FracFace: Breaking the Visual Clues—Fractal-Based Privacy-Preserving Face Recognition

Wanying Dai[1,4], Beibei Li[1 *], Naipeng Dong[2 *], Guangdong Bai[3], Jin Song Dong[4]
[1]Sichuan University [2]The University of Queensland
[3]City University of Hong Kong [4]National University of Singapore
daiwanying@stu.scu.edu.cn, libeibei@scu.edu.cn, n.dong@uq.edu.au
g.bai@cityu.edu.hk, dcsdjs@nus.edu.sg

Abstract

Face recognition is essential for identity authentication, but the rich visual clues in facial images pose significant privacy risks, highlighting the critical importance of privacy-preserving solutions. For instance, numerous studies have shown that generative models are capable of effectively performing reconstruction attacks that result in the restoration of original visual clues. To mitigate this threat, we introduce **FracFace**, a **frac**tal-based privacy-preserving **face** recognition framework. This approach effectively weakens the visual clues that can be exploited by reconstruction attacks by disrupting the spatial structure in frequency domain features, while retaining the vital visual clues required for identity recognition. To achieve this, we craft a Frequency Channels Refining module that reduces sparsity in the frequency domain. It suppresses visual clues that could be exploited by reconstruction attacks, while preserving features indispensable for recognition, thus making these attacks more challenging. More significantly, we design a Frequency Fractal Mapping module that obfuscates deep representations by remapping refined frequency channels into a fractal-based privacy structure. By leveraging the self-similarity of fractals, this module enhances both recognition performance and defense strength, thereby significantly improving the overall robustness of the protection scheme. Experiments conducted on multiple public face recognition benchmarks demonstrate that the proposed FracFace significantly reduces the visual recoverability of facial features, while maintaining high recognition accuracy, as well as the superiorities over state-of-the-art privacy protection approaches.

1 Introduction

Face recognition (FR) leverages distinct facial features for biometric identification and is increasingly integrated into security applications such as mobile unlocking, access control, and border security. With the growing deployment of face recognition systems, privacy concerns have intensified, as intricate visual details in facial images may serve as rich clues for potential attackers. To tackle these issues, privacy-preserving face recognition (PPFR) has emerged as a solution. PPFR protects personal data while retaining the essential functionality of recognition systems by modifying facial data to prevent the reconstruction of the original image, thus balancing privacy with practical utility.

Existing PPFR schemes are generally classified into two categories: cryptographic approaches [13, 19, 23, 25, 8, 52], and non-cryptographic approaches [43, 27, 3, 14, 42, 21]. Cryptographic methods aim to secure facial data by encrypting features or performing recognition within the encrypted data. While these methods have a solid theoretical foundation, they are often hindered

[*] Corresponding Author. [1]The School of Cyber Science and Engineering, Sichuan University.

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

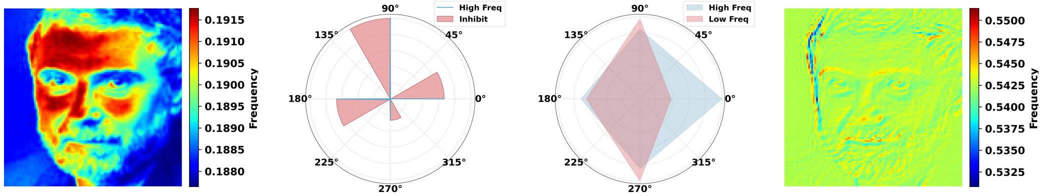
by high computational costs, limited data usability, and challenges in scaling effectively for real-world applications [44, 57]. Recent transformation-based, non-cryptographic PPFR methods achieve low latency and computational efficiency by suppressing facial visual details [29]. However, the inherent connection between identity and appearance features makes it challenging to balance recognition accuracy with privacy protection, often leaving residual visual clues that can be exploited in reconstruction attacks, with privacy risks remaining [15, 50]. To mitigate these issues, the precise selection of frequency domain channels has become essential. Existing approaches typically retain only those most strongly correlated with identity recognition, aiming to mitigate the suppression of high-frequency features by low-frequency channels and to emphasize the contribution of high-frequency components [28, 15]. However, we observe that the root of this issue lies in the inherent sparsity of the frequency domain, where identity-related information appears only in a few dominant components, sparsely scattered across both low and high frequency channels [20]. If this sparsity is not adequately addressed, there remains a risk of inadvertently leaving behind exploitable visual clues for potential attackers.

To overcome these challenges, we propose FracFace, a novel privacy-preserving face recognition framework built on two core components designed to refine frequency domain processing and disrupt potential visual cues. To address privacy issues arising from frequency domain sparsity and residual visual cue leakage, we present the Frequency Channel Refining (FCR) module, which selectively attenuates frequency bands to significantly diminish visual cues unrelated to identity while preserving key features essential for recognition. Furthermore, to enhance defense against reconstruction attacks, we propose the Frequency Fractal Mapping (FFM) module, which remaps the refined frequency representation to a fractal structure space, disrupting the continuity between spatial and frequency channels. By introducing structured perturbations rather than random noise, FFM fundamentally obfuscates the visual cues. The combined effect of these two modules reduces visual sparsity and enhances resilience against reconstruction attacks by generative networks, strengthening privacy protection. This paper makes the following contributions:

- First, we propose a novel fractal based framework for privacy preserving face recognition that disrupts spatial regularities in the frequency domain, thereby suppressing visual clues that can be exploited by reconstruction attacks.
- Second, we present a systematic frequency channel refining method that reduces sparsity and suppresses non-identity features, and introduce an innovative use of fractal structures to disrupt frequency continuity and enhance reconstruction resistance.
- Third, FracFace improves attack resistance by 15% to 60% under both white box and black box reconstruction scenarios compared to existing privacy preserving methods.

2 Related Works

Face Reconstruction Attacks. Face reconstruction attacks, especially those using deep learning, present a significant privacy challenge. These attacks are typically classified into optimization-based [36, 37], and deep learning-based methods[35, 34]. While optimization-based attacks iteratively refine inputs using feedback from face recognition systems, deep learning-based attacks improve efficiency by learning inverse mappings from facial features to images. U-Net [38] has become a widely used model for image reconstruction, attributed to its effective encoder-decoder structure and high reconstruction quality. Early studies by Zhmoginov et al. [56] and others utilized deep neural networks to convert facial embeddings into high-quality images, while more recent approaches like those by Cole et al. [4] and Dosovitskiy et al. [7] focused on generating images directly from these embeddings. In deep learning-based attacks, StyleGAN [17, 18] with its stylized generator architecture, enables precise control over facial feature reconstruction. StyleGAN's ability to generate face images nearly indistinguishable from the original poses significant privacy risks. Applications like TediGAN [47] and LAFITE [58] further enhance face generation and editing by manipulating the latent space, demonstrating high fidelity in reconstruction. Beyond GAN-based inversion, diffusion models [49, 16] provide powerful priors that enable adversaries to plausibly recover identity-bearing details. Specially, PGDiff [49] enforces high-quality structural and color priors during reverse diffusion, enabling robust reconstructions under complex or composite corruptions and thereby amplifying privacy risks. These advances highlight the growing efficiency of deep learning attacks and the pressing need for stronger privacy protection solutions.



(a) Unrefined Frequency Spectrum

(b) Frequency Channel Interference

(c) Refined Frequency Channel

(d) Refined Frequency Spectrum

Figure 1: Refining targeted frequency channels: An illustration

Privacy-preserving Face Recognition. FR has become a predominant biometric modality for identity authentication. While state-of-the-art methods [48, 6, 40] achieve remarkable accuracy, the increasing risk of privacy leakage has brought growing attention to privacy-preserving face recognition [28]. The development of deep neural networks [30] and generative adversarial models [7] has greatly exacerbated the risk of reconstruction attacks. Recent advances in PPFR have shown promising progress, with existing methods broadly categorized into cryptographic and non-cryptographic approaches. *1) Cryptographic techniques*, such as homomorphic encryption (HE) [12], differential privacy (DP) [33], secure multi-party computation (MPC) [1], and cancelable biometrics (CB) [5] offer robust theoretical privacy guarantees. However, their real-world applicability is hindered by substantial computational overheads, particularly when deployed on resource-constrained edge devices [27]. The substantial computational demands of these methods introduce significant delays, rendering them impractical for real-time face recognition systems where both speed and efficiency are critical [15]. Consequently, while these cryptographic solutions promise strong privacy protection, their practicality and scalability in real-world applications remain a significant challenge. *2) Non-cryptographic approaches*, especially transformation-based techniques, have been actively explored as lightweight alternatives to cryptographic schemes. These methods apply deliberate transformations in the image or feature domain to obscure sensitive attributes while preserving cues that are essential for distinguishing identities. To achieve privacy protection, Wang et al. [43] employed a channel-wise shuffling and mixing strategy in the frequency domain. Mi et al. [27] extended this view, emphasizing the complementary role of diverse frequency bands and advocating their joint use. Building on this, Mi et al. [28] randomly selected and filtered a subset of frequency-domain channels to reduce visual cues. Later, they proposed a subtraction-based method [29], which generates privacy-preserving face images by reconstructing frequency features from the residual between the original and generated frequency domains. Ji et al. [14] introduced a learnable privacy budget for adaptive trade-offs between privacy and utility, while Yuan et al. [50] devised an obfuscation method operating in the frequency domain that balances attribute suppression and visual interpretability. Most recently, Jin et al. [15] retained two key frequency channels and employed gradient-based reconstruction to synthesize structurally complex features resistant to inversion.

Nevertheless, despite their practical effectiveness, these methods remain vulnerable to advanced reconstruction attacks. Zhang et al. [51] suggested that many approaches rely on unrealistic privacy assumptions, which compromise their reliability. Shahreza et al. [32] further demonstrated that generative models can learn mappings from facial templates to latent spaces of pretrained generators, enabling high-resolution face reconstruction. In addition to this, our evaluations show that models like U-Net [38] and StyleGAN [18] can still recover realistic images even with partial frequency corruption (see Sec. 4.3 for more details). This stems from a core issue: although visual cues can be explicitly suppressed, the inherent sparsity of frequency-domain features often preserves global structures and local consistencies, allowing generative models to reconstruct the obscured content by exploiting residual information and frequency patterns (see Fig. 1 and Sec. 4.3 for more details).

3 Methodology

This section presents the technical design of FracFace, that fundamentally reduce privacy leakage in frequency-based face representations by targeting not only visual cues but also the structured patterns embedded in frequency embeddings. The overall pipeline is illustrated in Fig. 2.

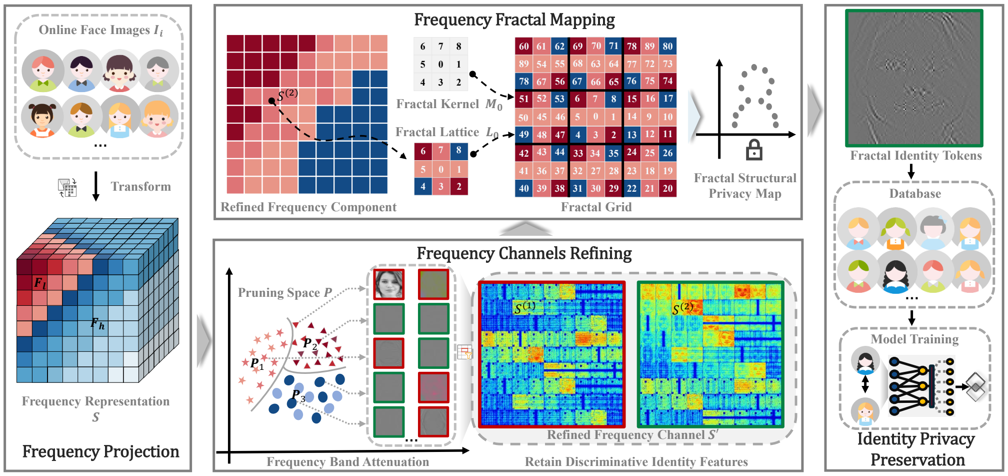


Figure 2: **The pipeline of proposed FracFace.** FracFace first projects input face images [3,H,W] into a frequency domain representation [192,H,W] via BDCT. Subsequently, band attenuation is applied to selectively suppress redundant frequency channels, obtaining the shape of [162, H, W]. These features are then refined through FCR to extract a privacy-protected subset of shape [81, H, W]. FFM is subsequently applied to disrupt spatial regularities, generating fractal identity tokens that are stored for recognition while enhancing resistance to reconstruction attacks.

3.1 Overview

Unlike previous work that mainly suppresses raw facial visual cues, FracFace introduces a novel transformation perspective that disrupts spatial regularities while preserving identity discriminative features. Such structural-level protection maintains reliable recognition performance and substantially mitigates the risk of visual reconstruction, thus achieving a strong balance between privacy preservation and recognition accuracy. We present two core modules. FCR *employs a refined selection of frequency domain channels to reduce feature sparsity and mitigate visual cue leakage, while preserving identity-discriminative information*. FFM *transforms frequency domain features into a fractal space characterized by self-similarity and local chaos, enhancing irreversibility without compromising recognizability*. These modules operate at two levels, channel selection and structural mapping, embedding frequency-domain features into a fractal-based framework to strike an optimal balance between privacy preservation and recognition accuracy under constrained channel conditions.

3.2 Frequency Projection

Given an input facial image $I \in \mathbb{R}^{3\times H\times W}$, we first normalize it into the $[0,1]$ range and apply an upsampling operation with a scaling factor r to obtain I'. The upsampled image I' is then transformed into the YCbCr color space, yielding $Y \in \mathbb{R}^{3\times rH\times rW}$, which better separates luminance and chrominance information. Next, Y is partitioned into non-overlapping 8×8 blocks, and a Block-based Discrete Cosine Transform (BDCT) is independently applied to each block, resulting in a localized frequency representation $S \in \mathbb{R}^{3\times 64\times \frac{rH}{8}\times \frac{rW}{8}}$. This transformation decomposes local spatial structures into 64 orthogonal frequency components, ordered from low to high frequency. For each channel, we denote the set of low-frequency components as $F_l \subset S$ and the high-frequency components as $F_h \subset S$, where $F_l \cup F_h = S$ and $F_l \cap F_h = \emptyset$. This decomposition into refined frequency bands lays the groundwork for subsequent privacy-preserving mechanisms.

3.3 Frequency Channels Refining

Frequency Band Attenuation After mapping the image to the frequency domain, the image features exhibit significant sparsity. As illustrated in Fig. 1a, the low-frequency channels contain the majority of the energy and visual information. However, this also results in components associated with low frequencies retaining substantial characteristics of the lower frequency bands, thereby suppressing the contribution of channels in the higher frequency range, as shown in Fig. 1b. To address this issue, we propose the Frequency Band Attenuation (FBA) mechanism, which selectively weakens the contribution of channels corresponding to low frequencies and reshapes the frequency features to better balance privacy protection and identity representation. Specifically, we adopt a three level pruning strategy, where frequency channels are grouped into sets P_1, P_2, P_3, each corresponding to a distinct frequency range. The final pruning space is defined as $P = P_1 \cup P_2 \cup P_3$. Next, we perform the channel removal operation on the frequency domain feature S, resulting in a refined frequency channel set $S' = \text{Prune}(S, P)$, where $S' \in \mathbb{R}^{3\times 162\times \frac{rH}{8}\times\frac{rW}{8}}$ represents the feature tensor after frequency band attenuation, with redundant frequency components removed, effectively reducing the risk of privacy leakage. The channel selection removes those associated with sensitive visual information, such as skin color, lighting, and expressions, while retaining those crucial for identity recognition with low visual sensitivity to ensure accuracy. After applying the FCR mechanism, as shown in Fig. 1c, the refined low-frequency channels no longer suppress high-frequency contributions, and instead, as seen in Fig. 1d, each channel independently highlights its unique contribution, leading to a more balanced and robust feature representation. We also utilize t-SNE [24] to evaluate the sparsity of frequency domain channels (see detailed analysis refer to Appendix A.1). Unlike Mi et al.'s PartialFace [28], our frequency band attenuation strategy balances identity recognizability and privacy protection, addressing frequency sparsity and weakening visual cues exploitable by reconstruction attacks.

Retain Discriminative Identity Features To further refine the frequency band structure, the attenuated frequency domain channels are partitioned into two groups: $(S^{(1)}, S^{(2)}) = \text{index}(S')$, where $S^{(1)}, S^{(2)} \in \mathbb{R}^{3\times 81\times \frac{rH}{8}\times\frac{rW}{8}}$. This strategy allows the assignment of different priorities to the frequency bands, enabling more granular privacy control (see Algorithm 1 for details).

3.4 Frequency Fractal Mapping

In our proposed FFM mechanism, the fractal kernel M_0 and the fractal lattice L_0 are randomly initialized, serving as the foundational components for subsequent fractal transformations. The fractal kernel M_0 and the fractal lattice L_0, both of size $m \times n$, are initialized such that M_0 is populated with random integers drawn from the range $[1, m \times n]$, introducing structural diversity, while L_0 defines the relative indexing order of the elements within the fractal kernel M_0. The initial fractal grid $\mathcal{F}_0$ is directly set as M_0, thus initializing the fractal mapping. The fractal transformation proceeds iteratively, where at each step, a new fractal mapping $\mathcal{F}_k$ is constructed by combining the previous mapping $\mathcal{F}_{k-1}$ with the foundational mapping M_0, scaled by a factor $\beta_k = 3^{2k}$. This scaling factor controls the expansion of sub-blocks within the fractal structure, and each element of the new mapping is updated as:

$$\mathcal{F}_k[i,j] = (M_0[i,j] - 1)\cdot \beta_k + \mathcal{F}_{k-1}[i,j], \tag{1}$$

where i and j index the mapping elements. This iterative process progressively refines the fractal structure, with each layer increasing in both size and complexity. To maintain structural coherence across scales, a correction $\mathcal{F}_k = \mathcal{F}_k - 1$ is applied at certain layers to mitigate misalignments introduced by scaling. The process continues for a predefined number of iterations, ultimately yielding a set of fractal mappings $\mathcal{F} = [\mathcal{F}_0, \mathcal{F}_1, \ldots, \mathcal{F}_{n-1}]$. The FFM constructs a multi-layered and complex fractal structure that effectively mitigates sparsity while preserving essential identity information. Building upon the recursive fractal construction, the proposed FFM forms a non-invertible, nonlinear index transformation through iterative integer-based perturbations, as formally defined below. Let $M_0 \in \mathbb{Z}^{n\times n}$ denote the initial index matrix. We define the k-th layer fractal mapping recursively as:

$$\mathcal{F}^{[k]} = M_0 + \sum_{i=1}^{k}(b_i - 1)\cdot\beta_i, \quad \beta_i = \prod_{j=1}^{i-1} b_j, \quad \beta_1 = 1, \tag{2}$$

where b_i is the expansion factor at the i-th fractal layer, and $\mathcal{F}^{[k]}$ denotes the positional encoding mapping for channel reordering. This recursive structure highlights the layered composition of FFM, in which discrete integer perturbations are progressively accumulated in a nonlinear fashion

Table 1: The performance of privacy protection methods in terms of face recognition accuracy. The space-time domain is denoted by S, the frequency domain by F_1, and the fractal domain by F_2, respectively. Green indicates the proportion at the given privacy-protection level (see Appendix A.3).

Method	LFW	CelebA	AgeDB	CFP-FP	CALFW	CPLFW	IJB-B		IJB-C		Domain	Protection	Venue
	(%)	(%)	(%)	(%)	(%)	(%)	1e-4	1e-5	1e-4	1e-5			
Arcface [6]	99.73	95.35	97.99	96.83	**95.89**	94.59	**94.81**	**91.98**	93.69	92.41	S	0%	CVPR-2019
Arcface-FD [48]	99.81	96.45	**98.27**	97.18	94.69	95.03	93.68	90.53	95.89	94.92	S	0%	CVPR-2020
PPFR-FD [43]	99.39	93.49	97.99	95.53	95.69	90.62	93.67	91.12	94.73	92.49	F_1	43%	AAAI-2022
Duetface [27]	99.81	92.13	96.17	93.24	95.18	92.19	92.63	90.32	95.28	94.16	F_1	45%	ACMMM-2022
PartialFace [28]	**99.82**	95.64	95.03	98.10	94.83	**95.61**	92.48	91.59	93.85	93.96	F_1	68%	ICCV-2023
Minusface [29]	99.79	95.89	96.03	96.94	95.93	92.89	93.89	93.51	**95.91**	**94.96**	F_1	85%	CVPR-2024
PRO-Face C [50]	99.29	91.69	93.79	95.63	89.44	90.65	88.38	83.27	90.89	89.94	F_1	40%	IEEE TIFS-2024
FaceObfuscator [15]	99.70	94.36	96.79	**98.82**	94.84	95.42	92.90	92.18	94.43	93.58	F_1	87%	USENIX-2024
FracFace (ours)	99.69	**95.91**	97.76	96.14	93.92	93.16	92.42	90.73	94.09	92.26	$\boldsymbol{S \rightarrow F_1 \rightarrow F_2}$	100%	NeruIPS-2025

through scaling and base multiplication. To establish the nonlinearity of the FFM transformation, we demonstrate in Appendix A.2 that it does not satisfy the fundamental properties of linearity, namely homogeneity and additivity, and the irreversibility. (see the Algorithm 2 for implementation details).

3.5 Identity Privacy Preservation

We present a privacy-preserving identity recognition framework where facial images are first processed through FFM and uploaded as irreversible obfuscated features. These features are structured via candidate feature sets to ensure uniqueness and non-replicability. While retaining sufficient identity-related cues for recognition, they are mathematically and structurally resistant to inversion, thus preventing facial image reconstruction even under full adversarial access. This design balances high recognition accuracy with strong privacy guarantees.

4 Experiments

4.1 Experimental Setup

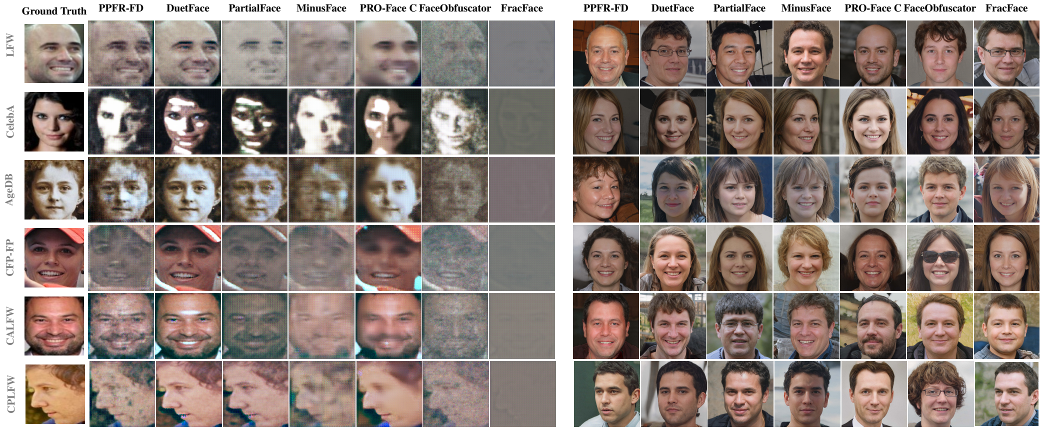
Models and Datasets To evaluate the FracFace[1] framework, we utilize various models and datasets. For recognition, we adopt the IR-50 [10] backbone, which offers a favorable trade-off between compactness and accuracy. The model is trained on the MS1Mv2 [9] dataset, which is widespread adoption as a standard benchmark in face recognition ensures fair and consistent comparisons with prior work [27], [29], [28]. To assess privacy robustness, we employ three of the deep learning-based adversaries: a lightweight U-Net [38] autoencoder for reconstruction-based attacks, a StyleGAN [18] generator for generative attacks, and a PGDiff[49] generator based on reverse diffusion process. These attackers rigorously test the irreversibility and security of FracFace transformations. We evaluate on standard benchmarks, including LFW [11], CelebA [22], AgeDB [31], CFP-FP [39], CALFW [55], CPLFW [54], IJB-B [46], and IJB-C [26].

Evaluation Metrics To evaluate the privacy-utility trade-off of FracFace, we employ five metrics. Pixel-level differences between reconstructed and original images are quantified by SSIM [45], MSE, PSNR, LPIPS [53], and IDS [41] where lower SSIM, higher MSE, higher LPIPS, and lower IDS indicate stronger privacy protection.These metrics comprehensively assess the effectiveness of FracFace in safeguarding privacy while maintaining utility. In addition, refer to Appendix A.7 for more details on the experiment.

4.2 Recognition Accuracy

We evaluate FracFace in comparison with two widely adopted baselines, ArcFace[6] and ArcFace-FD[48] (both without privacy protection), as well as six representative privacy-preserving methods. Despite introducing privacy protection, FracFace achieves recognition accuracy comparable to the baselines while significantly enhancing identity security. Tab. 1 reveal that while frequency domain methods like PPFR-FD and DuetFace incur 3% - 5% accuracy loss, approaches such as MinusFace, FaceObfuscator, and PartialFace still retain 13%, 15%, and 32% identity information, indicating notable privacy risks (refer to Sec. 4.3 for more details). In contrast, FracFace employs FCR and FFM

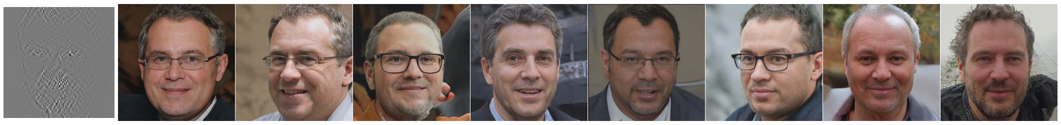
[1] Code is available at https://anonymous.4open.science/r/FracFace.



(a) U-Net-based Attacks

(b) StyleGAN-based Attacks

Figure 3: Evaluation of facial reconstruction vulnerabilities under U-Net and StyleGAN attacks.



(a) Target (b) $\mathcal{G}$ 1 (c) $\mathcal{G}$ 2 (d) $\mathcal{G}$ 3 (e) $\mathcal{G}$ 4 (f) $\mathcal{G}$ 5 (g) $\mathcal{G}$ 6 (h) $\mathcal{G}$ 7 (i) $\mathcal{G}$ 8

Figure 4: Comparative attack results on CelebA targets with StyleGAN. (a) shows the target image of the adversary attack; (b)-(i) shows the attack outputs.

to restructure the feature space, effectively suppressing low-frequency dominance while promoting the contribution of high-frequency components to identity representation. This design delivers strong privacy protection with only a marginal drop in accuracy compared to ArcFace-FD, achieving state-of-the-art privacy performance and an optimal trade-off between utility and security.

4.3 Empirical Evaluation of Privacy Protection

4.3.1 Effectiveness

Reconstruction Vulnerability Assessment In practical attacks, facial data leakage typically enables two reconstruction paths: (1) mapping protected features to original images via models like U-Net when identity embeddings are exposed; (2) leveraging generative models (e.g., StyleGAN and PGDiff) to reconstruct identities from intercepted features without requiring an explicit degradation model. We evaluate the robustness of FracFace against reconstruction attacks on six benchmark datasets and compare its performance with several representative privacy-preserving methods. As illustrated in Fig. 3, most existing approaches exhibit limited resistance to reconstruction attacks. For instance, PPFR-FD, Pro-Face, and DuetFace lack dedicated defense mechanisms, allowing adversaries to recover significant facial details from identity tokens. We further employ StyleGAN to reconstruct faces from these identity features and observe that PPFR-FD and Pro-Face C tend to preserve visually identifiable traits like hairstyle, while DuetFace enables more accurate reconstruction of facial geometry. Although PartialFace, MinusFace, and FaceObfuscator demonstrate stronger privacy protection, StyleGAN is still able to extract structural cues, resulting in reconstructions that retain notable similarity to the original identities in terms of facial proportions. In contrast, reconstructions from FracFace-preserved representations exhibit no clear identity similarity. Building on FCR, this outcome benefits from the local obfuscation introduced by FFM. It is further supported by quantitative results from our subsequent privacy reconstruction evaluation (see Tab. 2 for more details).

Robustness to StyleGAN We evaluate the reconstruction robustness of FracFace under adversarial attacks driven by StyleGAN, as illustrated in Fig. 4 and Fig. 5. Given limited identity tokens (Fig. 4a), the attacker employs StyleGAN to synthesize candidate faces (Fig. 4b–Fig. 4i), selecting the most similar ones as style vectors to enhance further attempts. Fig. 5 details the reconstruction process. While StyleGAN can reconstruct global attributes such as pose, hairstyle, and beard, it fails to

Table 2: Benchmarking the privacy utility tradeoff under U-Net and StyleGAN based reconstruction attacks, evaluated by SSIM, LPIPS, MSE, PSNR, and IDS across two leakage scenarios.

Metric	Method	U-Net-based Face Reconstruction Attack						StyleGAN-based Face Reconstruction Attack					
		LFW	CelebA	AgeDB	CFP-FP	CALFW	CPLFW	LFW	CelebA	AgeDB	CFP-FP	CALFW	CPLFW
SSIM ↓	Arcface	0.9642	0.9883	0.9436	0.9351	0.9555	0.9188	0.9906	0.9827	0.9548	0.9820	0.9408	0.9307
	Arcface-FD	0.9623	0.9097	0.9242	0.9172	0.9544	0.9021	0.9761	0.9438	0.9342	0.9411	0.9618	0.9461
	PPFR-FD	0.4231	0.1488	0.3079	0.3720	0.3557	0.5204	0.3116	0.2681	0.2404	0.3523	0.3323	0.1696
	Duetface	0.4963	0.2570	0.4280	0.4965	0.5158	0.4249	0.3367	0.2303	0.2873	0.2397	0.3245	0.1663
	PartialFace	0.4953	0.2927	0.2404	0.3592	0.3715	0.4423	0.2845	0.2683	0.2544	0.3314	0.2487	0.1729
	Minusface	0.3864	**0.1461**	0.2319	0.3407	**0.2878**	0.3263	0.3421	0.2266	0.2409	0.2962	0.2587	0.1699
	PRO-Face C	0.5517	0.3684	0.5135	0.4737	0.4665	0.4685	0.3535	0.2676	0.2339	0.3217	0.2946	0.1656
	FaceObfuscator	**0.3771**	0.1984	0.3477	0.3428	0.3468	**0.3189**	0.3654	0.2585	0.2882	0.2960	0.2595	0.1395
	FracFace (ours)	0.3997	0.2195	**0.2045**	**0.2749**	0.3317	0.4357	**0.2836**	**0.2019**	**0.2278**	**0.2264**	**0.2305**	**0.1045**
LPIPS ↑	Arcface	0.0141	0.0708	0.0436	0.0330	0.0301	0.0824	0.0163	0.0192	0.0139	0.0117	0.0131	0.0167
	Arcface-FD	0.0175	0.0676	0.0418	0.0487	0.0312	0.0831	0.0180	0.0127	0.0133	0.0128	0.0133	0.0172
	PPFR-FD	0.5433	0.4522	0.5198	0.5430	0.6683	0.5059	0.7206	0.5443	0.6596	0.6022	0.5910	0.6916
	Duetface	0.5264	0.4350	0.3328	0.3442	0.3458	0.4249	0.7378	0.5461	0.6842	0.6007	0.6412	0.6208
	PartialFace	0.5197	0.5056	0.6536	0.5592	0.6952	0.6502	0.7558	0.5652	0.6733	0.5715	0.6604	0.6911
	Minusface	0.6809	0.6790	0.6607	0.6720	0.6305	0.6675	0.7313	0.5894	0.6732	0.6322	0.7253	0.6768
	PRO-Face C	0.5018	0.4341	0.4173	0.4812	0.4645	0.5403	0.7091	0.6004	0.6522	0.5749	0.6335	0.6719
	FaceObfuscator	0.6512	0.6289	0.5790	0.6332	0.6364	0.6012	0.7419	0.5320	0.6891	0.6218	0.6614	0.6535
	FracFace (ours)	**0.6907**	**0.6834**	**0.7389**	**0.7796**	**0.6958**	**0.6990**	**0.8307**	**0.6354**	**0.6935**	**0.6412**	**0.6655**	**0.6935**
MSE ↑	Arcface	0.0002	0.0058	0.0001	0.0015	0.0012	0.0021	0.0011	0.0018	0.0021	0.0029	0.0016	0.0014
	Arcface-FD	0.0002	0.0054	0.001	0.0030	0.0012	0.0024	0.0014	0.0024	0.0025	0.0023	0.0019	0.0021
	PPFR-FD	0.0170	0.0453	0.0390	0.0475	0.0372	0.0164	0.0514	0.0466	0.0340	0.0613	0.0686	0.0462
	Duetface	0.0249	0.0474	0.0253	0.0235	0.0224	0.0263	0.0621	0.0583	0.0452	0.0645	0.0631	0.0728
	PartialFace	0.0251	0.0415	0.0872	0.0389	0.0532	0.042	0.0156	0.0613	0.0695	0.0637	0.0793	0.0405
	Minusface	0.0619	0.0425	0.0754	0.0549	0.0537	0.0418	0.0729	0.0675	0.0711	0.0646	0.0593	0.0637
	PRO-Face C	0.0018	0.0256	0.0127	0.0209	0.0149	0.0171	0.0567	0.0480	0.0633	0.0583	0.0635	0.0495
	FaceObfuscator	0.0418	0.0466	0.0578	0.0512	0.0409	0.0545	0.0769	0.0795	0.0635	0.0698	0.0646	0.0794
	FracFace (ours)	**0.0921**	**0.0839**	**0.1694**	**0.0855**	**0.0591**	**0.0643**	**0.0869**	**0.0993**	**0.0909**	**0.0750**	**0.0753**	**0.0831**
PSNR ↓	Arcface	28.3762	26.3394	28.0351	28.1658	28.9046	26.6827	27.3627	23.9351	29.9237	26.5816	20.3843	25.7364
	Arcface-FD	26.1864	26.5831	28.5249	27.9832	25.1352	26.1258	28.2943	25.1971	29.5851	27.6278	23.4935	27.1539
	PPFR-FD	16.6922	15.2175	14.0937	13.2350	14.3056	17.8451	10.6539	13.3151	10.7322	11.4767	11.6374	10.3582
	Duetface	16.0382	14.2463	13.1962	16.2930	16.1359	15.7981	10.3639	11.0696	10.0926	10.7314	11.9891	11.6206
	PartialFace	13.0542	10.9401	10.5918	12.9035	12.7318	13.7378	9.9318	12.1278	10.1875	11.9616	11.3478	10.5448
	Minusface	11.2158	**9.3362**	11.2309	12.0672	11.9623	13.7816	11.3744	11.7088	11.4857	10.7283	12.2753	10.1364
	PRO-Face C	17.8753	15.9239	18.9573	16.7931	16.2451	16.7569	10.1446	11.6173	11.4588	10.9503	11.9773	11.3907
	FaceObfuscator	10.3351	10.9748	12.3641	10.6841	**10.7845**	12.6315	10.6150	10.7325	10.1433	10.6668	11.9029	10.9551
	FracFace (ours)	**8.6099**	9.9682	**9.5171**	**10.0953**	11.7843	**10.0827**	**9.7742**	**10.0369**	**10.0421**	**10.0562**	**10.9270**	**10.1239**
IDS ↓	Arcface	0.9932	0.9966	0.9989	0.9904	0.9928	0.9991	0.9968	0.9834	0.9910	0.8973	0.9627	0.9624
	Arcface-FD	0.9927	0.9939	0.9969	0.9918	0.9919	0.9982	0.9915	0.9620	0.9837	0.8993	0.9639	0.9728
	PPFR-FD	0.5699	0.6549	0.8402	0.8829	0.7968	0.6982	0.7587	0.8319	0.6915	0.6512	0.8250	0.6983
	Duetface	0.5830	0.6172	0.7921	0.8786	0.6217	0.6826	0.7388	0.8239	0.6074	0.6288	0.8116	0.6799
	PartialFace	0.4670	0.4572	0.7308	0.6353	0.5384	0.5204	0.7317	0.8043	0.6391	0.5962	0.7298	0.6194
	Minusface	0.3946	0.4007	0.4147	0.5428	0.4218	0.2600	0.7267	0.8062	0.5950	0.5649	0.7329	0.5481
	PRO-Face C	0.4124	0.5028	0.7978	0.7548	0.8325	0.7271	0.8299	0.8501	0.7064	0.6294	0.8562	0.7158
	FaceObfuscator	0.3830	0.4094	0.4972	0.4565	0.46247	0.3412	0.7187	0.7374	0.5785	0.6218	0.7253	0.6101
	FracFace (ours)	**0.0057**	**0.0081**	**0.0003**	**0.0011**	**0.0018**	**0.0024**	**0.6705**	**0.7334**	**0.5242**	**0.5239**	**0.6150**	**0.5458**

faithfully recover fine grained structures like the eyes, nose, and mouth, leading to perceptible distortions in identity. This confirms that FracFace effectively disrupts the coherence of identity-critical features through FFM, offering strong resistance to clue and style-based inversion attacks. A more detailed analysis of how FCR and FFM influence the visualization of frequency domain channels is provided in Appendix A.4.

Robustness to Diffusion Model We evaluate FracFace under an adaptive white-box threat model against diffusion-based reconstruction (Fig. 6). Using a PGDiff model pretrained on the public CelebA dataset, the reconstructions in Figs. 6b to 6i recover only coarse attributes (e.g., pose, hairstyle) while failing to reproduce identity-critical details (eyes, nose, mouth) and they exhibit noticeable texture artifacts. These results indicate that the FCR module disperses energy and reduces sparsity in low-frequency channels, whereas FFM introduces localized perturbations that confound mid/high-frequency cues. Consequently, FracFace impedes faithful inversion, retaining only coarse attributes and thereby enhancing privacy.

Quantitative Comparison To comprehensively evaluate the privacy-utility tradeoff of our method under strong reconstruction threats, we conducted experiments on six public face recognition benchmarks, as summarized in Tab. 2. FracFace was compared with six state-of-the-art face privacy protection methods under two representative reconstruction attacks: a discriminative model based on U-Net and a generative model based on StyleGAN. Evaluation was conducted using SSIM, LPIPS,

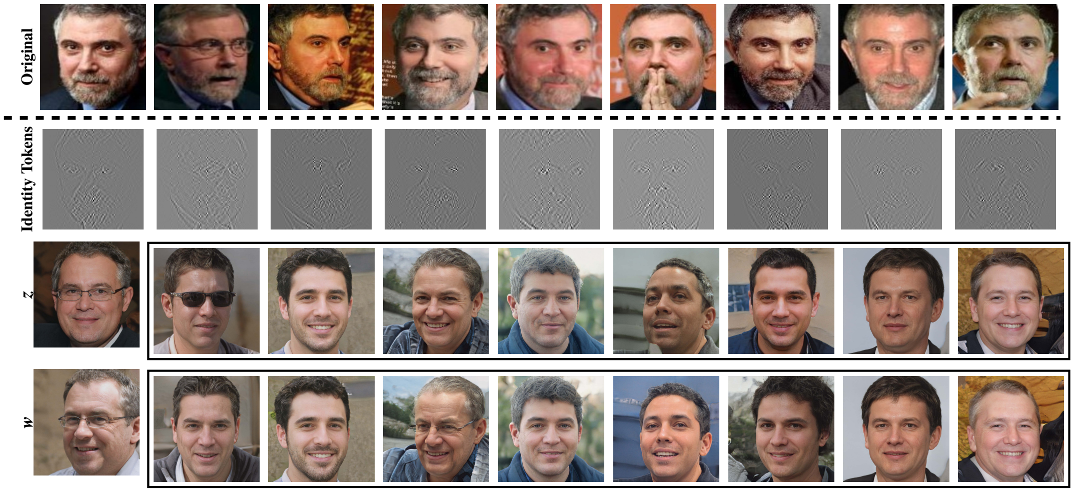
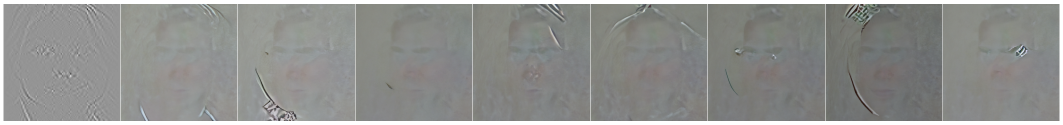


Figure 5: Visual analysis of StyleGAN vulnerabilities.



(a) Target (b) $\mathcal{D}_1$ (c) $\mathcal{D}_2$ (d) $\mathcal{D}_3$ (e) $\mathcal{D}_4$ (f) $\mathcal{D}_5$ (g) $\mathcal{D}_6$ (h) $\mathcal{D}_7$ (i) $\mathcal{D}_8$

Figure 6: Comparative reconstruction results on CelebA targets with PGDiff. (a) shows the target image of the adversary attack; (b)-(i) shows the reconstruction outputs.

MSE, PSNR, and identity similarity. FracFace consistently demonstrates stronger robustness across both attack types. It achieves significantly lower SSIM values, indicating reduced structural similarity between protected and reconstructed faces. Under StyleGAN attacks, FracFace attains SSIM scores as low as 0.2264 on CFP-FP and 0.1045 on CPLFW, suggesting the original structure is effectively obfuscated. In terms of perceptual dissimilarity, FracFace achieves the highest LPIPS scores, reaching 0.8307 on LFW and 0.7389 on AgeDB. Compared to structurally-aware baselines such as PartialFace, FaceObfuscator, and MinusFace, FracFace achieves superior performance, particularly under challenging generative attacks that often exploit residual structural cues. This strength stems from the synergy between our FCR and FFM modules, which jointly mitigate frequency domain sparsity and limit the reconstructive capacity of generative models.

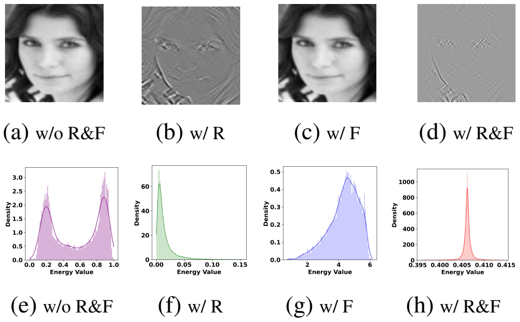
4.4 Ablation Study

4.4.1 Effectiveness of FCR and FFM

This section conducts ablation studies on the FCR and FFM modules in FracFace to evaluate their respective and combined impacts on recognition accuracy and privacy preservation. As shown in Fig. 7, when the two modules exist independently (as shown in Fig. 7b, Fig.7c, Fig. 7f and Fig. 7g), the obtained frequency histogram shows obvious sparsity, which exposes the system to potential statistical attacks. Introducing either FCR or FFM alone fails to sufficiently mitigate this sparsity, as the histograms still reveal distinct, reconstructable patterns. By contrast, integrating both modules leads to a significantly more compact and uniform frequency distribution, effectively reducing statistical leakage and enhancing privacy robustness, see Fig. 7d and Fig. 7h. As shown in Tab. 3, while applying no module or only FFM preserves high recognition accuracy, it offers little privacy protection due to insufficient disruption of frequency channels. FCR alone enhances privacy but degrades accuracy. It is the joint application of FCR and FFM that ensures both improved recognition performance and robust privacy protection, achieving a balanced compromise between accuracy and security. This indicates that FCR and FFM are capable of jointly resisting reconstruction attacks, while leveraging fractal self similarity and local confusion to preserve features relevant to identity.

4.4.2 Effectiveness of Fractal Depth k

To examine how the fractal depth k in FFM affects privacy and recognition accuracy, we performed a systematic ablation across varying k values. As shown in Tab. 4 (in Appendix A.9), larger k values lead to sustained and empirically consistent privacy gains (LPIPS rising from 0.5291 to 0.8357 and SSIM falling from 0.5227 to 0.2580), consistent with stronger obfuscation; however, recognition accuracy progressively deteriorates beyond $k = 2$. In particular, $k = 2$ emerges as the optimal operating point, achieving a 20% improvement in LPIPS relative to $k = 1$ with only a 0.02% loss in accuracy, thereby preserving near-perfect recognition while materially strengthening privacy. Accordingly, we employ $k = 2$ by default and view it as the sweet spot, beyond which the incremental privacy improvement is minor relative to the escalating drop in utility.



(a) w/o R&F (b) w/ R (c) w/ F (d) w/ R&F

(e) w/o R&F (f) w/ R (g) w/ F (h) w/ R&F

Figure 7: Visual comparisons on the impact of modules FCR (R) and FFM (F).

Table 3: Ablation study on the joint effect of FCR and FFM on recognition accuracy

Method			**LFW**	**AgeDB**
FCR	FFM	Protection		
✗	✗	○	99.71	97.72
✗	✓	○	99.30	97.28
✓	✗	◐	84.53	76.84
✓	✓	●	99.59	96.35

4.4.3 The Sensitivity of FBA Strength

We quantified how frequency-domain attenuation shapes the privacy–utility trade-off by performing a sensitivity analysis on the FBA attenuation strength (see Tab. 5 in Appendix A.9). We observe that as refinement strength increases, with more channels removed, privacy consistently improves (e.g., LPIPS: 0.32 → 0.86), whereas recognition utility declines, with a pronounced drop beyond 50%. Notably, 50% pruning yields the best balance, increasing LPIPS by 35% over the 20% case while maintaining near-perfect recognition accuracy. Beyond that point, privacy remains approximately constant, whereas accuracy continues to decrease. These results indicate that moderate refining (40-50%) constitutes an effective operating range, improving privacy with minimal impact on utility.

5 Conclusion

This work introduces FracFace, a novel privacy preserving face recognition framework designed to combat the critical threat of reconstruction attacks by targeting the vulnerability of visual cues in the frequency domain. Specifically, FCR disrupts spatial continuity (sparsity) and selectively refines identity relevant frequency channels, allowing FracFace to effectively reduce visual cues susceptible to reconstruction. Additionally, the FFM remaps optimized frequency features into a complex fractal structure space, substantially complicating reverse recovery. By decoupling identity from easily exploitable structural cues, FracFace weakens the implicit mapping between visual information and identity that many attacks rely on. Extensive evaluations on public benchmarks demonstrate that FracFace not only disrupts visual recoverability but also maintains high recognition accuracy, establishing it as an effective and secure solution for privacy sensitive face recognition applications. These findings suggest that fractal-guided frequency transformation may offer a viable path toward reconciling security with interpretability in future face recognition systems.

6 Broader Impacts

Notably, none of the datasets used in this work involve private or non-consented data collection. Licensing terms and access conditions for each dataset are documented in [2] (please refer to this Supplementary Material for details). Our work centers on privacy-preserving face recognition: we transform face images into frequency-domain representations that are not trivially invertible and empirically reduce identity leakage. Therefore, the proposed method does not target or infer sensitive attributes, and we are aware of no negative societal impacts within the scope of this research.

Acknowledgements

This research is supported by the National Natural Science Foundation of China (Grant No. 62372313); the Institutional Research Fund of Sichuan University (No. 2024SCUQTTX034); the China Scholarship Council (Award No. 202406240205); and the National Research Foundation, Singapore, under the programme "Enhancing Large Language Models with Rigorous Reasoning" (Award No. A-8003116-00-00 NAII). We express our deep gratitude to the program chairs and reviewers for their recognition of, and dedication to, our work. We also thank the NeurIPS committee for granting us the NeurIPS 2025 Scholar Award, which supported our participation in the conference.

References

[1] Renwan Bi, Jinbo Xiong, Changqing Luo, Jianting Ning, Ximeng Liu, Youliang Tian, and Yan Zhang. Communication-efficient privacy-preserving neural network inference via arithmetic secret sharing. *IEEE Transactions on Information Forensics and Security*, 19:6722–6737, 2024.

[2] Fadi Boutros, Meiling Fang, Marcel Klemt, Biying Fu, and Naser Damer. Cr-fiqa: face image quality assessment by learning sample relative classifiability. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5836–5845, 2023.

[3] M.A.P. Chamikara, P. Bertok, I. Khalil, D. Liu, and S. Camtepe. Privacy preserving face recognition utilizing differential privacy. *Computers & Security*, 97:101951, 2020.

[4] Forrester Cole, David Belanger, Dilip Krishnan, Aaron Sarna, Inbar Mosseri, and William T. Freeman. Synthesizing normalized faces from facial identity features. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3386–3395, 2017.

[5] Wanying Dai, Beibei Li, Qingyun Du, Ziqing Zhu, and Ao Liu. Chaos-based index-of-min hashing scheme for cancellable biometrics security. *IEEE Transactions on Information Forensics and Security*, 19:8982–8997, 2024.

[6] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4685–4694, 2019.

[7] Alexey Dosovitskiy and Thomas Brox. Inverting visual representations with convolutional networks. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4829–4837, 2016.

[8] Ovgu Ozturk Ergun. Privacy preserving face recognition in encrypted domain. In *Proc. IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*, pages 643–646, 2014.

[9] Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In *Proc. European Conference on Computer Vision (ECCV)*, pages 87–102. Springer, 2016.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[11] Gary B Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In *Proc. Workshop on Faces in Real-Life Images: Detection, Alignment, and Recognition*, 2008.

[12] Hai Huang and Luyao Wang. Efficient privacy-preserving face identification protocol. *IEEE Transactions on Services Computing*, 16(4):2632–2641, 2023.

[13] Yangsibo Huang, Zhao Song, Kai Li, and Sanjeev Arora. Instahide: Instance-hiding schemes for private distributed learning. In *Proc. International Conference on Machine Learning (ICML)*, pages 4507–4518, 2020.

[14] 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 *Proc. European Conference on Computer Vision (ECCV)*, pages 475–491. Springer, 2022.

[15] Shuaifan Jin, He Wang, Zhibo Wang, Feng Xiao, Jiahui Hu, Yuan He, Wenwen Zhang, Zhongjie Ba, Weijie Fang, Shuhong Yuan, and Kui Ren. FaceObfuscator: Defending deep learning-based privacy attacks with gradient descent-resistant features in face recognition. In *Proc. 33rd USENIX Security Symposium (USENIX Security)*, pages 6849–6866, 2024.

[16] Tero Karras, Miika Aittala, Tuomas Kynkäänniemi, Jaakko Lehtinen, Timo Aila, and Samuli Laine. Guiding a diffusion model with a bad version of itself. *Advances in Neural Information Processing Systems (NeurIPS)*, 37:52996–53021, 2024.

[17] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4401–4410, 2019.

[18] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8107–8116, 2020.

[19] Lamyanba Laishram, Muhammad Shaheryar, Jong Taek Lee, and Soon Ki Jung. Toward a privacy-preserving face recognition system: A survey of leakages and solutions. *ACM Computing Surveys*, 57(6):1–38, 2025.

[20] Florian Lieb and Hans-Georg Stark. Audio inpainting: Evaluation of time-frequency representations and structured sparsity approaches. *Signal Processing*, 153, 2018.

[21] Shuofeng Liu, Mengyao Ma, Minhui Xue, and Guangdong Bai. Modifier unlocked: Jailbreaking text-to-image models through prompts. In *Proc. IEEE Symposium on Security and Privacy (SP)*, pages 355–372, 2025.

[22] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proc. of the IEEE International Conference on Computer Vision (ICCV)*, pages 3730–3738, 2015.

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

[24] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9:2579–2605, 2008.

[25] Guangcan Mai, Kai Cao, Xiangyuan Lan, and Pong C. Yuen. Secureface: Face template protection. *IEEE Transactions on Information Forensics and Security*, 16:262–277, 2021.

[26] Brianna Maze, Jocelyn Adams, James A Duncan, Nathan Kalka, Tim Miller, Charles Otto, Anil K Jain, W Tyler Niggel, Janet Anderson, Jordan Cheney, et al. Iarpa janus benchmark-c: Face dataset and protocol. In *Proc. International Conference on Biometrics (ICB)*, pages 158–165, 2018.

[27] 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 *Proc. ACM International Conference on Multimedia*, page 6755–6764, 2022.

[28] 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 *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 19616–19627, 2023.

[29] Yuxi Mi, Zhizhou Zhong, Yuge Huang, Jiazhen Ji, Jianqing Xu, Jun Wang, Shaoming Wang, Shouhong Ding, and Shuigeng Zhou. Privacy-preserving face recognition using trainable feature subtraction. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 297–307, 2024.

[30] Pranab Mohanty, Sudeep Sarkar, and Rangachar Kasturi. From scores to face templates: A model-based approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(12):2065–2078, 2007.

[31] Stylianos Moschoglou, Athanasios Papaioannou, Christos Sagonas, Jiankang Deng, Irene Kotsia, and Stefanos Zafeiriou. Agedb: The first manually collected, in-the-wild age database. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1997–2005, 2017.

[32] Hatef Otroshi Shahreza and Sébastien Marcel. Face reconstruction from facial templates by learning latent space of a generator network. In *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, pages 12703–12720, 2023.

[33] Lu Ou, Shaolin Liao, Shihui Gao, Guandong Huang, and Zheng Qin. Rdp: Ranked differential privacy for facial feature protection in multi-scale sparsified subspaces. *IEEE Internet of Things Journal*, 2025.

[34] Binhang Qi, Hailong Sun, Xiang Gao, Hongyu Zhang, Zhaotian Li, and Xudong Liu. Reusing deep neural network models through model re-engineering. In *Proc. IEEE/ACM International Conference on Software Engineering (ICSE)*, pages 983–994, 2023.

[35] Binhang Qi, Hailong Sun, Hongyu Zhang, and Xiang Gao. Reusing convolutional neural network models through modularization and composition. *ACM Trans. Softw. Eng. Methodol.*, 33(3), 2024.

[36] Anton Razzhigaev, Klim Kireev, Edgar Kaziakhmedov, Nurislam Tursynbek, and Aleksandr Petiushko. Black-box face recovery from identity features. In *Proc. European Conference on Computer Vision (ECCV)*, pages 462–475. Springer, 2020.

[37] Anton Razzhigaev, Klim Kireev, Igor Udovichenko, and Aleksandr Petiushko. Darker than black-box: Face reconstruction from similarity queries. *arXiv preprint arXiv:2106.14290*, 2021.

[38] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241, 2015.

[39] Soumyadip Sengupta, Jun-Cheng Chen, Carlos Castillo, Vishal M. Patel, Rama Chellappa, and David W. Jacobs. Frontal to profile face verification in the wild. In *Proc. IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–9, 2016.

[40] 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 *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5265–5274, 2018.

[41] Jiayu Wang, Kang Zhao, Yifeng Ma, Shiwei Zhang, Yingya Zhang, Yujun Shen, Deli Zhao, and Jingren Zhou. Facecomposer: A unified model for versatile facial content creation. In *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, pages 13467–13479, 2023.

[42] Tao Wang, Wenying Wen, Xiangli Xiao, Zhongyun Hua, Yushu Zhang, and Yuming Fang. Beyond privacy: Generating privacy-preserving faces supporting robust image authentication. *IEEE Transactions on Information Forensics and Security*, 20:2564–2576, 2025.

[43] Yinggui Wang, Jian Liu, Man Luo, Le Yang, and Li Wang. Privacy-preserving face recognition in the frequency domain. In *Proc. AAAI Conference on Artificial Intelligence (AAAI)*, pages 2558–2566, 2022.

[44] 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 *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8212–8221, 2023.

[45] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[46] Cameron Whitelam, Emma Taborsky, Austin Blanton, Brianna Maze, Jocelyn Adams, Tim Miller, Nathan Kalka, Anil K Jain, James A Duncan, Kristen Allen, et al. Iarpa janus benchmark-b face dataset. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 90–98, 2017.

[47] Weihao Xia, Yujiu Yang, Jing-Hao Xue, and Baoyuan Wu. Tedigan: Text-guided diverse face image generation and manipulation. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2256–2265, 2021.

[48] Kai Xu, Minghai Qin, Fei Sun, Yuhao Wang, Yen-Kuang Chen, and Fengbo Ren. Learning in the frequency domain. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1737–1746, 2020.

[49] Peiqing Yang, Shangchen Zhou, Qingyi Tao, and Chen Change Loy. Pgdiff: Guiding diffusion models for versatile face restoration via partial guidance. *Advances in Neural Information Processing Systems (NeurIPS)*, 36:32194–32214, 2023.

[50] Lin Yuan, Wu Chen, Xiao Pu, Yan Zhang, Hongbo Li, Yushu Zhang, Xinbo Gao, and Touradj Ebrahimi. Pro-face c: Privacy-preserving recognition of obfuscated face via feature compensation. *IEEE Transactions on Information Forensics and Security*, 19:4930–4944, 2024.

[51] Hui Zhang, Xingbo Dong, YenLung Lai, Ying Zhou, Xiaoyan Zhang, Xingguo Lv, Zhe Jin, and Xuejun Li. Validating privacy-preserving face recognition under a minimum assumption. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12205–12214, 2024.

[52] Peng-Fei Zhang, Guangdong Bai, Hongzhi Yin, and Zi Huang. Proactive privacy-preserving learning for cross-modal retrieval. *ACM Transactions on Information Systems*, 41(2):1–23, 2023.

[53] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 586–595, 2018.

[54] Tianyue Zheng and Weihong Deng. Cross-pose lfw: A database for studying cross-pose face recognition in unconstrained environments. *Beijing University of Posts and Telecommunications, Tech. Rep*, 5:7, 2018.

[55] Tianyue Zheng, Weihong Deng, and Jiani Hu. Cross-age lfw: A database for studying cross-age face recognition in unconstrained environments. *arXiv preprint arXiv:1708.08197*, 2017.

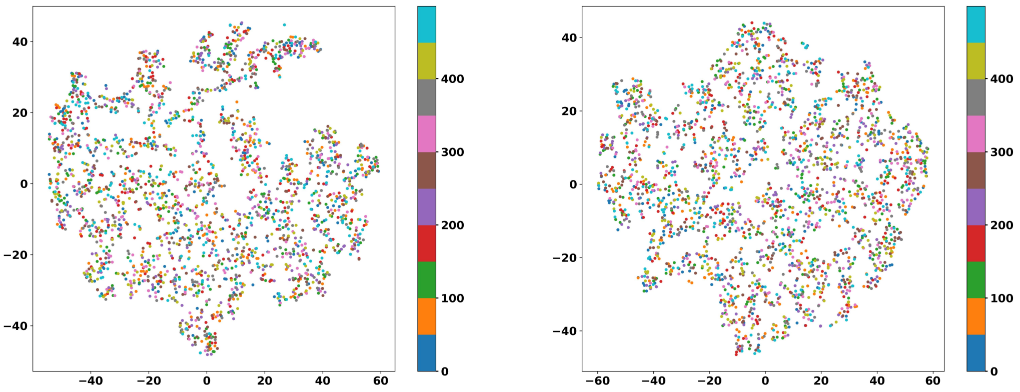
[56] Andrey Zhmoginov and Mark Sandler. Inverting face embeddings with convolutional neural networks, 2016. https://arxiv.org/abs/1606.04189.

[57] Zhizhou Zhong, Yuxi Mi, Yuge Huang, Jianqing Xu, Guodong Mu, Shouhong Ding, Jingyun Zhang, Rizen Guo, Yunsheng Wu, and Shuigeng Zhou. Slerpface: Face template protection via spherical linear interpolation. In *Proc. the Association for the Advancement of Artificial Intelligence (AAAI)*, pages 10698–10706, 2025.

[58] Yufan Zhou, Ruiyi Zhang, Changyou Chen, Chunyuan Li, Chris Tensmeyer, Tong Yu, Jiuxiang Gu, Jinhui Xu, and Tong Sun. Towards language-free training for text-to-image generation. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17886–17896, 2022.

A Appendix

A.1 Sparsity Analysis in Frequency Domain



(a) Frequency Channel Interference

(b) Refined Frequency Channel

Figure 8: Analyzing the sparsity of frequency domain channels through t-SNE.

As shown in Fig. 8a, low frequency channels still exert a strong influence on their high frequency counterparts. When low frequency components are fully expressed, the interclass distances increase, resulting in a sparser overall distribution. In contrast, Fig. 8b illustrates the refined frequency channels, where the contribution of high frequency components becomes more prominent. This leads to a more uniform distribution and a noticeable reduction in sparsity. These observations further highlight the necessity of regulating sparsity in the frequency domain by refining frequency channels to encourage more compact and balanced distributions, ultimately improving representation quality.

A.2 Nonlinearity Proof of FFM

We aim to demonstrate that the Fractal Feature Mapping (FFM) is **nonlinear**, by showing that it violates both homogeneity and additivity, the two essential conditions for linearity.

Proof (Homogeneity). Let $\lambda \in \mathbb{R}$ be a scalar and let the FFM mapping be defined as:

$$f(M_0) = M_0 + C, \tag{3}$$

where the constant offset $C \in \mathbb{R}^{n \times m}$ is given by:

$$C = \sum_{i=1}^{k} (b_i - 1) \cdot \beta_i. \tag{4}$$

Then applying f to a scaled input yields:

$$f(\lambda M_0) = \lambda M_0 + C. \tag{5}$$

However, if f were linear, we would expect:

$$f(\lambda M_0) = \lambda f(M_0) = \lambda (M_0 + C) = \lambda M_0 + \lambda C. \tag{6}$$

Clearly, unless $\lambda = 1$ or $C = 0$, we have:

$$f(\lambda M_0) \neq \lambda f(M_0). \tag{7}$$

Since in practice $C \neq 0$ is intentionally introduced to enhance privacy by perturbing identity-irrelevant components, and $\lambda \neq 1$ holds for general scaling operations, the homogeneity condition is violated. Thus, FFM is not a homogeneous transformation.

We further justify the validity of the conditions $\lambda \neq 1$ and $C \neq 0$ as follows:

- The constant offset C is constructed to be non-zero in order to perturb identity-irrelevant features and preserve privacy. If $C = 0$, the transformation degenerates to the identity map $f(M_0) = M_0$, which provides no privacy protection.
- The scalar $\lambda \neq 1$ represents any general scaling factor different from the identity scaling, and is commonly used in evaluating homogeneity.

Hence, under typical and intended design conditions, FFM violates the homogeneity property, confirming its nonlinear behavior.

Proof (Additivity). Let $M_a, M_b \in \mathbb{R}^{n\times m}$ be two arbitrary inputs. The FFM mapping is defined as:

$$f(M) = M + \sum_{k=1}^{K} (\beta_k(M) - 1) \cdot \gamma_k, \tag{8}$$

where $\beta_k(\cdot)$ denotes a nonlinear feature mapping and $\gamma_k \in \mathbb{R}^{n\times m}$ are fixed perturbation matrices.

Then,

$$f(M_a) = M_a + \sum_{k=1}^{K} (\beta_k(M_a) - 1) \cdot \gamma_k, \quad f(M_b) = M_b + \sum_{k=1}^{K} (\beta_k(M_b) - 1) \cdot \gamma_k. \tag{9}$$

If f were additive, we would have:

$$f(M_a + M_b) = f(M_a) + f(M_b). \tag{10}$$

However,

$$f(M_a + M_b) = M_a + M_b + \sum_{k=1}^{K} (\beta_k(M_a + M_b) - 1) \cdot \gamma_k, \tag{11}$$

while

$$f(M_a) + f(M_b) = M_a + M_b + \sum_{k=1}^{K} [(\beta_k(M_a) - 1) + (\beta_k(M_b) - 1)] \cdot \gamma_k. \tag{12}$$

These are not equal unless

$$\beta_k(M_a + M_b) = \beta_k(M_a) + \beta_k(M_b), \tag{13}$$

which is generally false due to the nonlinearity of $\beta_k(\cdot)$. Therefore,

$$f(M_a + M_b) \neq f(M_a) + f(M_b), \tag{14}$$

proving that FFM does not satisfy additivity either.

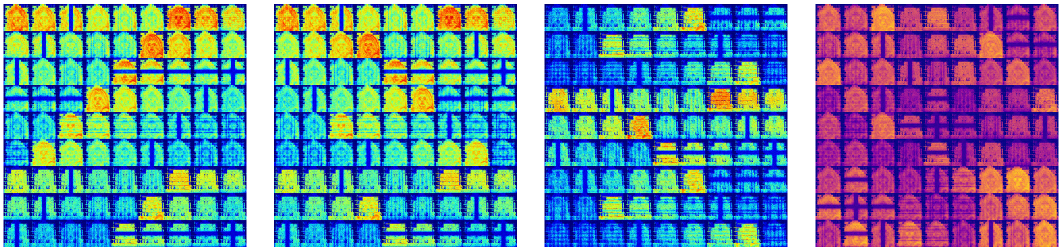
In summary, the FFM transformation violates both homogeneity and additivity, it is not a linear transformation. In fact, it is an affine mapping (due to the additive constant C), and its nonlinearity is intentionally designed to obfuscate identity-irrelevant features and ensure privacy preservation.

Non-invertibility. We also consider a threat model where an adversary aims to reconstruct and identify faces from obfuscated images, assuming knowledge of FracFace parameters (e.g., fractal depth and expansion schedule). While visual cues cannot be fully removed due to recognition needs, FracFace substantially reduces them compared to prior work. In detail, two components—the initial fractal kernel $M_0 \in \mathbb{Z}^{n\times n}$ and the index lattice $L_0 \in \mathbb{Z}^{n\times n}$—are secret and randomized per deployment, ensuring unique, secure instances. In addition, the fractal index mapping is defined recursively as

$$\mathcal{F}^{[k]}[i,j] = M_0[i,j] \; + \; \sum_{\ell=1}^{k} \big(b_\ell - 1\big) \cdot \beta_\ell, \qquad \beta_1 = 1, \quad \beta_\ell = \prod_{s=1}^{\ell-1} b_s. \tag{15}$$

This is then projected modulo the channel dimension C:

$$\psi[i,j] \; = \; \mathcal{F}^{[k]}[i,j] \bmod C. \tag{16}$$



(a) Vanilla Frequency Domain Channel Visualization

(b) Refined Frequency Domain Channel Visualization (Part 1)

(c) Filtered Frequency Domain Channel Visualization (Part 2)

(d) Fractal Frequency Domain Channel Visualization

Figure 9: Visual comparison of refined frequency domain channels.

Because $\mathcal{F}^{[k]}$ grows exponentially with k, many different positions collide to the same index under ψ.

$$\exists\,(i_1, j_1) \neq (i_2, j_2) \;\; \text{s.t.} \;\; \psi[i_1, j_1] = \psi[i_2, j_2], \quad \text{but} \quad \mathcal{F}^{[k]}[i_1, j_1] \neq \mathcal{F}^{[k]}[i_2, j_2]. \tag{17}$$

This shows that ψ is non-injective by design, and no inverse mapping exists. Even if the adversary learns the full mapping rule ψ, they cannot resolve the original positions or channel correspondences without knowing the randomized M_0 and L_0:

$$\psi^{-1} : \mathbb{Z}_C \to \mathbb{Z}^{n \times n} \tag{18}$$

These design choices ensure that even subtle visual cues are unrecoverable under strong white-box attacks. The non-invertibility of ψ and the randomness from M_0 and L_0 jointly establish a robust privacy boundary, making FracFace highly resistant to reconstruction.

A.3 The Definition and Computation of Protection(%)

Definition Protection (%) represents the share of frequency-domain channels that are either (i) filtered out by FCR or (ii) structurally disrupted by FFM, expressed as a percentage of the total number of channels.

$$\text{Protection (\%)} \;=\; \frac{|P_{\text{mask}}| + |P_{\text{remap}}|}{P_{\text{total}}}, \tag{19}$$

where, $P_{\text{total}}$is the total number of DCT frequency channels (e.g., 192 for 12 × 16 DCT), P_{mask} is the number of low-energy channels pruned by FCR, and P_{remap} is the number of remaining channels remapped by FFM.

A.4 Visual Privacy Protection in Face Features

We begin by visualizing the frequency domain distribution extracted from raw facial images, as illustrated in Fig.9a. The distribution exhibits a clear dominance of low frequency components and an overall sparse structure. These characteristics raise potential privacy risks: sparsity may allow adversaries to reconstruct identity revealing features; meanwhile, the dominance of low frequency components suppresses high frequency details, limiting the representational capacity of the frequency space. To mitigate these risks, we apply FCR to perform band attenuation, as shown in Fig.9b. This process selectively preserves features that are identity relevant yet privacy preserving, while filtering out noisy frequency components (see Fig.9c). Then, a fractal frequency domain channel is generated via FFM (Fig. 9d), mitigating sparsity and preserving self similarity for recognition, while its inherent randomness obfuscates identity cues to reduce reconstruction risk and enhance privacy protection.

A.5 The Algorithm of FCR

Algorithm 1 Frequency Channel Refining (FCR)

```
Require: Grid size M, N
Ensure: Two groups $S_1$, and $S_2$
$F \leftarrow$ Create matrix (M,N)
freq_list $\leftarrow$ [ ]
for $i \leftarrow 0$ to $M - 1$ do
    row $\leftarrow F[i]$
    if $i \bmod 2 = 0$ then
        Extend (freq_list, row)
    else
        Extend (freq_list, Reverse (row))
    end if
end for
$S_1 \leftarrow$ freq_list[0:80]
$S_2 \leftarrow$ freq_list[80:161]
return $S_1, S_2$ =0
```

A.6 The Algorithm of FFM

Algorithm 2 Frequency Fractal Mapping (FFM)

```
Require: Number of iterations $K$, fractal kernel size $m \times n$
Ensure: Fractal mappings $\mathcal{F} = [\mathcal{F}_0, \mathcal{F}_1, \ldots, \mathcal{F}_{K-1}]$
Initialize $M_0 \in \mathbb{Z}^{m \times n}$ with random integers in $[1, m \cdot n]$
Initialize $L_0 \in \mathbb{Z}^{m \times n}$ as indexing order
Set $\mathcal{F}_0 = M_0$
for $k = 1$ to $K - 1$ do
    Compute scaling factor: $\beta_k = 3^{2k}$
    for $i = 1$ to $m$ do
        for $j = 1$ to $n$ do
            $\mathcal{F}_k[i, j] = (M_0[i, j] - 1) \cdot \beta_k + \mathcal{F}_{k-1}[i, j]$
        end for
    end for
    if correction required at layer $k$ then
        $\mathcal{F}_k = \mathcal{F}_k - 1$
    end if
end for
return $\mathcal{F}$ =0
```

A.7 More Implementation Details

Data Preprocessing of Input Images All face images were first resized to 112×112 and normalized as RGB tensors, then transformed into the frequency domain via BDCT. To retain identity-relevant information while reducing redundancy, we applied Frequency Channels Refining (FCR) to extract 81 informative channels. These were further refined using Frequency Fractal Mapping (FFM), which enhances feature consistency across both spatial and scale dimensions. The final 81-channel representations were saved as .npy files. Throughout preprocessing, we ensured numerical stability by checking for NaNs and infinities, and accelerated the pipeline using 8-worker parallel loading. This procedure was uniformly applied across training and evaluation datasets, including MS1M-ArcFace, LFW, and AgeDB.

Training Details We trained the FracFace model on the MS1Mv2 dataset using PyTorch with two RTX 6000 GPUs (49 GB VRAM each). Training the FracFace model on MS1Mv2 took about 8 days for a total of 50 epochs. During training, the peak memory usage per GPU was about 45GB. The input comprised 81-channel feature maps produced by the FracFace pipeline, incorporating DCT

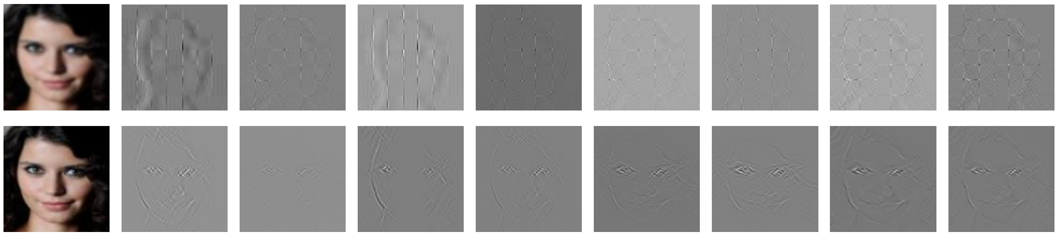


Figure 10: Resilient identity learning despite frequency degradation

transformation, Frequency Channel Refinement (FCR), and Frequency Fractal Mapping (FFM). An IR-50 backbone with ArcMargin loss was employed to enhance identity discrimination. Optimization was performed using AdamW (lr=0.001, weight decay=1e-4) with a cosine annealing scheduler. Training leveraged automatic mixed precision (AMP) and gradient clipping (max norm 5.0) for stability. We monitored performance via TensorBoard and validated on a held-out set after each epoch. Data loading was parallelized with 8 workers and prefetching, and all experiments used `torch.backends.cudnn.benchmark=True` for optimal GPU performance.

A.8 Limitations

As described in Sec. 3, FracFace performs frequency channel refinement via the Frequency Channels Refining (FCR) module after applying the BDCT transform. The retained identity-relevant frequency components are then mapped into a fractal structural space through the Frequency Fractal Mapping (FFM) process. As illustrated in Fig. 10, when the input images are of low resolution, the high-frequency bands may carry limited or unstable identity cues, making it challenging for the model to extract robust representations. This degradation is likely caused by the loss of discriminative patterns in the high-frequency spectrum under adverse imaging conditions. Nonetheless, as shown in Fig. 10, when the training data preserve sufficient frequency fidelity, FracFace can still effectively learn identity-aware representations while maintaining strong privacy protection guarantees.

A.9 Ablation Study

Table 4: Fractal depth k

k	Accuracy ↑	SSIM ↓	LPIPS ↑
1	99.71	0.5227	0.5291
2	99.69	0.4015	0.6353
3	96.46	0.3729	0.7925
4	92.13	0.2580	0.8357

Table 5: FBA pruning strength

Ratio	Accuracy ↑	SSIM ↓	LPIPS ↑
20%	99.83	0.7857	0.3184
40%	99.71	0.6291	0.4833
50%	99.69	0.3012	0.6839
60%	89.26	0.3109	0.7294
80%	87.24	0.2793	0.8605

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The main claims presented in the abstract and introduction accurately reflect the paper's core contributions and scope. The abstract and introduction provide a clear overview of the proposed approach and its significance. The method is elaborated in detail in Sec. 3, while Sec. 4.2, Sec. 4.3, and Sec. 4.4 present extensive experimental results and ablation studies that substantiate the claims made earlier. In addition, we also mentioned the novelty of our proposed solution in the key contribution section. Compared with other work in related work, we proposed to use fractal to solve the current PPFR problem.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We discuss the limitation in Appendix A.8. In addition, we tested our proposed FracFace on 6 different datasets, which contain facial images of different people (based on public datasets). Due to page limitations, we have placed the relevant content in the Appendix. However, if necessary, we will also accept placing it in the main text to explain the issue.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.

- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: The theoretical foundations of our approach are thoroughly established. All relevant assumptions are explicitly presented, and Appendix A.2 contains the full and correct proofs, ensuring the soundness and transparency of the theoretical claims. In addition, our Sec. 3 also describes in great detail the calculation of each data and the flow of the algorithm, and all steps are explained.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The paper provides sufficient information to enable reproduction of the main experimental results that support the central claims. We include the core implementation of the FracFace framework, along with access to the training dataset and code necessary to replicate the training procedure. These materials are made available through the link referenced in Sec. 4.1. In addition, we presents the essential pseudocode that demonstrates how FracFace can be implemented (refer more for Algorithm. 2).

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.

- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
    (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
    (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
    (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
    (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: The paper specifies all the necessary training and testing details. These details are thoroughly outlined in Sec. 4.1 of the paper. Additionally, we provide the core pseudocode of FracFace, offering a detailed implementation process. We provide open access to both the data and the source code (Code is available at https://anonymous.4open.science/r/FracFace) , along with comprehensive instructions to ensure the faithful reproduction of the main experimental results. Detailed implementation guidelines can be found in Appendix. A.7 of the paper.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: The paper specifies all the necessary training and testing details. These details are thoroughly outlined in Sec. 4.1 and Appendix A.7 of the paper. (Code is available at https://anonymous.4open.science/r/FracFace)

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [No]

Justification: We adhere to the established experimental protocols from prior PPFR works and present the best achieved results.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: The paper offers comprehensive details regarding the computational resources necessary for both training and inference. These specifications are meticulously outlined in Appendix A.7, ensuring the full reproducibility of the experiments.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: The research reported in this paper complies with the NeurIPS Code of Ethics and adheres to all relevant guidelines and standards. All datasets used are publicly available; their licensing terms and access conditions are summarized in Sec. 6.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

   Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

   Answer: [Yes]

   Justification: We discuss the broader impacts in Sec. 6. FracFace aims to enhance facial privacy protection while preserving identity information for secure face recognition, which could have societal impacts in contexts such as privacy-preserving authentication, surveillance minimization, and responsible biometric data usage (see Sec. 4.3 for more details). The method may help mitigate risks of facial data misuse, especially in publicly deployed face recognition systems. All experiments were conducted on publicly available datasets strictly for academic research, and no personally identifiable data was used beyond the scope of these datasets.

   Guidelines:

   - The answer NA means that there is no societal impact of the work performed.
   - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
   - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
   - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
   - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
   - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

   Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

   Answer: [NA]

Reason: This paper does not have such risks. In addition, all the data sets used in our experiments are public data sets. For an introduction to the data sets, please refer to Sec. 4.1

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: All external assets used in this work, including code, data, and models, are properly credited, with their respective licenses and terms of use clearly acknowledged and fully respected. All data and code employed in this paper were acquired with the required legal permissions.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: The paper does not release new assets.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: This paper does not involve crowdsourcing or research involving human subjects, therefore, such details are not applicable.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

    Justification: The paper does not involve the usage of Large Language Models (LLMs) as a core component of the methodology. LLMs were not integral to the core scientific processes, which did not affect the originality, rigor, or methodology of the research.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (https://neurips.cc/Conferences/2025/LLM) for what should or should not be described.