Neural Implicit Morphing of Face Images
- Supplementary Material -

1. Morphing attack detection performance

In addition to the quality metrics of morphings, we present our results from the perspective of morphing attack detection (MAD) performance metrics. This experiment was conducted under the same protocol defined in Sec. 4.2 in the main work. The morphing type is different for protocols (but follow the same identity pairing list) while the original images remain the same. The morphed images were tested using the MAD approach proposed in [4].

The primary metric for assessing the detection performance of single image morphing relies on the relationship between the Bona fide Presentation Classification Error Rate (BPCER) and the Attack Presentation Classification Error Rate (APCER)(according to ISO/IEC 30107-3 [2]), which may be plotted as a Detection Error Trade-off (DET) curve. We report the performance by $BPCER@APCER = (0.1, 0.01)$ (see Table 1) and DET curve (see Fig 2).

Table 1. APCER@BPCER = (0.1, 0.01) of MAD method [4] in different protocols. Lower means better detector performance against the morphing method.

Morphing Type	$BPCER@APCER = \delta$	
	$\delta = 0.1$	$\delta = 0.01$
OpenCV	0.117	0.951
StyleGAN3	0.264	0.666
DiffAE	0.745	0.951
Ours (Linear)	0.176	0.519
Ours (S. Clone)	0.137	0.500
Ours (S. Mix)	0.421	0.824
Ours (diffAE)	0.754	0.950

By evaluating the MAD performance, we can gauge the detection complexity inherent in our morphs. Note that these results come with a disclaimer: the effectiveness of the chosen detector is not universal and may be influenced by certain unknown biases inherent to the detector itself [4].

Based on the results, the "Cloning" method demonstrates comparable performance to "OpenCV". The "Mix" morphs are harder to detect than other interpolation based morphs and they appear to be more challenging than StyleGAN3. DiffAE morphs are challenging to detect and indeed represent an unseen type of attack for this detector instance.

2. Ablation Studies

We present experiments showing the effects of each loss term (Sec. 2.1) and network width (Sec. 2.2). For these experiments, we employed the landmarks detected using DLib [3, 5], and images from the provided with the FRLL dataset [1]. Additionally, we used the network initialization procedure proposed in [6].

2.1. Loss function study

In this section, we measure the impact of each constraint of our loss function in the warping/blending. Recall that our loss function $\mathscr{L}(\theta) = \mathscr{W}(\theta) + \mathscr{D}(\theta) + \mathscr{T}(\theta)$ (Eq. 2 in the paper), has three terms: Warping ($\mathscr{W}$), data ($\mathscr{D}$), and thin-plate energy ($\mathscr{T}$). The warping constraint is composed by the identity and inverse constraints (Eq. 3 in the paper):

$$\mathscr{W}(\theta) = \underbrace{\int_{\mathbb{R}^2} \|\mathbf{T}(x,0) - x\|^2 dx}_{\text{Identity constraint}} + \underbrace{\int_{\mathbb{R}^2 \times \mathbb{R}} \|\mathbf{T}(\mathbf{T}(x,t),-t) - x\|^2 dx dt}_{\text{Inverse constraint}}. \quad (1)$$

Table 2 presents the FID and LPIPS ($I_0 \rightarrow I$, and $I \rightarrow I_1$) for these experiments. We used linear blending for all cases, and hidden-layer width of 128 neurons.

Table 2. FID and LPIPS for our warping with linear blending with different loss terms deactivated.

Experiment	**FID** ↓	**LPIPS** (I_0, I) ↓	**LPIPS** (I, I_1) ↓
No $\mathscr{D}$	113.599	0.215	0.246
No $\mathscr{T}$	499.289	0.594	0.607
No Id.	118.201	0.306	0.314
No Inv.	113.224	0.278	0.285
Original	31.950	0.158	0.164

From Table 2 we conclude that $\mathscr{T}$ has the most impact in our results, followed by the identity constraint (Id.), $\mathscr{D}$, and the inverse constraint (Inv.). $\mathscr{T}$ regularizes the loss, ensuring that the alignment is smooth during training, while Id. ensures that the warping will match the original landmark positions at $t = 0$ and $t = 1$. Additionally, $\mathscr{D}$ ensures that the source and target landmarks match for every t.

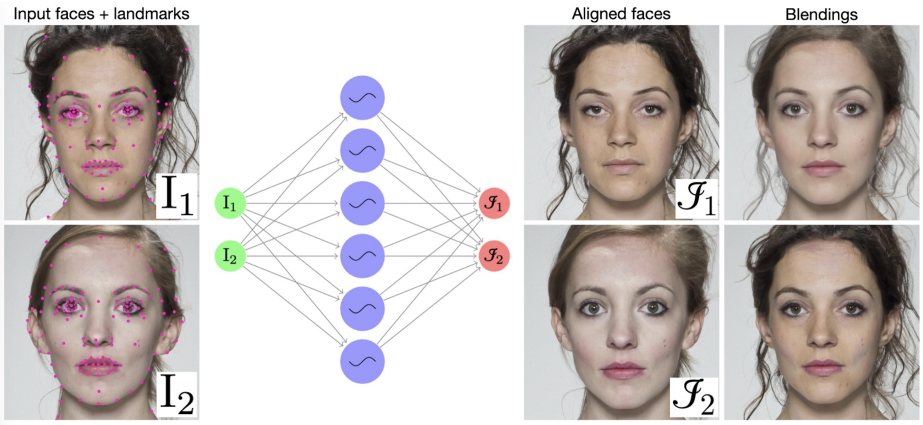

Figure 1. Overview of the proposed morphing approach. Let I_1 and I_2 be two face images represented as sinusoidal MLPs. We mark a series landmarks (show as magenta points) and pass these as input to train a warping network, shown in the second column, also a (shallow) sinusoidal MLP. The result of this step are two aligned and warped images $\mathcal{I}_1$ and $\mathcal{I}_2$, shown in the third column, encoded in the network itself. Afterwards, the warpings may be used to create different blendings. We've focused on linear, Poisson and generative blendings.

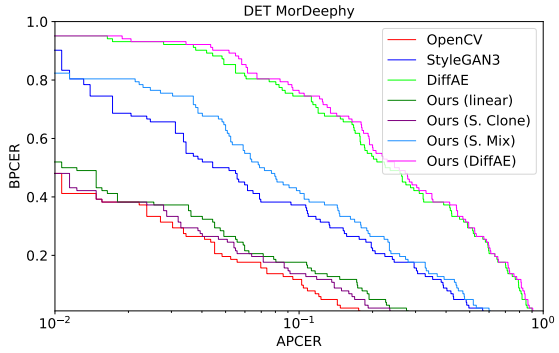

Figure 2. MAD performance of MorDeephy approach [4] on various types of face morphing.

Figure 3. Samples of results drawn with varying network widths. The left and right columns show I_0 and I_1 respectively. The second column shows a morphing from I_0 to I_1 at $t = 0.4$, while the third column shows morphing from I_1 to I_0 at $t = 0.6$.

2.2. Network width

For the neural warping, we vary the network width of its single hidden layer by 32 and 64 neurons. Fig 3 shows reconstructions using mixed cloning with varying values of t and morphing directions (either I_0 to I_1 or I_1 to I_0).

Table 3 shows the FID and LPIPS scores for varying hidden-layer widths. The result quality improves dramatically as the network size increases as well. However, at 128 neurons, the results reach a good trade-off between quality and training speed. Thus, we chose this size for the experiments in the main paper.

Table 3. FID and LPIPS for our warping with linear blending using 32, 64 or 128 neurons for the hidden layer. The 128 neuron results were reproduced from Tab. 1 in the main paper.

Net. Width	**FID** ↓	**LPIPS** (I_0, I) ↓	**LPIPS** (I, I_1) ↓
32	101.034	0.211	0.222
64	112.270	0.215	0.224
128	31.950	0.158	0.164

3. Additional Results

We show additional variations of ethnicities and genders to illustrate the flexibility of our approach (Sec. 3.1). Furthermore, we show experiments using automatic face landmark detection via DLib (Sec. 3.2) and how the morphings may be improved with additional manual adjustments to these landmarks (Sec. 3.2.1).

3.1. Variations of Gender and Ethnicity

One of the main limitations of the current state-of-the-art methods is generating credible morphings of people of different genders and ethnicities. Fig 4 shows results of these morphings using our approach. We employed the landmarks provided with the FRLL dataset for the neural warping.

Figure 4. Morphings between people of varying ethnicities and genders. The left column shows the source image (I_0), followed by the morphing of I_0 to I_1 at $t = 0.6$, morphing of I_1 to I_0 at $t = 0.4$, and the target image (I_1).

3.2. Employing Automatic Landmark Detection

The landmarks used for the experiments were provided with the dataset, however, other approaches for facial landmark detection may be employed and attain equally good results. For the experiments below, we used DLib with the 68 landmark model [3, 5]. Fig 5 shows the landmarks overlaid on sample images of the FRLL dataset, while Fig 6 shows the results of morphings using our method with said landmarks.

3.2.1 Manual Doctoring of Landmarks

In addition to automatic landmark detection, we may also manually edit landmarks to constrain other regions during the warping, such as clothing, and hairline. This improves the morphing on areas other than the face. Fig 7 shows a visual comparison between morphing using only automatic landmark detection and manual landmark adjustments.

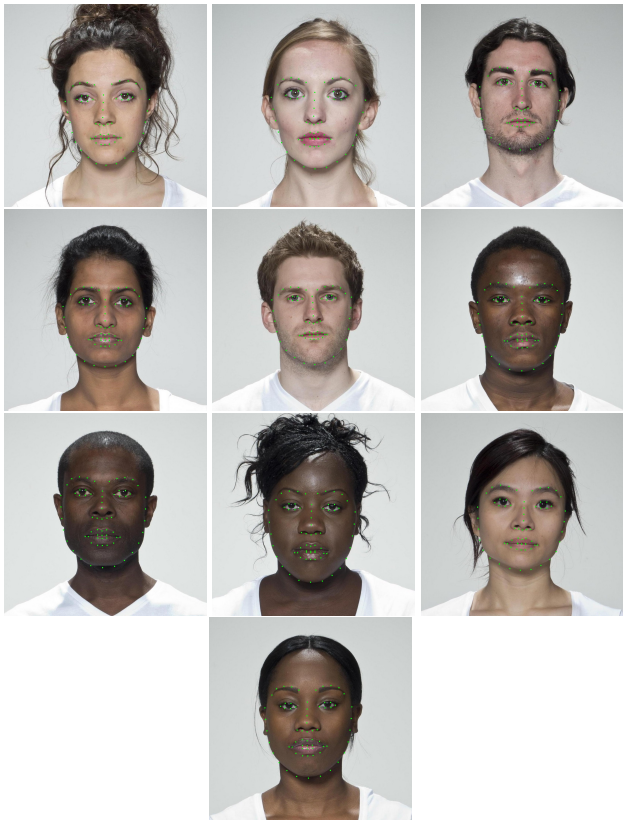

Figure 5. Landmarks found by DLib 68 landmarks overlaid on sample faces of the FRLL dataset.

Figure 6. Morphings created using our neural warping approach with landmarks obtained with DLib 68 landmarks model. From left to right: Source image (I_0), followed by the morphing of I_0 to I_1 at $t = 0.6$, morphing of I_1 to I_0 at $t = 0.4$, and target image (I_1).

Figure 7. Morphing with automatic and manually placed landmarks. Left and right columns show I_0 and I_1. The second column shows a linear blending at $t = 0.6$ with manually positioned landmarks in addition to the ones detected by DLib. The third column shows a linear blending after our neural warping using only DLib landmarks. Notice how, besides the face, other areas such as neck and hairline are aligned on the second column, compared to the third.

References

[1] L. DeBruine and B. Jones. Face research lab london set, May 2017. 1

[2] International Organization for Standardization. ISO/IEC 30107–3:2017. Information Technology—Biometric Presentation Attack Detection — Part 3: Testing and Reporting. ISO/IEC JTC 1/SC 37 Biometrics, 09 2017. 1

[3] V. Kazemi and J. Sullivan. One millisecond face alignment with an ensemble of regression trees. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014. 1, 3

[4] I. Medvedev, F. Shadmand, and N. Gonçalves. Mordeephy: Face morphing detection via fused classification. In *Proceedings of the 12th International Conference on Pattern Recognition Applications and Methods - Volume 1: ICPRAM,*, pages 193–204. INSTICC, SciTePress, 2023. ISBN 978-989-758-626-2. doi: 10.5220/0011606100003411. 1, 2

[5] C. Sagonas, E. Antonakos, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic. 300 faces in-the-wild challenge: database and results. *Image and Vision Computing*, 47:3–18, 2016. ISSN 0262-8856. doi: https://doi.org/10.1016/j.imavis.2016.01.002. URL https://www.sciencedirect.com/science/article/pii/S0262885616000147. 300-W, the First Automatic Facial Landmark Detection in-the-Wild Challenge. 1, 3

[6] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein. Implicit neural representations with periodic activation functions. *Advances in Neural Information Processing Systems*, 33, 2020. 1