WaFusion: A Wavelet-Enhanced Diffusion Framework for Face Morph Generation

Seyed Rasoul Hosseini, Omid Ahmadieh, Jeremy Dawson and Nasser Nasrabadi
West Virginia University
Morgantown, USA
sh00111, oa00023@mix.wvu.edu, jeremy.dawson@mail.wvu.edu, nnasrab1@gmail.com

Abstract

Biometric face morphing poses a critical challenge to identity verification systems, undermining their security and robustness. To address this issue, we propose WaFusion, a novel framework combining wavelet decomposition and diffusion models to generate high-quality, realistic morphed face images efficiently. WaFusion leverages the structural details captured by wavelet transforms and the generative capabilities of diffusion models, producing face morphs with minimal artifacts. Experiments conducted on FERET, FRGC, FRLL, and WVU Twin datasets demonstrate WaFusion's superiority over state-of-the-art methods, producing high-resolution morphs with fewer artifacts. Our framework excels across key biometric metrics, including the Attack Presentation Classification Error Rate (APCER), Bona Fide Presentation Classification Error Rate (BPCER), and Equal Error Rate (EER). This work sets a new benchmark in biometric morph generation, offering a cutting-edge and efficient solution to enhance biometric security systems.

1. Introduction

Owning to non-intrusive nature, ease of use, and broad acceptance [22], Facial Recognition Systems (FRS) are widely adopted in security-related applications, particularly as the primary biometric for electronic Machine-Readable Travel Documents (eMRTD) [28, 62, 69]. However, the increased reliance on FRS exposes them to face morphing attacks, *i.e*. meticulously blending the facial features of two individuals to create an image resembling both identities [16, 53]. These morphed images undermine FRS by increasing the False Acceptance Rates (FAR), deceiving human verifiers, and risking unauthorized access by malicious actors to secure facilities [13, 45]. As morphing tools become more accessible, even individuals with limited technical skills can generate high-quality morphs, exacerbating these risks [4].

While efforts to counter face morphing often focus on detection methods integrated into biometric pipelines that identify morphed images [59], these approaches struggle against high-quality morphs, notably, those generated by advanced deep learning techniques [44, 65]. This highlights the urgent need for improved morph generation techniques to evaluate and intensify detection algorithms.

Morphing methods typically fall into two categories: 1) landmark-based approaches [14] and 2) deep learning-based techniques [8]. Landmark-based methods rely on a structured pipeline involving landmark detection, warping, and blending [6, 30, 40]. While effective, these methods often require manual adjustments and are prone to obvious visual inconsistencies. Conversely, deep learning-based approaches, such as Generative Adversarial Networks (GANs), generate morphs in an end-to-end manner [24]. GANs have become a prominent choice due to their ability to produce high-quality images through adversarial training [18]. However, GAN-based methods face challenges such as mode collapse, geometric distortion, and sensitivity to dataset quality, often leading to artifacts or vulnerability to adversarial-style variations [1, 49, 63].

Recently, Diffusion Probabilistic Models (DPMs) have emerged as a robust alternative to GANs for image synthesis, offering enhanced diversity and fidelity [20]. These models iteratively refine noisy data to model complex distributions, enabling the generation of high-quality images [19, 26]. The Denoising Diffusion Probabilistic Model (DDPM), introduced by Ho et al. [20], achieves remarkable image quality through iterative noise reduction. Building on this, the Denoising Diffusion Implicit Model (DDIM) improves efficiency by reducing the number of sampling steps while maintaining high synthesis quality [57]. Nevertheless, diffusion models face notable challenges, requiring substantial computational resources and extended runtimes. They also struggle with fine-grained control over morphological traits, making them less practical for specific tasks [12, 47, 63]. These limitations underscore the need for approaches combining diffusion models' strengths while ad-

dressing their challenges, such as improving computational efficiency and feature control.

This work aims to generate high-quality face morphs while preserving critical image details efficiently. Despite advancements in state-of-the-art (SOTA), including GANs and diffusion models, existing methods face significant limitations. To overcome these challenges, we need a hybrid approach that integrates complementary techniques.

To this end, we propose a novel hybrid framework that combines Discrete Wavelet Transform (DWT) and diffusion models. Wavelet decomposition excels in preserving essential image textures by analyzing images at multiple frequency bands [9, 29], while diffusion models ensure high-fidelity synthesis [20, 57]. In our approach, the Low-low (LL) sub-band, which captures critical image features, is extracted through wavelet decomposition and processed using diffusion models to generate realistic morphed components. The Inverse Wavelet Transform (IWT) then combines the morphed LL sub-band with the high-frequency sub-bands from the original image, producing the final morph. By using the complementary strengths of these techniques, our method effectively addresses the limitations of existing approaches, enabling the generation of high-quality morphs with enhanced efficiency.

The primary contributions of this study are as follows:

- An innovative approach using wavelet decomposition for biometric morphing, enhancing image quality while preserving essential features.
- The ability to generate high-resolution morphs without increasing computational costs compared to conventional baseline methods.
- A comprehensive evaluation of the proposed framework against SOTA techniques across diverse datasets, demonstrating its superior quality, efficiency, and robustness.

2. Related Works

The risks posed by face morphing attacks are highlighted by Ferrara *et al.* [13], who manually created morphed images using GIMP, an open-source image editor. These images could deceive Automated Border Control (ABC) systems despite showing minimal artifacts, but their manual nature lacked scalability for large-scale dataset production. Landmark-based algorithms such as Facemorpher [40] and OpenCV [30] subsequently automated morph generation, using warping and splicing to produce morphs efficiently. However, these methods often introduce artifacts in high-frequency regions like the iris and facial contours [66], limiting their effectiveness for realistic morph generation.

Recent developments in GANs have significantly improved morph quality by automating synthesis and minimizing perceptual inconsistencies [7, 17, 51]. Damer *et al.* [8] introduced MorGAN, one of the first GAN-based morphing methods, which blends two identities into a single morphed identity. Subsequent architectures such as StyleGAN2 [50] and MIPGAN-II [66] further enhanced morph quality and reduced artifacts. These advancements have driven further research in the field of face morph generation [39, 60]. However, GAN-based approaches face persistent challenges, such as mode collapse and dataset sensitivity, often resulting in artifacts [46].

DPMs have emerged as a promising alternative to GANs, offering superior diversity and fidelity [56]. Ho *et al.* [20] demonstrated the potential of DPMs for generating highly detailed images through iterative denoising and effectively addressing mode collapse. Diffusion autoencoders extended these capabilities by disentangling semantic and random data, allowing for more controlled image generation [38]. Building on this, recent methods have applied diffusion models specifically to face morphing. DiffMorpher [67] enables full-image semantic interpolation but incurs high computational costs. Fast-DiM [3] improves efficiency by reducing the number of sampling steps but does not enhance morph quality or structure preservation. Blasingame *et al.* [2] proposed a strong morphing attack pipeline based on diffusion, yet it lacks control over fine-grained facial features. Despite these advances, computational demands and limited structural fidelity remain open challenges. In contrast, WaFusion introduces a hybrid wavelet-diffusion approach that selectively processes only low-frequency components, achieving efficient morphs while maintaining critical facial structure.

Wavelet-based methods have proven effective in enhancing both the quality and robustness of face morphs. O'Haire *et al.* [33] employed wavelet transformations to generate morphed faces, while Huang *et al.* [21] applied them for image restoration in morphing contexts. More recently, Hybrid approaches such as Phung *et al.* [37] have combined wavelet decomposition with diffusion models, leveraging texture analysis for high-fidelity synthesis with reduced computational cost. These advancements demonstrate the adaptability of wavelet-based frameworks in addressing prominent challenges in morph generation.

As face morphing techniques evolve, Morphing Attack Detection (MAD) has become a critical research area. Early works [41, 52] emphasized the challenge of distinguishing morphs from bona fide images. More recent approaches leverage diffusion models to detect morphs as out-of-distribution samples [23], deep face embeddings from models like MagFace [25], and multispectral imaging for enhanced detection [44]. These works demonstrate how progress in morph generation drives advances, reflecting the co-evolution of attack and defense strategies.

While landmark-based methods offer simplicity and con-

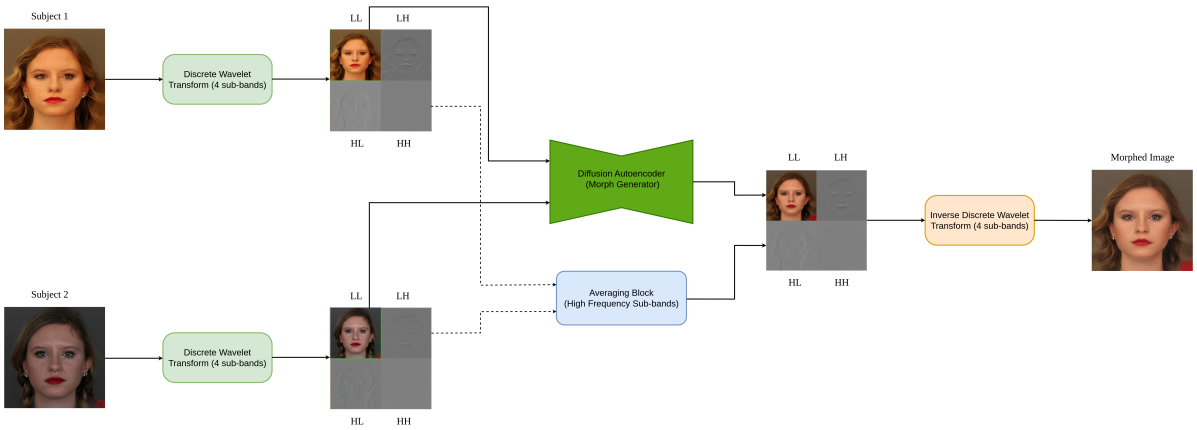


Figure 1. Overview of the WaFusion framework for face morphing. The input images, Subject 1 and Subject 2, are first aligned and decomposed into four sub-bands (LL, LH, HL, HH) using wavelet decomposition. The LL sub-bands, highlighted in the green dotted box, are processed through the diffusion model to generate the morphed LL sub-band. The remaining sub-bands, shown by orange dots, are averaged. The final morph is reconstructed using the inverse wavelet transform, combining both morphed and averaged components.

trol, they often fail to capture complex facial structures and expressions. GAN- and diffusion-based techniques significantly improve realism and diversity, yet they remain susceptible to artifacts and high computational cost [59]. Recent hybrid approaches that integrate wavelet decomposition with generative models show potential in addressing these limitations. Building on this direction, our approach combines wavelet decomposition with diffusion autoencoders to selectively enhance structural fidelity while reducing generation overhead.

3. Methodology

3.1. Wavelet Transform

The wavelet transform is a foundational component of our model, enabling efficient decomposition of images into their low-frequency approximations and high-frequency details [29]. This hierarchical process captures coarse structure in the Low-Low (LL) sub-band, while encoding finer details , *e.g.*, vertical, horizontal, and diagonal edges in the remaining sub-bands, namely Low-High (LH), High-Low (HL), and High-High (HH). Wavelet decomposition preserves essential structural and textural information while reducing the spatial dimensions, facilitating efficient downstream processing.

We use the Haar wavelet for its simplicity and computational efficiency. The decomposition employs low-pass (L) and high-pass (H) filters, resulting in four sub-bands: X_{ll} represents a coarse, low-frequency approximation for the image, capturing its overall structure, X_{lh} encodes high-frequency details in the horizontal direction, highlighting fine horizontal edges, X_{hl} encodes high-frequency details in the vertical direction, focusing on vertical edges, and X_{hh} encodes high-frequency details in the diagonal direction, representing texture and diagonal edges.

These sub-bands are represented in Fig. 1, where the LL sub-band is highlighted as the primary source of structural information for morph generation, while the high-frequency sub-bands contribute fine-grained textures. The Haar wavelet transform of an image $X \in \mathbb{R}^{H \times W}$ is computed as:

$$X_{pq} = F_p^\top \cdot X \cdot F_q, \qquad p, q \in \{L, H\}, \tag{1}$$

yielding four $\frac{H}{2} \times \frac{W}{2}$ sub-bands. To reconstruct the original image, the IWT recombines these sub-bands, ensuring no loss of fidelity [29].

In our framework, the LL sub-band is processed using a diffusion model to generate the morphed component. The high-frequency sub-bands (LH, HL, and HH) are averaged and fused with the morphed LL sub-band using the IWT to produce the final high-resolution morph. The wavelet analysis was implemented using the PyWavelets package [27], ensuring efficient computation and detailed texture preservation.

3.2. Diffusion Model

Diffusion-based generative models progressively transform a noise map, initialized from a standard Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{I})$, into a clean image through a series of denoising iterations. First introduced by Ho *et al.* [20], these models rely on a learned denoising function, $\epsilon_\theta(x_t, t)$, which predicts the noise added to an image at step t. By iteratively removing this noise, the models reconstruct the orig-

[1]Note: The subjects shown in Figure 2 are identical twins from the WVU Twin dataset, which explains the high visual similarity between the bona fide images. This characteristic applies to all image pairs generated from the WVU Twin dataset.

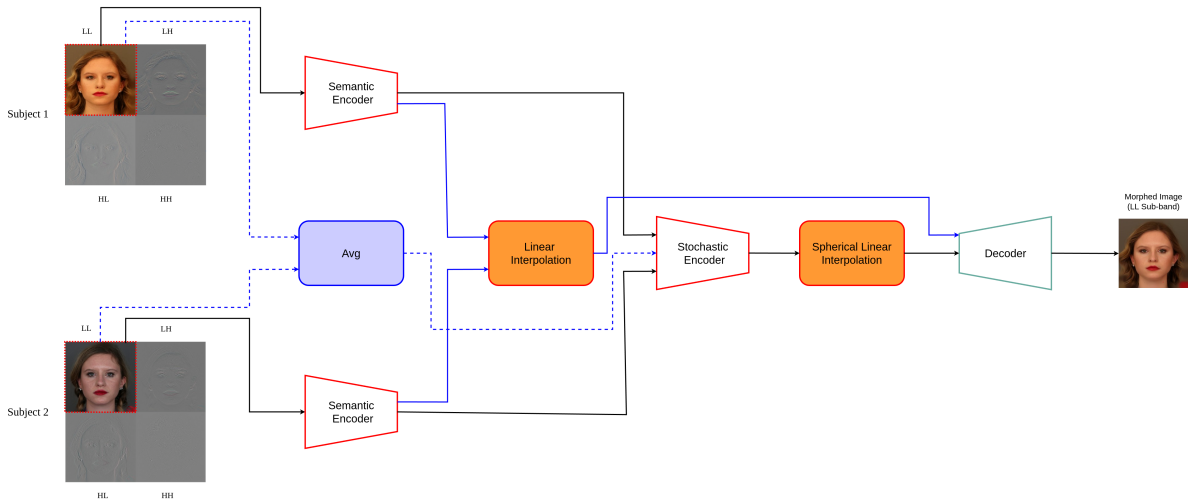


Figure 2. Architecture for diffusion-based morph generation using Diffusion Autoencoders. LL sub-bands (red dashed lines) from Subject 1 and Subject 2 are processed through semantic encoders and an averaging block. Outputs are passed through stochastic encoders, followed by interpolation. Semantic features undergo linear interpolation, while stochastic features use spherical linear interpolation. The combined features are decoded to generate the morphed LL sub-band.[1]

inal image x_0, starting from a fully noisy sample x_t. This denoising function is implemented using a UNet architecture [48], enabling high-quality image generation. Moreover, the approach simplifies the variational lower bound on the data's marginal log-likelihood, a technique that has gained significant adoption in the field [57, 32].

In the forward process, Gaussian noise is added to x_0 over T steps, modeled as:

$$q(x_t \mid x_{t-1}) = \mathcal{N}((\sqrt{1-\beta_t})x_{t-1}, \beta_t \mathbf{I}), \tag{2}$$

where β_t controls the noise schedule. The noisy image x_t at step t can also be expressed directly in terms of x_0:

$$q(x_t \mid x_0) = \mathcal{N}(\sqrt{\alpha_t}x_0, (1-\alpha_t)\mathbf{I}), \tag{3}$$

with $\alpha_t = \prod_{s=1}^{t}(1-\beta_s)$. The reverse process aims to remove this noise, reconstruct x_0, modeled as:

$$p(x_{t-1} \mid x_t) = \mathcal{N}(\mu_\theta(x_t, t), \sigma_t), \tag{4}$$

where μ_θ is a learned mean function and σ_t is predefined [20]. This reverse process enables the stepwise reconstruction of x_0 from x_T.

Building on this foundation, Song *et al.* [57] introduced the DDIM, which modifies the reverse process to be deterministic. DDIM generates samples using the following update rule:

$$x_{t-1} = \sqrt{\alpha_{t-1}}\left(\frac{x_t - \sqrt{1-\alpha_t}\epsilon_\theta^t(x_t)}{\sqrt{\alpha_t}}\right) + \sqrt{1-\alpha_{t-1}}\epsilon_\theta^t(x_t), \tag{5}$$

allowing faster sampling and more precise control over the generative process without altering the marginal distribution. DDIM further introduces an inference distribution:

$$q(x_{t-1} \mid x_t, x_0) = \mathcal{N}\left(\sqrt{\alpha_{t-1}}x_0 + \sqrt{1-\alpha_{t-1}} \cdot \frac{x_t - \sqrt{\alpha_t}x_0}{\sqrt{1-\alpha_t}}, \mathbf{0}\right), \tag{6}$$

which retains the core principles of DDPM while enabling more efficient and deterministic sample generation.

Despite their advantages, DPMs have limitations. The latent variable x_T, representing the starting point of the reverse process, lacks high-level semantic information, making it less interpretable for applications requiring meaningful feature manipulation, such as face morphing. Preechakul *et al.* [38] explored this limitation and proposed methods to disentangle semantic and random data, highlighting the importance of enriching the latent variable's representation for downstream applications.

Our hybrid approach builds on the foundational strengths of DPMs while addressing their inherent limitations. By integrating structural precision with detailed feature control, this methodology enables the generation of high-quality, realistic face morphs, while enhancing both interpretability and computational efficiency.

3.3. Proposed Method

The proposed WaFusion framework for face morph generation is depicted in Fig. 1, which provides an overview of its key components and workflow. The process be-

gins with the decomposition of two bona fide input images, *i.e*. distinct identities, into four wavelet sub-bands: LL, LH, HL, and HH. This decomposition, performed using a single-level Haar transform, which separates low-frequency approximations (LL) from high-frequency details (LH, HL, HH). The LL sub-bands, which capture the structural essence of the images, are then processed by the morph generation block, implemented using diffusion autoencoders [38], to generate the morphed LL sub-band. Finally, the inverse wavelet transform fuses the morphed LL sub-band with the averaged high-frequency sub-bands from subject 1 and subject 2, reconstructing the final morphed image with high fidelity and detail.

Our diffusion autoencoders, as illustrated in Fig. 2, employ a dual-encoder structure consisting of semantic and stochastic encoders. The semantic encoder focuses on preserving structural attributes, such as facial feature alignment, ensuring the morph retains the essential characteristics of both input identities. In contrast, the stochastic encoder captures fine-grained details, such as textures, hair direction, and clothing, enriching the morph's visual fidelity while maintaining its identity relevance. Together, this dual-encoder structure provides a balance between maintaining overall structure and incorporating finer details. During the interpolation process, the semantic encoder streams are blended linearly to maintain consistent facial landmarks. Simultaneously, the stochastic encoder outputs are interpolated using spherical linear interpolation, allowing smooth and natural blending of fine-grained appearance features such as skin texture, hair style, and clothing patterns. This dual-path interpolation mechanism ensures the structural coherence and realistic textural details of generated morphed. To facilitate the effective blending of attributes, three fundamental functions are incorporated into the framework: 1) the image space preprocessing function ξ, 2) the image space interpolation function $\ell_{\mathcal{X}}$, and 3) the latent space interpolation function $\ell_{\mathcal{Z}}$. The image space preprocessing function $\xi : \mathcal{X} \times \mathcal{X} \rightarrow \mathcal{X}$, aligns and enhances key features from the input images before encoding, ensuring consistency in structural attributes. The image space interpolation function, $\ell_{\mathcal{X}}(u, v; \gamma) = \gamma u + (1 - \gamma)v$, blends the two input images linearly, where $\gamma \in [0, 1]$ determines the blending ratio and controls the influence of each image. Finally, the latent space interpolation function, as defined in [57], handles non-linear relationships in the latent space:

$$\ell_{\mathcal{Z}}(u, v; \gamma) = \frac{\sin((1-\gamma)\theta)}{\sin\theta} u + \frac{\sin(\gamma\theta)}{\sin\theta} v, \qquad (7)$$

where $\theta = \frac{\arccos(u \cdot v)}{|u|,|v|}$. This function enhances the realism of the generated morphs by addressing the complexity of interpolating stochastic codes and preserving semantic coherence.

While WaFusion builds on validated techniques such as diffusion autoencoders and wavelet transforms, its novelty lies in combining these domains for face morphing. By selectively applying generative processing only on low-frequency components while maintaining high-frequency details through efficient averaging, WaFusion uniquely balances computational efficiency and morph quality, which has not been explored in previous morphing methods. This hybrid approach results in high-quality, realistic face morphs that are robust against detection systems.

4. Experiment and Results

4.1. Datasets

We use four datasets: WVU Twin [2] [34], FRLL [11], FRGC [35], and FERET [36]. WVU Twin contains frontal images with neutral expressions and plain backgrounds. It includes 2,268 unique identities, with some subjects appearing multiple times in collections spanning different years. This results in morph pairs with naturally high visual similarity. The images have resolutions ranging from 2848×4288 to 5760×3840.

The other datasets contain passport-style frontal images under ideal lighting: FRLL (102 IDs, 413×531), FRGC (765 IDs subset, 1704×2272), and FERET (1,199 IDs, 512×768). We generate 2,971 WVU Twin-based, 1,122 FRLL-based, 964 FRGC-based, and 529 FERET-based morphs.

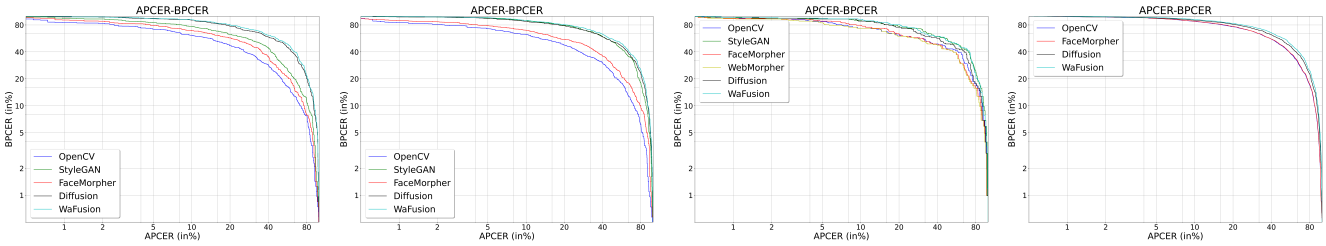
Baseline comparisons include FaceMorpher [40], OpenCV [30], StyleGAN [24], WebMorpher [10] (on FRLL only), and Diffusion Autoencoders [38, 2]. Morphing protocols follow Neubert *et al.* for the AMSL dataset applied to FRLL [31], and Scherhag *et al.* for FERET and FRGC [54]. We also include an LL-only baseline (LL-morphs), where only the low-frequency sub-band is morphed and high-frequency sub-bands are averaged.

4.2. Implementations Details

All the images are resized to 512×512, and aligned via facial landmarks. A single-level Haar wavelet transform is applied, resulting in four 256×256 sub-bands. The LL sub-band is input into the morph generator, while high-frequency sub-bands are averaged. The inverse wavelet transform reconstructs the final 512×512 morph. Morph generation employs pre-trained Diffusion Autoencoders [38] using DDIM with 100 steps. All experiments are conducted on an NVIDIA Titan RTX GPU with 24 GB of VRAM.

For verification, we use FaceNet [55], a widely adopted framework for facial recognition, with an Inception backbone [58], pre-trained on the VGGFace2 [5]. FaceNet is

[2]The dataset is available upon request. To gain access, please contact Dr. Jeremy Dawson at jeremy.dawson@mail.wvu.edu.



(a) (b) (c) (d)

Figure 3. APCER-BPCER curves for FaceNet verifier across (a) FERET, (b) FRGC, (c) FRLL, and (d) WVU Twin datasets.

leveraged to quantify look-alikes by generating compact feature embeddings for input images. It uses a triplet loss, where the Euclidean distance (L_2) for embeddings of the same identity are positive examples, and differing identities are considered negative examples [55].

4.3. Evaluation Metrics

To evaluate the effectiveness of WaFusion, we use several key metrics: Area Under the Curve (AUC), Attack Presentation Classification Error Rate (APCER), Bona Fide Presentation Classification Error Rate (BPCER), and Equal Error Rate (EER). These metrics follow the ISO/IEC 30107-3 framework [15] and are commonly used in morphing attack detection.

The AUC is derived from the Receiver Operating Characteristic (ROC) curve, which plots the true positive rate against the false positive rate at various thresholds [64]. The EER represents the point where the FAR equals the False Rejection Rate (FRR) [42]. APCER measures the ratio of morph attack samples incorrectly classified as bona fide presentations, while BPCER quantifies the percentage of genuine images misclassified as morphs [43, 59].

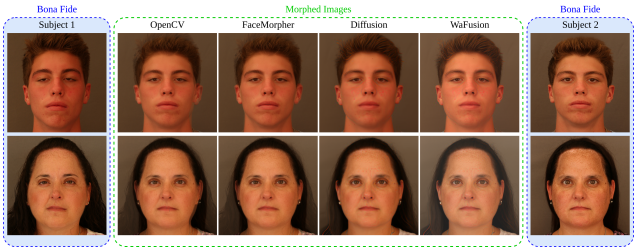


Figure 4. Visual comparison of morph images generated by OpenCV, FaceMorpher, Diffusion, and WaFusion using the WVU Twin dataset. Bona fide images are shown in blue boxes and morphs are shown in green boxes.

4.4. Vulnerability Analysis

This section evaluates the proposed WaFusion framework using the FRGC, FERET, FRLL, and WVU Twin datasets. The WVU Twin dataset provides controlled and high-similarity morphs. To further evaluate across broader appearance variations and conditions, we use FRLL, FERET, and FRGC. This multi-dataset evaluation ensures generalization beyond highly similar subjects. The evaluation includes visual comparisons, quantitative metrics, and an ablation study to assess individual component contributions.

A visual comparison of morph generation methods is presented in Fig. 4, using the WVU Twin dataset. WaFusion generates morphs at 512×512, surpassing the 256×256 resolution of the Diffusion approach. This demonstrates WaFusion's efficiency in achieving higher-quality morphs without increasing computational costs. Additionally, WaFusion benefits from targeting the LL sub-band, which captures most of the structural content, while averaging the high-frequency sub-bands to reduce unnecessary computation.

WaFusion morphs exhibit superior realism and detail, particularly in finer features such as facial textures and hair. In contrast, Diffusion morphs lack resolution, and traditional methods like OpenCV and FaceMorpher produce visible artifacts in facial contours and hairlines. The WVU Twin dataset, characterized by highly similar identities, poses challenges for all methods, yet WaFusion consistently minimizes artifacts and delivers visually consistent results. This highlights the framework's robustness even under extreme inter-subject similarity.

The metrics in Table 1 provide a quantitative comparison of WaFusion with OpenCV, FaceMorpher, StyleGAN, and Diffusion across various datasets, using AUC, APCER, BPCER, and EER. WaFusion stands out with the lowest EER and competitive results across most datasets, highlighting its effectiveness in generating realistic and diverse morphs. On the FRGC and FERET, WaFusion outperforms other methods, achieving high-quality morphs with minimal artifacts. On the FRLL, StyleGAN slightly outperforms WaFusion in some metrics due to the dataset's limited diversity, but WaFusion's morphs remain visually more realistic. For WVU Twin, WaFusion achieves the lowest AUC and EER, demonstrating its effectiveness even with high-similarity identities.

Fig. 3 and Fig. 5 illustrate the APCER-BPCER and ROC curves, respectively, using the FaceNet, further validating

Table 1. Performance of FaceNet verifier across both high-similarity (WVU Twin) and diverse-appearance datasets (FRLL, FERET, FRGC) evaluated using AUC, APCER, BPCER, and EER.

Datasets	Methods	AUC ↓	APCER@BPCER (%) ↑			BPCER@APCER(%) ↑			EER (%) ↑
			5%	10%	30%	5%	10%	30%	
FRGC	OpenCV [30]	0.7267	71.294	61.351	35.647	84.631	71.339	40.291	34.145
	FaceMorpher [40]	0.6735	77.861	69.231	46.341	91.069	80.166	47.04	37.523
	StyleGAN [24]	0.5073	94.746	88.555	68.292	94.496	88.681	71.131	50.281
	Diffusion [38]	0.4955	94.559	86.679	68.292	96.053	91.381	73.624	50.656
	WaFusion	**0.4843**	**95.717**	**89.453**	**71.422**	**96.148**	**92.487**	**75.861**	**52.593**
FERET	OpenCV [30]	0.7295	71.455	60.302	36.862	85.037	72.727	37.31	32.703
	FaceMorpher [40]	0.6798	77.315	68.241	45.557	88.825	76.515	43.56	37.618
	StyleGAN [24]	0.6295	83.742	76.37	52.741	91.477	82.765	50.568	41.398
	Diffusion [38]	0.501	93.383	90.359	67.296	97.348	90.909	69.696	50.85
	WaFusion	**0.4923**	**94.424**	**90.276**	**70.473**	**97.299**	**92.247**	**73.163**	**51.856**
FRLL	OpenCV [30]	0.5741	85.294	73.529	56.862	95.983	91.147	65.573	46.078
	FaceMorpher [40]	0.5848	88.235	77.45	56.862	95.331	88.615	61.179	44.117
	WebMorpher [10]	0.6	80.392	73.529	52.941	94.098	86.803	60.163	43.137
	StyleGAN [24]	**0.4793**	93.137	88.235	**71.568**	**97.788**	**93.529**	74.856	**51.96**
	Diffusion [38]	0.5149	93.137	88.235	61.764	95.61	89.64	70.676	49.019
	WaFusion	0.4909	**94.403**	**90.662**	65.177	97.496	91.891	**75.891**	50.798
WVU Twin	OpenCV [30]	0.5349	93.886	86.876	65.342	93.797	88.246	65.197	47.507
	FaceMorpher [40]	0.5315	93.964	87.5	65.537	94.001	88.517	65.678	47.507
	Diffusion [38]	0.4829	94.859	89.369	71.495	95.847	91.729	73.335	50.506
	WaFusion	**0.4715**	**96.296**	**91.583**	**73.983**	**96.934**	**93.274**	**75.218**	**52.681**

WaFusion's effectiveness. These curves demonstrate our method's ability to generate more challenging morphs for verifiers, particularly at stricter thresholds, achieving a favorable balance between attack success rates and bona fide misclassification rates across all datasets.

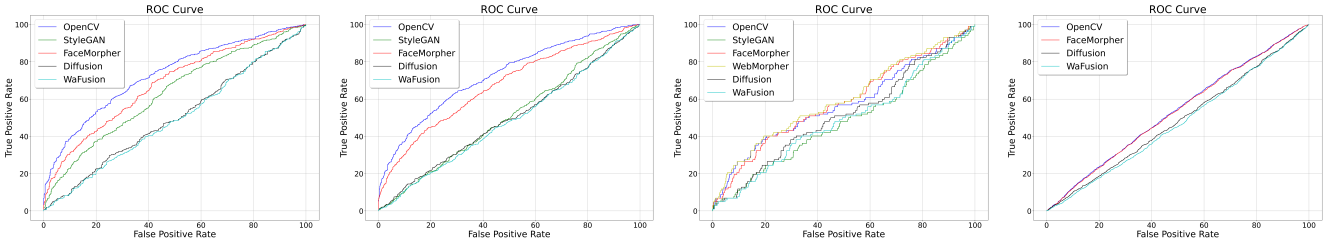
A broader visual comparison, shown in Fig. 6, highlights WaFusion's superior realism and consistency across morphs generated for the FRGC and FERET. While StyleGAN produces high-quality images, its morphs lack realism due to GAN-specific limitations such as mode collapse. Diffusion-based morphs suffer from resolution constraints, while traditional methods like OpenCV and FaceMorpher generate artifacts in facial features. WaFusion consistently delivers detailed, realistic morphs, reinforcing its adaptability and scalability across datasets with varying characteristics.

4.5. Ablation Study

The ablation study in Table 2 compares two variations of WaFusion: morphing all four sub-bands (LL, LH, HL, HH) versus morphing only the LL sub-band while averaging the high-frequency sub-bands. Results using Learned Perceptual Image Patch Similarity (LPIPS) [68] and Structural Similarity Index Measure (SSIM) [61] show that LL-only morphing achieves comparable performance, as the LL sub-band captures most structural information, while high-frequency sub-bands mainly represent edge details.

Table 2. Comparison of morph generation methods on the Twin dataset, using LPIPS and SSIM metrics.

Method	LPIPS ↓	SSIM ↑
WaFusion (All sub-bands)	0.243	0.792
WaFusion (LL sub-band)	**0.239**	**0.798**



(a) (b) (c) (d)

Figure 5. ROC curves for FaceNet verifier across (a) FERET, (b) FRGC, (c) FRLL, and (d) WVU Twin datasets.

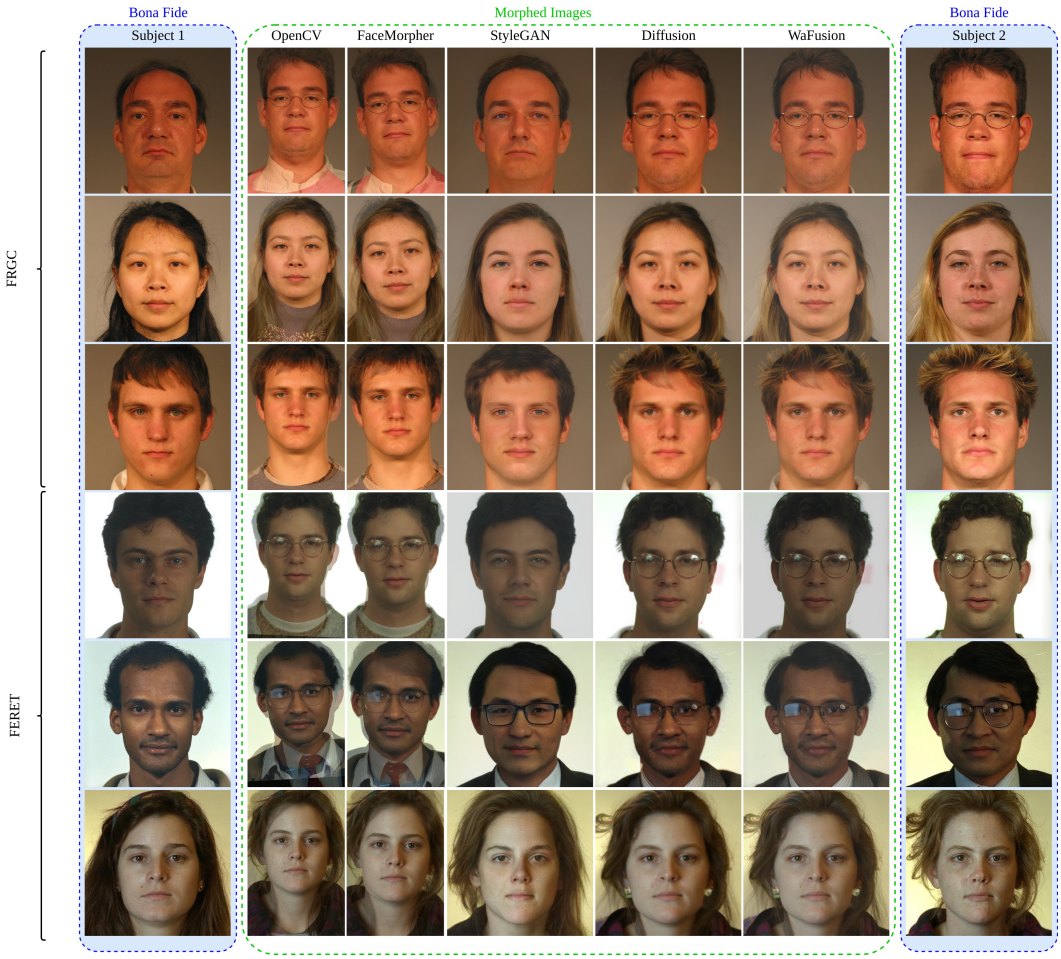


Figure 6. Morph images generated by different methods on the FRGC and FERET datasets, showing bona fide images (blue boxes) and morphs (green boxes).

Moreover, processing all four sub-bands requires approximately four times more computation than morphing, confirming the design's efficiency without compromising morph quality.

4.6. Future Work

Future work will focus on improving the scalability, generalization, and performance of WaFusion. While this study uses a single-level Haar wavelet for simplicity, future versions will explore multi-level decompositions and alternative bases (e.g., Daubechies) to capture fine-grained morphological details. We will also investigate selective processing of high-frequency sub-bands at higher wavelet levels to better balance morph quality and computation. Another direction involves extending WaFusion to real-time scenarios, such as video-based biometric authentication. These enhancements will further position WaFusion as a robust and adaptable framework for biometric security.

5. Conclusion

In this study, we introduce WaFusion, a hybrid framework combining wavelet decomposition and diffusion models to generate high-quality facial morphs. WaFusion preserves structural integrity and fine texture while minimizing perceptual artifacts, all without additional computational cost. Extensive evaluations on FRGC, FERET, FRLL, and WVU Twin demonstrate WaFusion's superiority over state-of-the-art methods across biometric metrics including APCER, BPCER, and EER. Its ability to produce realistic and challenging morphs underscores its potential to advance morph generation and strengthen biometric security systems.

References

[1] K. Ashwini, D. Nagajyothi, C. Ramakrishna, and V. Jyothi. Generation of high-quality realistic faces with StyleGAN. In

2023 4th IEEE Global Conference for Advancement in Technology (GCAT), pages 1–7. IEEE, 2023.

[2] Z. Blasingame and C. Liu. Leveraging diffusion for strong and high quality face morphing attacks. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2024.

[3] Z. W. Blasingame and C. Liu. Fast-dim: Towards fast diffusion morphs. *IEEE Security & Privacy*, 2024.

[4] K. W. Bowyer. Face recognition technology: security versus privacy. *IEEE Technology and society magazine*, 23(1):9–19, 2004.

[5] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman. Vggface2: A dataset for recognising faces across pose and age. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, pages 67–74. IEEE, 2018.

[6] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transactions on pattern analysis and machine intelligence*, 23(6):681–685, 2001.

[7] N. Damer, K. Raja, M. Süßmilch, S. Venkatesh, F. Boutros, M. Fang, F. Kirchbuchner, R. Ramachandra, and A. Kuijper. Regenmorph: Visibly realistic GAN generated face morphing attacks by attack re-generation. In *Advances in Visual Computing: 16th International Symposium, ISVC 2021, Virtual Event, October 4-6, 2021, Proceedings, Part I*, pages 251–264. Springer, 2021.

[8] N. Damer, A. M. Saladie, A. Braun, and A. Kuijper. MorGAN: Recognition vulnerability and attack detectability of face morphing attacks created by generative adversarial network. In *2018 IEEE 9th international conference on biometrics theory, applications and systems (BTAS)*, pages 1–10. IEEE, 2018.

[9] I. Daubechies. *Ten lectures on wavelets*. SIAM, 1992.

[10] L. DeBruine. debruine/webmorph: Beta release 2. *Zenodo https://doi. org/10*, 5281, 2018.

[11] L. DeBruine and B. Jones. Face research lab london set. *Psychol. Methodol. Des. Anal*, 3, 2017.

[12] P. Dhariwal and A. Nichol. Diffusion models beat GANs on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

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

[14] M. Ferrara, A. Franco, and D. Maltoni. Decoupling texture blending and shape warping in face morphing. In *2019 international conference of the biometrics special interest group (BIOSIG)*, pages 1–5. IEEE, 2019.

[15] I. O. for Standardization. ISO/IEC 30107-3:2023: Information technology – Biometric presentation attack detection. Technical report, ISO, 2023.

[16] S. R. Godage, F. Løvåsdal, S. Venkatesh, K. Raja, R. Ramachandra, and C. Busch. Analyzing human observer ability in morphing attack detection—where do we stand? *IEEE Transactions on Technology and Society*, 4(2):125–145, 2022.

[17] M. Gomez-Barrero, K. B. Raja, C. Rathgeb, A. F. Sequeira, M. Todisco, L. Colbois, and S. Marcel. On the detection of morphing attacks generated by GANs. *2022 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–5, 2022.

[18] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

[19] M. Hamza, S. Tehsin, M. Humayun, M. F. Almufareh, and M. Alfayad. A comprehensive review of face morph generation and detection of fraudulent identities. *Applied Sciences*, 12(24):12545, 2022.

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

[21] Y. Huang, J. Huang, J. Liu, Y. Dong, J. Lv, and S. Chen. Wavedm: Wavelet-based diffusion models for image restoration. *arXiv preprint arXiv:2305.13819*, 2023.

[22] D. ICAO. 9303-machine readable travel documents-part 9: Deployment of biometric identification and electronic storage of data in emrtds. *International Civil Aviation Organization (ICAO)*, 123, 2015.

[23] M. Ivanovska and V. Štruc. Face morphing attack detection with denoising diffusion probabilistic models. In *2023 11th International Workshop on Biometrics and Forensics (IWBF)*, pages 1–6. IEEE, 2023.

[24] T. Karras, S. Laine, and T. Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019.

[25] R. Kessler, K. Raja, J. Tapia, and C. Busch. Towards minimizing efforts for morphing attacks—deep embeddings for morphing pair selection and improved morphing attack detection. *Plos one*, 19(5):e0304610, 2024.

[26] J. Kim, C. Oh, H. Do, S. Kim, and K. Sohn. Diffusion-driven gan inversion for multi-modal face image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10403–10412, 2024.

[27] G. R. Lee, R. Gommers, F. Waselewski, K. Wohlfahrt, and A. O'Leary. Pywavelets: A python package for wavelet analysis. *Journal of Open Source Software*, 4(36):1237, 2019.

[28] A. Makrushin, D. Siegel, and J. Dittmann. Simulation of border control in an ongoing web-based experiment for estimating morphing detection performance of humans. In *Proceedings of the 2020 ACM Workshop on Information Hiding and Multimedia Security*, pages 91–96, 2020.

[29] S. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11:674–693, 1989.

[30] S. Mallick. Face morph using opencv—c++/python. *LearnOpenCV*, 1(1), 2016.

[31] T. Neubert, A. Makrushin, M. Hildebrandt, C. Kraetzer, and J. Dittmann. Extended stirtrace benchmarking of biometric and forensic qualities of morphed face images. *Iet Biometrics*, 7(4):325–332, 2018.

[32] A. Q. Nichol and P. Dhariwal. Improved denoising diffusion probabilistic models. In *International conference on machine learning*, pages 8162–8171. PMLR, 2021.

[33] K. O'Haire, S. Soleymani, B. Chaudhary, P. Aghdaie, J. M. Dawson, and N. M. Nasrabadi. Adversarially perturbed wavelet-based morphed face generation. *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, pages 01–05, 2021.

[34] K. O'Haire, S. Soleymani, B. Chaudhary, J. Dawson, and N. M. Nasrabadi. Identical twins face morph database generation. In *2022 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–9. IEEE, 2022.

[35] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 947–954. IEEE, 2005.

[36] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss. The feret database and evaluation procedure for face-recognition algorithms. *Image and vision computing*, 16(5):295–306, 1998.

[37] H. Phung, Q. Dao, and A. Tran. Wavelet diffusion models are fast and scalable image generators. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10199–10208, 2023.

[38] K. Preechakul, N. Chatthee, S. Wizadwongsa, and S. Suwajanakorn. Diffusion autoencoders: Toward a meaningful and decodable representation. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10609–10619, 2021.

[39] S. Price, S. Soleymani, and N. M. Nasrabadi. Landmark enforcement and style manipulation for generative morphing. In *2022 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–10. IEEE, 2022.

[40] A. Quek. Facemorpher, 2019.

[41] R. Raghavendra, K. B. Raja, and C. Busch. Detecting morphed face images. *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–7, 2016.

[42] K. Raja, M. Ferrara, A. Franco, L. Spreeuwers, I. Batskos, F. De Wit, M. Gomez-Barrero, U. Scherhag, D. Fischer, S. K. Venkatesh, et al. Morphing attack detection-database, evaluation platform, and benchmarking. *IEEE transactions on information forensics and security*, 16:4336–4351, 2020.

[43] R. Ramachandra and C. Busch. Presentation attack detection methods for face recognition systems: A comprehensive survey. *ACM Computing Surveys (CSUR)*, 50(1):1–37, 2017.

[44] R. Ramachandra, S. Venkatesh, N. Damer, N. Vetrekar, and R. S. Gad. Multispectral imaging for differential face morphing attack detection: A preliminary study. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 6185–6193, 2024.

[45] D. J. Robertson, R. S. Kramer, and A. M. Burton. Fraudulent ID using face morphs: Experiments on human and automatic recognition. *PLoS One*, 12(3):e0173319, 2017.

[46] D. Roich, R. Mokady, A. H. Bermano, and D. Cohen-Or. Pivotal tuning for latent-based editing of real images. *ACM Transactions on graphics (TOG)*, 42(1):1–13, 2022.

[47] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.

[48] O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.

[49] M. S. E. Saadabadi, S. R. Malakshan, S. R. Hosseini, and N. M. Nasrabadi. Boosting unconstrained face recognition with targeted style adversary. In *2024 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–11. IEEE, 2024.

[50] E. Sarkar, P. Korshunov, L. Colbois, and S. Marcel. Vulnerability analysis of face morphing attacks from landmarks and generative adversarial networks. *arXiv preprint arXiv:2012.05344*, 2020.

[51] E. Sarkar, P. Korshunov, L. Colbois, and S. Marcel. Are GAN-based morphs threatening face recognition? In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2959–2963. IEEE, 2022.

[52] U. Scherhag, L. Debiasi, C. Rathgeb, C. Busch, and A. Uhl. Detection of face morphing attacks based on PRNU analysis. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 1(4):302–317, 2019.

[53] U. Scherhag, A. Nautsch, C. Rathgeb, M. Gomez-Barrero, R. N. Veldhuis, L. Spreeuwers, M. Schils, D. Maltoni, P. Grother, S. Marcel, et al. Biometric systems under morphing attacks: Assessment of morphing techniques and vulnerability reporting. In *2017 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–7. IEEE, 2017.

[54] U. Scherhag, C. Rathgeb, J. Merkle, and C. Busch. Deep face representations for differential morphing attack detection. *IEEE transactions on information forensics and security*, 15:3625–3639, 2020.

[55] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.

[56] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pages 2256–2265. PMLR, 2015.

[57] J. Song, C. Meng, and S. Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.

[58] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[59] S. Venkatesh, R. Ramachandra, K. Raja, and C. Busch. Face morphing attack generation and detection: A comprehensive survey. *IEEE transactions on technology and society*, 2(3):128–145, 2021.

[60] S. Venkatesh, H. Zhang, R. Ramachandra, K. Raja, N. Damer, and C. Busch. Can GAN generated morphs

threaten face recognition systems equally as landmark based morphs?-vulnerability and detection. In *2020 8th International Workshop on Biometrics and Forensics (IWBF)*, pages 1–6. IEEE, 2020.

[61] Z. 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.

[62] G. Wolberg. Image morphing: a survey. *The visual computer*, 14(8-9):360–372, 1998.

[63] Z. Xiao, K. Kreis, and A. Vahdat. Tackling the generative learning trilemma with denoising diffusion GANs. *arXiv preprint arXiv:2112.07804*, 2021.

[64] T. Yang and Y. Ying. Auc maximization in the era of big data and ai: A survey. *ACM Computing Surveys*, 55(8):1–37, 2022.

[65] H. Zhang, R. Ramachandra, K. Raja, and C. Busch. Generalized single-image-based morphing attack detection using deep representations from vision transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1510–1518, 2024.

[66] H. Zhang, S. Venkatesh, R. Ramachandra, K. Raja, N. Damer, and C. Busch. MIPGAN—generating strong and high quality morphing attacks using identity prior driven GAN. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 3(3):365–383, 2021.

[67] K. Zhang, Y. Zhou, X. Xu, B. Dai, and X. Pan. Diffmorpher: Unleashing the capability of diffusion models for image morphing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7912–7921, 2024.

[68] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018.

[69] B. Zope and S. B. Zope. A survey of morphing techniques. *International Journal of Advanced Engineering, Management and Science*, 3(2):239773, 2017.