Supplementary Material for QuantArt: Quantizing Image Style Transfer Towards High Visual Fidelity

Siyu Huang[1]* Jie An[2]* Donglai Wei[3] Jiebo Luo[2] Hanspeter Pfister[1]
[1]Harvard University [2]University of Rochester [3]Boston College
{huang,pfister}@seas.harvard.com {jan6,jluo}@cs.rochester.edu donglai.wei@bc.edu

In this supplementary material, we show additional quantitative (Section A) and qualitative (Section B) comparisons with the state-of-the-art image style transfer methods for different task settings. Section C presents more examples and user study results of our Artistic Style Transfer Confusion Test.

A. More Quantitative Comparisons

We quantitatively compare our method with the state-of-the-art methods for three different settings: artwork-to-artwork (Table A), photo-to-photo (Table B), and artwork-to-photo (Table C). Tables A and B show that our QuantArt(1,1) framework achieves the best FID and ArtFID scores on both settings, which is consistent with the quantitative result of the photo-to-artwork task shown in the paper. The consistent results in different settings reinforce our finding that feature quantization can lead to a higher visual fidelity in image style transfer. QuantArt(0,1) has a lower Gram loss than QuantArt(1,1) because the visual and style fidelities are orthogonal transfer directions in some cases. To this end, we introduce a stronger control ability to the proposed framework, where the user can easily control the trade-off between visual, style, and content fidelities by adjusting the parameters α and β.

For the artwork-to-photo task, following Art2Real [14], we do not adopt the style reference, such that the task is formulated as an unpaired image translation problem. We train an SGA module that takes the feature of an artwork image as the content input. The second attention block in SGA is replaced with a self-attention block. We compare our method with CycleGAN [17], CUT [12], DRIT [6] and Art2Real [14], where CycleGAN [17], CUT [12] and DRIT [6] are benchmark methods for unpaired image translation, and Art2Real [14] is the algorithm specially designed for artwork-to-photo translation. Table C shows that our QuantArt framework can achieve decent performance without special modification for the artwork-to-photo task, since it directly fetches photorealism patch representations from the learned photo codebook.

B. More Qualitative Results

Figs. A and B show more examples of artwork-to-artwork image style transfer. Fig. C shows more examples of photo-to-photo image style transfer. Fig. D shows more examples of artwork-to-photo image style transfer. Figs. E, F, G, H, I show more examples of photo-to-artwork image style transfer.

C. Examples of Artistic Style Transfer Confusion Test

We present all 36 examples and corresponding user feedbacks of our Artistic Style Transfer Confusion Test in Figs. J, K, L, M, N, and O. The ratios of correct selections with respect to individual examples range from 19.4% to 76.5% with an overall average of 51.6%. The results indicate that our QuantArt method can generate highly realistic artistic images, and most of the generations are difficult for humans to identify from the real artworks.

References

[1] Jie An, Siyu Huang, Yibing Song, Dejing Dou, Wei Liu, and Jiebo Luo. Artflow: Unbiased image style transfer via reversible neural flows. In *CVPR*, pages 862–871, 2021. 3, 4, 7, 8, 9, 10, 11

Table A. Quantitative comparison of the state-of-the-art methods for **artwork-to-artwork** image style transfer. The **best** and <u>second best</u> results are highlighted, respectively.

Metric ↓	AdaIN	WCT	LinearWCT	ArtFlow	EFDM	StyleSwap	AvatarNet	SANet	AdaAttN	StyTR2	Ours (α, β)		
											$(0,0)$	$(0,1)$	$(1,1)$
LPIPS ↓	0.600	0.695	0.591	0.522	0.630	0.616	0.725	0.622	0.587	0.494	*0.264*	**0.407**	<u>0.437</u>
Gram loss $(\times 10^3)$ ↓	0.159	0.267	0.160	**0.105**	0.395	1.367	0.712	<u>0.123</u>	0.242	0.384	-	0.677	0.965
FID ↓	31.668	73.170	42.248	25.642	47.264	66.375	50.234	25.721	24.827	<u>14.363</u>	-	15.763	**10.973**
ArtFID ↓	52.266	125.719	68.793	40.544	78.663	108.878	88.369	43.343	40.980	<u>22.959</u>	-	23.587	**17.209**

Table B. Quantitative comparison of the state-of-the-art methods for **photo-to-photo** image style transfer.

Metric ↓	PhotoWCT	WCT2	LinearWCT	DSTN	Ours (α, β)		
					$(0,0)$	$(0,1)$	$(1,1)$
LPIPS ↓	0.509	**0.348**	0.353	0.415	*0.158*	0.494	0.388
Gram loss $(\times 10^3)$ ↓	0.650	0.372	0.499	0.460	-	**0.355**	1.114
FID ↓	15.462	8.946	17.775	20.583	-	19.651	**4.261**
ArtFID ↓	24.848	13.409	25.402	30.536	-	30.847	**7.300**

Table C. Quantitative comparison of the state-of-the-art methods for **artwork-to-photo** image style transfer.

Metric ↓	CycleGAN	CUT	DRIT	Art2Real	Ours	
					$\alpha = 0$	$\alpha = 1$
LPIPS ↓	**0.226**	0.427	0.573	0.280	0.457	0.457
FID ↓	42.899	35.237	41.419	34.608	31.002	**20.864**
ArtFID ↓	53.827	51.724	66.717	45.585	46.641	**31.847**

[2] Tian Qi Chen and Mark Schmidt. Fast patch-based style transfer of arbitrary style. *arXiv preprint arXiv:1612.04337*, 2016. 3, 4, 7, 8, 9, 10, 11

[3] Yingying Deng, Fan Tang, Weiming Dong, Chongyang Ma, Xingjia Pan, Lei Wang, and Changsheng Xu. Stytr2: Image style transfer with transformers. In *CVPR*, pages 11326–11336, 2022. 3, 4, 7, 8, 9, 10, 11

[4] Kibeom Hong, Seogkyu Jeon, Huan Yang, Jianlong Fu, and Hyeran Byun. Domain-aware universal style transfer. In *ICCV*, pages 14609–14617, 2021. 5

[5] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *CVPR*, pages 1501–1510, 2017. 3, 4, 7, 8, 9, 10, 11

[6] Hsin-Ying Lee, Hung-Yu Tseng, Jia-Bin Huang, Maneesh Singh, and Ming-Hsuan Yang. Diverse image-to-image translation via disentangled representations. In *ECCV*, pages 35–51, 2018. 1, 6

[7] Xueting Li, Sifei Liu, Jan Kautz, and Ming-Hsuan Yang. Learning linear transformations for fast image and video style transfer. In *CVPR*, pages 3809–3817, 2019. 3, 4, 5, 7, 8, 9, 10, 11

[8] Yijun Li, Chen Fang, Jimei Yang, Zhaowen Wang, Xin Lu, and Ming-Hsuan Yang. Universal style transfer via feature transforms. In *NIPS*, pages 386–396, 2017. 3, 4, 7, 8, 9, 10, 11

[9] Yijun Li, Ming-Yu Liu, Xueting Li, Ming-Hsuan Yang, and Jan Kautz. A closed-form solution to photorealistic image stylization. In *ECCV*, pages 453–468, 2018. 5

[10] Songhua Liu, Tianwei Lin, Dongliang He, Fu Li, Meiling Wang, Xin Li, Zhengxing Sun, Qian Li, and Errui Ding. Adaattn: Revisit attention mechanism in arbitrary neural style transfer. In *ICCV*, pages 6649–6658, 2021. 3, 4, 7, 8, 9, 10, 11

[11] Dae Young Park and Kwang Hee Lee. Arbitrary style transfer with style-attentional networks. In *CVPR*, 2019. 3, 4, 7, 8, 9, 10, 11

[12] Taesung Park, Alexei A Efros, Richard Zhang, and Jun-Yan Zhu. Contrastive learning for unpaired image-to-image translation. In *ECCV*, pages 319–345, 2020. 1, 6

[13] Lu Sheng, Ziyi Lin, Jing Shao, and Xiaogang Wang. Avatar-net: Multi-scale zero-shot style transfer by feature decoration. In *CVPR*, pages 8242–8250, 2018. 3, 4, 7, 8, 9, 10, 11

[14] Matteo Tomei, Marcella Cornia, Lorenzo Baraldi, and Rita Cucchiara. Art2real: Unfolding the reality of artworks via semantically-aware image-to-image translation. In *CVPR*, pages 5849–5859, 2019. 1, 6

[15] Jaejun Yoo, Youngjung Uh, Sanghyuk Chun, Byeongkyu Kang, and Jung-Woo Ha. Photorealistic style transfer via wavelet transforms. In *ICCV*, pages 9036–9045, 2019. 5

[16] Yabin Zhang, Minghan Li, Ruihuang Li, Kui Jia, and Lei Zhang. Exact feature distribution matching for arbitrary style transfer and domain generalization. In *CVPR*, pages 8035–8045, 2022. 3, 4, 7, 8, 9, 10, 11

[17] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, pages 2223–2232, 2017. 1, 6

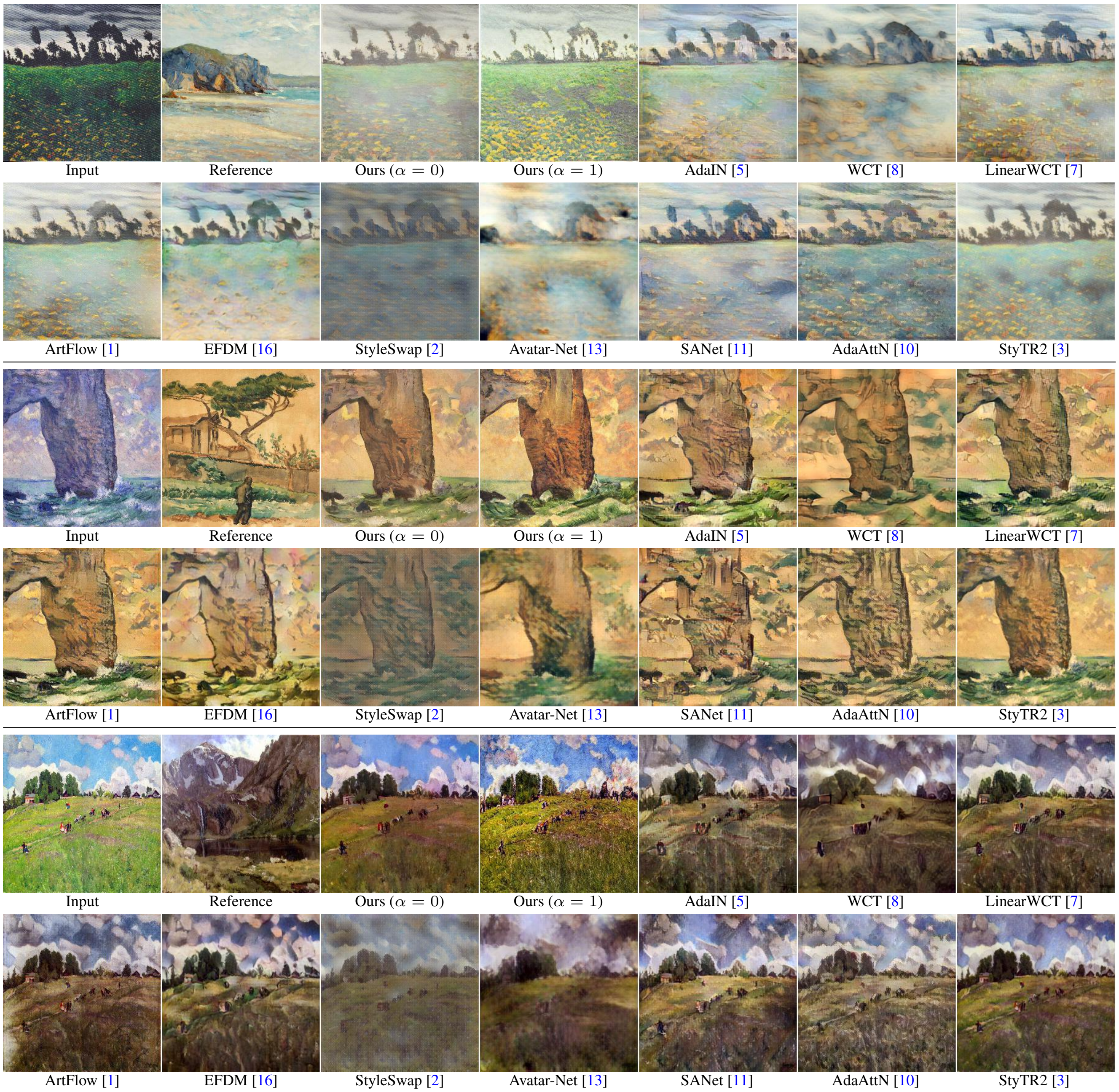


Figure A. More **artwork-to-artwork** style transfer results by the state-of-the-art methods.

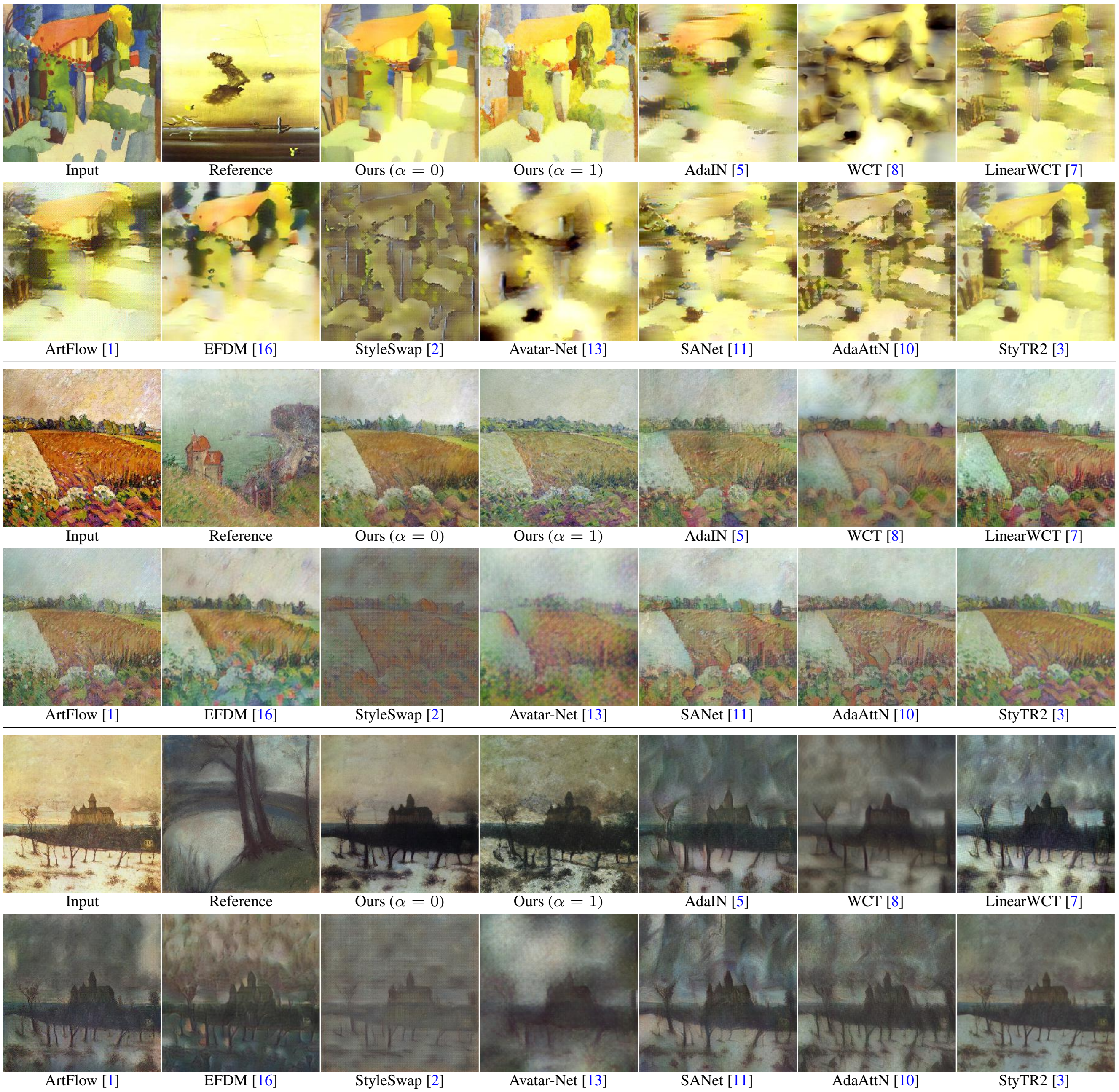
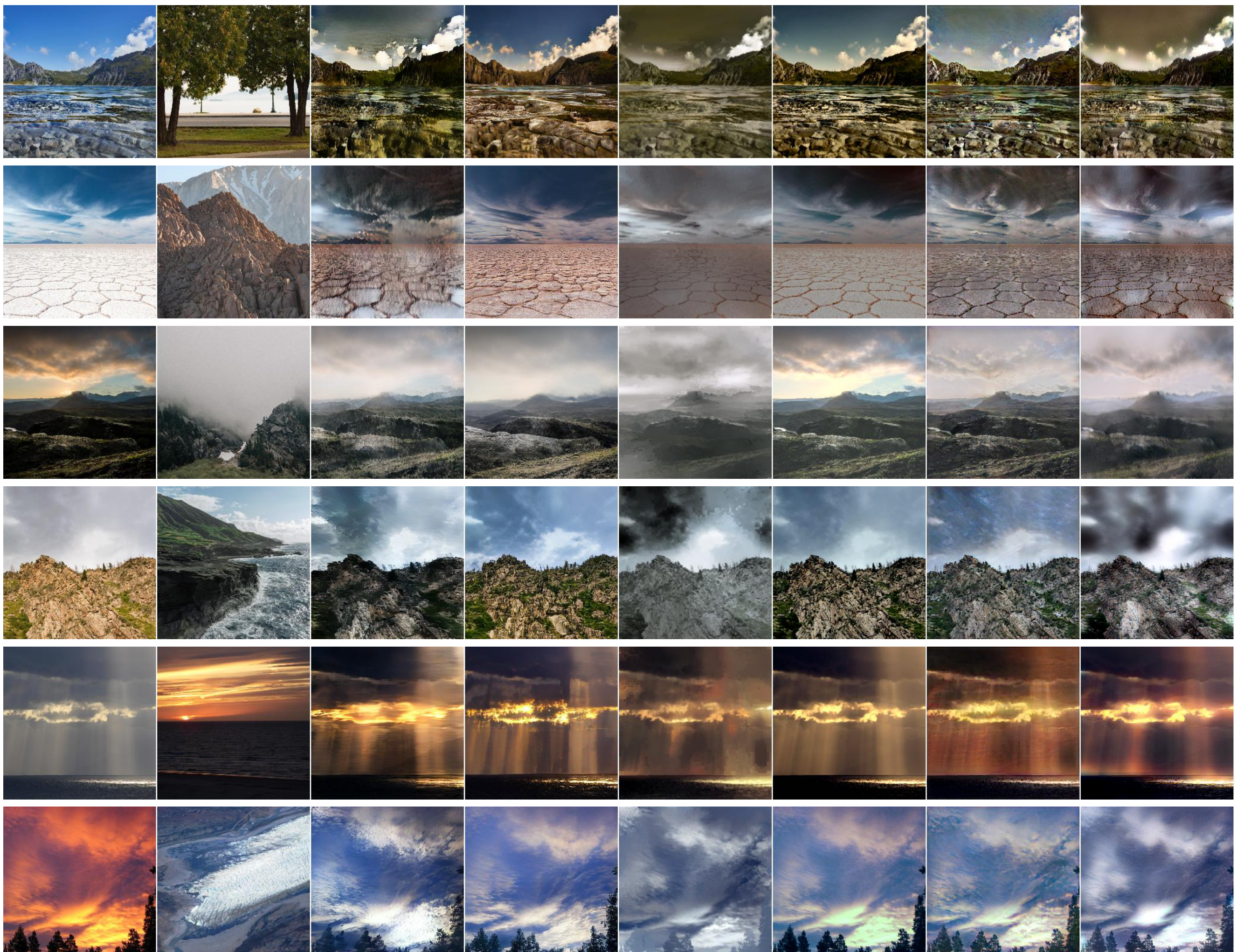


Figure B. More **artwork-to-artwork** style transfer results by the state-of-the-art methods.



Input | Reference | Ours ($\alpha = 0$) | Ours ($\alpha = 1$) | PhotoWCT [9] | WCT^2 [15] | LinearWCT [7] | DSTN [4]

Figure C. More **photo-to-photo** style transfer results by the state-of-the-art methods.

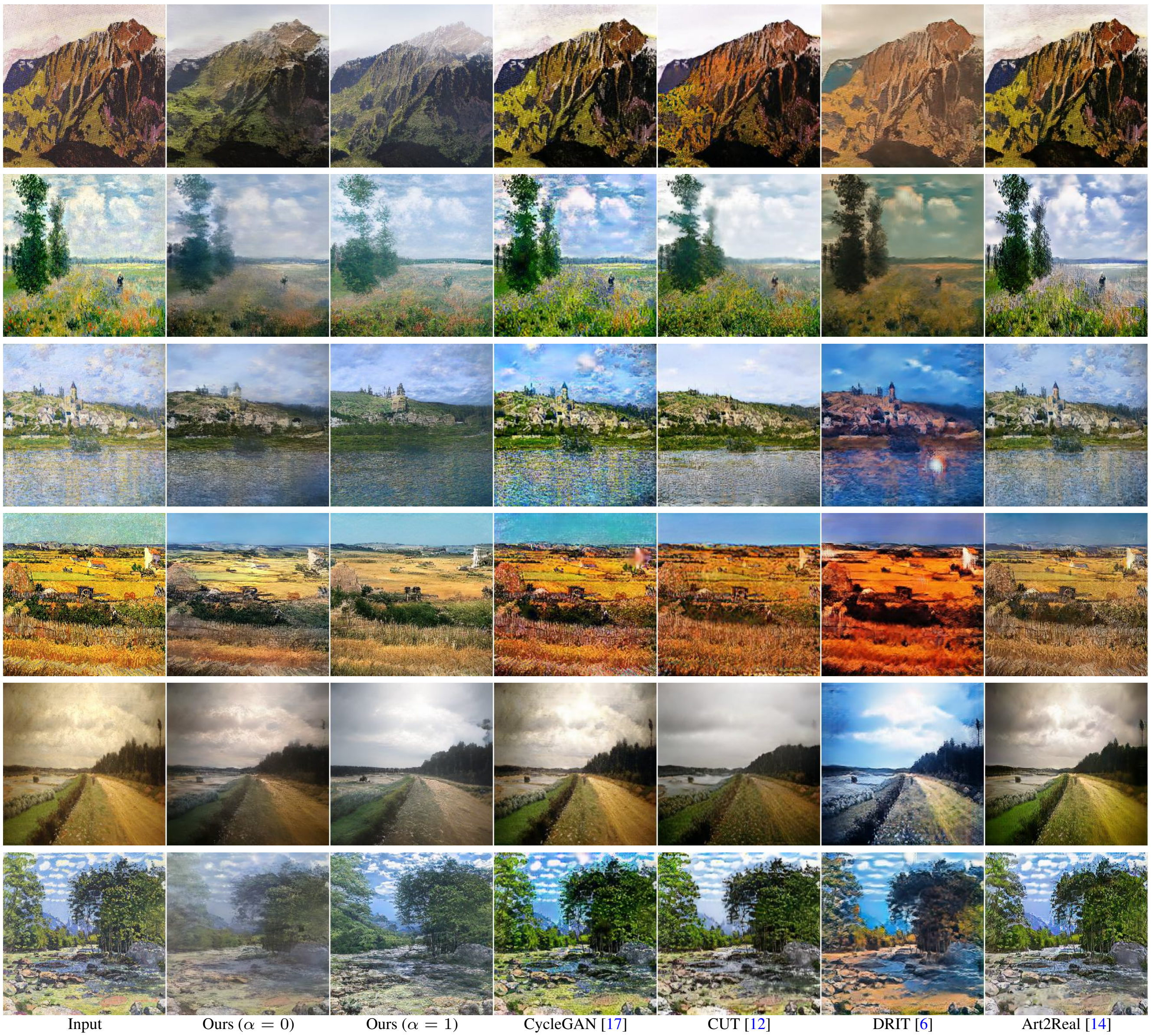


Figure D. More **artwork-to-photo** style transfer results by the state-of-the-art methods.

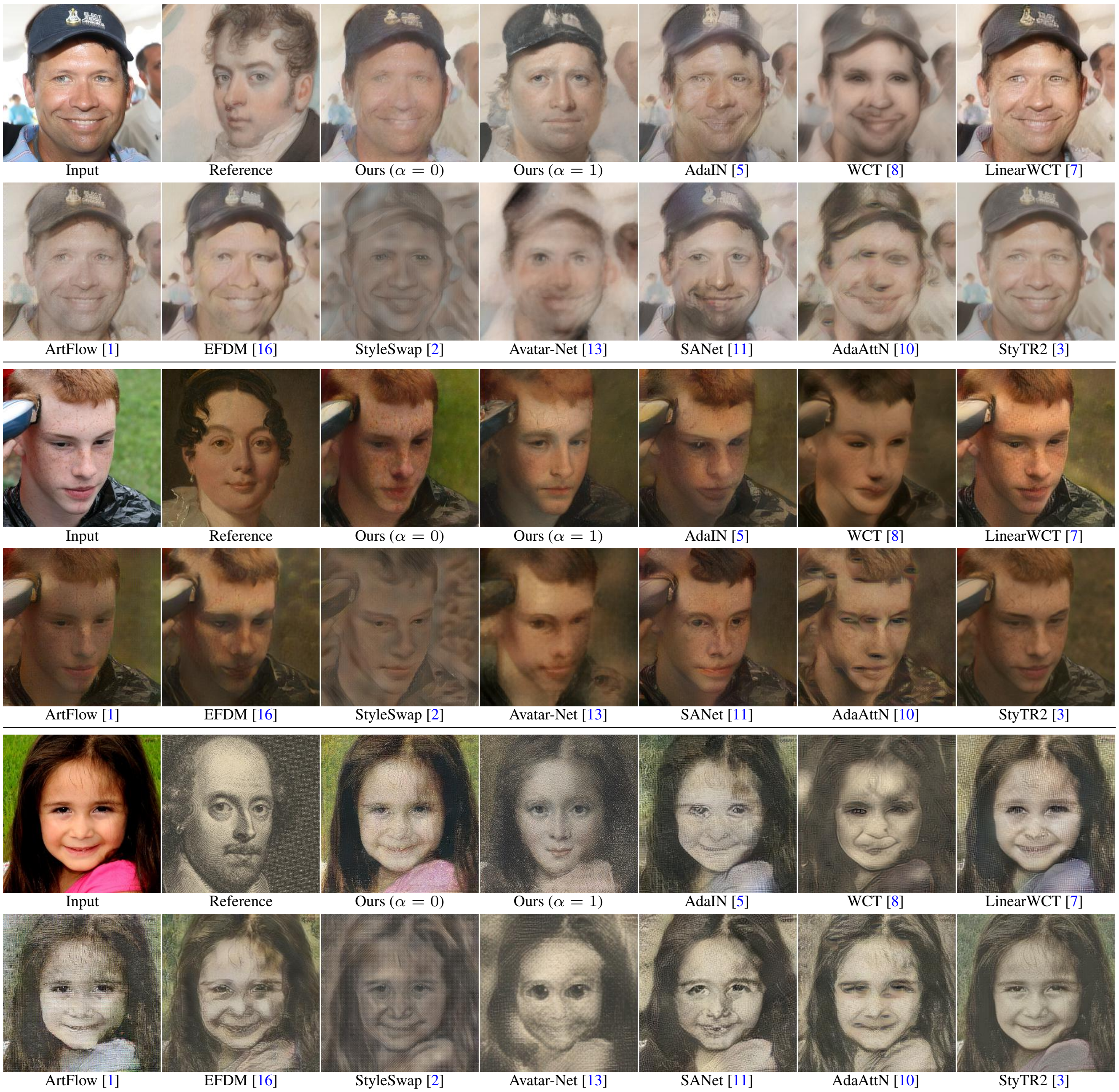


Figure E. More **face-to-artwork** style transfer results by the state-of-the-art methods.

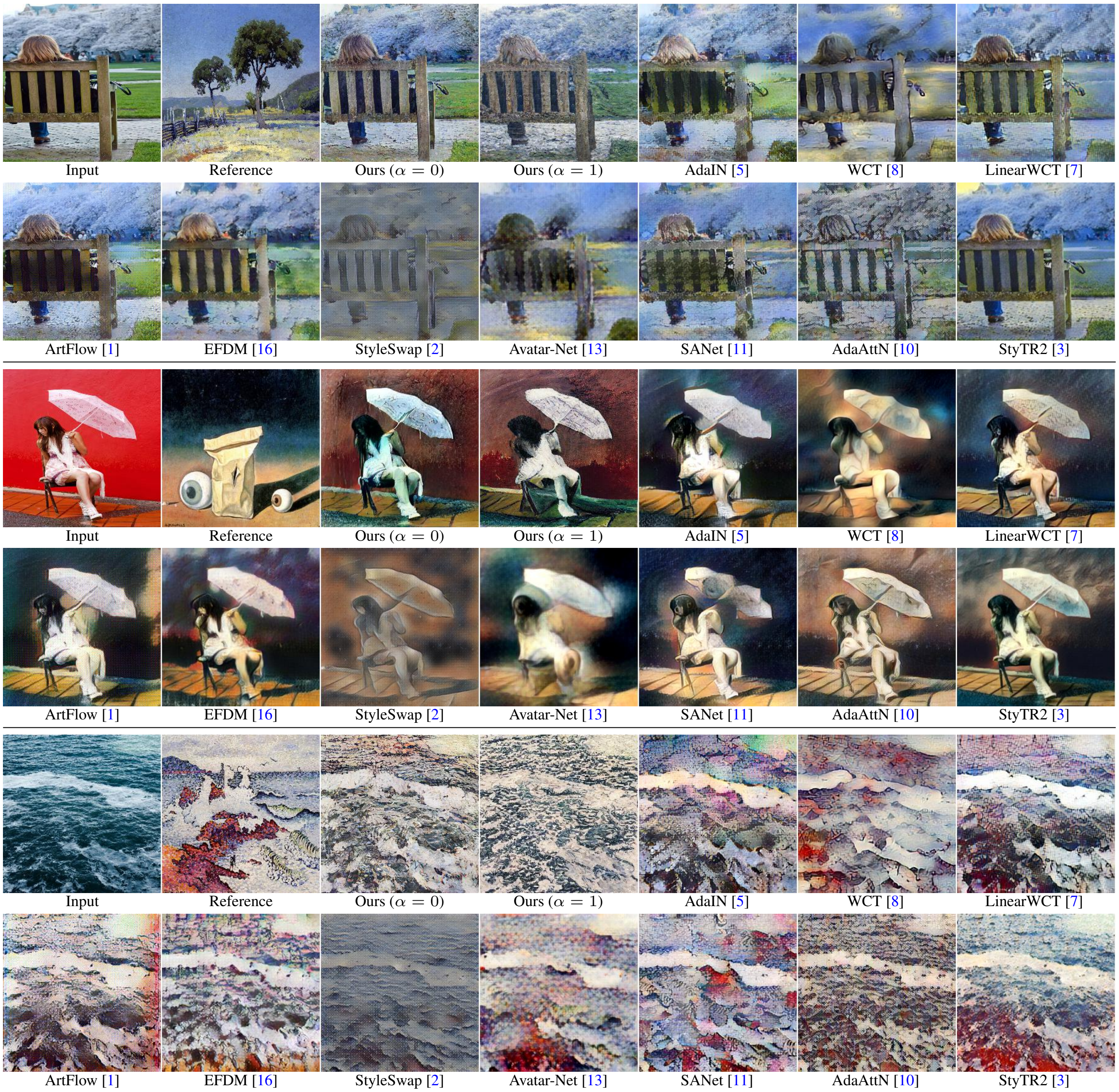


Figure F. More **photo-to-artwork** style transfer results by the state-of-the-art methods.

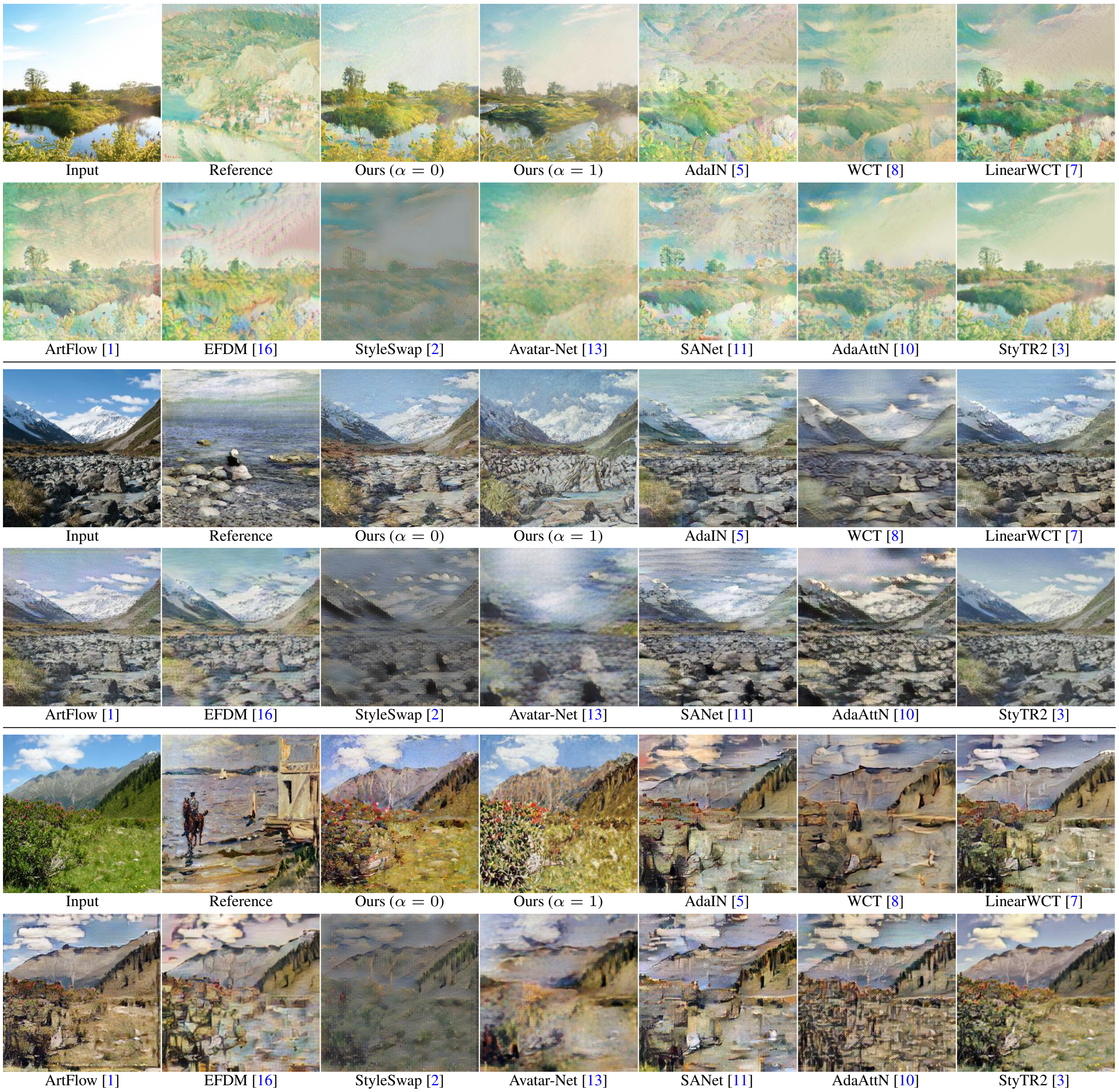


Figure G. More **photo-to-artwork** style transfer results by the state-of-the-art methods.

Figure H. More **photo-to-artwork** style transfer results by the state-of-the-art methods.

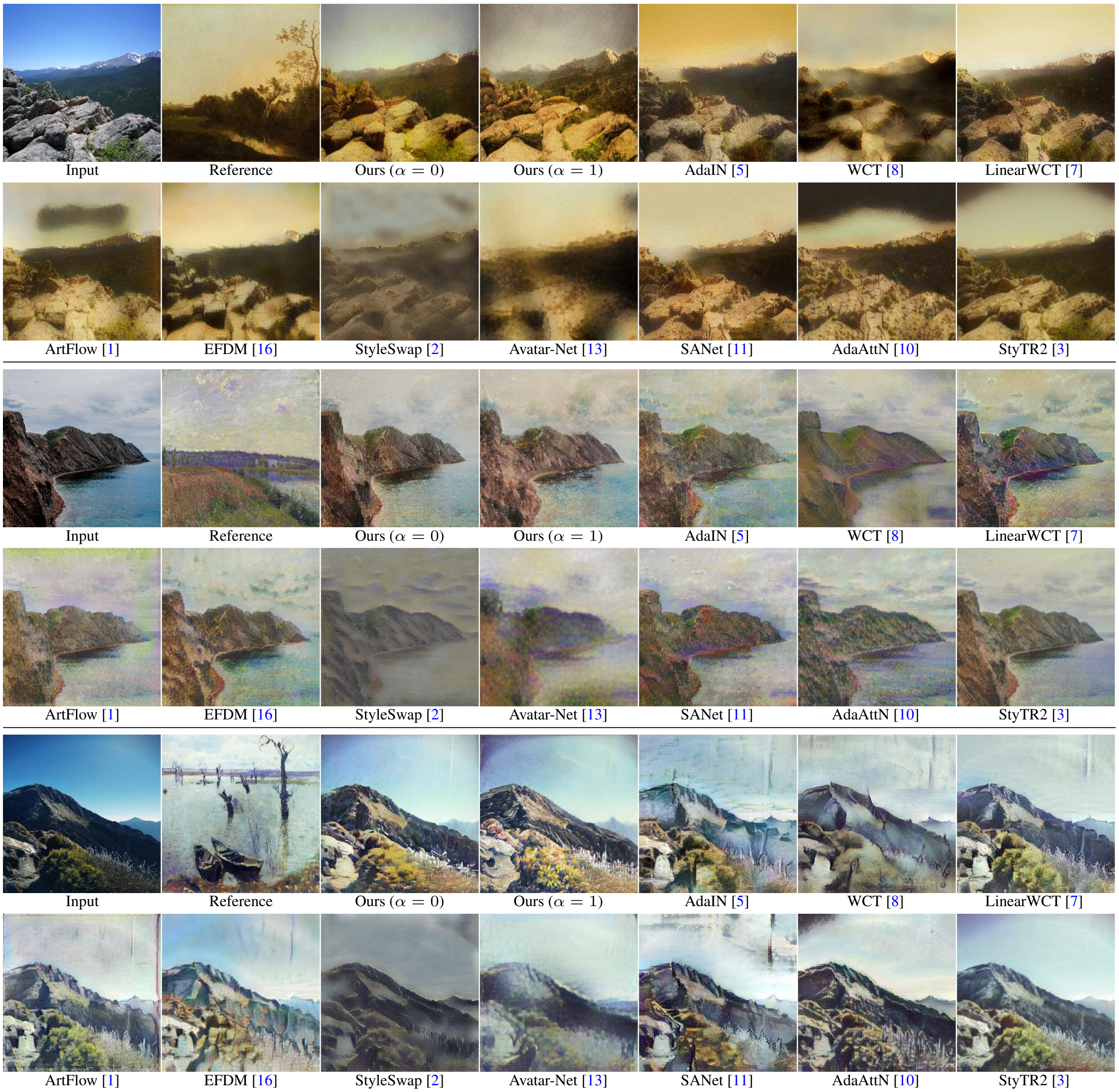


Figure I. More **photo-to-artwork** style transfer results by the state-of-the-art methods.

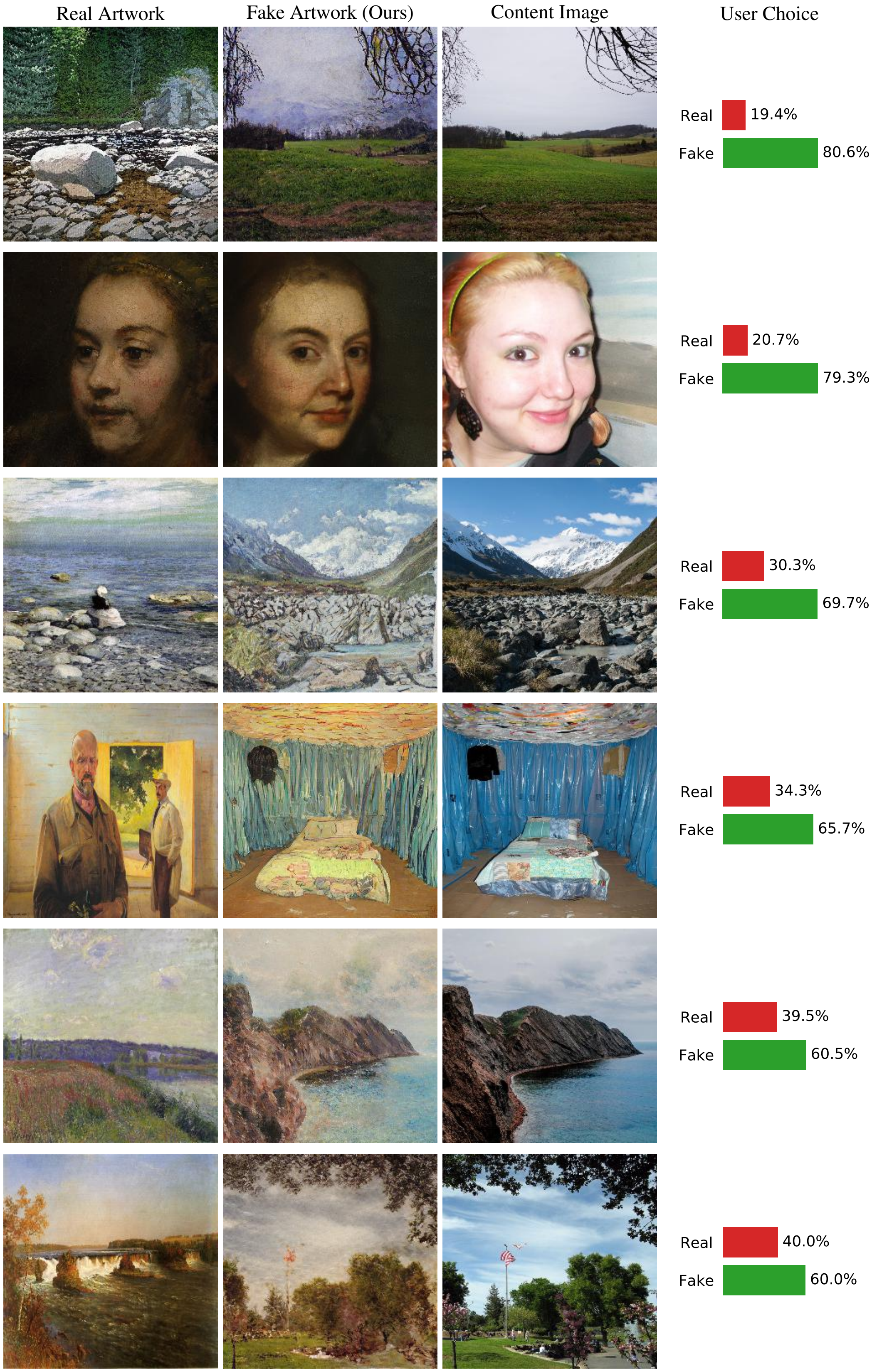


Figure J. Examples of artistic confusion test (1/6). The figures are sorted in an ascending order of the ratios of users choosing the real artwork.

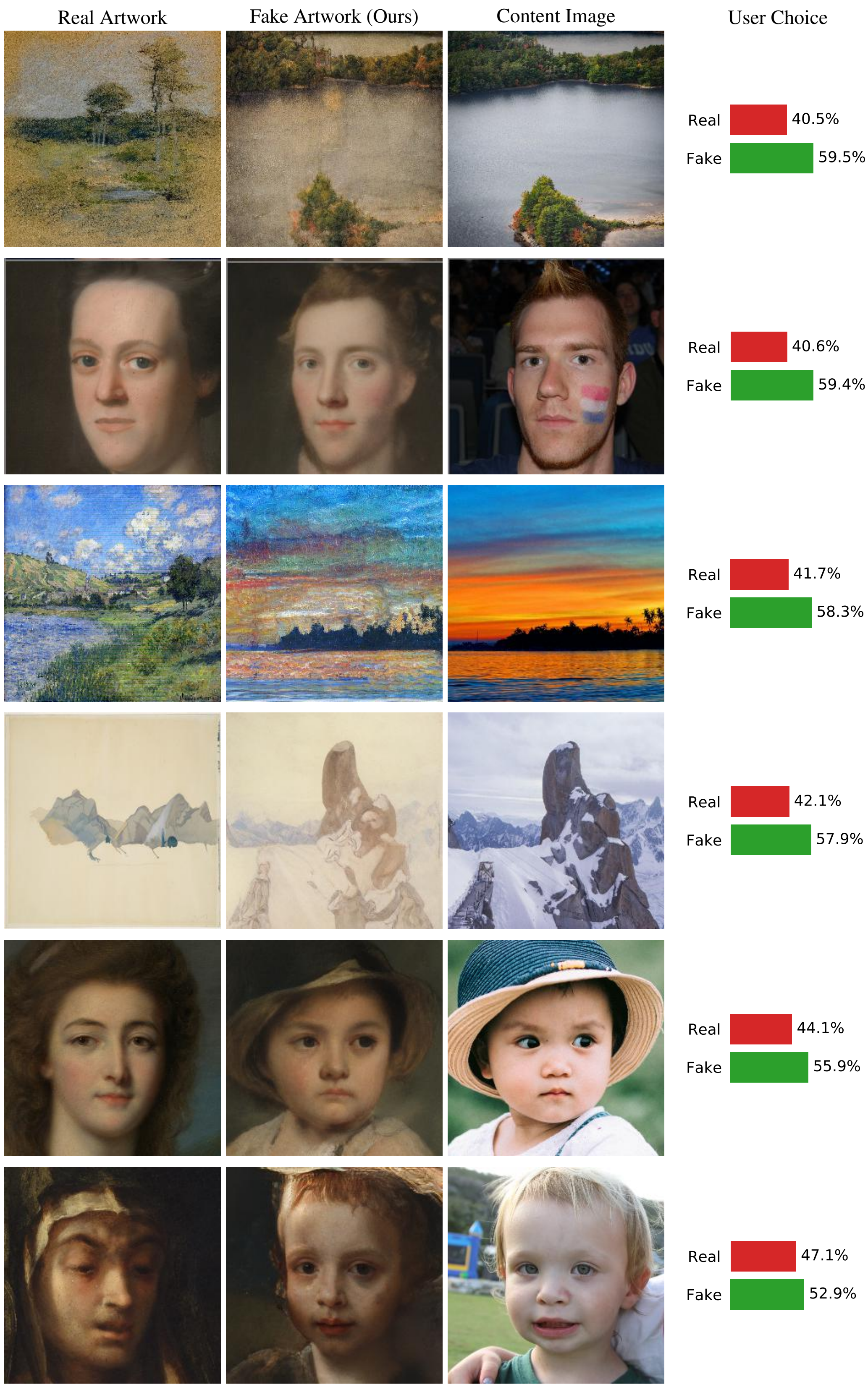


Figure K. Examples of artistic confusion test (2/6). The figures are sorted in an ascending order of the ratios of users choosing the real artwork.

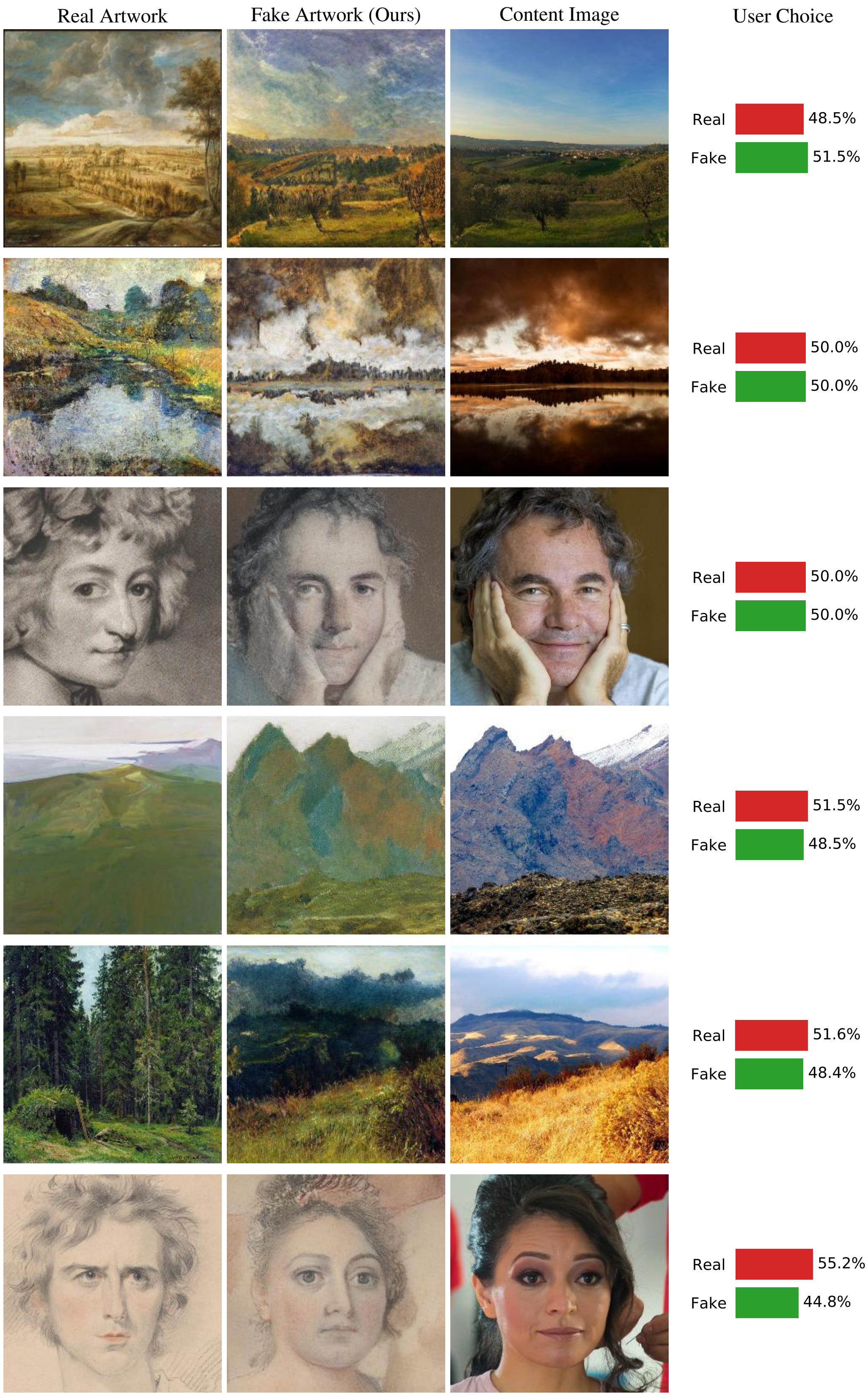


Figure L. Examples of artistic confusion test (3/6). The figures are sorted in an ascending order of the ratios of users choosing the real artwork.

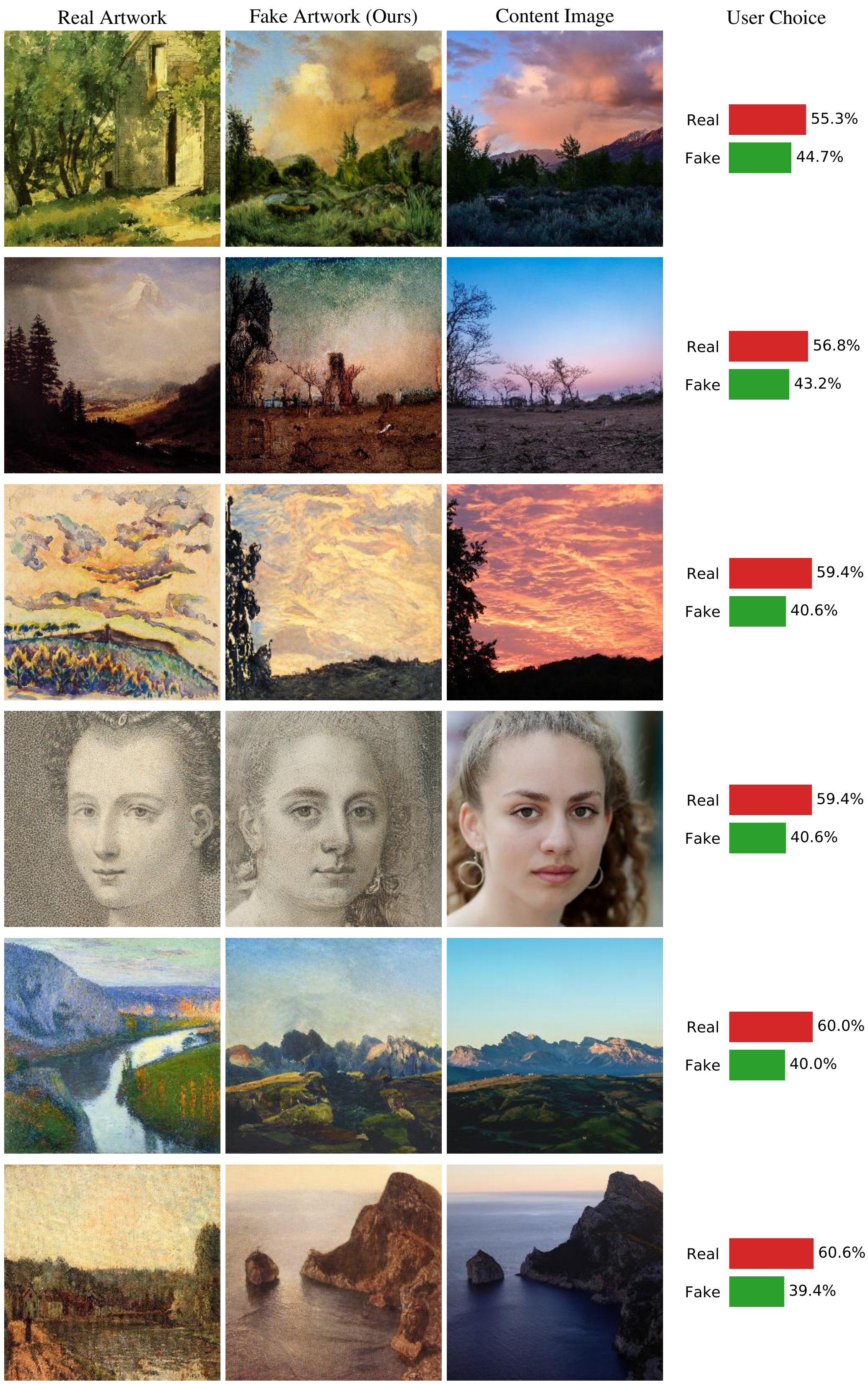


Figure M. Examples of artistic confusion test (4/6). The figures are sorted in an ascending order of the ratios of users choosing the real artwork.

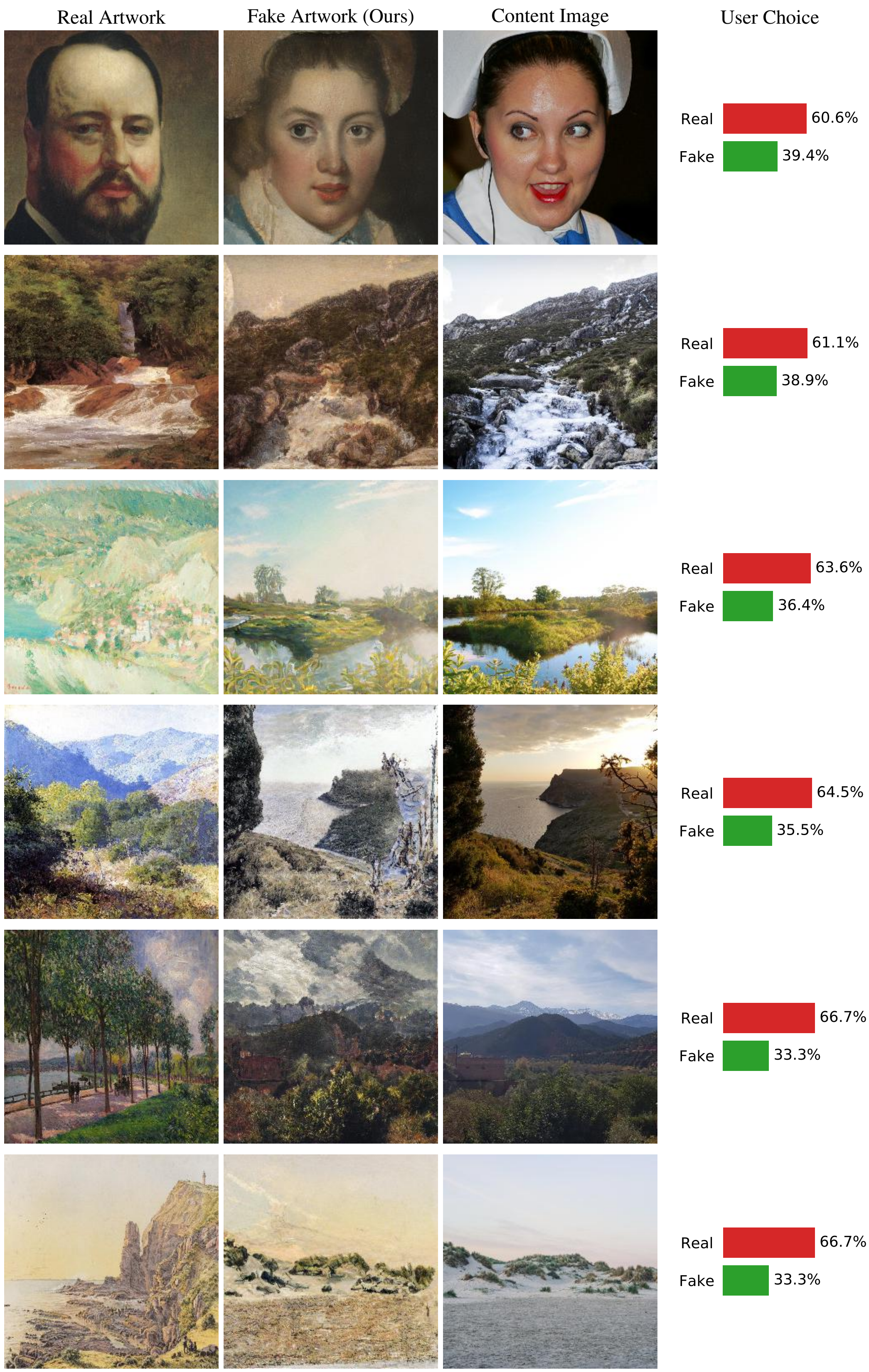


Figure N. Examples of artistic confusion test (5/6). The figures are sorted in an ascending order of the ratios of users choosing the real artwork.

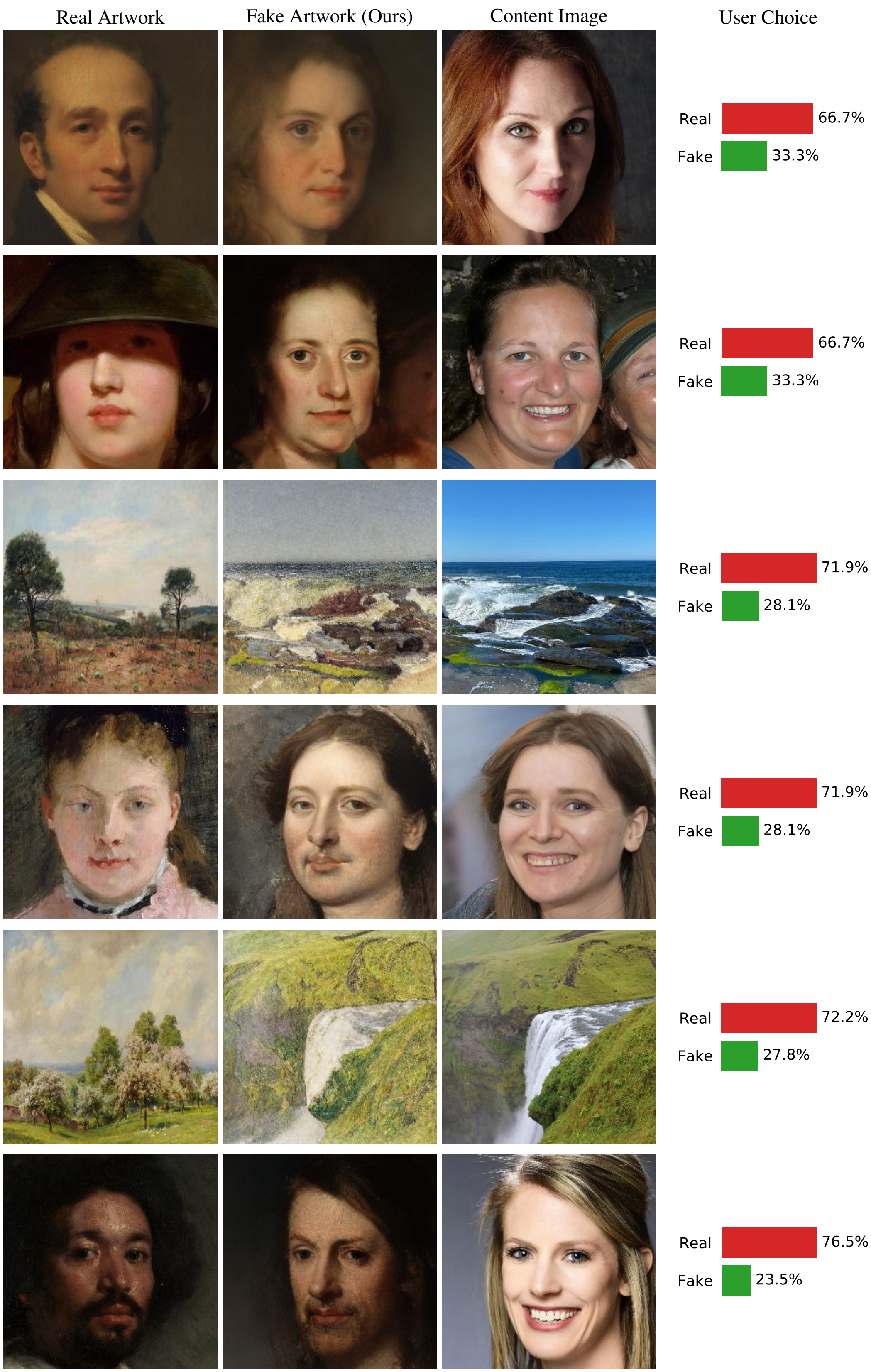


Figure O. Examples of artistic confusion test (6/6). The figures are sorted in an ascending order of the ratios of users choosing the real artwork.