Diffusion for the Generation of Face Morphs

Zander W. Blasingame and Chen Liu

05.06.2024

Department of Electrical and Computer Engineering
Clarkson University
8 Clarkson Ave, Potsdam NY, 13699, USA

Face Morphing

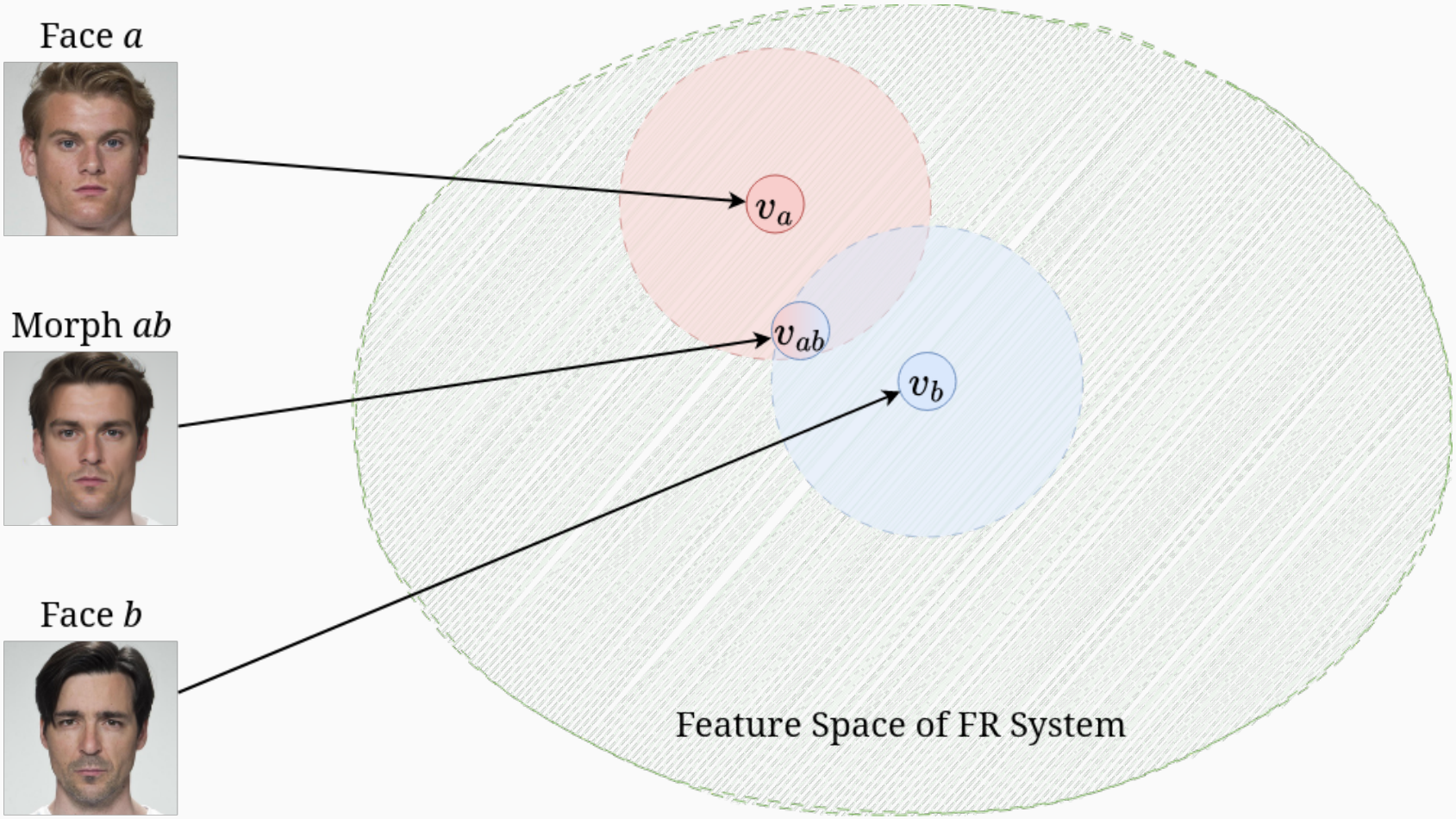


Figure 1: Images from FRLL dataset. Morph generated by us.

Generative AI Morph Creation Pipeline

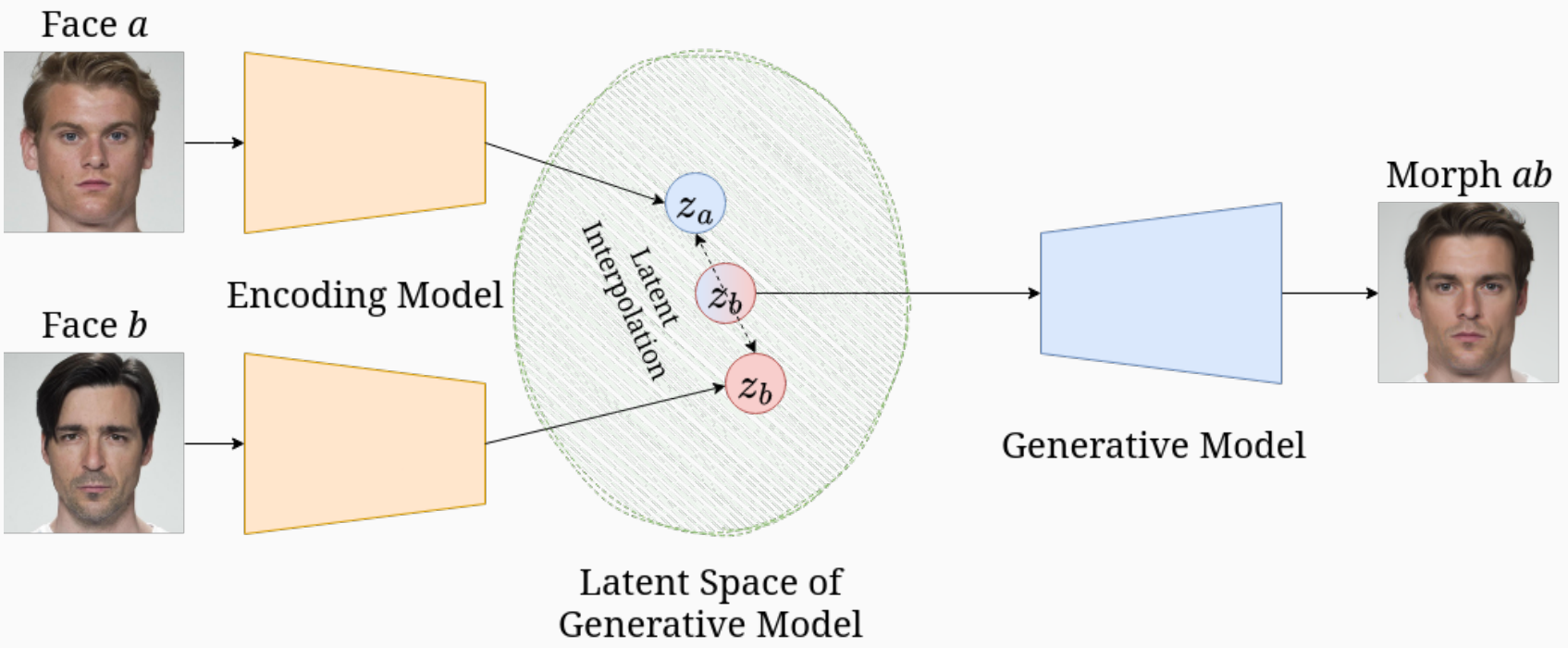


Figure 2: General morph creation pipeline using generative AI models.

Diffusion Models

$$\mathrm{d}\mathbf{x}_t = \mu(\mathbf{x}_t, t)\mathrm{d}t + \sigma(t)\mathrm{d}\mathbf{w}_t$$

Figure 3: Forward Diffusion Process

Reverse Diffusion Process

$$\mathrm{d}\mathbf{x}_t = \mu(\mathbf{x}_t, t)\mathrm{d}t + \sigma(t)\mathrm{d}\mathbf{w}_t$$

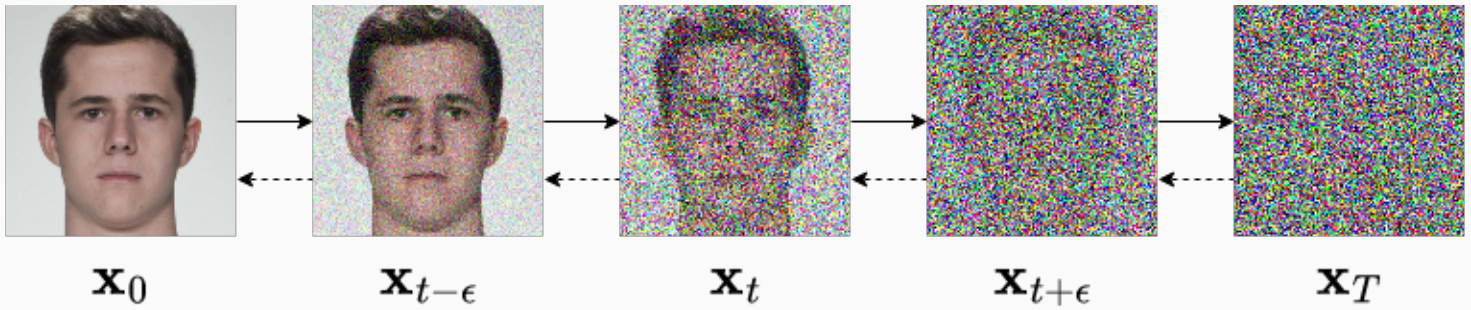
$$\mathrm{d}\mathbf{x}_t = [\mu(\mathbf{x}_t, t) - \sigma^2(t)\nabla_{\mathbf{x}} \log p_t(\mathbf{x}_t)]\mathrm{d}t + \sigma(t)\mathrm{d}\tilde{\mathbf{w}}_t$$

- Reverse SDE has an associated ODE[1]

$$\frac{\mathrm{d}\mathbf{x}_t}{\mathrm{d}t} = \mu(\mathbf{x}, t) - \frac{1}{2}\sigma^2(t)\nabla_{\mathbf{x}} \log p_t(\mathbf{x}_t) \tag{1}$$

[1] Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-based generative modeling through stochastic differential equations. In International Conference on Learning Representations, 2021.

Learning The Reverse Diffusion SDE



- Often $\mu(\mathbf{x}_t, t) = 0$ and $\sigma(t) = \sqrt{2t}$ so forward sampling is

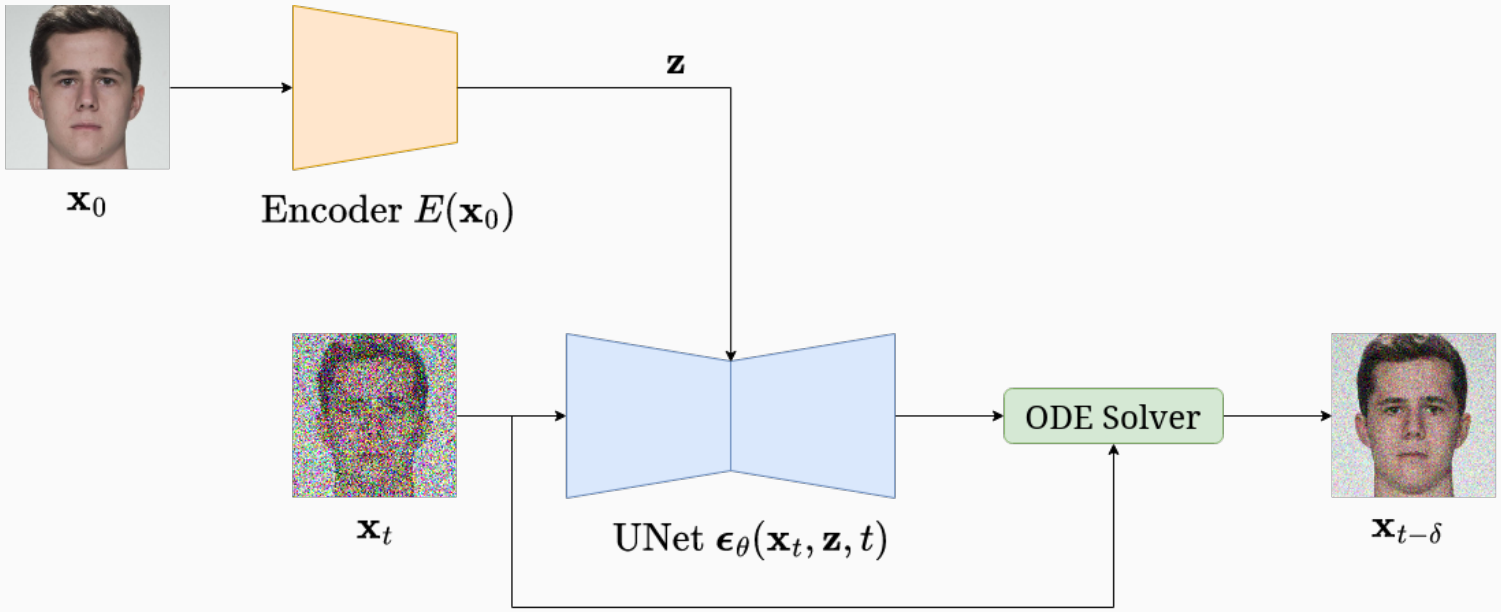
$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$
$$\mathbf{x}_t = \mathbf{x}_0 + t \cdot \boldsymbol{\epsilon} \tag{2}$$

- Learning the score $\nabla_{\mathbf{x}} \log p_t(\mathbf{x}_t)$ amounts to learning $\boldsymbol{\epsilon}$
- Train a UNet, $\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)$, to learn the added noise
- I.e., solve

$$\hat{\theta} = \arg\min_\theta d(\boldsymbol{\epsilon}, \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)) \tag{3}$$

for some distance metric $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, e.g., ℓ^2 distance

Conditional Generation



- Learn an encoder $\mathbf{z} = E(\mathbf{x}_0)$
- Condition the noise prediction model on $\mathbf{z}$
- To get a consistent $\mathbf{x}_T$ run ODE solver in reverse from $\mathbf{x}_0$[2]

[2] K. Preechakul, N. Chatthee, S. Wizadwongsa, and S. Suwajanakorn, "Diffusion autoencoders: Toward a meaningful and decodable representation," in Proceedings of CVPR, June 2022

Face Morphing with Diffusion

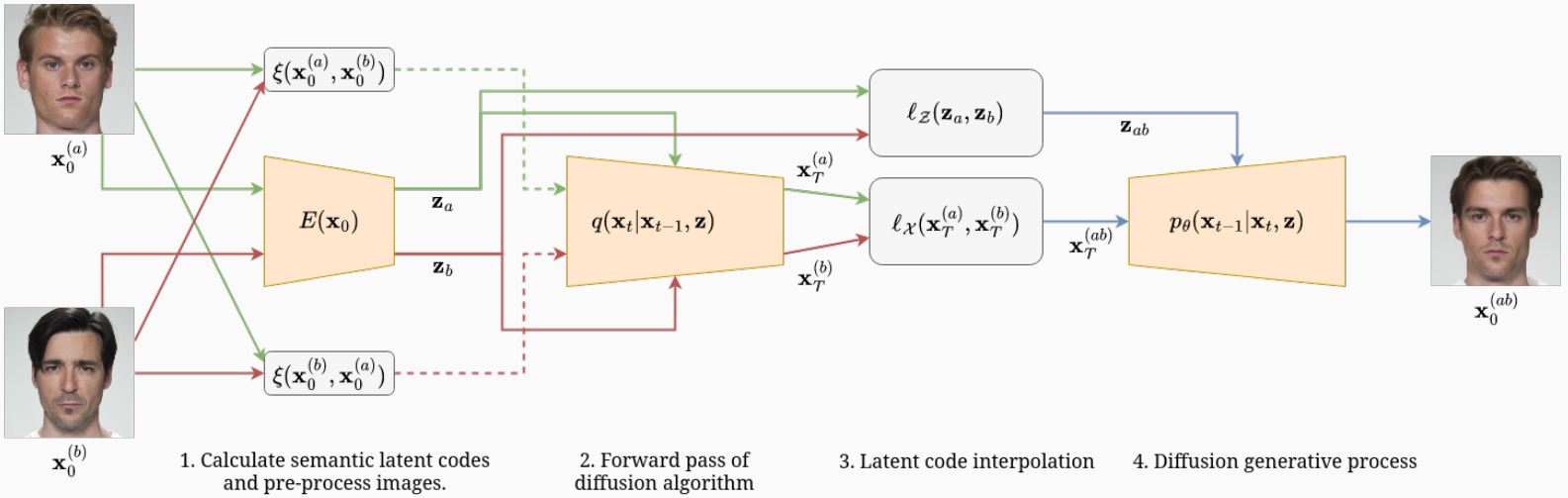


Figure 4: Face morphing pipeline

- Encode bona fide images with E
- Pre-morph bona fide images with ξ
- Encode pre-morphed images by running ODE solver backwards
- Morph encoded images and latents, $\ell_{\mathcal{X}}$ and $\ell_{\mathcal{Z}}$
- Run ODE solver to get morphed image

Impact of Interpolation Strategies

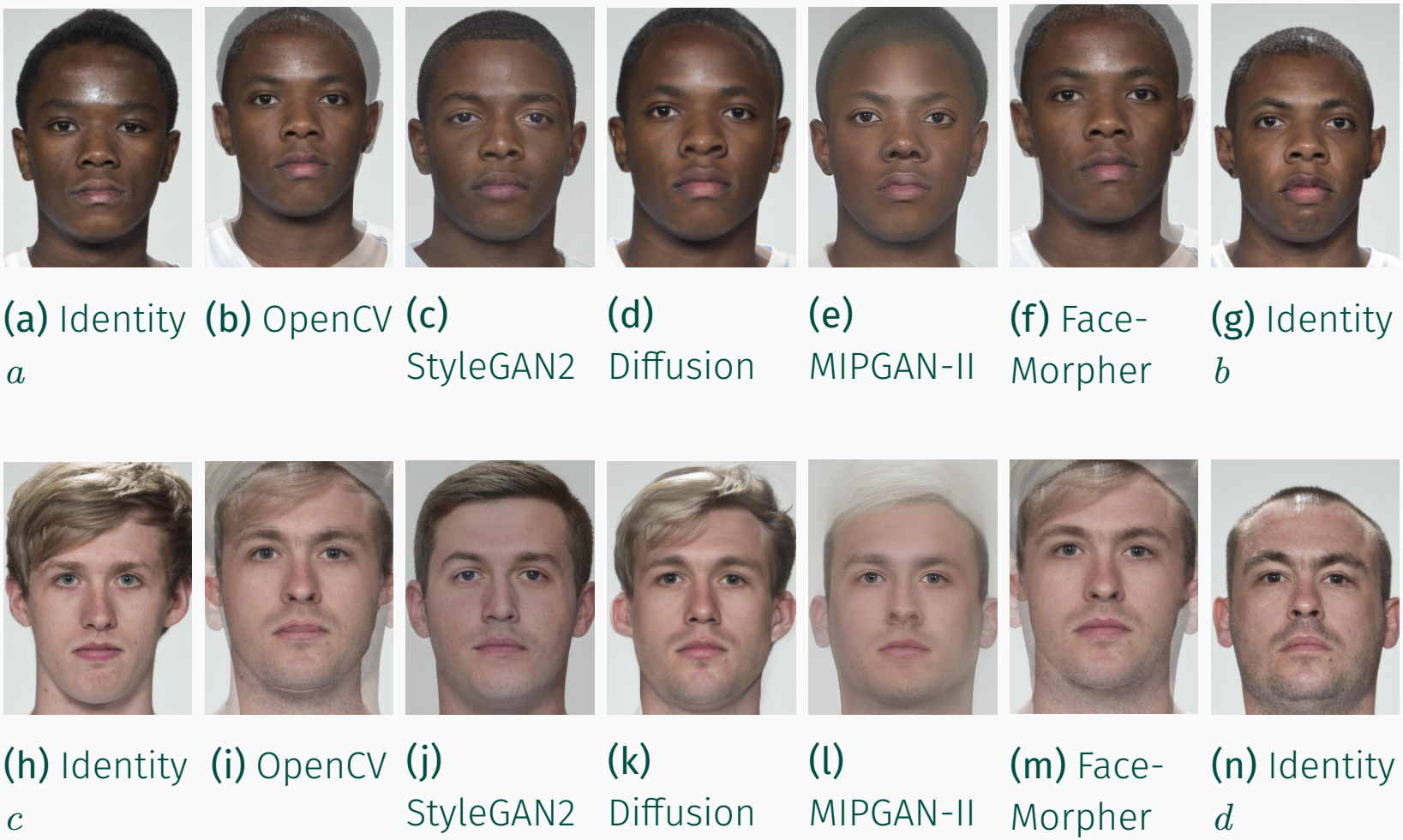
(a) No pre-morph, slerp for $\mathbf{x}_T$

(b) No pre-morph, lerp for $\mathbf{x}_T$

(c) Pixel-wise pre-morph, slerp for $\mathbf{x}_T$

Figure 5: Morphed image generated by different Diffusion attack variants on FRLL.

Visual Comparison to Other Morphing Attacks



(a) Identity a (b) OpenCV (c) StyleGAN2 (d) Diffusion (e) MIPGAN-II (f) Face-Morpher (g) Identity b

(h) Identity c (i) OpenCV (j) StyleGAN2 (k) Diffusion (l) MIPGAN-II (m) Face-Morpher (n) Identity d

Figure 6: Comparison across different morphing algorithms of two identity pairs from the FRLL dataset.

Evaluation of Visual Fidelity

Table 1: FID across different morphing attacks. Lower is better.

Morphing Attack	FRLL	FRGC	FERET
StyleGAN2	45.19	86.41	**41.91**
FaceMorpher	91.97	88.14	79.58
OpenCV	85.71	100.02	91.94
MIPGAN-II	66.41	115.96	70.88
Diffusion	**42.63**	**64.16**	50.45

- Fréchet Inception Distance (FID) is a widely used metric for image synthesis task

Quantitative Comparison

Table 2: MMPMR at FMR = 0.1% across different morphing attacks.

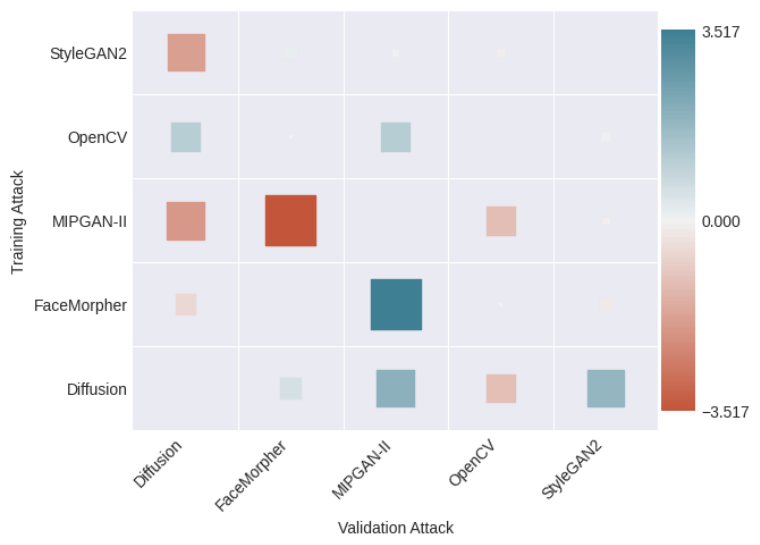
	FRLL			FRGC		
Morphing Attack	**FaceNet**	**VGGFace2**	**ArcFace**	**FaceNet**	**VGGFace2**	**ArcFace**
StyleGAN2	4.69	6.05	19.89	0.18	0.85	5.49
FaceMorpher	11.26	36.4	45.03	0.51	9.15	41.28
OpenCV	17.34	**40.93**	47.7	0.14	**12.16**	3.99
MIPGAN-II	**30.96**	26.74	56.52	**3.12**	7.94	33.54
Diffusion	28.14	35.37	**88.09**	2.68	8.47	**46.74**

- The ProdAvg Mated Morphed Presentation Match Rate (MMPMR)

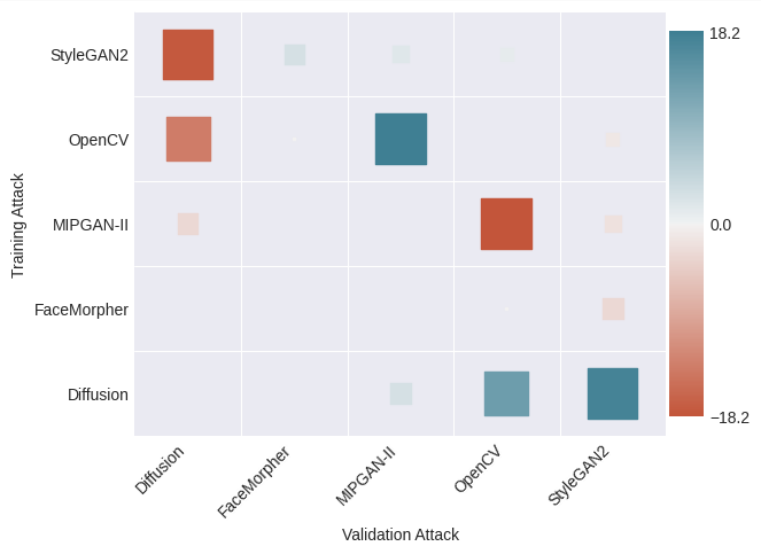
$$M(\gamma) = \mathbb{E}_{x_{ab} \sim \mathbb{P}_M} \left[\prod_{k \in \{a,b\}} \mathbb{E}_{x \sim \mathbb{P}_{k \setminus x_{ab}}} \left[\| F(x_{ab}) - F(x) \|_2 < \gamma \right] \right] \quad (4)$$

where γ is the acceptance threshold, $\mathbb{P}_M$ is the distribution of morphs, $\mathbb{P}_k$, is the distribution of bona fide images for identity k, and $F : \mathcal{X} \to V$ is the FR system

Strength of Diffusion Morphing Attacks



(a) RSM on FRGC



(b) RSM on FRLL

- The transferability of morphing attack α to β is defined as

$$T(\alpha, \beta) = P(f^{\alpha}(X^{\beta}) = 1 \mid f^{\alpha}(X^{\alpha}) = 1) \tag{5}$$

  where X^{α}, X^{β} are morphs created by α, β and f^{α} is a detector
- The relative strength metric (RSM) from α to β is:

$$\Delta(\alpha \| \beta) = \log\left(\frac{T(\alpha, \beta)}{T(\beta, \alpha)}\right) \tag{6}$$

Ablation Study

Table 3: Ablation study on validation accuracy.

	Training Attack					Validation Attack				
Dataset	Diffusion	FaceMorpher	MIPGAN-II	OpenCV	StyleGAN2	Diffusion	FaceMorpher	MIPGAN-II	OpenCV	StyleGAN2
FERET	✗	✓	✓	✓	✓	72.73	99.23	100	99.95	99.33
FERET	✓	✗	✓	✓	✓	99.9	76.39	100	99.85	99.64
FERET	✓	✓	✗	✓	✓	99.69	99.38	100	99.95	99.54
FERET	✓	✓	✓	✗	✓	99.74	99.48	100	99.74	99.43
FERET	✓	✓	✓	✓	✗	99.74	98.56	99.9	99.74	87.89
FRGC	✗	✓	✓	✓	✓	75.89	99.98	99.97	99.9	99.93
FRGC	✓	✗	✓	✓	✓	99.95	99.48	100	99.9	99.95
FRGC	✓	✓	✗	✓	✓	99.83	99.85	99.82	99.8	99.85
FRGC	✓	✓	✓	✗	✓	99.93	100	100	99.23	99.93
FRGC	✓	✓	✓	✓	✗	99.93	99.93	99.94	99.88	97.83
FRLL	✗	✓	✓	✓	✓	13.96	99.58	99.32	99.65	99.65
FRLL	✓	✗	✓	✓	✓	99.23	99.09	98.91	99.37	99.44
FRLL	✓	✓	✗	✓	✓	99.09	98.95	98.24	99.02	99.09
FRLL	✓	✓	✓	✗	✓	99.51	99.44	99.19	99.16	99.58
FRLL	✓	✓	✓	✓	✗	99.93	99.86	99.86	99.93	95.02

Summary of Diffusion for Face Morphing

- Advantages
  1. Better visual fidelity
  2. Can achieve a stronger attack
  3. Hard to detect as a novel attack
  4. More flexible generation as generation parameters can change per iteration
- Disadvantages
  1. Slower inference speed due to multiple iterations
  2. Greater computational requirements

Conclusions

- Face morphs generated via Diffusion are a powerful threat to FR systems
- Diffusion-based morphs have visual fidelity which make them harder to detect
- Morph detectors trained on this attack can be more resilient
- Our article ”Leveraging Diffusion For Strong and High Quality Face Morphing Attacks” was recently accepted in IEEE TBIOM, *https://ieeexplore.ieee.org/document/10381591*

?