DCMorph: Face Morphing via Dual-Stream Cross-Attention Diffusion

Anonymous CVPR submission

Paper ID 21

Abstract

Advancing face morphing attack techniques is crucial to anticipate evolving threats and develop robust defensive mechanisms for identity verification systems. This work introduces DCMorph, a dual-stream diffusion-based morphing framework that simultaneously operates at both identity conditioning and latent space levels. Unlike image-level methods suffering from blending artifacts or GAN-based approaches with limited reconstruction fidelity, DCMorph leverages identity-conditioned latent diffusion models through two mechanisms: (1) decoupled cross-attention interpolation that injects identity-specific features from both source faces into the denoising process, enabling explicit dual-identity conditioning absent in existing diffusion-based methods, and (2) DDIM inversion with spherical interpolation between inverted latent representations from both source faces, providing geometrically consistent initial latent representation that preserves structural attributes. Vulnerability analyses across four state-of-the-art face recognition systems demonstrate that DCMorph achieves the highest attack success rates compared to existing methods at both operational thresholds, while remaining challenging to detect by current morphing attack detection solutions.

1. Introduction

Face recognition (FR) systems, despite their high accuracy, remain vulnerable to face morphing attacks, which create images that can be verified as belonging to multiple identities. Such attacks pose significant security threats in identity verification scenarios, potentially enabling unauthorized access for multiple people with the same document [19]. Understanding and developing advanced morphing techniques is essential to anticipate evolving attacks and strengthen defensive mechanisms against increasingly sophisticated threats.

Existing morphing approaches operate at two levels: image-level or representation-level. Image-level methods interpolate facial landmarks and blend textures, achieving strong identity preservation but suffering from visible blending artifacts [20, 36]. Representation-level methods, primarily GAN-based [10, 51], avoid such artifacts but exhibit limited reconstruction fidelity due to constrained latent space capacity. Recent diffusion-based morphing methods [2, 14] achieves higher visual fidelity through latent space interpolation but operates without explicit identity-aware conditioning during the generation process, limiting its ability to precisely control identity characteristics in the synthesized morphs.

The emergence of identity-conditioned latent diffusion models (DM) [6, 30], which leverage cross-attention mechanisms to inject identity information into the denoising process, combined with DDIM inversion [46] that enables high-fidelity latent recovery, provides an opportunity to fundamentally advance face morphing. These developments enable simultaneous manipulation at both the conditioning pathway and the latent space, facilitating more effective identity blending while maintaining structural consistency.

This work introduces DCMorph, a dual-stream morphing framework that extends identity-conditioned DMs for face morphing. DCMorph combines decoupled cross-attention interpolation, which merges identity-specific features from both source faces during the denoising process, with spherical interpolation of DDIM-inverted latents, which provides geometrically consistent initialization for the denoising process. Our main contributions are:

- A novel dual-stream morphing framework that simultaneously leverages identity conditioning via decoupled cross-attention and latent space interpolation via DDIM inversion, enabling more effective identity blending than existing approaches.
- Comprehensive vulnerability and detectability analyses demonstrating that DCMorph produces morphing attacks that are very effective against state-of-the-art FR systems in comparison to a set of both traditional and recent diffusion-based morphing methods, while posing challenging detectability characteristics.

2. Related Work

Morphing Attack Generation Methods: Face morphing attacks create images verifiable as multiple identi-

CVPR #21

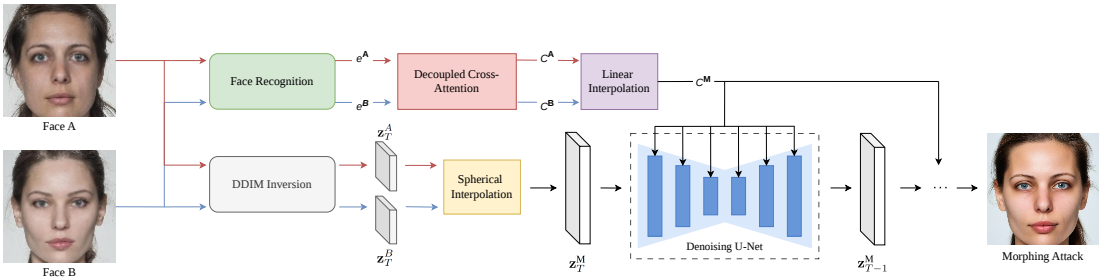

Figure 1. Overview of the DCMorph dual-stream framework. Identity embeddings $\mathbf{e}^A$ and $\mathbf{e}^B$ are extracted from input faces and injected via decoupled cross-attention layers to produce $\mathbf{C}^A$ and $\mathbf{C}^B$, which are linearly interpolated to obtain $\mathbf{C}^{\mathrm{M}}$. Simultaneously, DDIM inversion recovers latent representations $\mathbf{z}_T^A$ and $\mathbf{z}_T^B$, which are spherically interpolated to produce $\mathbf{z}_T^{\mathrm{M}}$. The U-Net then performs iterative denoising ($\mathbf{z}_T^{\mathrm{M}} \rightarrow \mathbf{z}_{T-1}^{\mathrm{M}} \rightarrow \cdots \rightarrow \mathbf{z}_0^{\mathrm{M}}$) conditioned on $\mathbf{C}^{\mathrm{M}}$, generating a high-fidelity morphing attack blending identity characteristics from both sources through dual-stream manipulation at the conditioning and latent space levels.

ties, posing security threats to face recognition systems [19]. Existing approaches are categorized as image-level or representation-level, each with distinct characteristics. **Image-level Morphing:** The earliest morphing attacks were created by detecting facial landmarks in source images, interpolating these landmarks, and blending texture information. Landmark-based morphs (LMA) [20, 36] have been widely studied, with early comparisons [40] showing certain approaches [9, 36] achieved strong identity preservation. More advanced techniques perform partial-region interpolation [35], producing morphs that are more challenging to detect. Despite their effectiveness, image-level morphs inherently suffer from blending artifacts due to pixel-level manipulation and geometric inconsistencies [51]. **Representation-level Morphing:** To overcome these limitations, representation-level approaches perform interpolation in learned latent spaces. MorGAN [10] pioneered this direction, later enhanced through cascaded quality improvement [11]. Building on this foundation, StyleGAN-based approaches [47] and MIPGAN I/II [51] advanced the field through identity-preserving losses. However, GAN-based methods still suffer from limited reconstruction fidelity due to constrained latent capacity [34], causing synthetic artifacts [51]. More recently, diffusion-based morphing has demonstrated improved fidelity, with MorDIFF [14] and Blasingame and Liu [2] performing interpolation in diffusion latent spaces to achieve superior reconstruction quality. Beyond core generation techniques, post-processing methods have been developed to improve morph realism. ReGenMorphs [12] combines image-level morphing with GAN refinement, while Borghi et al. [3] and Di Domenico et al. [17] proposed attention-based and face restoration techniques for artifact retouching. Exploring alternative domains, Singh and Ramachandra [43, 44] investigated 3D morphing via geometric interpolation and non-rigid registration, while MorCode [32] introduced codebook-based morphing using discrete latent representations. Closely related to morph generation are demorphing methods, which attempt to recover the contributing identities from a morphed image. Shukla and Ross [41] proposed identity-preserving decomposition using generative frameworks, later introducing dc-GAN [42] for dual-conditioned reconstruction of individual source identities.

Diffusion Models for Face Generation: Diffusion probabilistic models (DPMs) [22, 45] achieve superior visual fidelity over GANs by reversing gradual noising. Building on this, Latent Diffusion Models (LDMs) [37] operate in compressed latent space for efficient high-resolution generation, with U-Net denoising networks learning to remove noise conditioned on text, class labels, or identity embeddings. **Identity-Conditioned Diffusion:** Recent methods [6, 8, 29, 30, 49] inject identity embeddings from pre-trained FR models into denoising networks via cross-attention, enabling dynamic identity adaptation and high-fidelity identity-consistent synthesis. **Diffusion-based Morphing:** Leveraging these advances, diffusion autoencoders [34] encode semantic and stochastic information for meaningful latent manipulations. MorDIFF [14] exploits this by interpolating these codes for high-fidelity morphing attacks. However, existing diffusion morphing performs only latent interpolation without explicit identity-aware conditioning during denoising.

DCMorph extends this work through dual-stream processing: decoupled cross-attention for dual-identity conditioning and DDIM inversion [46] with spherical interpolation for geometrically consistent latent initialization, enabling effective identity blending with high visual fidelity.

3. Methodology

This section presents Dual-Stream Cross-Attention Morphing, a dual-identity conditioning framework for controlled identity morphing built upon pre-trained identity-conditional latent diffusion models (LDMs).

We begin this section by formalizing the LDM denoising process and the standard approach to identity-conditioned generation via cross-attention injection, which serves as the generative backbone of our morphing framework. We then introduce our two complementary morphing streams: De-

CVPR #21

coupled Identity Cross-Attention Interpolation, which injects identity-specific attention features from both source faces within the denoising network, U-Net, of LDM to guide the conditional generation toward a morphed identity representation, and Spherical Inverted Latents Interpolation, which applies DDIM inversion to each source image and performs spherical linear interpolation between the recovered latents to provide a geometrically consistent initialization for reverse diffusion. An overview of the full pipeline is illustrated in Figure 1.

3.1. Latent DM Preliminaries

LDM [37] improves the efficiency of diffusion-based image generation by operating in a compressed latent space rather than directly in pixel space. Given a training image $\mathbf{x}_0 \in \mathbb{R}^{H\times W\times C}$, a pre-trained encoder $\mathcal{E} : \mathbb{R}^{H\times W\times C} \to \mathbb{R}^d$ maps it to a latent representation $\mathbf{z}_0 = \mathcal{E}(\mathbf{x}_0)$.

To learn the generative process, LDMs adopt the denoising diffusion probabilistic model (DDPM) framework [22], where the latent code $\mathbf{z}_0$ is gradually perturbed by Gaussian noise over T discrete timesteps:

$$q(\mathbf{z}_t|\mathbf{z}_{t-1}) = \mathcal{N}(\mathbf{z}_t; \sqrt{1-\beta_t}\mathbf{z}_{t-1}, \beta_t\mathbf{I}), \quad (1)$$

where $\{\beta_t\}_{t=1}^T$ is a predefined variance schedule. After sufficient noising steps, the latent $\mathbf{z}_T$ approximates an isotropic Gaussian distribution, i.e., $\mathbf{z}_T \sim \mathcal{N}(0, \mathbf{I})$.

The reverse process is modeled by a neural network θ, typically implemented as a U-Net [37], which is trained to denoise the latent variable by predicting the noise component added at each step. The network takes as input the noisy latent $\mathbf{z}_t$, the timestep t, and an optional conditioning signal $\mathbf{c} \in \mathbb{R}^d$, such as an identity embedding from a pre-trained face recognition model. The reverse process is thus formulated as:

$$p_\theta(\mathbf{z}_{t-1}|\mathbf{z}_t, \mathbf{c}) = \mathcal{N}(\mathbf{z}_{t-1}; \mu_\theta(\mathbf{z}_t, t, \mathbf{c}), \Sigma_\theta(\mathbf{z}_t, t, \mathbf{c})). \quad (2)$$

After the denoising trajectory is complete, the final latent $\hat{\mathbf{z}}_0$ is passed through a pre-trained decoder $\mathcal{D} : \mathbb{R}^d \to \mathbb{R}^{H\times W\times C}$, to reconstruct image $\hat{\mathbf{x}}_0 = \mathcal{D}(\hat{\mathbf{z}}_0)$.

Classifier-Free Guidance (CFG): Classifier-Free Guidance (CFG) [21] is applied to reinforce the impact of the condition (e.g., identity) during sampling [30]. It is effective in a wide range of conditional generation tasks, including both image- and text-conditioned synthesis [1, 30, 48].

At inference time, the final noise prediction is obtained by linearly combining the conditional and unconditional outputs:

$$\hat{\epsilon} = (1+\omega)\hat{\epsilon}_\theta(\mathbf{z}_t, t, \mathbf{c}) - \omega\hat{\epsilon}_\theta(\mathbf{z}_t, t), \quad (3)$$

where ω denotes the guidance strength. Increasing ω amplifies the contribution of the identity-conditioned prediction, leading to stronger adherence to the specified identity and improved identity consistency in the generated samples.

Identity-conditional LDMs: To generate identity-conditioned face images using an LDM, the common approach [6, 29, 30, 49] is to utilize identity representation $\mathbf{c}$ extracted from a pretrained FR model as a condition for LDM. The identity representation $\mathbf{c}$ is injected into the U-Net via cross-attention layers [37]. This mechanism projects the identity context from a single source into the denoising network's intermediate feature representations, allowing the generative process to adapt dynamically to the given identity [6, 37]. In contrast to single-identity conditioning, identity morphing requires simultaneous conditioning on two source images, enabling the generative process to synthesize a representation that integrates identity characteristics from both inputs. The next section describes our morphing approach that injects identity-representation from two source images into the denoising process.

3.2. DCMorphing: Dual-Stream Cross-Attention Morphing

This section introduces **Dual-Stream Cross-Attention Morphing (DCMorphing)**, a dual-stream morphing framework that extends identity-conditioned LDMs by jointly combining identity information in both the conditioning pathway and the latent space. The two streams operate at complementary levels: cross-attention interpolation merges identity features during denoising, while latent interpolation defines a geometrically meaningful generative trajectory.

Decoupled Identity Cross-Attention Interpolation: Let $\mathbf{x}^A, \mathbf{x}^B$ denote two input face images to be morphed. Identity embeddings are extracted as:

$$\mathbf{c}^A = f(\mathbf{x}^A), \quad \mathbf{c}^B = f(\mathbf{x}^B). \quad (4)$$

Each embedding is injected into the denoising network, U-Net, via decoupled cross-attention layers. For identity A, the output of the cross-attention is defined as:

$$\mathbf{C}^A = \text{Attn}(Q, K^A, V^A) = \text{softmax}\Big(\frac{Q(K^A)^\top}{\sqrt{d}}\Big)V^A, \quad (5)$$

where the query (Q), key (K), and value (V) matrices are computed as $Q = FW_q$, $K^A = \mathbf{c}^A W_k$, and $V^A = \mathbf{c}^A W_v$, with F denoting the query features from the U-Net, and W_q, W_k, W_v being the weight matrices of the trainable linear projection layers.

Similarly, we compute $\mathbf{C}^B = \text{Attn}(Q, K^B, V^B)$ using its respective identity embedding $\mathbf{c}^B$:

$$\mathbf{C}^B = \text{Attn}(Q, K^B, V^B) = \text{softmax}\Big(\frac{Q(K^B)^\top}{\sqrt{d}}\Big)V^B. \quad (6)$$

Noting that the W_q, W_k, W_v weights are shared between the two cross-attention layers. These cross-attention outputs are then linearly interpolated to produce a mixed identity repre-

CVPR #21

Figure 2. Samples of the DCMorph (right most column) attacks, the baseline attacks (created by FaceMorpher, OpenCV, WebMorph, MorDIFF, MIPGANI and II), and the bona fide images that were morphed to creat the attack (two left-most columns). The image-level morphs (FaceMorpher, OpenCV, WebMorph) show the traditional blending artifacts, while the representation-level morphs (MIPGAN-I and II) show typical streaking GAN artifacts. Diffusion-based approaches, such as ours and MorDIFF, exhibit substantially fewer generative artifacts compared to GAN-based methods.

sentation for downstream generation. We denote linear interpolation as $r_l(\lambda; A, B)$, where $\lambda \in [0, 1]$ is the interpolation coefficient between the anchors A and B. The final formulation of the decoupled interpolated cross-attention $\mathbf{C}^{\mathrm{M}}$ is therefore defined as:

$$\mathbf{C}^{\mathrm{M}} = r_l(\lambda; \mathbf{C}^A, \mathbf{C}^B) = \lambda \mathbf{C}^A + (1 - \lambda)\mathbf{C}^B, \quad (7)$$

where λ controls the relative contribution of the two identity-specific cross-attention outputs. In our experiments, we set $\lambda = 0.5$, assigning equal weight to both components. As a result, the synthesized representation reflects an equal contribution of identity-specific features from both sources. This balanced interpolation in the attention space encourages the generated output to retain characteristic attributes of each identity, while ensuring structural and semantic consistency in the resulting morph.

Spherical Inverted Latents Interpolation: In addition to performing decoupled identity cross-attention interpolation, we also operate in the latent space by applying DDIM inversion [46] to each input image to recover their corresponding latent representations, and subsequently perform spherical interpolation between these latents to generate an intermediate representation that effectively captures a balanced combination of the original face images. For input face images $\mathbf{x}^A$ and $\mathbf{x}^B$ corresponding to identities A and B, we denote the DDIM inversion process by $d(\cdot)$, which maps an image to its latent representation by iteratively adding noise:

$$\mathbf{z}_T^A = \mathrm{d}(\mathbf{x}^A), \quad \mathbf{z}_T^B = \mathrm{d}(\mathbf{x}^B), \quad \mathbf{z}_T^A, \mathbf{z}_T^B \sim \mathcal{N}(0, \mathbf{I}). \quad (8)$$

Ideally, direct sampling from $\mathbf{z}_T^A$ and $\mathbf{z}_T^B$ without any further manipulation should reconstruct an image that closely approximates $\mathbf{x}^A$ and $\mathbf{x}^B$, respectively. To interpolate between the two identities in latent space, we perform spherical linear interpolation (slerp) [27] between $\mathbf{z}_T^A$ and $\mathbf{z}_T^B$. Spherical interpolation is specifically employed to maintain the Gaussian statistics of the latent space [38], ensuring that the interpolated latent $\mathbf{z}_T^{\mathrm{M}}$ remains consistent with the distribution of valid latents. We denote spherical interpolation as $r_s(\lambda; A, B)$, respectively, where $\lambda \in [0, 1]$ represents the interpolation coefficient, and A and B denote the interpolation anchors. With $\theta = arccos\frac{\mathbf{z}_T^A \cdot \mathbf{z}_T^B}{\|\mathbf{z}_T^A\| \|\mathbf{z}_T^B\|}$, the interpolated latent $\mathbf{z}_T^{\mathrm{M}}$ is defined as:

$$\mathbf{z}_T^{\mathrm{M}} = r_s(\lambda; \mathbf{z}_T^A, \mathbf{z}_T^B) = \frac{\sin(1-\lambda)\theta}{\sin\theta}\mathbf{z}_T^A + \frac{\sin(\lambda\theta)}{\sin\theta}\mathbf{z}_T^B, \quad (9)$$

This procedure produces an intermediate latent representation that is intended to mix the identity-specific features

CVPR #21

of both inputs, providing a suitable starting point for subsequent generation or morphing. The reverse diffusion process is then applied to $\mathbf{z}_T^{\mathrm{M}}$ to generate the corresponding interpolated image, which is expected to reflect characteristics from both identities in a balanced manner.

The interpolation strategies are designed to contribute distinct but synergistic benefits to the morphing process. Interpolating identities via decoupled cross-attention interpolation aims to influence the model's conditional feature aggregation, guiding the denoising process toward a representation that contains features from both identities during generation, consistent with how cross-attention enables flexible conditioning in LDMs [37]. In addition, spherical interpolation in latent space preserving high-level structural and semantic attributes encoded in the latent space, this approach provides a stable initialization for the diffusion process, providing a geometrically consistent trajectory in the generative prior, preserving global structure and semantic coherence. Collectively, these mechanisms aim to produce interpolated images that uphold both realistic generation quality and balanced representation of the source identities.

4. Experimental Setup

Morph Generation Protocol: The DCMorph dataset extends over the SYN-MAD 2022 competition [25] and uses the same morphing pairs to enable a comparable dataset. Both are based on the Face Research Lab London (FRLL) dataset [15]. The FRLL contains images of 102 different individuals and provides high-quality frontal images created in a controlled scenario with a wide range of different ethnicities. All individuals present in the dataset signed consent for their images to be used in lab-based and web-based studies in their original or altered forms and to illustrate research. For the morph generation, we limited the data to the frontal images of the dataset, following SYN-MAD 2022. The pairs are defined in SYN-MAD 2022 by splitting the frontal images of the FRLL depending on the provided gender and expression (neutral or smiling). ElasticFace-Arc [5] was then used to generate embeddings of the images and these embeddings are then compared with cosine similarity within the split to find the most similar faces. The 250 most similar face pairs are selected, resulting in 250 female neutral pairs, 250 female smiling pairs, 250 male neutral pairs, and 250 male smiling pairs, for a total of 1000 attack images and the 204 bona fide images of the SYN-MAD 2022.

Benchmark Datasets: We use the SYN-MAD 2022 [25] benchmark, containing morphed images from 5 different approaches, three image-level (FaceMorpher (commercial-of-the-shelf), OpenCV [31] and Webmorph (online tool [1]) and two representation-level GAN-based (MIPGAN I [51] and MIPGAN II [51]). Additionally, we add the more recent MorDIFF [14] a representation-level diffusion-based approach. We follow the same morph pair selection protocol and add our DCMorph attacks to the benchmark, which will be publicly released.

Vulnerability of FR systems: We evaluated the vulnerability of four FR systems to the DCMorph attacks in comparison to six different attacks. The FR systems are AdaFace [28], ArcFace [16], ElasticFace (ElasticFace-Arc) [5] and CurricularFace [24]. All considered models are based on ResNet-100 architectures and have 55.52M parameters with 24192.51 MFLOPs. All the FR models are the official releases by the respective authors. We present the vulnerability results by reporting the Mated Morphed Presentation Match Rate (MMPMR) [39], evaluated at decision thresholds corresponding to false match rates (FMR) of 1% and 0.1%, denoted as MMPMR100 and MMPMR1000, respectively. The FMR decision thresholds were calculated on the LFW [23] benchmark, following [25].

Detectability of Morphing Attacks: The MAD performance (detectability) is presented by the Attack Presentation Classification Error Rate (APCER), i.e. the proportion of attack images incorrectly classified as bona fide samples, at a fixed (1%, 10%, and 20%) Bona fide Presentation Classification Error Rate (BPCER), i.e. the proportion of bona fide images incorrectly classified as attack samples, as defined in the ISO/IEC 30107-3 [26]. Additionally, the Detection Equal Error Rate (EER), i.e. the value of APCER or BPCER at the decision threshold where they are equal, is reported. We use three MAD systems to evaluate the morphing attack detectability of the proposed dataset. MAD-PromptS [7] utilizes multiple prompt aggregation to exploit the full potential of zero-shot learning MAD in foundation models. SPL-MAD [18] is an unsupervised method based on self-paced learning trained for anomaly detection. The public released supervised MixFaceNet-MAD [13] uses the efficient MixFaceNet [4] architecture (originally developed for FR) trained with the binary cross-entropy loss function to detect morphing attacks.

Base Diffusion Model: The generative backbone of DCMorph is built upon Stable Diffusion XL (SDXL) [33], fine-tuned via IP-Adapter [50] to support conditional face generation [2]. Specifically, IP-Adapter extends the pre-trained SDXL model by introducing a lightweight adapter module that enables image-prompt conditioning via dual cross-attention layers, for text and image conditions, allowing face-identity embeddings to be injected into the denoising U-Net without modifying the original model weights. We eliminated the cross-attention layer of the text condition and duplicated the face identity condition layer, enabling dual conditioning from both source identities. The conditional face embeddings used to guide the generation processes are extracted from a pre-trained ArcFace [16] model.

[1] https://webmorph.org/

[2] https://huggingface.co/h94/IP-Adapter-FaceID

Table 1. Ablation study analyzing the vulnerability of four FR systems to different morphing strategies. The table compares embedding interpolation alone, cross-attention interpolation alone, embedding interpolation with DDIM inversion, and DCMorph (cross-attention with DDIM inversion). Embedding interpolation operates in the feature space of pre-trained FR models, cross-attention injects dual-identity conditioning into the denoising process, and DDIM inversion with spherical interpolation provides geometrically consistent latent initialization. DCMorph achieves the highest MMPMR values (bold) across all FR systems and both operational thresholds (FMR 1% and 0.1%), demonstrating the synergistic benefit of combining identity-conditioned generation with latent space interpolation.

Morphing Technique	Method			ElasticFace [5]		CurricularFace [24]		AdaFace [28]		ArcFace [16]	
	Embedding Interpolation	Cross-Attention	DDIM-Inversion	MMPMR100	MMPMR1000	MMPMR100	MMPMR1000	MMPMR100	MMPMR1000	MMPMR100	MMPMR1000
Embedding Interpolation	X			0.992	0.936	0.995	0.960	1.000	0.986	0.998	0.966
Cross-attention Interpolation		X		0.990	0.916	0.992	0.956	0.999	0.978	0.993	0.963
Embedding Interpolation + DDIM	X		X	0.997	0.959	0.996	0.969	0.999	0.991	0.993	0.963
DCMorph		X	X	1.000	0.965	0.999	0.981	1.000	0.995	0.998	0.982

Figure 3. Qualitative comparison of morphing approaches. Bona fide images (left columns) and the corresponding DCMorph attacks (right column), generated using decoupled cross-attention identity interpolation combined with spherical latent-space interpolation via DDIM inversion, are shown alongside the alternative approaches considered in this work. DDIM inversion preserves high-level latent semantics, enabling a geometrically consistent reverse diffusion process that maintains global structure and attributes such as head pose.

5. Results

We first conduct an ablation study to systematically evaluate the individual and combined contributions of embedding interpolation, cross-attention interpolation, and DDIM inversion to morph generation effectiveness. The results show that both interpolation strategies alone already produce highly vulnerable morphs, while incorporating DDIM inversion further strengthens identity preservation under stricter verification settings. The full DCMorph framework consistently achieves the strongest vulnerability across all FR systems. From the detectability perspective, the findings reveal that although some intermediate variants are more easily detected by certain MAD approaches, the complete DCMorph pipeline significantly increases detection difficulty for the considered MAD approaches. We then compare DCMorph with established baseline attack methods, analyzing both vulnerability and detectability across multiple evaluation scenarios. Figure 2 provides a qualitative comparison of DCMorph attacks with other morphing methods. The combined detectability and vulnerability results show that DCMorph achieves a favorable trade off by maximizing FR vulnerability while remaining difficult to detect. Even the zero shot foundation model MADPromptS demonstrates only moderate detection capability, with DCMorph remaining harder to detect than most attacks.

5.1. DCMorph Component Analysis

In this section, we assess the vulnerability and detectability of different morphing strategies to understand the individual and combined contributions of DCMorph's components. We evaluate four variants: (1) embedding interpolation alone, (2) cross-attention interpolation alone, (3) embedding interpolation combined with DDIM inversion, and (4) DCMorph, which combines cross-attention interpolation with DDIM inversion.

Vulnerability Analysis: Table 1 presents the vulnerability of four state-of-the-art (SOTA) FR systems to different morphing techniques. Embedding interpolation alone achieves high MMPMR values across all FR systems, ranging from 0.992 to 1.000 at MMPMR100 and 0.936 to 0.986 at MMPMR1000, demonstrating that direct interpolation of face embeddings can already generate highly effective morphing attacks. Cross-attention interpolation performs comparably, with MMPMR100 values between 0.990 and 0.999, though showing slightly lower MMPMR1000 scores (0.916 to 0.978), suggesting minor effectiveness reductions at the more restrictive operational threshold. Combining embedding interpolation with DDIM inversion yields improvements across most FR models, demonstrating that spherical interpolation of inverted latents provides better identity preservation at stricter thresholds. However, DCMorph, which combines cross-attention interpolation with DDIM inversion, achieves the highest vulnerability across all evaluation settings. DCMorph scores perfect or near-perfect MMPMR100 values (0.998–1.000) and the highest

MMPMR1000 values (0.965–0.995) for all four FR systems. Notably, on AdaFace and ElasticFace, DCMorph achieves perfect scores (1.000) at MMPMR100, while on CurricularFace and ArcFace, it reaches MMPMR100 values of 0.999 and 0.998 respectively, outperforming all other variants. The superior performance of DCMorph demonstrates the synergistic benefit of jointly operating at both the conditioning pathway and latent space levels. Figure 3 provides qualitative comparisons showing how DDIM inversion preserves high-level latent semantics, maintaining global structure and attributes more consistently than other approaches, such as slight head pose inconsistency that might be seen more clearly in the third row of Figure 3.

Detectability Analysis: Table 2 presents the detectability of different morphing techniques across three MAD systems, presented in Section 4. The results reveal complex patterns depending on the MAD architecture and detection strategy. For MADPromptS [7], a zero-shot foundation model approach, morphing attacks generated through embedding interpolation and cross-attention interpolation alone are poorly detected, with EERs of 41.90% and 44.00% respectively, and APCER values exceeding 84% at 10% BPCER. In contrast, DCMorph shows improved detectability from the MAD perspective, with a lower EER of 27.60% and APCER of 48.90% at 10% BPCER, though still representing a challenging detection scenario. For SPL-MAD [18], an unsupervised anomaly detection method, attacks generated with embedding interpolation remains moderately challenging to detect (EER 38.30%), while cross-attention interpolation becomes substantially easier to detect. However, DCMorph becomes extremely difficult to detect by this MAD, achieving an EER of 77.60% with APCER values of 100% across all BPCER operating points. Similarly, for MixFaceNet-MAD [13], cross-attention interpolation alone is relatively well-detected (EER 6.40%), while DCMorph becomes highly challenging to detect (EER 69.20%, APCER 100% at both 1% and 10% BPCER). These results demonstrate that while DCMorph's dual-stream approach maximizes FR vulnerability, it also creates detection challenges for current MAD solutions, particularly those based on supervised learning with limited exposure to such attacks.

5.2. Comparison with SOTA Methods

Vulnerability Analysis: We evaluated the vulnerability of four FR systems to DCMorph attacks compared against six different approaches: three image-level methods (OpenCV, FaceMorpher, WebMorph) and three representation-level methods (MIPGAN-I, MIPGAN-II, MorDIFF). Table 3 presents comprehensive vulnerability results across all attack types and FR systems. On all four FR systems at FMR 1%, DCMorph demonstrates superior attack effectiveness compared to all representation-level baseline attacks. At FMR 1%, DCMorph achieves attack effectiveness that is competitive with the strongest representation level methods and, for some systems such as AdaFace, attains the highest MMPMR values. While WebMorph yields slightly higher scores in certain cases, DCMorph remains consistently among the top performing approaches across all FR systems. When compared to the diffusion-based MorDIFF, DCMorph shows consistent improvements, achieving an MMPMR at FMR 0.1% of 0.965 vs. 0.948 on ElasticFace, 0.981 vs. 0.968 on CurricularFace, 0.995 vs. 0.962 on AdaFace, and 0.982 vs. 0.917 on ArcFace.

Table 2. Detectability analysis of different morphing strategies across three MAD systems. The table presents EER and APCER at three BPCER operating points (1%, 10%, 20%). Lower EER and APCER indicate easier detection (better MAD performance). DCMorph shows varied detectability patterns: moderately challenging for MADPromptS, but extremely challenging for SPL-MAD and MixFaceNet-MAD.

Method	Test data	EER (%)	APCER (%) @ BPCER (%)		
			1.00	10.00	20.00
MADPromptS [7]	Embedding Interpolation	41.90	99.90	84.70	68.50
	Cross-attention Interpolation	44.00	99.50	84.20	71.70
	Embedding Interpolation + DDIM	32.40	96.30	60.40	44.60
	DCMorph	27.60	92.50	48.90	35.70
SPL-MAD [18]	Embedding Interpolation	38.30	99.00	77.90	66.70
	Cross-attention Interpolation	11.00	75.70	12.30	6.40
	Embedding Interpolation + DDIM	32.40	96.30	60.40	44.60
	DCMorph	77.60	100.0	100.0	100.0
MixFaceNet-MAD [13]	Embedding Interpolation	17.50	99.70	38.20	11.60
	Cross-attention Interpolation	6.40	85.40	3.70	0.30
	Embedding Interpolation + DDIM	92.70	100.00	100.00	100.00
	DCMorph	69.20	100.00	100.00	99.80

Detectability Analysis: Table 4 presents the detectability of DCMorph attacks compared to six different morphing methods across three MAD systems, presented in Section 4. The evaluation follows a realistic cross-dataset protocol where all MADs are tested on attacks generated from datasets different from their training data, representing practical deployment scenarios. The detectability results reveal varied patterns across different MAD architectures. For MADPromptS, a zero-shot learning approach based on foundation models, DCMorph shows the highest EER (27.60%) among all attack types, making it the most challenging to detect. At 10% BPCER, DCMorph achieves an APCER of 48.90%, substantially higher than image-level methods (21.14–30.60%) and comparable to MorDIFF (52.20%). For SPL-MAD, an unsupervised anomaly detection method, DCMorph presents extreme detection challenges with an EER of 77.60% and APCER of 100% across all BPCER operating points. This represents the poorest detection performance across all attack types evaluated, far exceeding the detection difficulty of other representation-level attacks (MIPGAN-I: 16.30% EER, MIPGAN-II: 11.01% EER, MorDIFF: 7.70% EER) and even image-level attacks (5.89–11.60% EER). The extremely high EER suggests that the unsupervised anomaly detector's learned representations fail to distinguish DCMorph attacks from bona fide samples. Similarly, for MixFaceNet-MAD, a supervised detection approach, DC-

Table 3. Vulnerability comparison of four SOTA FRs to DCMorph and six baseline morphing attacks. Attacks are categorized as image-level (OpenCV, FaceMorpher, WebMorph) or representation-level (MIPGAN-I, MIPGAN-II, MorDIFF, DCMorph). Higher MMPMR values indicate stronger attacks and greater FR vulnerability. DCMorph achieves the highest MMPMR values (bold) across all FR systems at both operational thresholds (MMPMR100 at FMR 1% and MMPMR1000 at FMR 0.1%), demonstrating superior effectiveness compared to all representation-level methods and matching or exceeding the best image-level approaches.

Morphing technique		ElasticFace [5]		CurricularFace [24]		AdaFace [28]		ArcFace [16]	
		MMPMR100	MMPMR1000	MMPMR100	MMPMR1000	MMPMR100	MMPMR1000	MMPMR100	MMPMR1000
Image level	OpenCV	**0.997**	0.980	**0.996**	0.986	**1.000**	**0.993**	**0.997**	0.979
	FaceMorpher	0.962	0.913	0.970	0.935	0.973	0.948	0.958	0.920
	WebMorph	0.990	**0.986**	0.988	**0.988**	0.988	0.988	0.988	**0.988**
Representation level	MIPGAN-I	0.980	0.845	0.962	0.890	0.971	0.902	0.961	0.853
	MIPGAN-II	0.953	0.778	0.953	0.832	0.965	0.868	0.953	0.818
	MorDIFF	0.990	0.948	0.995	0.968	0.991	0.962	0.985	0.917
	DCMorph (our)	**1.000**	**0.965**	**0.999**	**0.981**	**1.000**	**0.995**	**0.998**	**0.982**

Table 4. Detectability comparison of DCMorph and six baseline morphing attacks across three MAD systems in a realistic cross-dataset evaluation protocol. The table shows EER and APCER at three BPCER operating points (1%, 10%, 20%). Lower values indicate easier detection (better MAD performance). DCMorph demonstrates the most challenging detectability profile across all three MADs: highest EER for MADPromptS, and extremely high EER for SPL-MAD and MixFaceNet-MAD, substantially exceeding all baseline attacks including the recent MorDIFF.

Method	Test data	EER (%)	APCER (%) @ BPCER (%)		
			1.00	10.00	20.00
MADPromptS [7]	FaceMorph	18.10	61.20	25.10	16.40
	MIPGAN_I	5.40	26.50	4.50	1.60
	MIPGAN_II	3.50	13.41	1.20	0.20
	OpenCV	16.06	66.97	21.14	10.98
	WebMorph	18.40	75.60	30.60	17.40
	MorDIFF	24.50	94.70	52.20	30.10
	DCMorph (our)	27.60	92.50	48.90	35.70
SPL-MAD [18]	FaceMorph	0.00	65.80	65.80	0.00
	MIPGAN_I	16.30	67.30	23.00	11.20
	MIPGAN_II	11.01	54.35	14.31	6.41
	OpenCV	5.89	20.53	3.15	1.32
	WebMorph	11.60	50.00	13.00	6.00
	MorDIFF	7.70	25.50	6.40	2.50
	DCMorph (our)	77.60	100.0	100.0	100.0
MixFaceNet-MAD [13]	FaceMorph	4.60	5.60	3.70	2.90
	MIPGAN_I	16.60	75.80	22.40	14.50
	MIPGAN_II	20.52	81.68	32.13	20.62
	OpenCV	8.33	36.48	6.50	3.86
	WebMorph	18.20	74.20	24.00	17.60
	MorDIFF	9.40	36.30	8.90	5.20
	DCMorph (our)	69.20	100.00	100.00	99.80

Morph demonstrates substantial detection challenges with an EER of 69.20% and APCER of 100% at both 1% and 10% BPCER. This is considerably higher than the detection difficulty of other attacks, including MorDIFF (9.40% EER), image-level methods (4.60–18.20% EER), and GAN-based methods (16.60–20.52% EER).

These detectability results, combined with the vulnerability analyses, demonstrate that DCMorph achieves a favorable trade-off from an attacker's perspective: maximizing FR vulnerability while remaining highly challenging to detect by current MAD solutions. The detection difficulty is particularly pronounced for learning-based MADs (SPL-MAD and MixFaceNet-MAD), suggesting that DCMorph's dual-stream approach produces morphs with characteristics that differ from the attack patterns these detectors have been trained to recognize. Only the zero-shot foundation model approach (MADPromptS) shows moderate detection capability, though DCMorph remains more challenging to detect than most analyzed attacks even for this detector.

6. Conclusion

This paper presented DCMorph, a dual-stream morphing framework addressing fundamental limitations of existing approaches by simultaneously operating at both identity conditioning and latent space levels. The framework contributes two key mechanisms: (1) decoupled cross-attention interpolation that injects identity-specific features from both source faces into the denoising process, enabling explicit dual-identity conditioning absent in prior diffusion-based methods, and (2) DDIM inversion with spherical interpolation between inverted latent representations, providing geometrically consistent initialization that preserves structural attributes. Comprehensive vulnerability analyses across four SOTA FRs demonstrated that DCMorph achieves MMPMR values ranging from 0.965 to 0.995 at FMR 0.1%, outperforming all representation-level baselines, including GAN-based methods (MIPGAN-I, MIPGAN-II) and the recent DM-based MorDIFF, while matching or exceeding the best image-level techniques (OpenCV, FaceMorpher, WebMorph) at restrictive thresholds. Ablation studies further illustrated that the dual-stream approach (combining both mechanisms) consistently outperforms single-stream variants across all FR systems. Detectability analyses across three morphing attack detection systems revealed that DCMorph presents substantial challenges for current MAD solutions, achieving EER values of 77.60% for SPL-MAD and 69.20% for MixFaceNet-MAD, indicating that the dual-stream approach produces morphs that evade conventional detection patterns. Cross-dataset evaluation demonstrated that DCMorph maintains consistent detectability characteristics compared to baseline methods.

References

[1] Yuanhao Ban, Ruochen Wang, Tianyi Zhou, Minhao Cheng, Boqing Gong, and Cho-Jui Hsieh. Understanding the impact of negative prompts: When and how do they take effect? In *European Conference on Computer Vision*, pages 190–206. Springer, 2024. 3

[2] Zander Blasingame and Chen Liu. Leveraging diffusion for strong and high quality face morphing attacks. *IEEE Trans. Biom. Behav. Identity Sci.*, 6(1):118–131, 2024. 1, 2

[3] Guido Borghi, Annalisa Franco, Gabriele Graffieti, and Davide Maltoni. Automated artifact retouching in morphed images with attention maps. *IEEE Access*, 9:136561–136579, 2021. 2

[4] Fadi Boutros, Naser Damer, Meiling Fang, Florian Kirchbuchner, and Arjan Kuijper. Mixfacenets: Extremely efficient face recognition networks. In *IJCB*, pages 1–8. IEEE, 2021. 5

[5] Fadi Boutros, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. Elasticface: Elastic margin loss for deep face recognition. In *CVPR Workshops*, pages 1577–1586. IEEE, 2022. 5, 6, 8

[6] Fadi Boutros, Jonas Henry Grebe, Arjan Kuijper, and Naser Damer. Idiff-face: Synthetic-based face recognition through fizzy identity-conditioned diffusion models. In *IEEE/CVF International Conference on Computer Vision, ICCV 2023, Paris, France, October 1-6, 2023*, pages 19593–19604. IEEE, 2023. 1, 2, 3

[7] Eduarda Caldeira, Fadi Boutros, and Naser Damer. Madprompts: Unlocking zero-shot morphing attack detection with multiple prompt aggregation. *CoRR*, abs/2508.08939, 2025. 5, 7, 8

[8] Eduarda Caldeira, Naser Damer, and Fadi Boutros. Negfacediff: The power of negative context in identity-conditioned diffusion for synthetic face generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5914–5924, 2025. 2

[9] Naser Damer, Viola Boller, Yaza Wainakh, Fadi Boutros, Philipp Terhörst, Andreas Braun, and Arjan Kuijper. Detecting face morphing attacks by analyzing the directed distances of facial landmarks shifts. In *GCPR*, pages 518–534. Springer, 2018. 2

[10] Naser Damer, Alexandra Mosegui Saladie, Andreas Braun, and Arjan Kuijper. Morgan: Recognition vulnerability and attack detectability of face morphing attacks created by generative adversarial network. In *9th IEEE International Conference on Biometrics Theory, Applications and Systems, BTAS 2018, Redondo Beach, CA, USA, October 22-25, 2018*, pages 1–10. IEEE, 2018. 1, 2

[11] Naser Damer, Fadi Boutros, Alexandra Mosegui Saladie, Florian Kirchbuchner, and Arjan Kuijper. Realistic dreams: Cascaded enhancement of gan-generated images with an example in face morphing attacks. In *10th IEEE International Conference on Biometrics Theory, Applications and Systems, BTAS 2019, Tampa, FL, USA, September 23-26, 2019*, pages 1–10. IEEE, 2019. 2

[12] Naser Damer, Kiran B. Raja, Marius Süßmilch, Sushma Venkatesh, Fadi Boutros, Meiling Fang, Florian Kirchbuchner, Raghavendra Ramachandra, and Arjan Kuijper. Regenmorph: Visibly realistic GAN generated face morphing attacks by attack re-generation. In *ISVC (1)*, pages 251–264. Springer, 2021. 2

[13] Naser Damer, César Augusto Fontanillo López, Meiling Fang, Noémie Spiller, Minh Vu Pham, and Fadi Boutros. Privacy-friendly synthetic data for the development of face morphing attack detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1606–1617, 2022. 5, 7, 8

[14] Naser Damer, Meiling Fang, Patrick Siebke, Jan Niklas Kolf, Marco Huber, and Fadi Boutros. Mordiff: Recognition vulnerability and attack detectability of face morphing attacks created by diffusion autoencoders. In *IWBF*, pages 1–6. IEEE, 2023. 1, 2, 5

[15] Lisa DeBruine and Benedict Jones. Face Research Lab London Set. 2021. 5

[16] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *CVPR*, pages 4690–4699. Computer Vision Foundation / IEEE, 2019. 5, 6, 8

[17] Nicolò Di Domenico, Guido Borghi, Annalisa Franco, and Davide Maltoni. Face restoration for morphed images retouching. In *12th International Workshop on Biometrics and Forensics, IWBF 2024, Enschede, The Netherlands, April 11-12, 2024*, pages 1–6. IEEE, 2024. 2

[18] Meiling Fang, Fadi Boutros, and Naser Damer. Unsupervised face morphing attack detection via self-paced anomaly detection. In *IJCB*, pages 1–11, 2022. 5, 7, 8

[19] Matteo Ferrara, Annalisa Franco, and Davide Maltoni. The magic passport. In *IJCB*, pages 1–7. IEEE, 2014. 1, 2

[20] Matteo Ferrara, Annalisa Franco, and Davide Maltoni. Face demorphing. *IEEE Trans. Inf. Forensics Secur.*, 13(4):1008–1017, 2018. 1, 2

[21] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *CoRR*, abs/2207.12598, 2022. 3

[22] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. 2, 3

[23] 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 *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, Marseille, France, 2008. Erik Learned-Miller and Andras Ferencz and Frédéric Jurie. 5

[24] Yuge Huang, Yuhan Wang, Ying Tai, Xiaoming Liu, Pengcheng Shen, Shaoxin Li, Jilin Li, and Feiyue Huang. Curricularface: Adaptive curriculum learning loss for deep face recognition. In *CVPR*, pages 5900–5909. Computer Vision Foundation / IEEE, 2020. 5, 6, 8

[25] Marco Huber, Fadi Boutros, Anh Thi Luu, Kiran B. Raja, Raghavendra Ramachandra, Naser Damer, Pedro C. Neto, Tiago Gonçalves, Ana F. Sequeira, Jaime S. Cardoso, João Tremoço, Miguel Lourenço, Sergio Serra, Eduardo

Cermeño, Marija Ivanovska, Borut Batagelj, Andrej Kronovsek, Peter Peer, and Vitomir Struc. SYN-MAD 2022: Competition on face morphing attack detection based on privacy-aware synthetic training data. In *IJCB*, pages 1–10. IEEE, 2022. 5

[26] International Organization for Standardization. ISO/IEC DIS 30107-3:2016 Information Technology – Biometric Presentation Attack Detection – Part 3: Testing and Reporting. International Standard ISO/IEC 30107-3:2016, ISO/IEC, 2017. 5

[27] Young Kyun Jang, Dat Huynh, Ashish Shah, Wen-Kai Chen, and Ser-Nam Lim. Spherical linear interpolation and text-anchoring for zero-shot composed image retrieval. In *Computer Vision - ECCV 2024 - 18th European Conference, Milan, Italy, September 29-October 4, 2024, Proceedings, Part XIX*, pages 239–254. Springer, 2024. 4

[28] Minchul Kim, Anil K. Jain, and Xiaoming Liu. Adaface: Quality adaptive margin for face recognition. In *CVPR*, pages 18729–18738. IEEE, 2022. 5, 6, 8

[29] Minchul Kim, Feng Liu, Anil K. Jain, and Xiaoming Liu. Dcface: Synthetic face generation with dual condition diffusion model. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023*, pages 12715–12725. IEEE, 2023. 2, 3

[30] Xiao Lin, Yuge Huang, Jianqing Xu, Yuxi Mi, Shuigeng Zhou, and Shouhong Ding. Uiface: Unleashing inherent model capabilities to enhance intra-class diversity in synthetic face recognition. In *ICLR*. OpenReview.net, 2025. 1, 2, 3

[31] Satya Mallick. Face morph using opencv — c++ / python. *LearnOpenCV*, 1(1), 2016. 5

[32] Aravinda Reddy P. N., Raghavendra Ramachandra, Krothapalli Sreenivasa Rao, and Pabitra Mitra. Morcode: Face morphing attack generation using generative codebooks. In *35th British Machine Vision Conference Workshop Proceedings, BMVC 2024 Workshops, Glasgow, UK, November 25-28, 2024*. BMVA Press, 2024. 2

[33] Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. SDXL: improving latent diffusion models for high-resolution image synthesis. *CoRR*, abs/2307.01952, 2023. 5

[34] Konpat Preechakul, Nattanat Chatthee, Suttisak Wizadwongsa, and Supasorn Suwajanakorn. Diffusion autoencoders: Toward a meaningful and decodable representation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022*, pages 10609–10619. IEEE, 2022. 2

[35] Le Qin, Fei Peng, Sushma Venkatesh, Raghavendra Ramachandra, Min Long, and Christoph Busch. Low visual distortion and robust morphing attacks based on partial face image manipulation. *IEEE Trans. Biom. Behav. Identity Sci.*, 3(1):72–88, 2021. 2

[36] Ramachandra Raghavendra, Kiran B. Raja, Sushma Venkatesh, and Christoph Busch. Face morphing versus face averaging: Vulnerability and detection. In *IJCB*, pages 555–563. IEEE, 2017. 1, 2

[37] Robin Rombach, A. Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10674–10685, 2021. 2, 3, 5

[38] Dvir Samuel, Rami Ben-Ari, Nir Darshan, Haggai Maron, and Gal Chechik. Norm-guided latent space exploration for text-to-image generation. In *NeurIPS*, 2023. 4

[39] Ulrich Scherhag, Andreas Nautsch, Christian Rathgeb, Marta Gomez-Barrero, Raymond N. J. Veldhuis, Luuk J. Spreeuwers, Maikel Schils, Davide Maltoni, Patrick Grother, Sébastien Marcel, Ralph Breithaupt, Ramachandra Raghavendra, and Christoph Busch. Biometric systems under morphing attacks: Assessment of morphing techniques and vulnerability reporting. In *BIOSIG*, pages 149–159. GI / IEEE, 2017. 5

[40] Ulrich Scherhag, Jonas Kunze, Christian Rathgeb, and Christoph Busch. Face morph detection for unknown morphing algorithms and image sources: a multi-scale block local binary pattern fusion approach. *IET Biom.*, 9(6):278–289, 2020. 2

[41] Nitish Shukla and Arun Ross. Facial demorphing via identity preserving image decomposition. In *IEEE International Joint Conference on Biometrics, IJCB 2024, Buffalo, NY, USA, September 15-18, 2024*, pages 1–10. IEEE, 2024. 2

[42] Nitish Shukla and Arun Ross. dc-gan: Dual-conditioned GAN for face demorphing from a single morph. In *19th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2025, Tampa/Clearwater, FL, USA, May 26-30, 2025*, pages 1–9. IEEE, 2025. 2

[43] Jag Mohan Singh and Raghavendra Ramachandra. 3d face morphing attack generation using non-rigid registration. In *18th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2024, Istanbul, Turkey, May 27-31, 2024*, pages 1–5. IEEE, 2024. 2

[44] Jag Mohan Singh and Raghavendra Ramachandra. 3-d face morphing attacks: Generation, vulnerability and detection. *IEEE Trans. Biom. Behav. Identity Sci.*, 6(1):103–117, 2024. 2

[45] Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, pages 2256–2265. JMLR.org, 2015. 2

[46] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *ICLR*. OpenReview.net, 2021. 1, 2, 4

[47] Sushma Venkatesh, Haoyu Zhang, Raghavendra Ramachandra, Kiran B. Raja, Naser Damer, and Christoph Busch. Can GAN generated morphs threaten face recognition systems equally as landmark based morphs? - vulnerability and detection. In *8th International Workshop on Biometrics and Forensics, IWBF 2020, Porto, Portugal, April 29-30, 2020*, pages 1–6. IEEE, 2020. 2

[48] Ruochen Wang, Ting Liu, Cho-Jui Hsieh, and Boqing Gong. On discrete prompt optimization for diffusion models. In *ICML*. OpenReview.net, 2024. 3

[49] Jianqing Xu, Shen Li, Jiaying Wu, Miao Xiong, Ailin Deng, Jiazhen Ji, Yuge Huang, Guodong Mu, Wenjie Feng, Shouhong Ding, and Bryan Hooi. Id^3: Identity-preserving-yet-diversified diffusion models for synthetic face recognition. In *NeurIPS*, 2024. 2, 3

[50] Hu Ye, Jun Zhang, Sibo Liu, Xiao Han, and Wei Yang. Ip-adapter: Text compatible image prompt adapter for text-to-image diffusion models. 2023. 5

[51] Haoyu Zhang, Sushma Venkatesh, Raghavendra Ramachandra, Kiran Bylappa Raja, Naser Damer, and Christoph Busch. MIPGAN - generating strong and high quality morphing attacks using identity prior driven GAN. *IEEE Trans. Biom. Behav. Identity Sci.*, 3(3):365–383, 2021. 1, 2, 5