FreeMorph: Tuning-Free Generalized Image Morphing with Diffusion Model
(– –*Supplementary Material*– –)

Yukang Cao[1*] Chenyang Si[2*‡] Jinghao Wang[3] Ziwei Liu[1†]
[1]S-Lab, Nanyang Technological University, [2]Nanjing University
[3]The Chinese University of Hong Kong
https://yukangcao.github.io/FreeMorph/

Contents

* Equal contributions, ‡ Project lead, † Corresponding author.
Jinghao was a Master student at NTU during this work.

List of Figures

A. Further Analysis

A.1. Usage of the Fast Fourier Transform (FFT)

In our approach, we employ the fast Fourier transform (FFT) to inject high-frequency Gaussian noise, which enhances flexibility. An alternative and straightforward variation involves replacing the FFT with the discrete cosine transform (DCT). To investigate this, we conducted experiments using both FFT and DCT, presenting the results in Fig. 1. The findings indicate that DCT performs comparably to FFT.

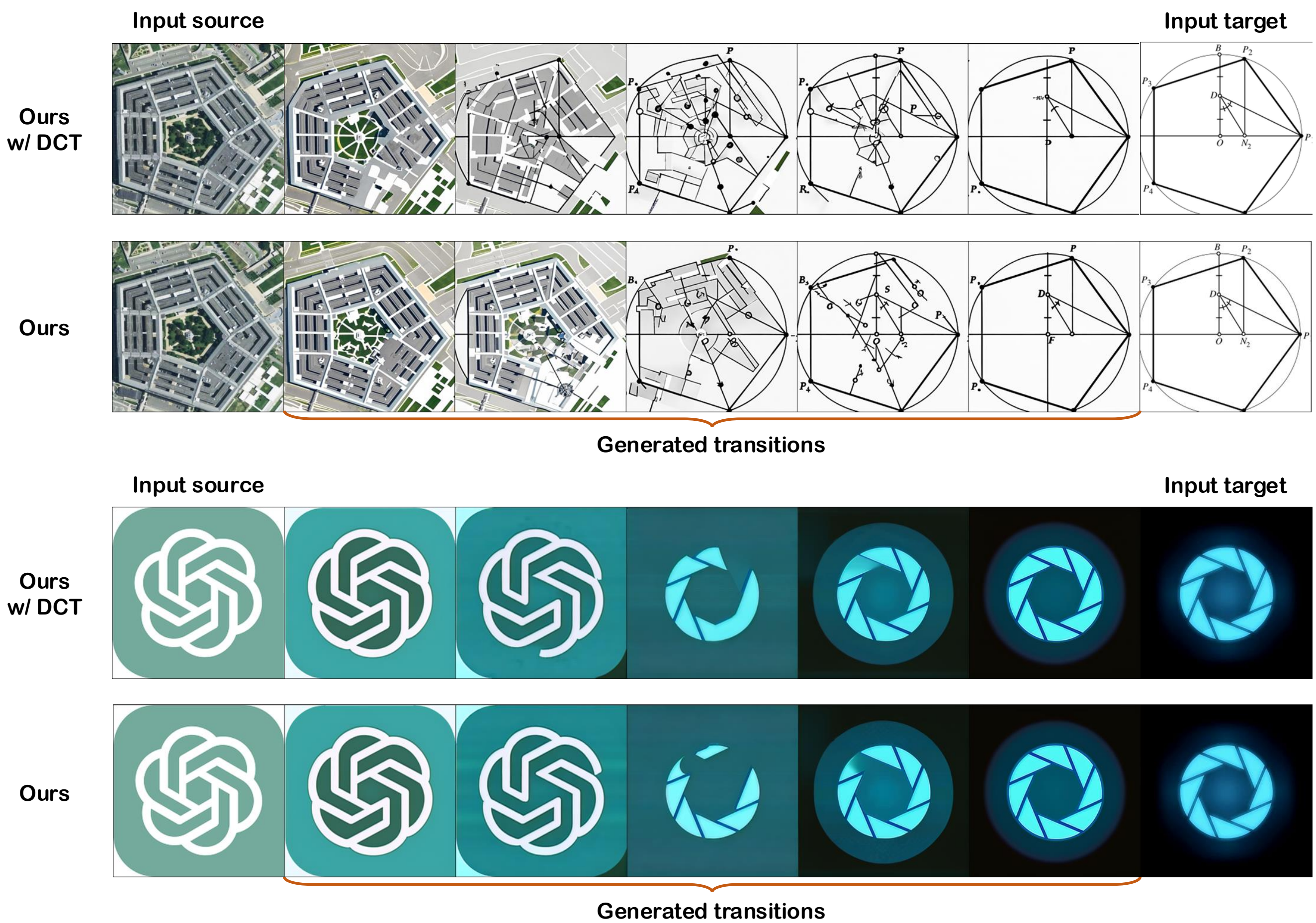


Figure 1. **Analysis of the usage of Fast Fourier Transform (FFT) over Discrete Cosine Transform (DCT).**

B. Qualitative Comparisons

B.1. Qualitative Comparisons with AID [3] and Smooth Diffusion [2]

In addition to the comparisons discussed in the main paper, we extend our evaluation to include AID [3] and Smooth Diffusion [2]. As illustrated in Fig. 2 and Fig. 3, the results demonstrate that both methods are limited to processing images with similar layouts and semantics, rendering them ineffective for inputs with different layouts or semantics. Beyond their qualitative shortcomings, it is worth noting that **(1)** AID relies on IP-Adapter for image morphing, which adversely affects training efficiency, and **(2)** Smooth Diffusion requires parameter tuning, making it slower and less efficient than our approach.

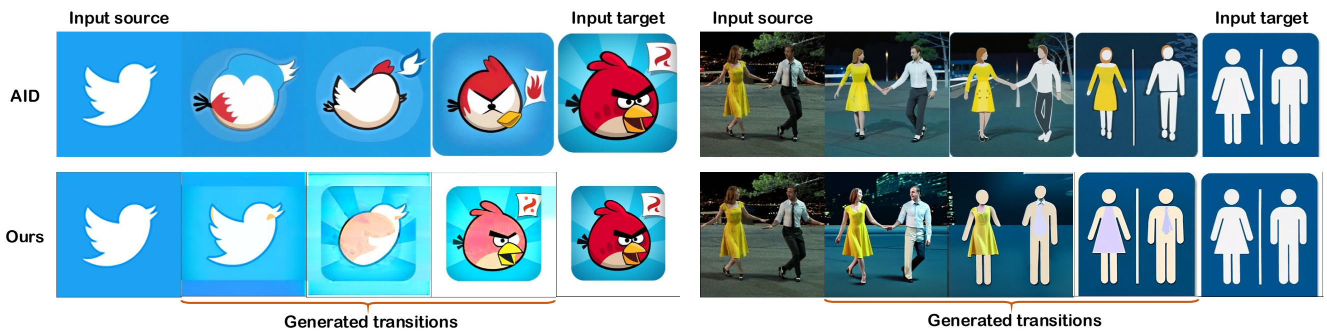


Figure 2. **Qualitative comparisons with AID [3].**

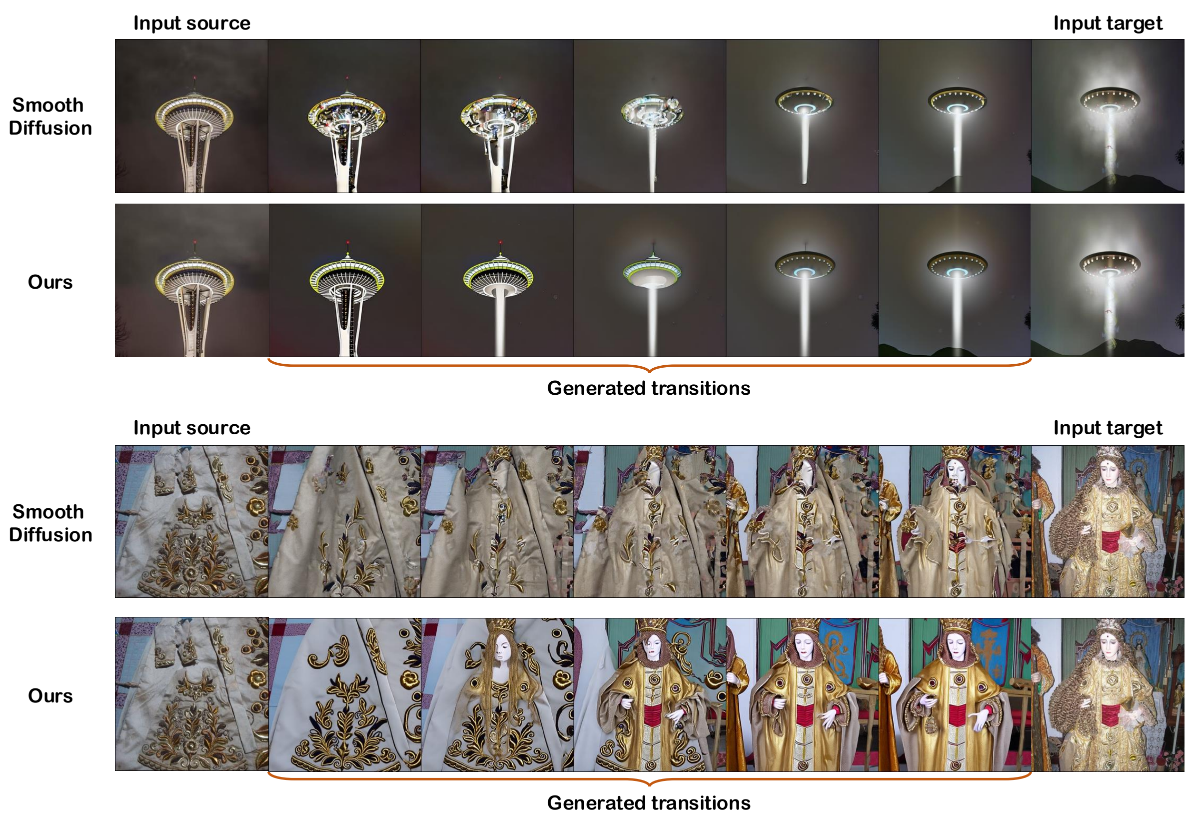


Figure 3. **Qualitative comparisons with Smooth Diffusion [2]**

B.2. Comparison with video generative models

Given the rapid development of video generative techniques. Methods like PixelDance [6] and SEINE [1] have been designed to achieve image morphing. We hereby provide more comparisons with these video generative models to demonstrate our performance. Considering PixelDance hasn't released code or an online demo, we ran FreeMorph on the examples from their webpages to perform qualitative comparisons (see Fig. 4 below). Surprisingly, our method performs similarly with PixelDance and outperforms SEINE in reducing ghost artifacts.

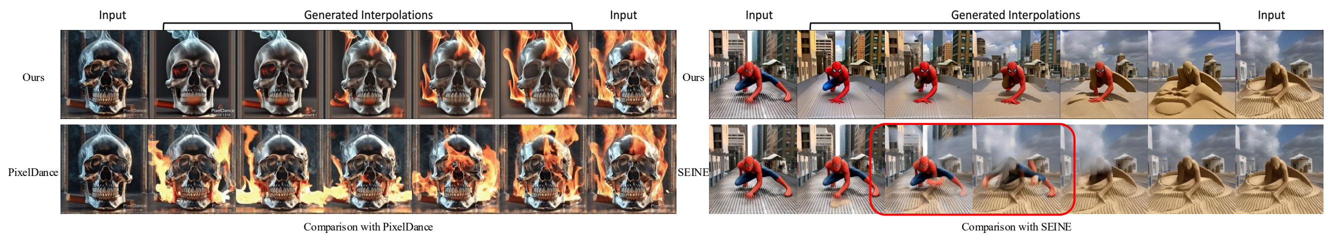


Figure 4. **Comparisons with video generative models.**

B.3. Comparison with GAN-based morphing methods

We further compare our method with the early GAN-based morphing method (Neural Crossbreed) to demonstrate the performance. The results, presented in Fig. 5, show superior image quality, identity preservation, and smoother transitions. Unlike GAN-based approaches, ours is training-free, is able to handle out-of-domain inputs, and remains robust to varying layouts and semantics. Additional evaluations and discussions will be included in the revised version.

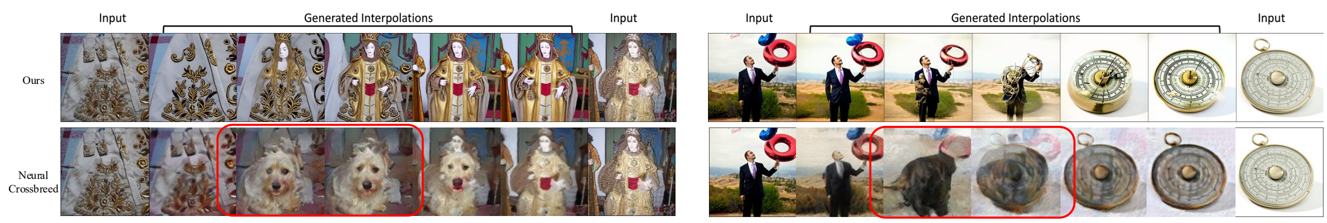
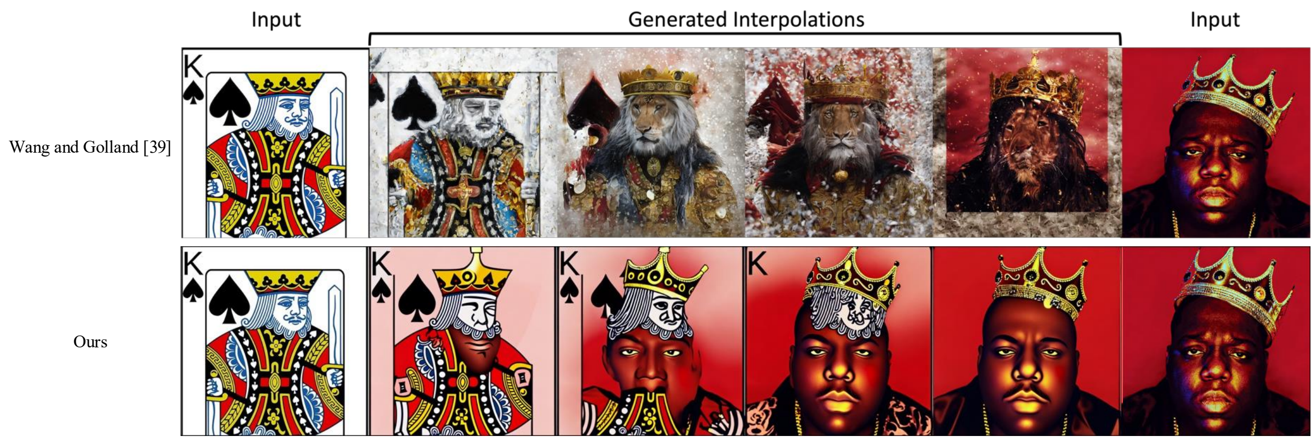


Figure 5. **Comparison with GAN-based morphing methods.**

B.4. Comparison with Wang and Golland [5]

We further compare with Wang and Golland [39] and present the results in Fig. 6. We can clearly observe that our method consistently show superior performance over it, both qualitatively and quantitatively.



Method	MorphBench			Morph4Data			Overall		
	$\text{LPIPS}_{\text{sum}} \downarrow$	$\text{FID}_{\text{mean}} \downarrow$	$\text{PPL}_{\text{sum}} \downarrow$	$\text{LPIPS}_{\text{sum}} \downarrow$	$\text{FID}_{\text{mean}} \downarrow$	$\text{PPL}_{\text{sum}} \downarrow$	$\text{LPIPS}_{\text{sum}} \downarrow$	$\text{FID}_{\text{mean}} \downarrow$	$\text{PPL}_{\text{sum}} \downarrow$
Wang and Golland [39]	145.4	225.57	3302.49	139.09	296.43	3144.05	284.49	252.75	6446.54
Ours	**84.91**	**141.32**	**2122.80**	**80.30**	**201.09**	**2007.52**	**162.99**	**152.88**	**4192.82**

Figure 6. **Comparison with Wang and Golland [5].**

B.5. Experiments with different poses/actions

We further present results for various poses and actions below (Fig. 7), using input images from the MorphBench dataset.

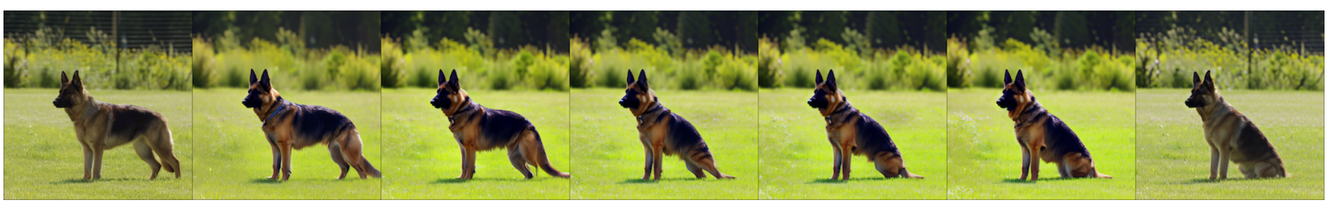


Figure 7. **Qualitative results with different poses/actions.**

B.6. Additional Qualitative Comparisons

We provide additional qualitative comparisons with three methods in Fig. 8–Fig. 15. These results reinforce the conclusions drawn in Sec. 4.2 of the main paper, offering further evidence of the superior performance of our FreeMorph method in achieving high-fidelity and smooth image morphing.

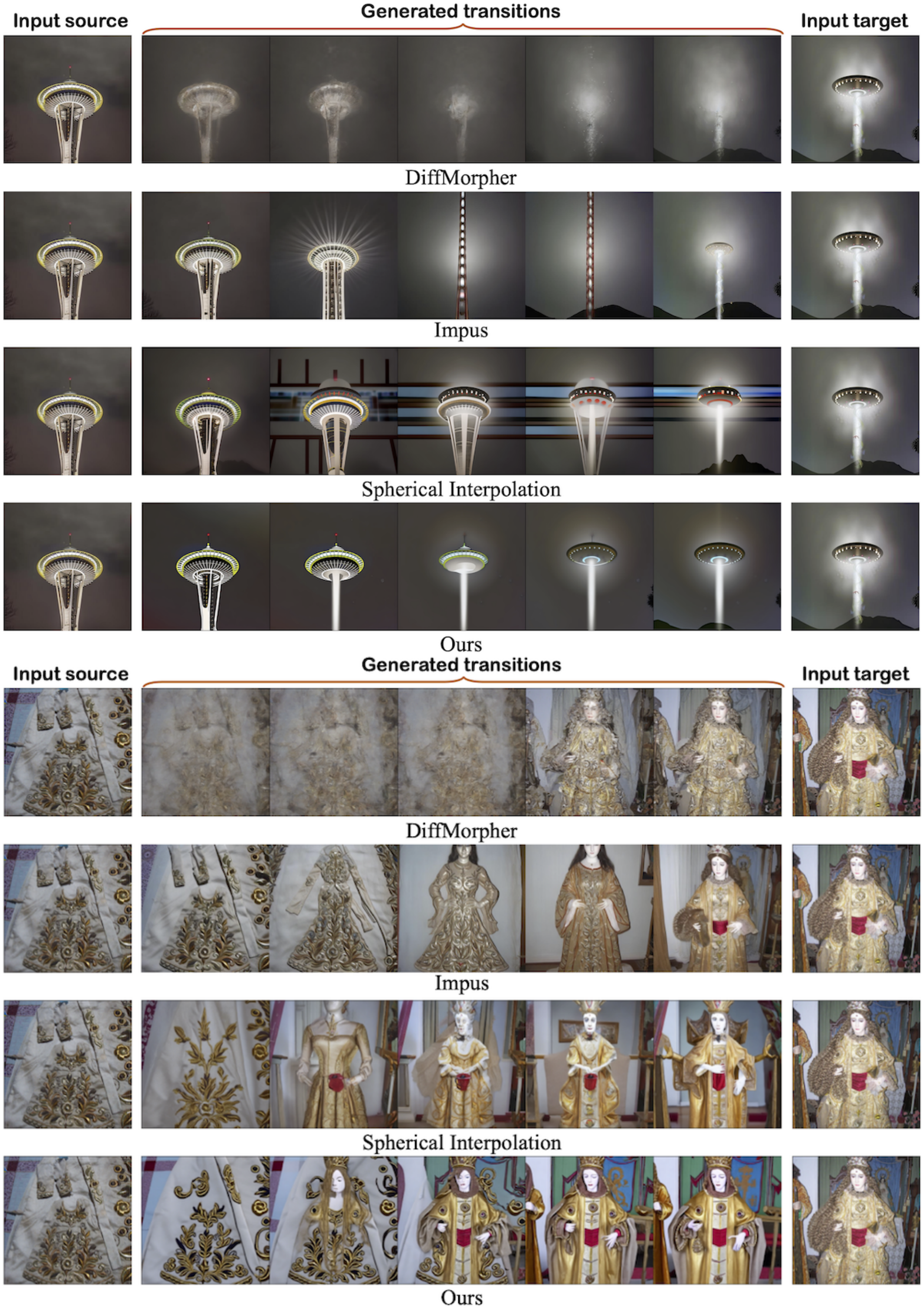


Figure 8. **More qualitative comparisons with existing techniques (Part I).**

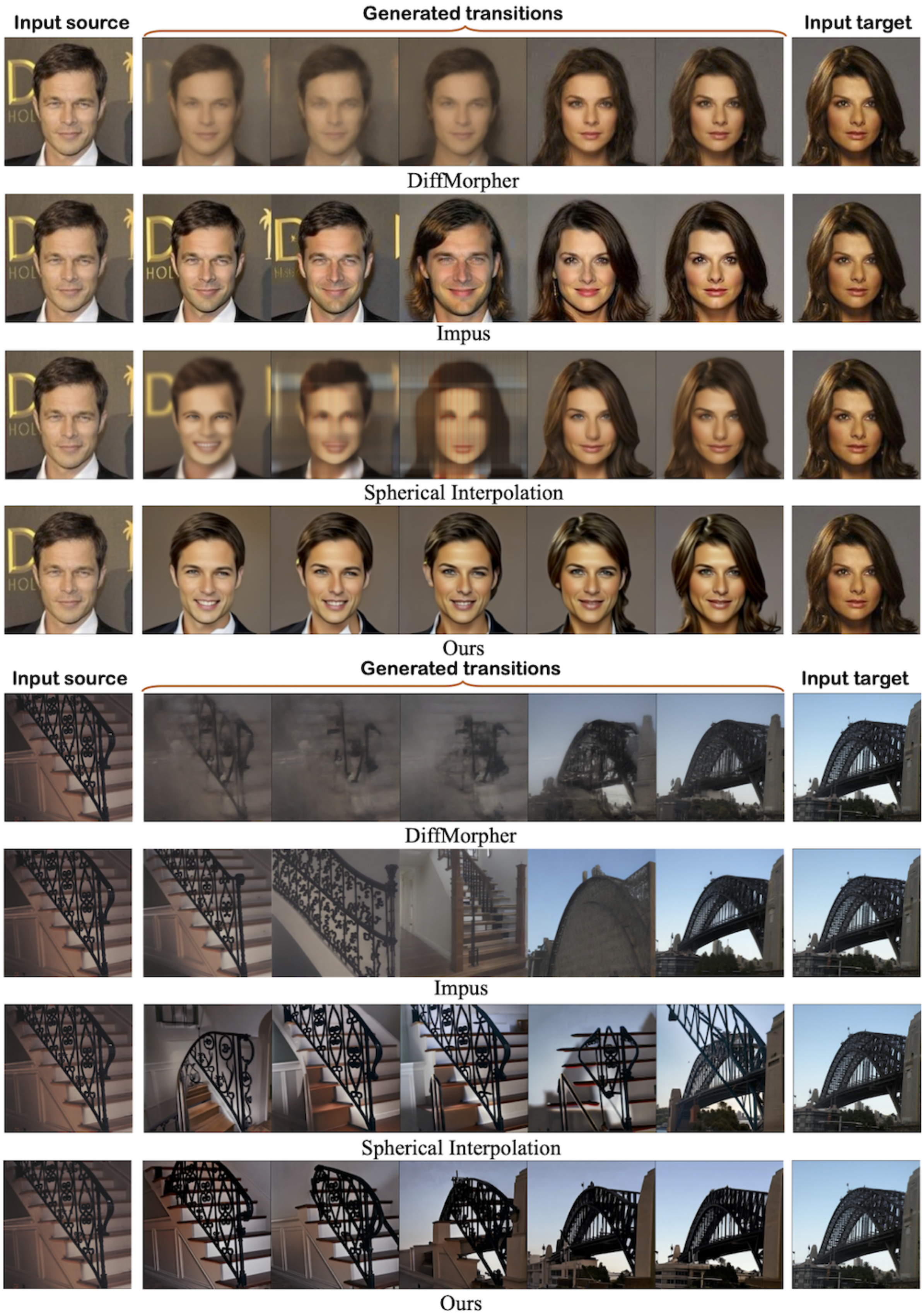


Figure 9. **More qualitative comparisons with existing techniques (Part II).**

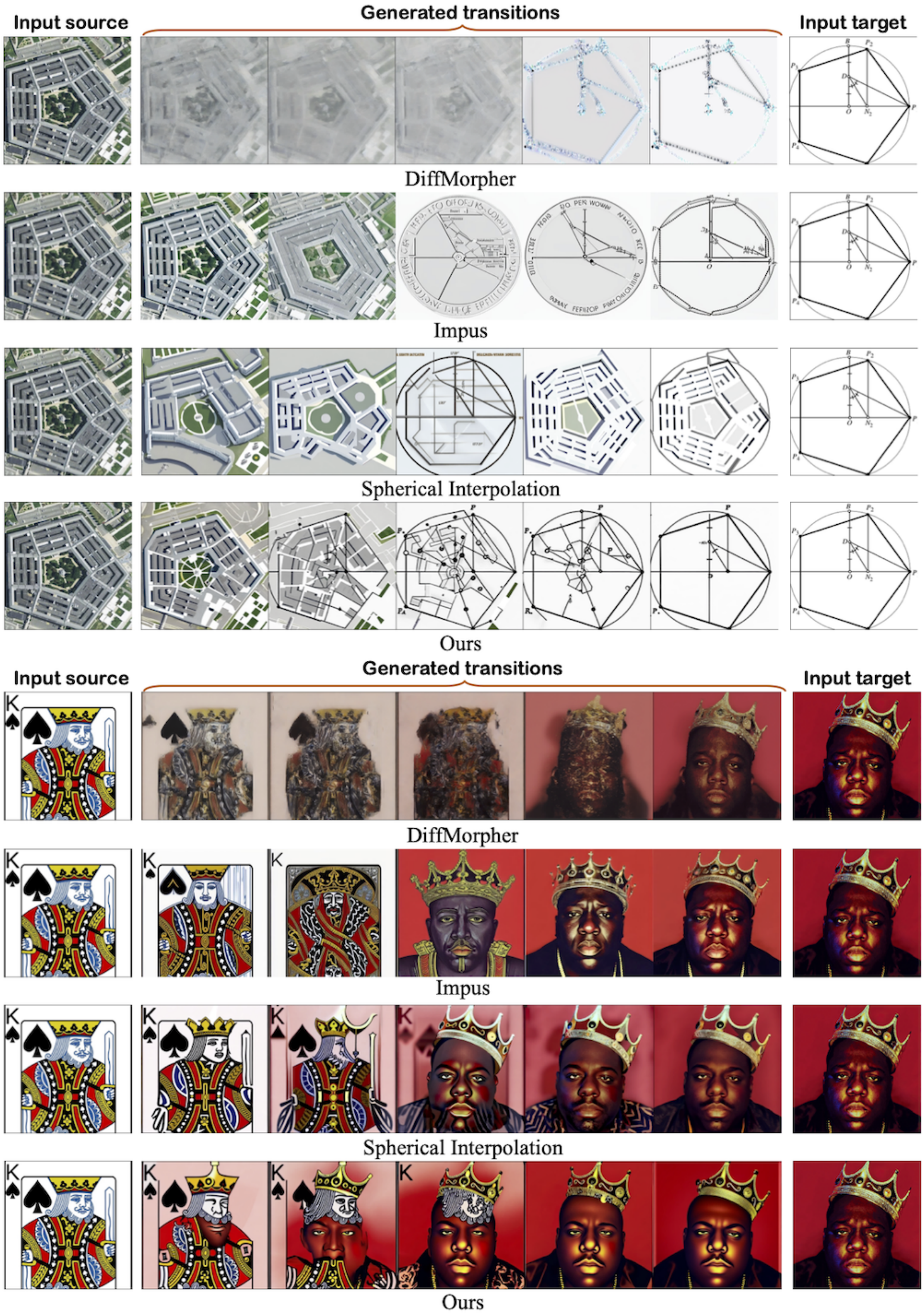


Figure 10. **More qualitative comparisons with existing techniques (Part III).**

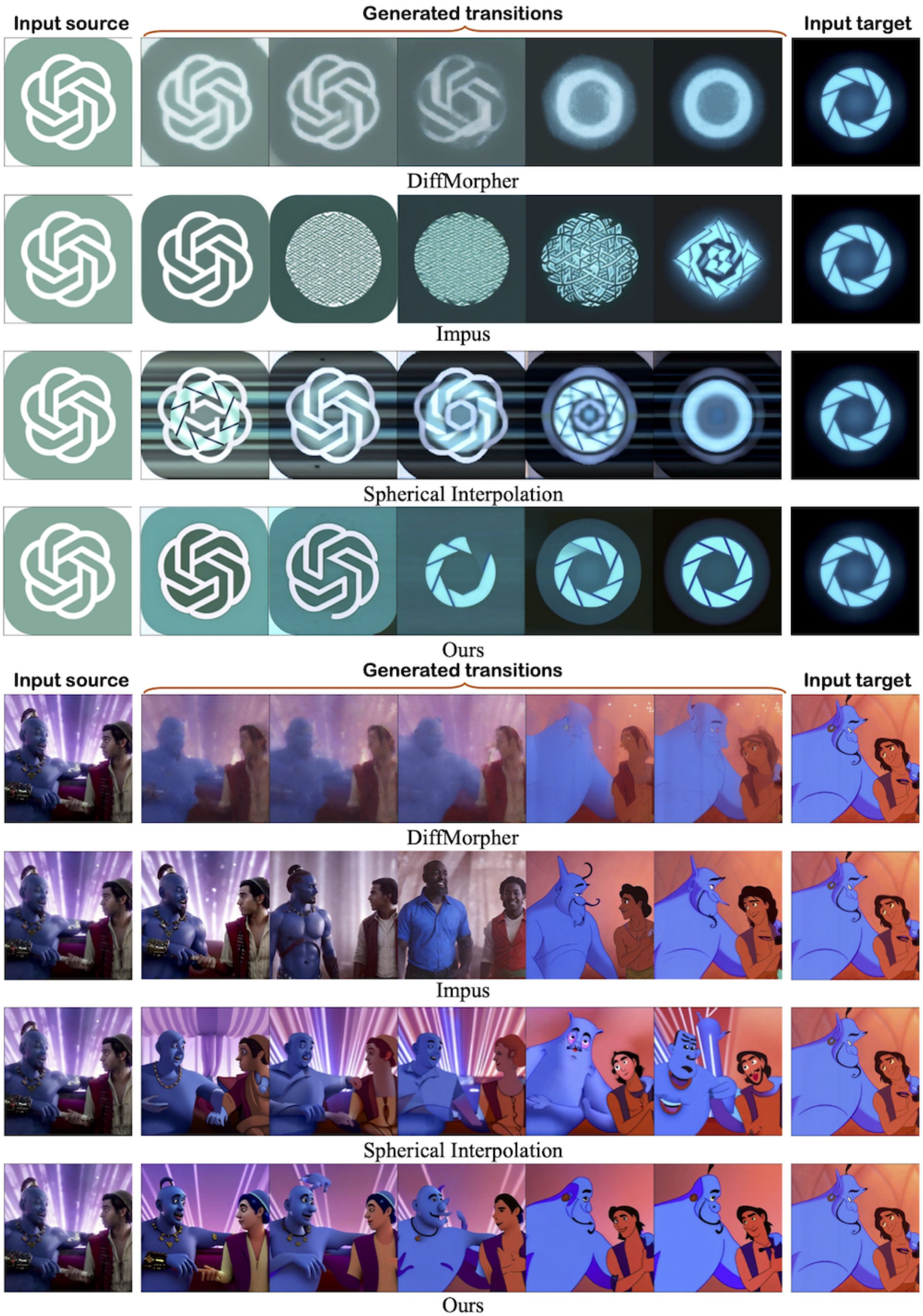


Figure 11. **More qualitative comparisons with existing techniques (Part IV).**

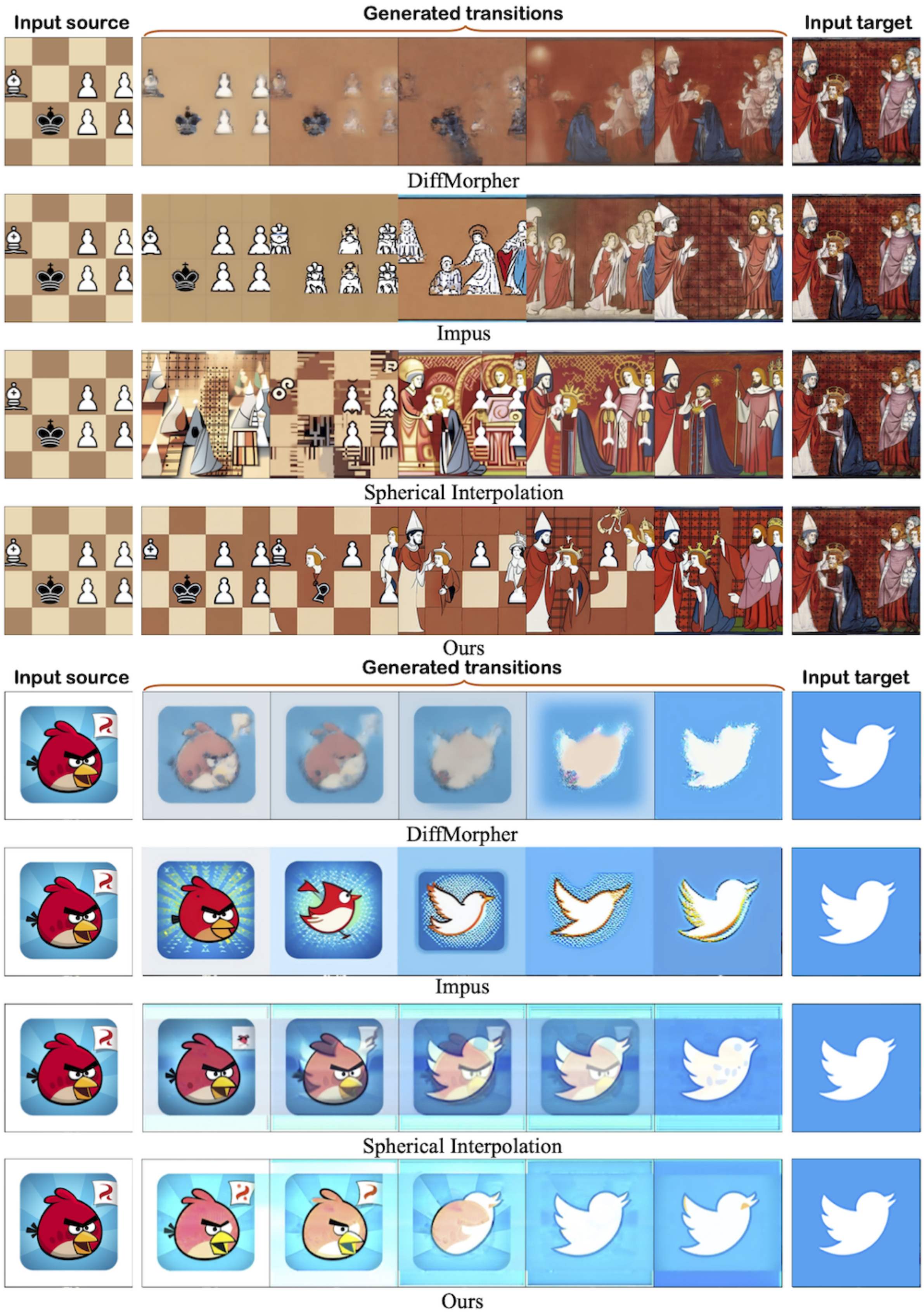


Figure 12. **More qualitative comparisons with existing techniques (Part V).**

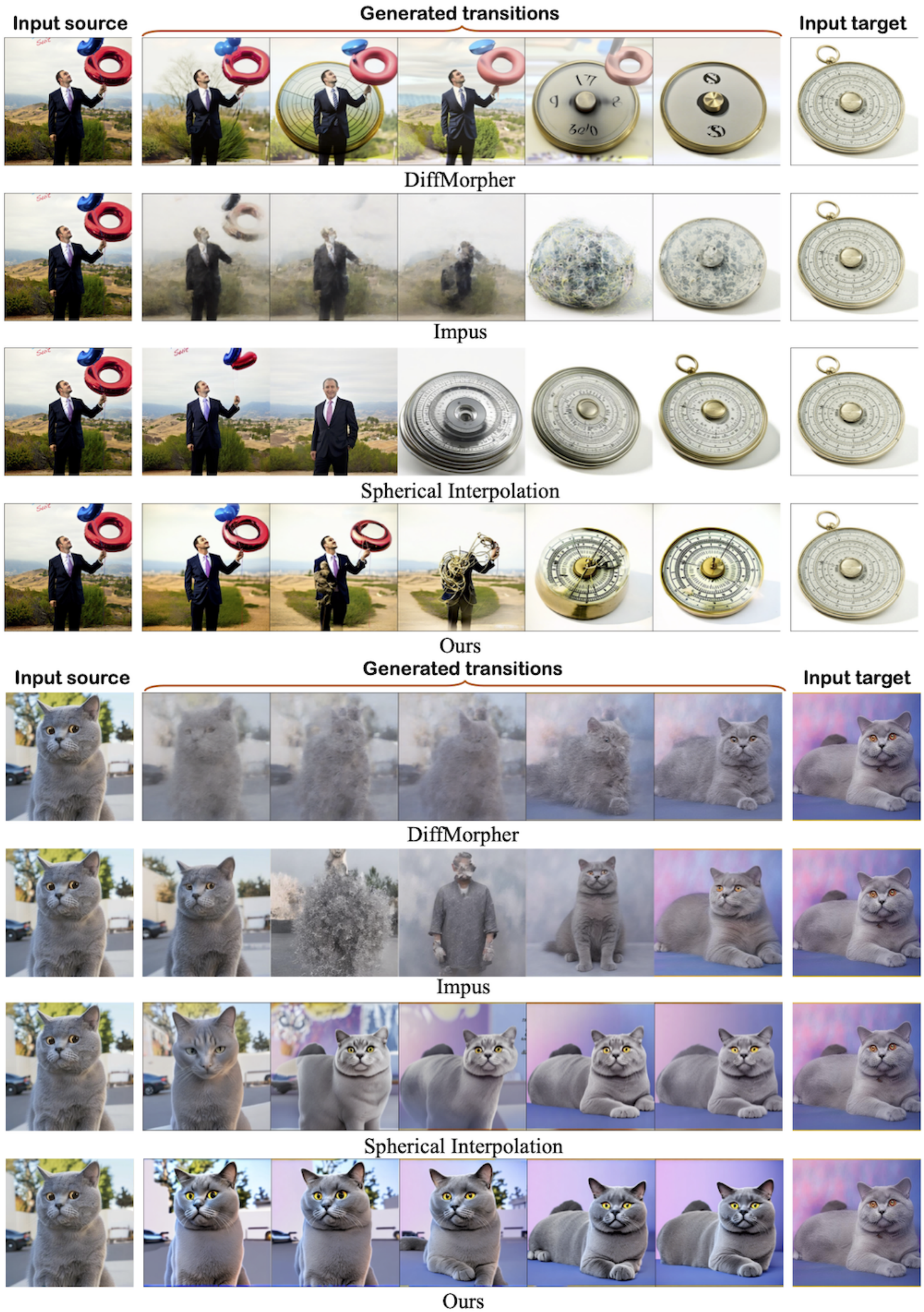


Figure 13. **More qualitative comparisons with existing techniques (Part VI).**

Figure 14. **More qualitative comparisons with existing techniques (Part VII).**

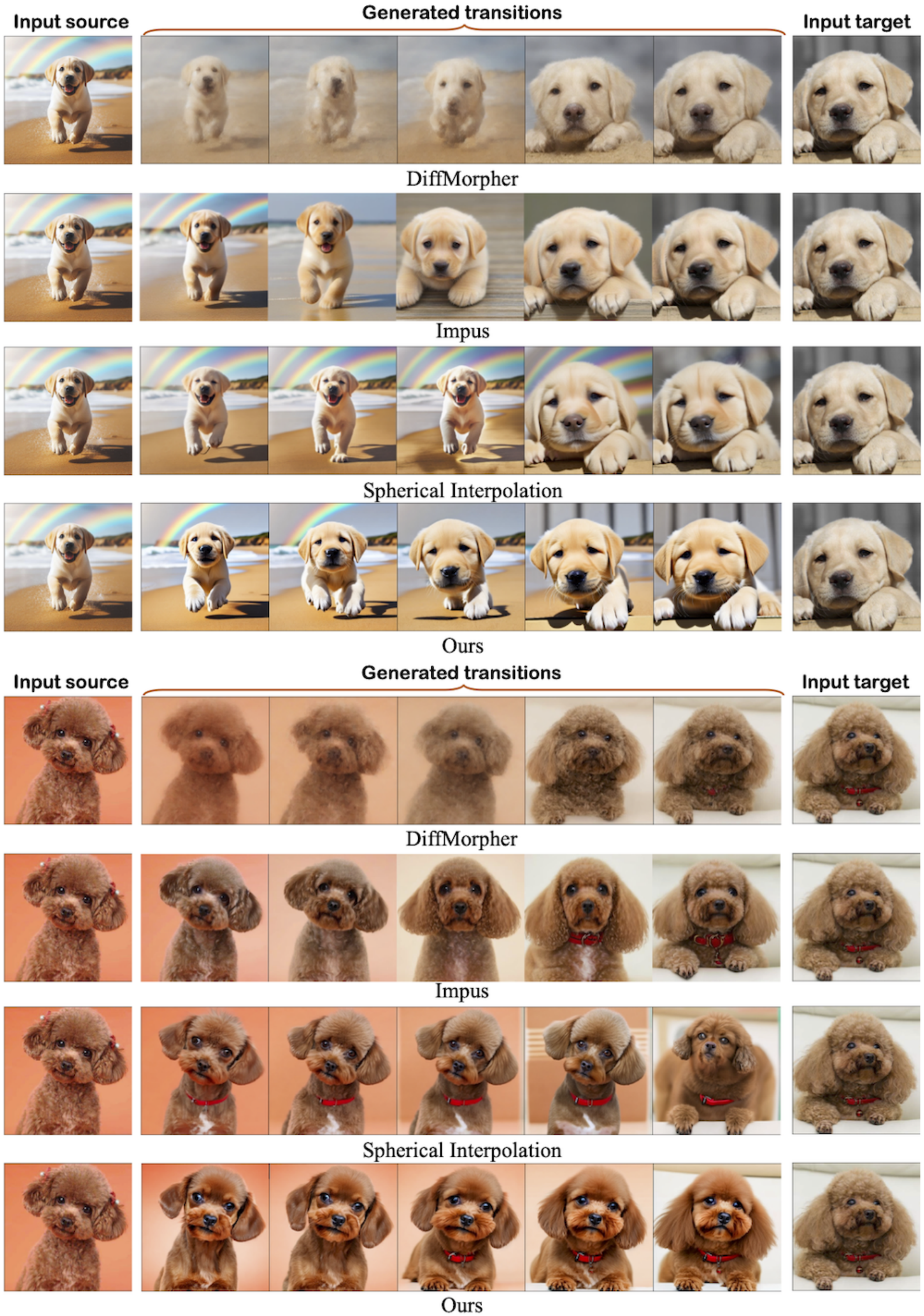


Figure 15. **More qualitative comparisons with existing techniques (Part VIII).**

C. More Qualitative Results

To provide a better understanding of the intermediate generated transitions, in addition to the animated videos, we also present generated images in Fig. 16–Fig. 19, **which correspond to the animated videos in the HTML file**.

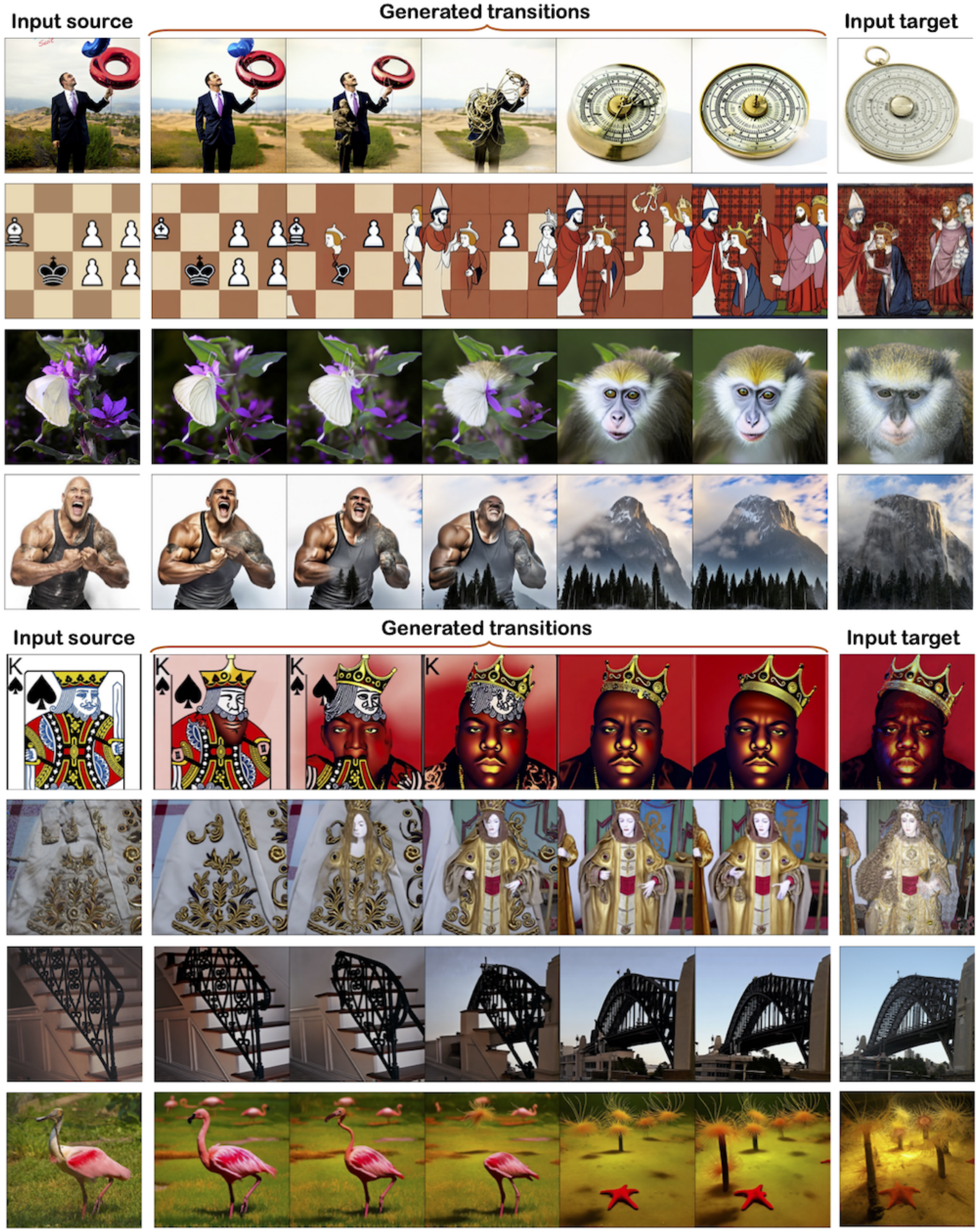


Figure 16. **Images with different semantics and different layouts.**

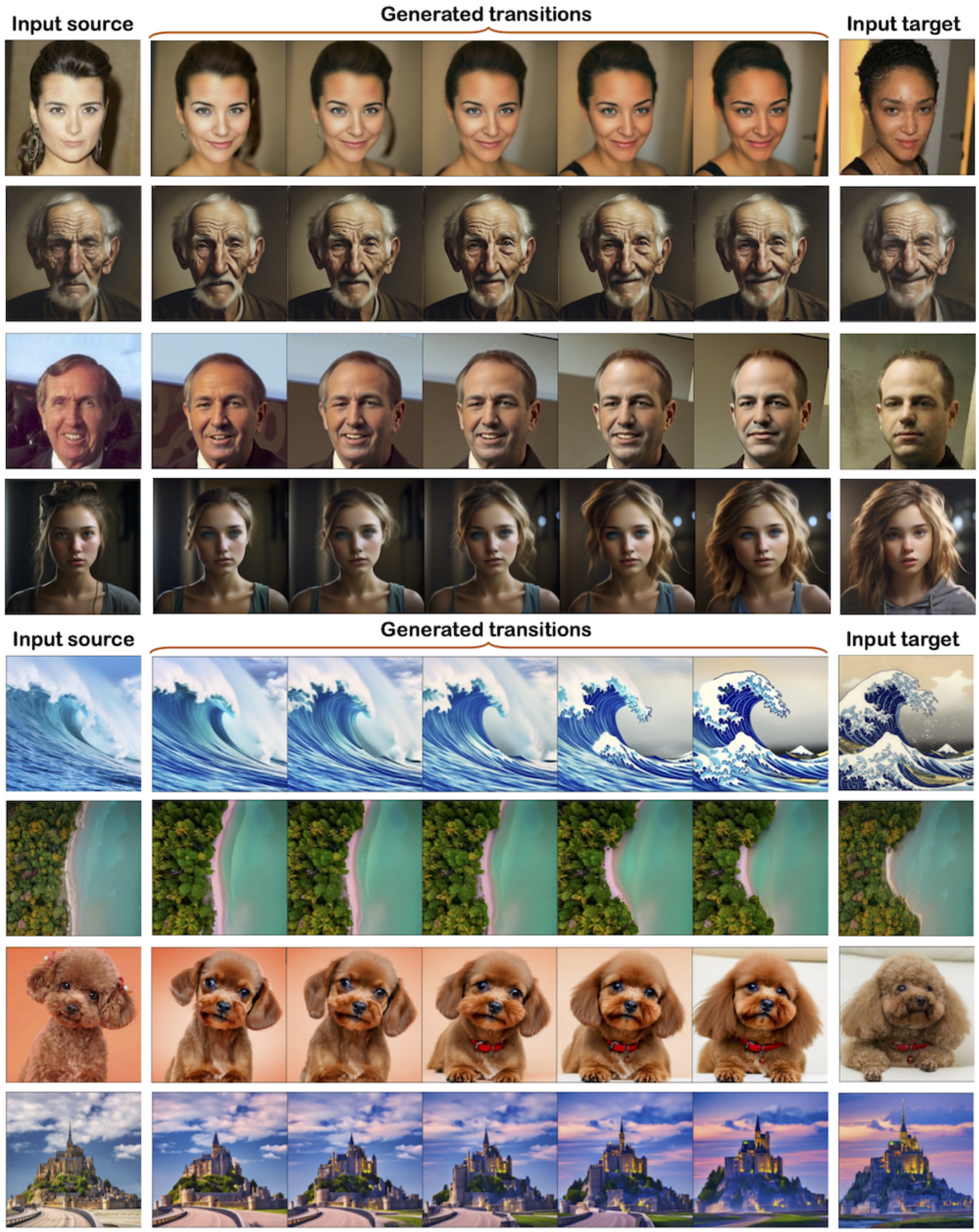


Figure 17. **Images with similar semantics and similar layouts.**

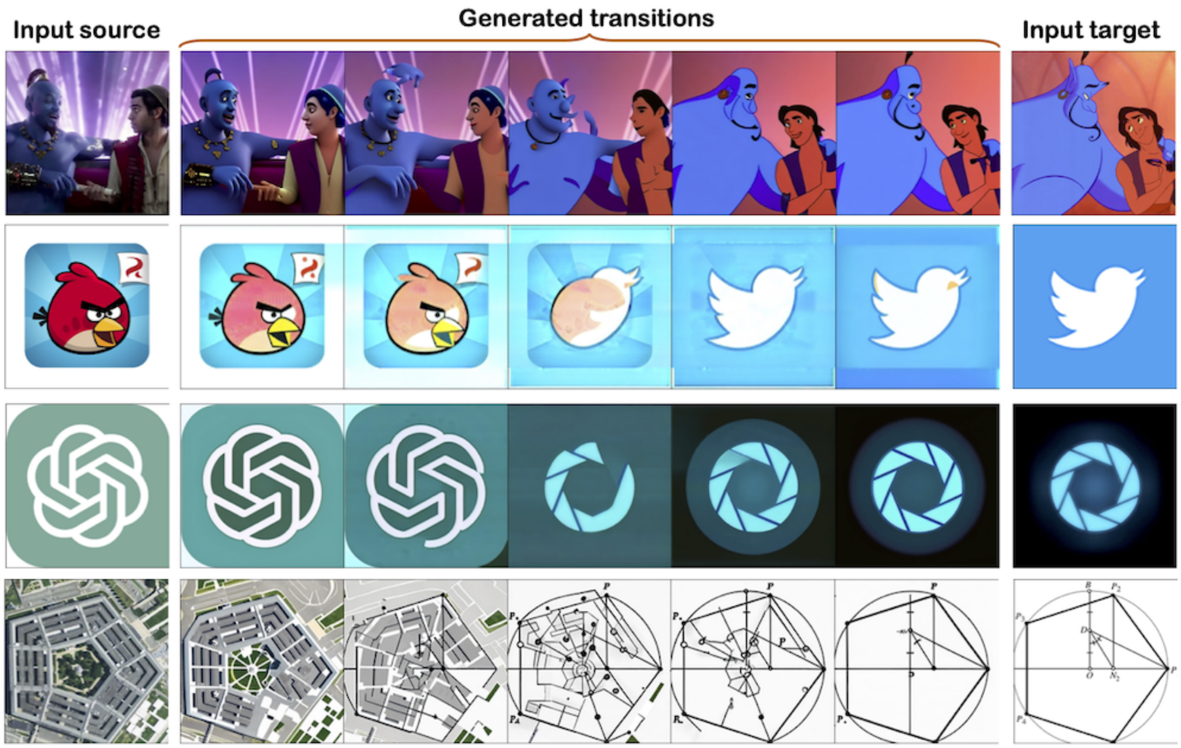


Figure 18. **Images with different semantics and similar layouts.**

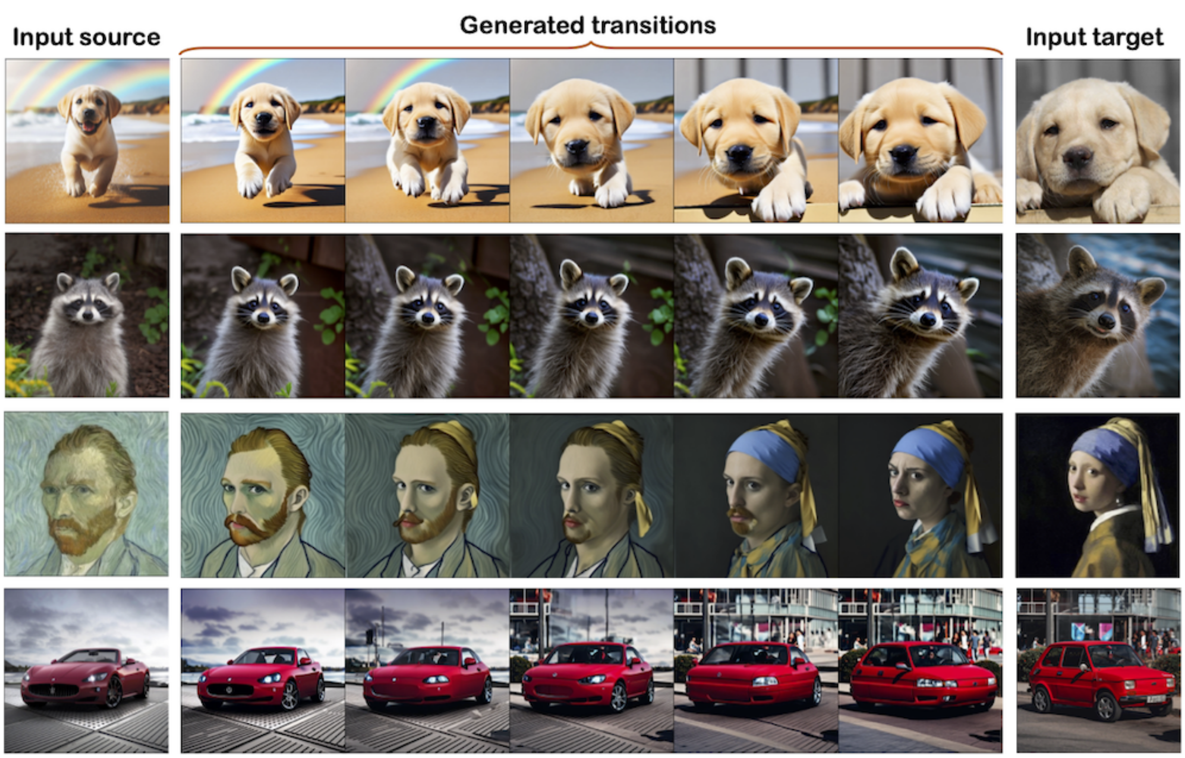


Figure 19. **Images with similar semantics and different layouts.**

D. Visualization of Morph4Data

We present a range of visualizations from our collected Morph4Data to enhance understanding of the dataset and the distinctions among its different classes.

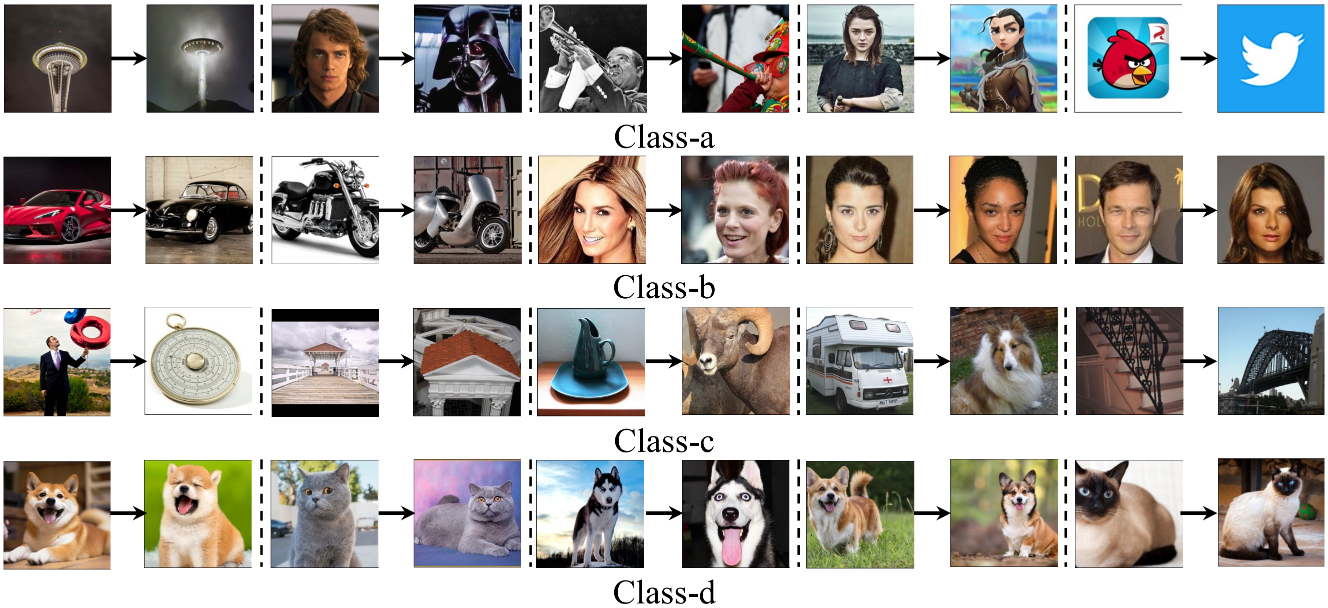


Figure 20. **Examples of 4 classes in Morph4Data.**

E. Applications

We highlight that our FreeMorph method can be adapted for image editing tasks. Specifically, this is accomplished by (1) using the same image as both the "input source" and "input target," and (2) employing different text prompts, where the first prompt describes the original image and subsequent prompts indicate the desired editing direction. An example is provided in Fig. 21. Notably, our method produces image editing results that align correctly with the text prompts, preserving the original identity while effectively generating smooth transitions throughout the editing process.

Figure 21. **Application of FreeMorph in image editing**

F. Limitations and Failure Cases

While our method establishes a new state-of-the-art, we acknowledge that it has certain limitations. We illustrate several failure cases in Fig. 22. Specifically: **1)** Although our model can achieve reasonable results when processing images with no semantic or layout similarity, the generated transitions may not be smooth, potentially leading to abrupt changes. **2)** Our method inherits biases from Stable Diffusion [4], resulting in difficulties in accurately transitioning images that model human limbs.

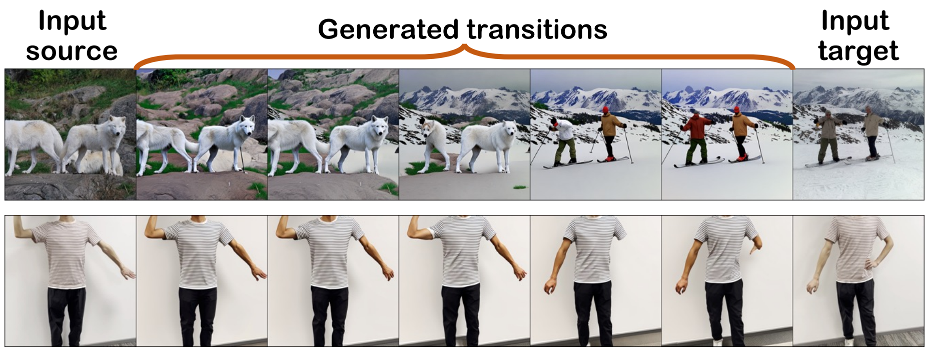


Figure 22. **Failure cases.**

G. Societal Impact

Our research advances the image morphing task across a range of semantics and layouts, establishing a more versatile pipeline. However, there is a risk of misuse, such as brands creating misleading advertisements that distort consumer perceptions and create unrealistic product expectations. This practice not only undermines consumer trust but also raises significant ethical concerns about the authenticity of marketing. Additionally, the complexities of copyright and consent are amplified, as manipulated images blur the lines of ownership and accountability. Therefore, we advocate for strict legal compliance and usage restrictions to regulate the application of image morphing techniques and derivative models.

References

[1] Xinyuan Chen, Yaohui Wang, Lingjun Zhang, Shaobin Zhuang, Xin Ma, Jiashuo Yu, Yali Wang, Dahua Lin, Yu Qiao, and Ziwei Liu. Seine: Short-to-long video diffusion model for generative transition and prediction. In *The Twelfth International Conference on Learning Representations*, 2023. 5

[2] Jiayi Guo, Xingqian Xu, Yifan Pu, Zanlin Ni, Chaofei Wang, Manushree Vasu, Shiji Song, Gao Huang, and Humphrey Shi. Smooth diffusion: Crafting smooth latent spaces in diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7548–7558, 2024. 1, 2, 4

[3] He Qiyuan, Jinghao Wang, Ziwei Liu, and Angela Yao. Aid: Attention interpolation of text-to-image diffusion. *Advances in Neural Information Processing Systems*, 2024. 1, 2, 4

[4] Stability.AI. Stable diffusion. `https://stability.ai/blog/stable-diffusion-public-release`, 2022. 19

[5] Clinton J Wang and Polina Golland. Interpolating between images with diffusion models. *arXiv preprint arXiv:2307.12560*, 2023. 1, 2, 6

[6] Yan Zeng, Guoqiang Wei, Jiani Zheng, Jiaxin Zou, Yang Wei, Yuchen Zhang, and Hang Li. Make pixels dance: High-dynamic video generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8850–8860, 2024. 5