Diffusion Morphs (DiM): *Leveraging Diffusion For Strong and High Quality Face Morphing Attacks*

Zander W. Blasingame   Chen Liu

Department of Electrical and Computer Engineering
Clarkson University
{blasinzw, cliu}@clarkson.edu

IEEE International Joint Conference on Biometrics

Motivation

(a) Identity a (b) Morphed image   (c) Identity b

Figure 1. Example morphs generated via DiM. Samples are from FRLL dataset [1].

- Face Recognition (FR) systems are vulnerable to face morphing attacks [2, 3].
- Two broad classes of morphing attacks:
  1. Landmark-based attacks
  2. Representation-based attacks
- Nearly all representation-based attacks are based on the GAN framework
- Diffusion models have been shown to outperform GANs [4]
- We propose a *novel family* of face morphing attacks known as Diffusion Morphs (DiM)

Methodology

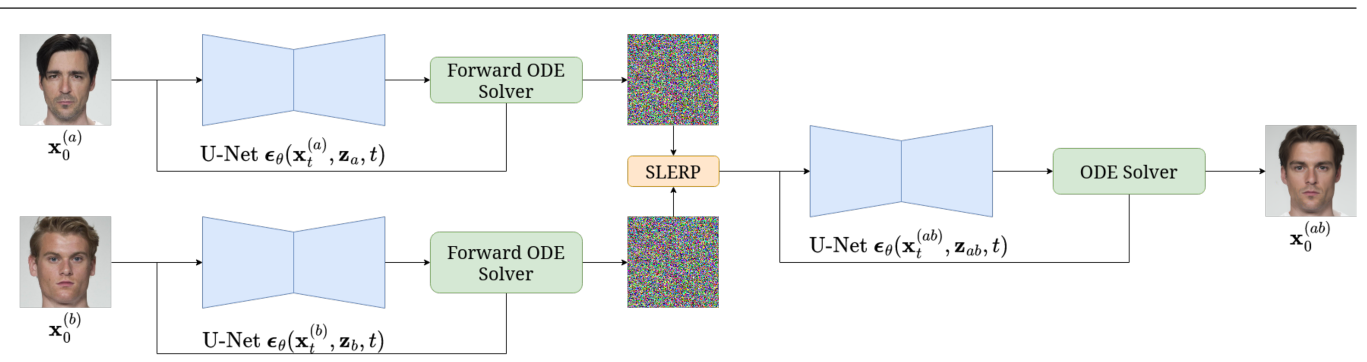


Figure 2. Overview of the DiM pipeline.

- The Variance Preserving (VP) type diffusion process is governed by an Itô SDE of the form

$$\mathrm{d}\mathbf{x}_t = f(t)\mathbf{x}_t\,\mathrm{d}t + g(t)\,\mathrm{d}\mathbf{w}_t \tag{1}$$

$$f(t) = \frac{\mathrm{d}\log\alpha_t}{\mathrm{d}t} \qquad g^2(t) = \frac{\mathrm{d}\sigma_t^2}{\mathrm{d}t} - 2\frac{\mathrm{d}\log\alpha_t}{\mathrm{d}t}\sigma_t^2 \tag{2}$$

with noise schedule $\alpha_t^2 + \sigma_t^2 = 1$ such that $\mathbf{x}_t = \alpha_t\mathbf{x}_0 + \sigma_t\boldsymbol{\epsilon}$ where $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ [5]

- Denote bona fide faces via $\mathbf{x}_0^{(a)}, \mathbf{x}_0^{(b)} \in \mathcal{X}$ and encode bona fide faces into a latent representations

$$\mathbf{z}_a = E(\mathbf{x}_0^{(a)}) \qquad \mathbf{z}_b = E(\mathbf{x}_0^{(b)}) \tag{3}$$

- Let $\Phi(\mathbf{x}_0, \mathbf{z}, \mathbf{h}_\theta, \{t_n\}_{n=1}^N) \to \mathbf{x}_T$ denote a numerical ODE solver to the PF ODE with
  1. Initial image $\mathbf{x}_0$
  2. Latent representation of $\mathbf{x}_0$, $\mathbf{z} = E(\mathbf{x}_0)$
  3. Noise prediction U-Net conditioned on $\mathbf{z}$, $\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, \mathbf{z}, t) \approx \boldsymbol{\epsilon}_t$
  4. The empirical PF ODE given by

$$\mathbf{h}_\theta(\mathbf{x}_t, \mathbf{z}, t) = f(t)\mathbf{x}_t + \frac{g^2(t)}{2\sigma_t}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, \mathbf{z}, t) \tag{4}$$

  5. N monotonically increasing timesteps $\{t_n\}_{n=1}^N \subseteq [0, T]$
- Encode images by solving the PF ODE as time runs *forwards*

$$\mathbf{x}_T^{(\{a,b\})} = \Phi(\mathbf{x}_0^{(\{a,b\})}, \mathbf{z}_{\{a,b\}}, \mathbf{h}_\theta, \{t_n\}_{n=1}^{N_F}) \tag{5}$$

with N_F encoding steps and $t_n < t_{n+1}$

- Morph the latent representations

$$\mathbf{x}_T^{(ab)} = \mathrm{slerp}(\mathbf{x}_T^{(a)}, \mathbf{x}_T^{(b)}; \gamma) \tag{6}$$

$$\mathbf{z}_{ab} = \mathrm{lerp}(\mathbf{z}_a, \mathbf{z}_b; \gamma) \tag{7}$$

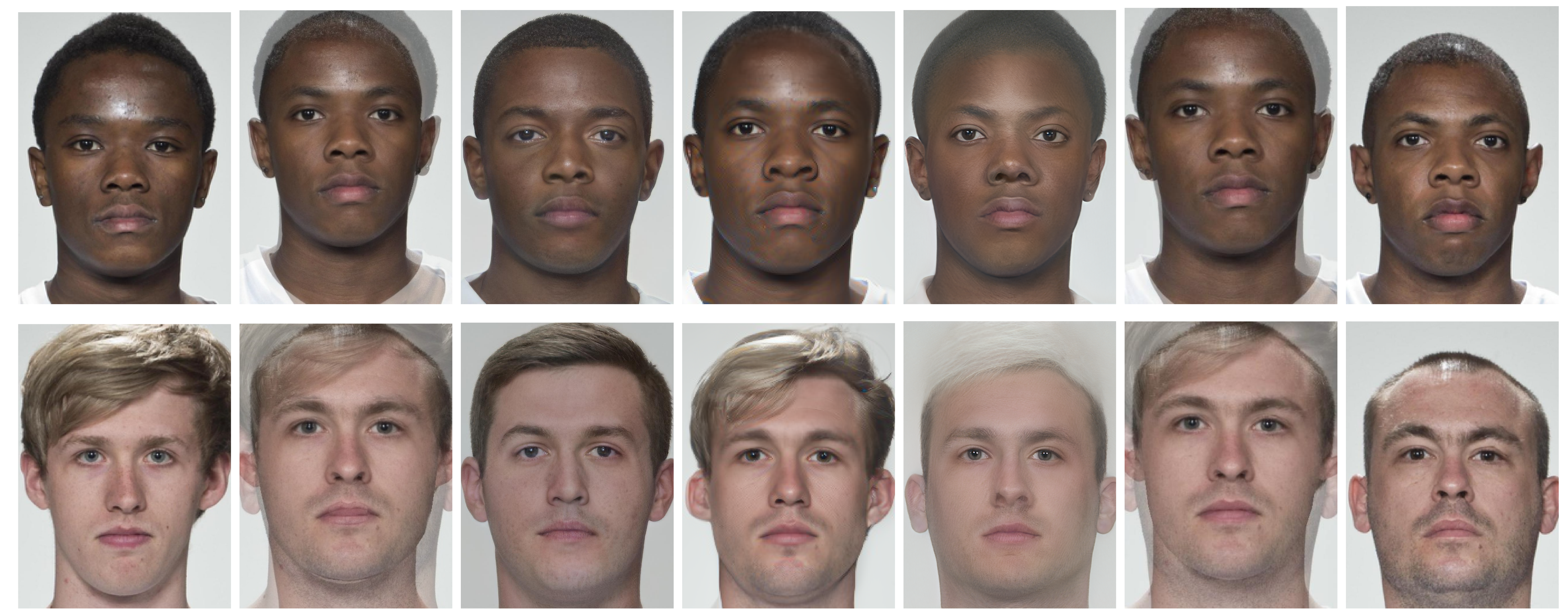
by a factor of $\gamma = 0.5$

- Create morph by solving the PF ODE as time runs *backwards*

$$\mathbf{x}_0^{(ab)} = \Phi(\mathbf{x}_T^{(ab)}, \mathbf{z}_{ab}, \mathbf{h}_\theta, \{\tilde{t}_n\}_{n=1}^N) \tag{8}$$

with N sampling steps and $\tilde{t}_n > \tilde{t}_{n+1}$

Highlighted Results



(a) Identity a (b) OpenCV   (c) StyleGAN2   (d) DiM   (e) MIPGAN-II   (f) FaceMorpher   (g) Identity b

Figure 3. Comparison across different morphing algorithms of two identity pairs from the FRLL dataset.

- The Mated Morph Presentation Match Rate (MMPMR) metric [6] is defined as

$$M(\delta) = \frac{1}{M}\sum_{n=1}^{M}\left\{\left[\min_{n\in\{1,\ldots,N_m\}} S_m^n\right] > \delta\right\} \tag{9}$$

where δ is the verification threshold, S_m^n is the similarity score of the n-th subject of morph m, N_m is the total number of contributing subjects to morph m, and M is the total number of morphed images

- We measure the vulnerability of an FR system w.r.t. a morphing attack using MMPMR

Table 1. Vulnerability of different FR systems across different morphing attacks on the SYN-MAD 2022 dataset [7]. FMR = 0.1%.

	MMPMR (↑)		
Morphing Attack	AdaFace [8]	ArcFace [9]	ElasticFace [10]
FaceMorpher [7]	89.78	87.73	89.57
OpenCV [7]	94.48	92.43	94.27
MIPGAN-I [11]	72.19	77.51	66.46
MIPGAN-II [11]	70.55	72.19	65.24
DiM [12]	92.23	90.18	93.05

- We preform an ablation study on the ability to detect morphing attacks
- We fine-tune a pre-trained SE-ResNeXt101-32x4d network on the Single image-based Morphing Attack Detection (S-MAD) problem
- The model is fine-tuned on all but *one* morphing attack using 5-fold cross validation
- We then report the detection accuracy on the studied morphing attacks

Table 2. Ablation study on the ability to detect morphing attacks.

	Included in the Training Set					Detection Accuracy (↓)				
Dataset	DiM	FaceMorpher	MIPGAN-II	OpenCV	StyleGAN2	DiM	FaceMorpher	MIPGAN-II	OpenCV	StyleGAN2
FERET [13]	✗	✓	✓	✓	✓	**72.73**	99.23	100	99.95	99.33
FERET [13]	✓	✗	✓	✓	✓	99.9	76.39	100	99.85	99.64
FERET [13]	✓	✓	✗	✓	✓	99.69	99.38	100	99.95	99.54
FERET [13]	✓	✓	✓	✗	✓	99.74	99.48	100	99.74	99.43
FERET [13]	✓	✓	✓	✓	✗	99.74	98.56	99.9	99.74	87.89
FRGC [14]	✗	✓	✓	✓	✓	**75.89**	99.98	99.97	99.9	99.93
FRGC [14]	✓	✗	✓	✓	✓	99.95	99.48	100	99.9	99.95
FRGC [14]	✓	✓	✗	✓	✓	99.83	99.85	99.82	99.8	99.85
FRGC [14]	✓	✓	✓	✗	✓	99.93	100	100	99.23	99.93
FRGC [14]	✓	✓	✓	✓	✗	99.93	99.93	99.94	99.88	97.83
FRLL [1]	✗	✓	✓	✓	✓	**13.96**	99.58	99.32	99.65	99.65
FRLL [1]	✓	✗	✓	✓	✓	99.23	99.09	98.91	99.37	99.44
FRLL [1]	✓	✓	✗	✓	✓	99.09	98.95	98.24	99.02	99.09
FRLL [1]	✓	✓	✓	✗	✓	99.51	99.44	99.19	99.16	99.58
FRLL [1]	✓	✓	✓	✓	✗	99.93	99.86	99.86	99.93	95.02

Relative Strength Metric

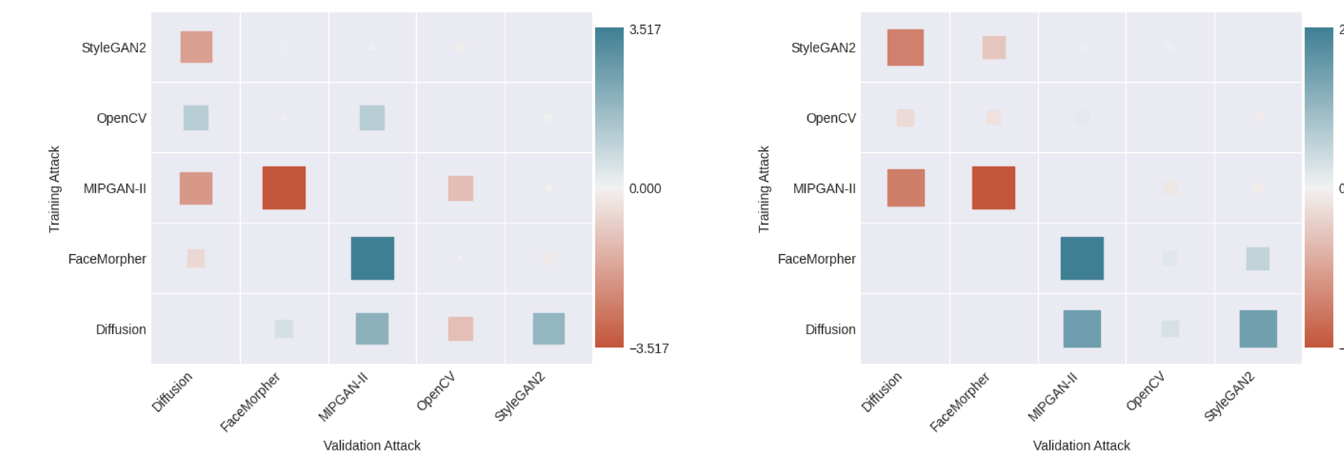
- We propose a metric to measure the relative strength between morphing attacks.
- The transferability of morphing attack α to β is defined as

$$T(\alpha, \beta) = P(f^\alpha(X^\beta) = 1 \mid f^\alpha(X^\alpha) = 1) \tag{10}$$

where X^α, X^β are morphs created by α, β and f^α is a detector trained on α.

- The relative strength metric (RSM) from α to β is:

$$\Delta(\alpha\|\beta) = \log\left(\frac{T(\alpha, \beta)}{T(\beta, \alpha)}\right) \tag{11}$$



(a) RSM on FRGC   (b) RSM on FERET

Figure 4. Blue indicates strong strength and red indicates weak strength.

Conclusion

- *First* morphing attack to use diffusion models
- Diffusion morphs are able to fool FR systems while retaining high visual fidelity
- Novel metric to compare the relative strength of morphing attacks
- Diffusion morphs are very difficult to detect if the detector is not trained against them

Related Works

Since our initial publication on DiM [12] several extensions to DiM have been proposed

Fast-DiM [15] High-order ODE solvers for faster sampling
Morph-PIPE [16] Brute force search for optimal γ w.r.t. an identity loss
Greedy-DiM [17] Greedy optimization for morphs with 100% MMPMR

References

[1] L. DeBruine and B. Jones, "Face Research Lab London Set," 5 2017.

[2] Z. Blasingame and C. Liu, "Leveraging adversarial learning for the detection of morphing attacks," *2021 IEEE International Joint Conference on Biometrics (IJCB)*, pp. 1–8, 2021.

[3] 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)*, pp. 2959–2963, 2022.

[4] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," in *Advances in Neural Information Processing Systems* (M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, eds.), vol. 34, pp. 8780–8794, Curran Associates, Inc., 2021.

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

[6] U. Scherhag, A. Nautsch, C. Rathgeb, M. Gomez-Barrero, R. N. J. Veldhuis, L. Spreeuwers, M. Schils, D. Maltoni, P. Grother, S. Marcel, R. Breithaupt, R. Ramachandra, and C. Busch, "Biometric systems under morphing attacks: Assessment of morphing techniques and vulnerability reporting," in *2017 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pp. 1–7, 2017.

[7] M. Huber, F. Boutros, A. T. Luu, K. Raja, R. Ramachandra, N. Damer, P. C. Neto, T. Gonçalves, A. F. Sequeira, J. S. Cardoso, J. Tremoço, M. Lourenço, S. Serra, E. Cermeño, M. Ivanovska, B. Batagelj, A. Kronovšek, P. Peer, and V. Štruc, "Syn-mad 2022: Competition on face morphing attack detection based on privacy-aware synthetic training data," in *2022 IEEE International Joint Conference on Biometrics (IJCB)*, pp. 1–10, 2022.

[8] M. Kim, A. K. Jain, and X. Liu, "Adaface: Quality adaptive margin for face recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

[9] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4690–4699, 2019.

[10] F. Boutros, N. Damer, F. Kirchbuchner, and A. Kuijper, "Elasticface: Elastic margin loss for deep face recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 1578–1587, June 2022.

[11] 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*, vol. 3, no. 3, pp. 365–383, 2021.

[12] Z. W. Blasingame and C. Liu, "Leveraging diffusion for strong and high quality face morphing attacks," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 6, no. 1, pp. 118–131, 2024.

[13] P. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The feret database and evaluation procedure for face-recognition algorithms," *Image Vis. Comput.*, vol. 16, pp. 295–306, 1998.

[14] P. Phillips, P. Flynn, T. Scruggs, K. 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)*, vol. 1, pp. 947–954 vol. 1, 2005.

[15] Z. W. Blasingame and C. Liu, "Fast-dim: Towards fast diffusion morphs," *IEEE Security & Privacy*, pp. 2–13, 2024.

[16] H. Zhang, R. Ramachandra, K. Raja, and B. Christoph, "Morph-pipe: Plugging in identity prior to enhance face morphing attack based on diffusion model," in *Norwegian Information Security Conference (NISK)*, 2023.

[17] Z. W. Blasingame and C. Liu, "Greedy-DiM: Greedy Algorithms for Unreasonably Effective Face Morphs," *arXiv e-prints*, p. arXiv:2404.06025, Apr. 2024.

Supported by CITeR and NSF under Grant # 1650503