbf{N}$. As properties necessary for later discussions, we require d, e together to be *invertible* and d alone to be *homomorphic*, *i.e.*,

$$\begin{cases} d\left(e(a)\right) = a & \forall a, \\ d(a_1 + a_2) = d(a_1) + d(a_2) & \forall a_1, a_2. \end{cases} \tag{3}$$

Also inspired by image compression, we choose *discrete cosine transform* (DCT) and its inverse (IDCT) as d, e, respectively. DCT is a linear transformation employed in JPEG [52] compression, that converts a $(3, H, W)$ image X into a $(192, H, W)$ high-dimensional x. We provide further details in the supplementary material. We opt for DCT, *wlog.*, for three main reasons: (1) It satisfies Eq. (3); (2) It produces x that preserves X's spatial structure and feature information: Study [61] shows models trained on x achieve similar performance as those on X; (3) It produces an x with 192 channels. The abundant channels later enhance privacy by shuffling their orders. Nonetheless, other d, e may be chosen provided at least Eq. (3) is satisfied.

Here, we describe producing r via high-dimensional feature subtraction. Figure 4(a) shows its pipeline. Note that all face figures here are experimentally obtained from MinusFace and illustrate their representations faithfully.

Specifically, we begin by encoding the high-dimensional representation of face image X as $x = e(X)$. Then, we regenerate $x' = g(x)$ using the model g, which is modified to accept a 192-channel input, and subsequently decode it into a spatial image as $X' = d(x')$. Similar to Eq. (1), g is trained by minimizing the $l1$-norm between X and X'.

Meanwhile, we can avoid a blank residue r by performing feature subtraction in high-dimension: We obtain the residue as $r = x - x' \neq 0$ and train the FR model f on r to help it acquire identity features, as previously discussed. The FR model f can be optimized using any SOTA FR loss; *Wlog.*, we opt for the popular ArcFace loss [8]:

$$\mathcal{L}_{fr} = l_{arc}(f(r), y). \tag{4}$$

The overall training objective of MinusFace is the combination of Eqs. (1) and (4), weighted by α, β:

$$\mathcal{L}_{minus} = \alpha \cdot \mathcal{L}_{gen} + \beta \cdot \mathcal{L}_{fr}. \tag{5}$$

We experimentally find both loss terms are optimized smoothly, and the produced residue r can be recognized by f with high accuracy, later shown in Sec. 4.6. Hence, by mapping X into high-dimension, we can acquire an r with identity features under feature subtraction. This satisfies our recognizability goal.

Before closing this section, further let $R' = d(r)$ be the decoding of r. Interestingly and crucially to the following discussion, we find $R' = R$, *i.e.*, R' equal to the spatial residue between X and X' that is *always blank*. The blankness of R' is contributed by the properties of d, e. Note that X can be rewritten by Eq. (3) as

$$X = d(e(X)) = d(x). \tag{6}$$

Combining Eq. (1) with d's homomorphism, it always holds

$$\begin{aligned}\mathcal{L}_{gen} &= \|X - X'\|_1 \\ &= \|d(x) - d(x')\|_1 = \|d(x - x')\|_1 \\ &= \|d(r)\|_1 = \|R'\|_1 .\end{aligned} \tag{7}$$

In Fig. 4(d), we exhibit sample R' experimentally generated, which is indeed blank. We use r and its mapping to a blank R' as key tools to produce the final protective X_p.

3.4. Random channel shuffling

The previous section creates a recognizable residue r. It is important to highlight that r cannot *directly* function as X_p since it lacks a guarantee of privacy: Feature subtraction only ensures removing visual features from R', but not necessarily from r. As exhibited in Fig. 4(c), subtle visual features in sample r persist, compromising its privacy.

To bridge the privacy gap, this section shows that a protective X_p can be simply derived as *perturbing then decoding* r, without further training. Specifically, we choose to perturb r by randomly shuffling its channels. Let $r_\triangle = s(r;\theta)$ represents the shuffling of r, where the channel order is determined by a *sample-wise* random seed θ. Thus, $X_p = d(s(r;\theta))$ serves as our final protective representation. The process is illustrated in Fig. 4(b). Following, we explain the motivations behind our design.

We first show shuffling will provably gain X_p with recognizability. Recall that r is primarily mapped to a blank $R' = d(r) \to 0$ devoid of any features, ensured by Eq. (7). Introducing a slight perturbation Δr as $r_\triangle = r + \Delta r$ plausibly results in a disrupted $R'_\triangle \neq R'$. Note that $R'_\triangle$ cannot be less informative than R' as the latter is already blank of

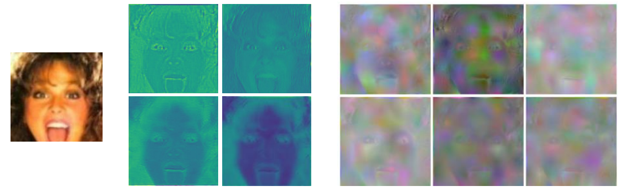


Figure 5. By randomly shuffling (b) channels of r, 192! distinct (c) X_p can be generated from (a) the same X. We exhibit some channels and X_p. Different X_p possess random texture patterns that obfuscate the recovery, by the nature of channel shuffling.

identity features. Conversely, it *acquires* features from the perturbation Δr, according to d's homomorphism:

$$d(r + \Delta r) = d(r) + d(\Delta r) \to d(\Delta r), \tag{8}$$

with believably $\|d(\Delta r)\|_1 > 0$ unless rare circumstances. Further to note that shuffling r' channels equals choosing

$$\Delta r = r - s(r;\theta) \neq 0. \tag{9}$$

Given that r preserves the identity features of the face image X, we anticipate that its shuffle $s(r;\theta)$ and their subtraction Δr will also be identity-descriptive of X. Consequently, X_p is able to assimilate the identity features of r through shuffling. In Sec. 4.2, we experimentally validate that the learning of identity features is robust, as FR on X_p attains satisfactory recognition accuracy.

We opt for random channel shuffling over other perturbations as it helps minimize privacy costs. Through perturbation, X_p is bound to unintentionally recover some visual features from r due to the intertwining of visual and identity features. In this context, shuffling demonstrates two-fold privacy benefits: *natural obfuscation* of visual features and *introduction of randomness* to X_p's representations.

To explain the natural obfuscation, we closely examine the sample channels of r in Fig. 5(b). We find these channels reveal consistent signals in structure (*e.g.*, positions for eyes and noses) but diverse ones in texture (*e.g.*, color depths). This phenomenon arises from the use of CNN-based g (and spatial-preserving d, e), wherein CNNs inherently preserve the spatial relations of images and generate distinct channel-wise signals through various convolutional kernels. Existing studies [6, 19, 35, 36, 45] suggest that structural signals play a pivotal role in FR models, while generative models (say, *the attacker's recovery model*) rely on both structural and texture signals. In our scenario, the structural signals consistent across channels prove more resistant to shuffling than the texture features with channel-wise variations. Figure 5(c) illustrates different samples of X_p generated from the same X under varied θ. These samples exhibit very subtle facial contours similar to that of X, facilitating recognition. In contrast, their facial texture is

Method	Venue	PPFR	LFW	CFP-FP	AgeDB	CPLFW	CALFW	IJB-B	IJB-C
ArcFace [8]	CVPR '19	No	99.77	98.30	97.88	92.77	96.05	94.13	95.60
PEEP [4]	CS '20	Yes	98.41	74.47	87.47	79.58	90.06	5.82	6.02
InstaHide [20]	ICML '20	Yes	96.53	83.20	79.58	81.03	86.24	61.88	69.02
Cloak [37]	WWW '21	Yes	98.91	87.97	92.60	83.43	92.18	33.58	33.82
PPFR-FD [56]	AAAI '22	Yes	99.69	94.85	97.23	90.19	95.60	92.93	94.07
DCTDP [24]	ECCV '22	Yes	99.77	96.97	97.72	91.37	96.05	93.29	94.43
DuetFace [35]	MM '22	Yes	99.82	97.79	97.93	92.35	96.10	93.66	95.30
PartialFace [36]	ICCV '23	Yes	99.80	97.63	97.79	92.03	96.07	93.64	94.93
ProFace [64]	MM '22	Yes	98.27	93.77	92.81	88.17	93.20	69.39	72.96
AdvFace [57]	CVPR '23	Yes	98.45	92.21	92.57	83.73	93.62	70.21	74.39
MinusFace	(ours)	Yes	99.78	96.92	97.57	91.90	95.90	93.37	94.70

Table 1. The performance comparison among MinusFace, an unprotected baseline, and PPFR SOTAs on face verification and identification tasks. MinusFace achieves on-par ($\pm 1\%$) performance with the best frequency-based SOTAs and outperforms the others.

transformed into meaningless color patches. This outcome of shuffling factually allows us to *selectively* obfuscate most visual features while preserving identity features, achieving an improved privacy-accuracy trade-off.

Privacy is further enhanced through the randomness of produced X_p. A successful recovery attack [9, 15, 30] necessitates training the attack model on consistent representations. Recall that r is a high-dimensional representation with a shape of $(192, H, W)$. Randomly shuffling its channels can produce 192! different X_p with random textures for the same X. The attacker can neither learn from X_p with random textures nor determine the seed θ for a specific X_p. Results in Secs. 4.4 and 4.5 show MinusFace completely nullifies SOTA recovery attacks.

3.5. Summary

To deploy MinusFace, the service provider first trains f, g under Eq. (5) that produces r. It discards f, as f does not serve as the final FR model. It shares frozen g with its clients. Capturing X, the clients obtain protective representation $X_p = F(X)$ with random θ, outsourcing it to the provider. The provider recognizes X_p on a newly trained FR model f_p. The same FR result is expected regardless of specific θ. The final privacy-preserving transformation of MinusFace is $F = d(s(r;\theta))$, where $r = e(X) - e(g(X))$.

4. Experiments

4.1. Experimental setup

Model and dataset. We employ a U-Net [42] autoencoder with reduced scale as g, and IR-50 [14] models as f, f_p. Training is carried out on the MS1Mv2 [13] dataset, which possesses 5.8M face images. We carry out evaluations on 5 regular-size datasets, LFW [17], CFP-FP [44], AgeDB [40], CPLFW [67] and CALFW [68]. We also use 2 large-scale datasets, IJB-B [59] and IJB-C [32]. We leave further experimental and training setup to the supplementary material.

4.2. Recognition accuracy

Compared methods. We compare MinusFace with an unprotected baseline and 9 transform-based PPFR methods. Specifically[1], (1) **ArcFace** [8], a non-privacy-preserving FR model trained on original face images; (2) **PEEP** [4], which obfuscates images using differential noise (privacy budget set to 5); (3) **InstaHide** [20], mixing the face image with 2 other images to conceal appearance; (4) **Cloak** [37], compressing the image's feature space (trade-off parameter set to 100); (5) **PPFR-FD** [56], shuffling and mixing frequency channels; (6) **DCTDP** [24], appending a frequency noise perturbation mask (privacy budget set to ϵ=1); (7) **DuetFace** [35], pruning frequency components and restoring accuracy via two-party collaboration; (8) **PartialFace** [36] exploiting a random subset of frequency channels for recognition; (9) **ProFace** [64], hiding the image's appearance through deep steganography; (10) **AdvFace** [57], perturbing the image by cyclically adding adversarial noise. These methods are divided into two branches by their means: the first three are early works that indiscriminately perturb all features, while the remaining selectively perturb the most visually informative features to better maintain accuracy.

Performance analysis. We evaluate MinusFace, baseline and compared methods on LFW, CFP-FP, AgeDB, CPLFW, and CALFW, and report results as recognition accuracy. We also evaluate them on IJB-B and IJB-C, and report TPR@FPR(1e-4). Results are summarized in Tab. 1.

We observe that early methods [4, 20, 37] experience a significant performance drop, especially on IJB-B/C, due to the compromise of identity features in indiscriminate obfuscation. Despite being designed to conceal mostly visual features, [57, 64] also exhibit considerable accuracy loss, sug-

[1] We found no open-source code for PPFR-FD and AdvFace. We reproduce them to our best effort, recognizing the possibility of inconsistencies.

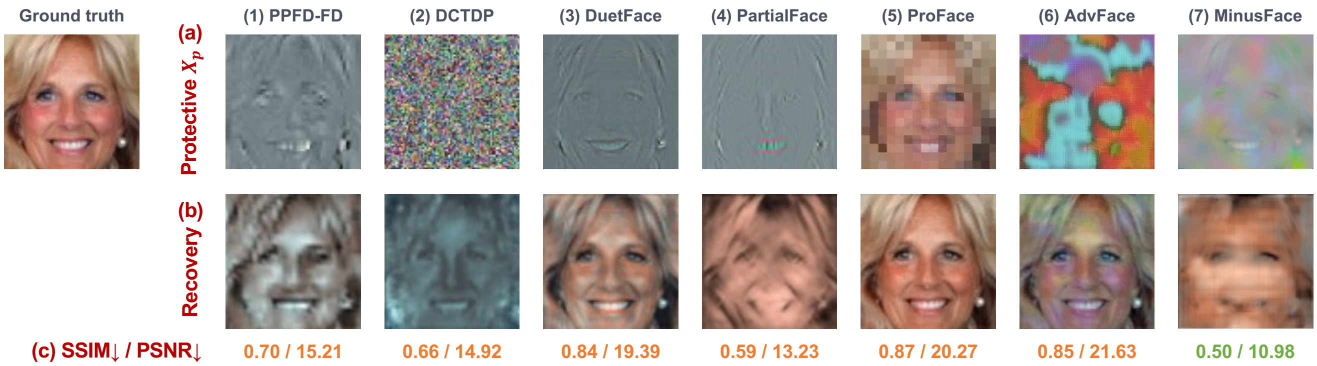


Figure 6. Privacy protection of MinusFace, compared with SOTAs. (a) MinusFace and most SOTAs successfully conceal the face image's visual appearance. (b) However, SOTAs fail to prevent recovery attacks. MinusFace outperforms all SOTAs as its recovered image is highly blurred and can hardly distinguish the face's existence. (c) Quantity results, where MinusFace exhibits the lowest SSIM and PSNR.

gesting inefficient trade-off between identity and visual features. Recently, frequency-based methods [24, 35, 36, 56] achieve notable accuracy by pruning visual appearance through removing low-frequency channels at a marginal utility cost. Their performance closely approaches the unprotected baseline. However, we later show that they can be susceptible to recovery attacks. MinusFace attains commendable performance, with a small gap ($\leq 2\%$) from the unprotected baseline. It is on par ($\pm 1\%$) with frequency-based methods and outperforms all other SOTAs. We argue that this slight accuracy trade-off is justified, as MinusFace offers significantly improved protection capability and efficiency, later discussed in Secs. 4.4 and 4.7.

4.3. Concealing of visual information

To evaluate MinusFace's privacy protection, recall our two-fold privacy goals: *visually concealing the face's appearance* and *hindering recovery attacks*. Here, we focus on the first goal and compare MinusFace with PPFR-FD, DCTDP, DuetFace, PartialFace, ProFace, and AdvFace. These SOTAs, similar to ours, treat visual and identity features discriminately. Specifically, we visualize their protective representations X_p to determine if visual appearances can be discerned. Note that X_p are not all in the form of images: Frequency-based methods produce X_p as frequency channels, which we convert back via a reverse transform; AdvFace generates feature maps, transformed into images using its shadow model; ProFace directly creates images.

Figure 6(a) displays X_p of each SOTA and MinusFace. Generally, all methods successfully conceal the face's appearance. DuetFace and ProFace provide slightly inferior protection, as their generated X_p reveal some discernible facial features. DCTDP and AdvFace better conceal visual appearance by applying noise and obfuscation. In Fig. 6a(7), MinusFace produces X_p that nearly eliminates the face's structural clues and completely conceals its texture details, effectively achieving the first privacy goal.

4.4. Protection against recovery

We here analyze the second privacy goal of protecting against recovery. We find MinusFace provides significantly better protection than SOTAs. We first describe the attack.

Threat model. We consider a white-box attacker who can query the PPFR framework and know its detailed protection mechanism. This attacker is typically envisioned as a malicious third-party wiretapping the transmission. While aware of the framework's general setup, such as hyper-parameters, the attacker does not know the specific sample-wise parameters (*e.g.*, θ in our case) used by the client to generate protective representations X_p. Assume the attacker has access to a training dataset of face images X. It can first obtain X_p by querying the PPFR framework. Then, it can train a recovery model f^{-1} to map X_p back to X, as $\arg\min_\delta \left\|f^{-1}(X_p;\delta), X\right\|_1$, and exploit f^{-1} to recover the client's shared X_p. We concretely use BUPT [54] of 1.3M images as the attacker's dataset, and employ a full-scale U-Net [42] as its f^{-1}.

Comparison with SOTAs. We train an attack model for MinusFace and each SOTA. Figure 6(b) displays examples of recovered images. We find that most prior arts provide insufficient protection against recovery. Specifically, ProFace is not designed to prevent recovery, resulting in a faithful recovered image from its protective X_p in Fig. 6b(5). Some prior arts [35, 56, 57] suggest resistance to the attack. However, in Fig. 6b(1)(3)(6), we find they can actually be recovered by a more powerful attacker, *e.g.*, training f^{-1} on a much larger dataset. For other methods, the faces' appearance can also be somewhat recovered. We attribute the SOTAs' shortcomings to their setback in ensuring adequate removal of visual features, especially facial textures. This leaves potential features that attackers can leverage. In Fig. 6b(7), MinusFace overcomes the drawback and exhibits strong protection to recovery, outperforming SOTAs.

Quantitative comparison. In Fig. 6(c), we quantify the quality of recovered images by their structural similarity

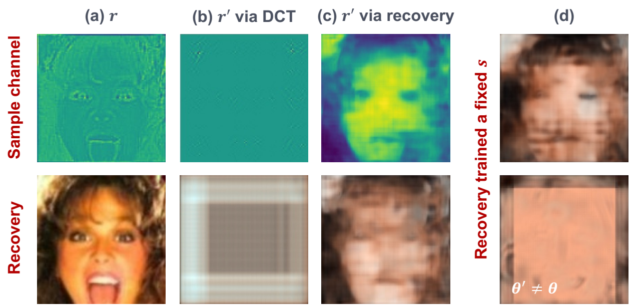


Figure 7. Left: Sample channels from r and the attacker's attempted inversions to reproduce r', together with their recovery. (a) r is not designed as privacy-preserving, hence can be recovered. (b-c) However, the attacker cannot obtain r or its correct inversion r', making recovery infeasible. Right: (d) Training recovery model on fixed θ does not pose an effective threat, as it fails entirely for $\theta' \neq \theta$, where θ has a random space of 192!.

(SSIM) and peak signal-to-noise ratio (PSNR) compared to the ground truth. Results are averaged on 10K IJB-C images. MinusFace exhibits the lowest SSIM and PSNR, indicating the best protection.

4.5. Protection against dedicated attacks

We further investigate two attacks dedicated to MinusFace's design that attempt to invert r or bypass X_p's randomness.
Inverting r. It is crucial to note that the high-dimensional residue r is not designed to be protective (although it produces a protective X_p) and can be easily recovered (Fig. 7(a)). Our design is safe as r is never shared. Yet, an attacker may attempt to further invert r from X_p and carry out recovery therefrom. However, we demonstrate that *inverting r is infeasible*. First, the attacker cannot invert an $r' = r$ by re-encoding it as $r' = e(X_p)$, even if it knows the specific θ (we omit θ for simplicity). Note that although Eq. (3) assures $d(e(X)) = X$, its opposite $e(d(x)) = x$ is not guaranteed to hold. In Fig. 7(b), re-encoding X_p produces inconsistent and uninformative r', further demolishing the attack. The attacker also cannot train a recovery model from X_p to r as it is essentially as difficult as training the previously discussed f^{-1}. Figure 7(c) demonstrates the unsuccessful r' and its recovery.
Bypassing randomness. The attacker is capable of generating X_p under a specific shuffle seed θ. It hence can train f^{-1} on X_p from the same θ, to bypass the randomness of representations. In Fig. 7(d), such trained f^{-1} produces slightly better recovery on X_p under the same θ, but fails entirely for any $\theta' \neq \theta$. As θ has a total random space of 192!, this attack does not impose any effective threat.

4.6. Ablation study

Recognition accuracy of r and R'. Recall r and R' represent the high-dimensional residue and its spatial decoding, respectively. We train FR models on them and show their

Method	CFP-FP	AgeDB	CPLFW
ArcFace	98.30	97.88	92.77
r	98.27	97.82	92.73
R'	53.85	54.71	51.20
X_p **(default)**	96.92	97.57	91.90

Table 2. Accuracy of models trained alternatively on r and R'.

Method	ours	[56]	[24]	[35]	[36]	[57]
Storage $\downarrow$	1	$\times$36	$\times$63	$\times$54	$\times$9	$\times$5.3

Table 3. Comparison of storage and transmission cost. $\times n$ indicates an n-time larger protective representation than MinusFace.

recognition accuracy. We expect the model trained on r (*i.e.*, f) to achieve high accuracy, as it indicates a lossless feature subtraction, favorable for MinusFace's overall utility. We also expect R''s model to experience a significant accuracy downgrade (close to a random guess of 50%), as R' should be fully removed of features. Results in Tab. 2 meet our expectations. We yield further ablation studies to the supplementary material, due to the limit of space.

4.7. Efficiency and compatibility

A practical PPFR framework is expected to have low inference latency and be efficient for storage and transmission. MinusFace is an ideal fit for these goals.
Latency. Testing on a personal laptop, MinusFace costs an average of 69 ms to turn an image into protective X_p, an order of magnitude smaller than typical communication time. It does not increase time costs for the service provider.
Storage and transmission. Many prior arts [24, 35, 36, 56, 57] produce X_p as high-dimensional feature channels (*e.g.*, DCTDP generates an $(189, H, W)$ X_p), resulting in additional storage and transmission costs. In contrast, MinusFace produces spatial images. As shown in Tab. 3, MinusFace's X_p requires far less size compared to SOTAs.
Compatibility. MinusFace is also compatible with different SOTA FR backbones and training objectives, as it neither modifies nor requires any specific design of FR architecture.

5. Conclusion

This paper investigates the privacy protection of face images. We present a new methodology, *feature subtraction*, to generate privacy-preserving face representations by capturing the residue between an original face and its regeneration. We further ensure the recognizability and privacy of residue via high-dimensional mapping and random channel shuffling, respectively. Our findings are concretized into a novel PPFR method, MinusFace. Extensive experiments demonstrate that it achieves satisfactory recognition accuracy and enhanced privacy protection.

References

[1] Michel Abdalla, Florian Bourse, Angelo De Caro, and David Pointcheval. Simple functional encryption schemes for inner products. In *Public-Key Cryptography - PKC 2015 - 18th IACR International Conference on Practice and Theory in Public-Key Cryptography, Gaithersburg, MD, USA, March 30 - April 1, 2015, Proceedings*, pages 733–751. Springer, 2015. 2

[2] Fadi Boutros, Marco Huber, Patrick Siebke, Tim Rieber, and Naser Damer. Sface: Privacy-friendly and accurate face recognition using synthetic data. In *2022 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–11. IEEE, 2022. 2

[3] Fadi Boutros, Jonas Henry Grebe, Arjan Kuijper, and Naser Damer. Idiff-face: Synthetic-based face recognition through fizzy identity-conditioned diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19650–19661, 2023. 2

[4] Mahawaga Arachchige Pathum Chamikara, Peter Bertók, Ibrahim Khalil, Dongxi Liu, and Seyit Camtepe. Privacy preserving face recognition utilizing differential privacy. *Comput. Secur.*, 97:101951, 2020. 2, 3, 6, 4

[5] Thee Chanyaswad, J. Morris Chang, Prateek Mittal, and Sun-Yuan Kung. Discriminant-component eigenfaces for privacy-preserving face recognition. In *26th IEEE International Workshop on Machine Learning for Signal Processing, MLSP 2016, Vietri sul Mare, Salerno, Italy, September 13-16, 2016*, pages 1–6. IEEE, 2016. 2

[6] Zhuangzhuang Chen, Jin Zhang, Zhuonan Lai, Jie Chen, Zun Liu, and Jianqiang Li. Geometry-aware guided loss for deep crack recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4703–4712, 2022. 5

[7] Jun Dan, Yang Liu, Haoyu Xie, Jiankang Deng, Haoran Xie, Xuansong Xie, and Baigui Sun. Transface: Calibrating transformer training for face recognition from a data-centric perspective. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 20642–20653, 2023. 2

[8] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pages 4690–4699. Computer Vision Foundation / IEEE, 2019. 2, 4, 6

[9] Alexey Dosovitskiy and Thomas Brox. Inverting visual representations with convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4829–4837, 2016. 2, 6

[10] Ovgu Ozturk Ergun. Privacy preserving face recognition in encrypted domain. In *2014 IEEE Asia Pacific Conference on Circuits and Systems, APCCAS 2014, Ishigaki, Japan, November 17-20, 2014*, pages 643–646. IEEE, 2014. 2

[11] Zekeriya Erkin, Martin Franz, Jorge Guajardo, Stefan Katzenbeisser, Inald Lagendijk, and Tomas Toft. Privacy-preserving face recognition. In *Privacy Enhancing Technologies, 9th International Symposium, PETS 2009, Seattle, WA, USA, August 5-7, 2009. Proceedings*, pages 235–253. Springer, 2009. 2

[12] Wenjing Gao, Jia Yu, Rong Hao, Fanyu Kong, and Xiaodong Liu. Privacy-preserving face recognition with multi-edge assistance for intelligent security systems. *IEEE Internet of Things Journal*, 2023. 2

[13] Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III*, pages 87–102. Springer, 2016. 6, 1

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 770–778. IEEE Computer Society, 2016. 6, 1

[15] Zecheng He, Tianwei Zhang, and Ruby B Lee. Model inversion attacks against collaborative inference. In *Proceedings of the 35th Annual Computer Security Applications Conference*, pages 148–162, 2019. 2, 6

[16] Katsuhiro Honda, Masahiro Omori, Seiki Ubukata, and Akira Notsu. A study on fuzzy clustering-based k-anonymization for privacy preserving crowd movement analysis with face recognition. In *7th International Conference of Soft Computing and Pattern Recognition, SoCPaR 2015, Fukuoka, Japan, November 13-15, 2015*, pages 37–41. IEEE, 2015. 2

[17] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, 2007. 6, 1

[18] Hai Huang and Luyao Wang. Efficient privacy-preserving face identification protocol. *IEEE Transactions on Services Computing*, 2023. 2

[19] Qidong Huang, Xiaoyi Dong, Dongdong Chen, Yinpeng Chen, Lu Yuan, Gang Hua, Weiming Zhang, and Nenghai Yu. Improving adversarial robustness of masked autoencoders via test-time frequency-domain prompting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1600–1610, 2023. 5

[20] Yangsibo Huang, Zhao Song, Kai Li, and Sanjeev Arora. Instahide: Instance-hiding schemes for private distributed learning. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, pages 4507–4518. PMLR, 2020. 2, 3, 6

[21] Yuge Huang, Yuhan Wang, Ying Tai, Xiaoming Liu, Pengcheng Shen, Shaoxin Li, Jilin Li, and Feiyue Huang. Curricularface: adaptive curriculum learning loss for deep face recognition. In *proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5901–5910, 2020. 2

[22] Ziqi Huang, Kelvin CK Chan, Yuming Jiang, and Ziwei Liu. Collaborative diffusion for multi-modal face generation and editing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6080–6090, 2023. 2

[23] Alberto Ibarrondo, Hervé Chabanne, Vincent Despiegel, and Melek Önen. Grote: Group testing for privacy-preserving face identification. In *Proceedings of the Thirteenth ACM Conference on Data and Application Security and Privacy*, pages 117–128, 2023. 2

[24] Jiazhen Ji, Huan Wang, Yuge Huang, Jiaxiang Wu, Xingkun Xu, Shouhong Ding, Shengchuan Zhang, Liujuan Cao, and Rongrong Ji. Privacy-preserving face recognition with learnable privacy budgets in frequency domain. In *Computer Vision - ECCV 2022 - 17th European Conference, Tel Aviv, Israel, October 23-27, 2022, Proceedings, Part XII*, pages 475–491. Springer, 2022. 1, 2, 3, 6, 7, 8, 4

[25] Tom A. M. Kevenaar, Geert Jan Schrijen, Michiel van der Veen, Anton H. M. Akkermans, and Fei Zuo. Face recognition with renewable and privacy preserving binary templates. In *Proceedings of the Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID 2005), 16-18 October 2005, Buffalo, NY, USA*, pages 21–26. IEEE Computer Society, 2005. 2

[26] Xiaoyu Kou, Ziling Zhang, Yuelei Zhang, and Linlin Li. Efficient and privacy-preserving distributed face recognition scheme via facenet. In *ACM TURC 2021: ACM Turing Award Celebration Conference - Hefei, China, 30 July 2021 - 1 August 2021*, pages 110–115. ACM, 2021. 2

[27] Yuancheng Li, Yimeng Wang, and Daoxing Li. Privacy-preserving lightweight face recognition. *Neurocomputing*, 363:212–222, 2019. 2

[28] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. Sphereface: Deep hypersphere embedding for face recognition. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 6738–6746. IEEE Computer Society, 2017. 2

[29] Zhuo Ma, Yang Liu, Ximeng Liu, Jianfeng Ma, and Kui Ren. Lightweight privacy-preserving ensemble classification for face recognition. *IEEE Internet Things J.*, 6(3):5778–5790, 2019. 2

[30] Guangcan Mai, Kai Cao, Pong C Yuen, and Anil K Jain. On the reconstruction of face images from deep face templates. *IEEE transactions on pattern analysis and machine intelligence*, 41(5):1188–1202, 2018. 2, 6

[31] Yunlong Mao, Jinghao Feng, Fengyuan Xu, and Sheng Zhong. A privacy-preserving deep learning approach for face recognition with edge computing. In *USENIX Workshop on Hot Topics in Edge Computing, HotEdge 2018, Boston, MA, July 10, 2018*. USENIX Association, 2018. 2

[32] Brianna Maze, Jocelyn C. Adams, James A. Duncan, Nathan D. Kalka, Tim Miller, Charles Otto, Anil K. Jain, W. Tyler Niggel, Janet Anderson, Jordan Cheney, and Patrick Grother. IARPA janus benchmark - C: face dataset and protocol. In *2018 International Conference on Biometrics, ICB 2018, Gold Coast, Australia, February 20-23, 2018*, pages 158–165. IEEE, 2018. 6, 1

[33] Blaz Meden, Peter Rot, Philipp Terhörst, Naser Damer, Arjan Kuijper, Walter J. Scheirer, Arun Ross, Peter Peer, and Vitomir Struc. Privacy-enhancing face biometrics: A comprehensive survey. *IEEE Trans. Inf. Forensics Secur.*, 16: 4147–4183, 2021. 2

[34] Pietro Melzi, Christian Rathgeb, Rubén Tolosana, Rubén Vera-Rodríguez, and Christoph Busch. An overview of privacy-enhancing technologies in biometric recognition. *CoRR*, abs/2206.10465, 2022. 2

[35] Yuxi Mi, Yuge Huang, Jiazhen Ji, Hongquan Liu, Xingkun Xu, Shouhong Ding, and Shuigeng Zhou. Duetface: Collaborative privacy-preserving face recognition via channel splitting in the frequency domain. In *MM '22: The 30th ACM International Conference on Multimedia, Lisboa, Portugal, October 10 - 14, 2022*, pages 6755–6764. ACM, 2022. 1, 2, 3, 5, 6, 7, 8, 4

[36] Yuxi Mi, Yuge Huang, Jiazhen Ji, Minyi Zhao, Jiaxiang Wu, Xingkun Xu, Shouhong Ding, and Shuigeng Zhou. Privacy-preserving face recognition using random frequency components. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19673–19684, 2023. 1, 2, 3, 5, 6, 7, 8

[37] Fatemehsadat Mireshghallah, Mohammadkazem Taram, Ali Jalali, Ahmed Taha Elthakeb, Dean M. Tullsen, and Hadi Esmaeilzadeh. Not all features are equal: Discovering essential features for preserving prediction privacy. In *WWW '21: The Web Conference 2021, Virtual Event / Ljubljana, Slovenia, April 19-23, 2021*, pages 669–680. ACM / IW3C2, 2021. 2, 6, 4

[38] Vahid Mirjalili, Sebastian Raschka, Anoop M. Namboodiri, and Arun Ross. Semi-adversarial networks: Convolutional autoencoders for imparting privacy to face images. In *2018 International Conference on Biometrics, ICB 2018, Gold Coast, Australia, February 20-23, 2018*, pages 82–89. IEEE, 2018. 2

[39] Vahid Mirjalili, Sebastian Raschka, and Arun Ross. Gender privacy: An ensemble of semi adversarial networks for confounding arbitrary gender classifiers. In *9th IEEE International Conference on Biometrics Theory, Applications and Systems, BTAS 2018, Redondo Beach, CA, USA, October 22-25, 2018*, pages 1–10. IEEE, 2018. 2

[40] Stylianos Moschoglou, Athanasios Papaioannou, Christos Sagonas, Jiankang Deng, Irene Kotsia, and Stefanos Zafeiriou. Agedb: The first manually collected, in-the-wild age database. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 1997–2005. IEEE Computer Society, 2017. 6, 1

[41] Lijing Ren, Denghui Zhang, et al. A privacy-preserving biometric recognition system with visual cryptography. *Advances in Multimedia*, 2022, 2022. 2

[42] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015 - 18th International Conference Munich, Germany, October 5 - 9, 2015, Proceedings, Part III*, pages 234–241. Springer, 2015. 6, 7, 1

[43] Ahmad-Reza Sadeghi, Thomas Schneider, and Immo Wehrenberg. Efficient privacy-preserving face recognition. In *Information, Security and Cryptology - ICISC 2009, 12th International Conference, Seoul, Korea, December 2-4, 2009, Revised Selected Papers*, pages 229–244. Springer, 2009. 2

[44] Soumyadip Sengupta, Jun-Cheng Chen, Carlos Domingo Castillo, Vishal M. Patel, Rama Chellappa, and David W. Jacobs. Frontal to profile face verification in the wild. In *2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016, Lake Placid, NY, USA, March 7-10, 2016*, pages 1–9. IEEE Computer Society, 2016. 6, 1

[45] Chenyang Si, Ziqi Huang, Yuming Jiang, and Ziwei Liu. Freeu: Free lunch in diffusion u-net. *arXiv preprint arXiv:2309.11497*, 2023. 5

[46] Athanassios Skodras, Charilaos Christopoulos, and Touradj Ebrahimi. The jpeg 2000 still image compression standard. *IEEE Signal processing magazine*, 18(5):36–58, 2001. 3

[47] Xin Sun, Chengliang Tian, Changhui Hu, Weizhong Tian, Hanlin Zhang, and Jia Yu. Privacy-preserving and verifiable src-based face recognition with cloud/edge server assistance. *Computers & Security*, 118:102740, 2022. 2

[48] Deyan Tang, Siwang Zhou, Hongbo Jiang, Haowen Chen, and Yonghe Liu. Gender-adversarial networks for face privacy preserving. *IEEE Internet of Things Journal*, 9(18): 17568–17576, 2022. 2

[49] Lucas Theis, Wenzhe Shi, Andrew Cunningham, and Ferenc Huszár. Lossy image compression with compressive autoencoders. *arXiv preprint arXiv:1703.00395*, 2017. 3

[50] Bo-Wei Tseng and Pei-Yuan Wu. Compressive privacy generative adversarial network. *IEEE Trans. Inf. Forensics Secur.*, 15:2499–2513, 2020. 3, 4

[51] Paul Voigt and Axel Von dem Bussche. The eu general data protection regulation (gdpr). *A Practical Guide, 1st Ed., Cham: Springer International Publishing*, 10(3152676):10–5555, 2017. 1

[52] Gregory K. Wallace. The JPEG still picture compression standard. *Commun. ACM*, 34(4):30–44, 1991. 3, 4, 1

[53] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. Cosface: Large margin cosine loss for deep face recognition. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pages 5265–5274. Computer Vision Foundation / IEEE Computer Society, 2018. 2

[54] Mei Wang and Weihong Deng. Mitigate bias in face recognition using skewness-aware reinforcement learning. arXiv, 2019. 7, 1

[55] Mei Wang and Weihong Deng. Deep face recognition: A survey. *Neurocomputing*, 429:215–244, 2021. 2

[56] Yinggui Wang, Jian Liu, Man Luo, Le Yang, and Li Wang. Privacy-preserving face recognition in the frequency domain. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*, pages 2558–2566. AAAI Press, 2022. 1, 2, 3, 6, 7, 8

[57] Zhibo Wang, He Wang, Shuaifan Jin, Wenwen Zhang, Jiahui Hu, Yan Wang, Peng Sun, Wei Yuan, Kaixin Liu, and Kui Ren. Privacy-preserving adversarial facial features. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8212–8221, 2023. 1, 2, 3, 6, 7, 8

[58] Yunqian Wen, Bo Liu, Ming Ding, Rong Xie, and Li Song. Identitydp: Differential private identification protection for face images. *Neurocomputing*, 501:197–211, 2022. 2

[59] Cameron Whitelam, Emma Taborsky, Austin Blanton, Brianna Maze, Jocelyn C. Adams, Tim Miller, Nathan D. Kalka, Anil K. Jain, James A. Duncan, Kristen Allen, Jordan Cheney, and Patrick Grother. IARPA janus benchmark-b face dataset. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 592–600. IEEE Computer Society, 2017. 6, 1

[60] Can Xiang, Chunming Tang, Yunlu Cai, and Qiuxia Xu. Privacy-preserving face recognition with outsourced computation. *Soft Comput.*, 20(9):3735–3744, 2016. 2

[61] Kai Xu, Minghai Qin, Fei Sun, Yuhao Wang, Yen-Kuang Chen, and Fengbo Ren. Learning in the frequency domain. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, pages 1737–1746. Computer Vision Foundation / IEEE, 2020. 4, 2

[62] Wencheng Yang, Song Wang, Hui Cui, Zhaohui Tang, and Yan Li. A review of homomorphic encryption for privacy-preserving biometrics. *Sensors*, 23(7):3566, 2023. 2

[63] Xiaopeng Yang, Hui Zhu, Rongxing Lu, Ximeng Liu, and Hui Li. Efficient and privacy-preserving online face recognition over encrypted outsourced data. In *IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), iThings/GreenCom/CPSCom/SmartData 2018, Halifax, NS, Canada, July 30 - August 3, 2018*, pages 366–373. IEEE, 2018. 2

[64] Lin Yuan, Linguo Liu, Xiao Pu, Zhao Li, Hongbo Li, and Xinbo Gao. Pro-face: A generic framework for privacy-preserving recognizable obfuscation of face images. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 1661–1669, 2022. 1, 2, 6

[65] Chen Zhang, Xiongwei Hu, Yu Xie, Maoguo Gong, and Bin Yu. A privacy-preserving multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. *Frontiers Neurorobotics*, 13:112, 2019. 2

[66] Shan Zhao, Lefeng Zhang, and Ping Xiong. Priface: a privacy-preserving face recognition framework under untrusted server. *Journal of Ambient Intelligence and Humanized Computing*, 14(3):2967–2979, 2023. 2

[67] T. Zheng and W. Deng. Cross-pose lfw: A database for studying cross-pose face recognition in unconstrained environments. Technical Report 18-01, Beijing University of Posts and Telecommunications, 2018. 6, 1

[68] Tianyue Zheng, Weihong Deng, and Jiani Hu. Cross-age LFW: A database for studying cross-age face recognition in unconstrained environments. arXiv, 2017. 6, 1