

***Art2Real*: Unfolding the Reality of Artworks via Semantically-Aware Image-to-Image Translation**

Supplementary Material

Matteo Tomei Marcella Cornia Lorenzo Baraldi Rita Cucchiara
University of Modena and Reggio Emilia
{name.surname}@unimore.it

In this document, we present additional material about our method. In particular, we provide a visualization of the segmentation maps, an analysis of the importance of multi-scale, and additional quantitative and qualitative results.

1. Segmentation maps

In Fig. 1, we show some qualitative examples of segmentation masks extracted on paintings and generated images, through the model from Hu *et al.* [3]. Each color represents a specific class label. Only the most relevant masks are shown for each image. It can be observed that the segmentation strategy extracts meaningful semantic regions from the input images, thus enforcing the retrieval of semantically correct patches on large portions of the image. Overall, we found that the use of semantic segmentation can greatly improve results. While some image regions might not be labelled, a sufficiently realistic appearance can still be recovered from the background memory bank.

2. Multi-scale importance

In order to evaluate the contribution of the multi-scale approach to the realism of the generation, we run a set of experiments without the multi-scale variant. We use a single scale, *i.e.* a patch size of 16 and a stride of 6, and train our model on Monet, landscape and portrait settings. The full objective, in this case, is that presented in Eq. 5 of the main paper. We then compare the FID [2] values obtained through our approach with and without the multi-scale variant. Results are presented in Table 1. As it can be seen, the multi-scale strategy effectively increases the realism of the generation, outperforming by a clear margin the single-scale baseline on almost all settings.

3. Additional experimental results

Here we present additional quantitative results, computing the FID [2] with different layers of Inception-v3 [7]

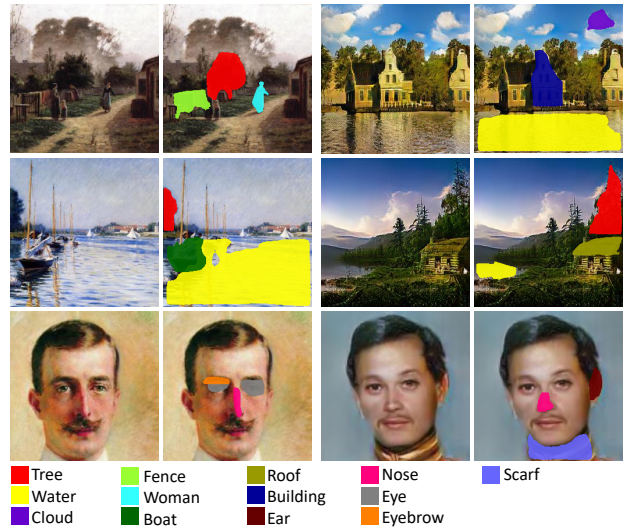


Figure 1: Segmentation masks visualization. The first two columns show original paintings, while the third and the fourth columns show generated images. Only some of the extracted masks are visible.

Method	Monet	Landscapes	Portraits	Mean
Single scale	46.28	35.88	34.74	38.97
Art2Real (multi-scale)	44.71	35.03	34.03	37.92

Table 1: Multi-scale importance analysis in terms of Fréchet Inception Distance [2].

and showing the distribution of ResNet-152 [1] features extracted from all the available settings.

Fréchet Inception Distance. In the main paper, we showed how our model is able to generate images which lower the FID [2] with respect to real images, fitting the Gaussians on the final average pooling layer features of Inception-v3 [7]

Method	Monet	Cezanne	Van Gogh	Ukiyo-e	Landscapes	Portraits	Mean
768 dimensions							
Original paintings	0.45	0.94	1.03	1.34	0.37	0.42	0.76
Style-transferred reals	0.58	0.94	1.12	1.23	0.56	0.36	0.80
DRIT [4]	0.41	0.54	0.56	0.60	0.37	0.28	0.46
UNIT [5]	0.30	0.43	0.44	0.35	0.25	0.25	0.34
Cycle-GAN [8]	0.29	0.37	0.36	0.43	0.24	0.16	0.31
Art2Real	0.21	0.30	0.35	0.31	0.17	0.19	0.26
192 dimensions							
Original paintings	0.95	1.67	3.96	1.86	0.49	0.22	1.53
Style-transferred reals	0.97	1.76	4.09	2.44	0.55	0.21	1.67
DRIT [4]	0.30	0.33	0.40	0.38	0.49	0.11	0.34
UNIT [5]	0.26	0.26	0.37	0.16	0.21	0.07	0.22
Cycle-GAN [8]	0.26	0.31	0.18	0.19	0.55	0.03	0.25
Art2Real	0.10	0.13	0.12	0.17	0.19	0.05	0.13

Table 2: Evaluation in terms of Fréchet Inception Distance [2].

(2048-d). In Table 2, we also show FID values obtained fitting the two Gaussians on the pre-auxiliary classifier layer features (768-d) and on the second max-pooling layer features (192-d). The FID value is computed for our model and for a number of competitors and again our model produces a lower FID on almost all the settings. Note that FID values computed at different layers have different magnitude and are not directly comparable.

Feature distributions visualization. Fig. 2 shows the feature distribution visualizations of our method and competitors computed on Monet, Cezanne, Van Gogh and Ukiyo-e images. As previously mentioned, for each considered setting, we extract image features from the average pooling layer of a ResNet-152 [1] and we use the t-SNE algorithm [6] to project them into a 2-dimensional space. Each plot reports the distributions of visual features extracted from real photos, original paintings and the corresponding translations generated by our model or by one of the competitors. Also for these settings, the distributions of visual features extracted from our generated images are very close to the distributions of real photos, thus further confirming a greater reduction of domain shift compared to that of competitors.

4. Additional qualitative results

In Fig. 3, we show some relevant failure cases generated by the model. Failures can be due to false positives in the segmentation (upper left: part of the tree is wrongly labelled as sky), excessive lack of realism of the source painting (upper right and bottom right), lack of real patches of a given semantic class (bottom left: the realistic set for portraits does not include hand patches).

Several other qualitative results are shown in the rest of the supplementary. Firstly, we report sample images generated by our model taking as input sample paintings depicting landscapes and portraits. Secondly, we show additional qualitative comparisons with respect to Cycle-GAN [8], UNIT [5], and DRIT [4] on all considered settings. Overall, the results demonstrate that our model is able to generate more realistic images, creating fewer artifacts and better preserving the original contents, facial expressions, and colors of original paintings.

References

- [1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.
- [2] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, Günter Klambauer, and Sepp Hochreiter. GANs trained by a two time-scale update rule converge to a Nash equilibrium. *Advances in Neural Information Processing Systems*, 2017.
- [3] Ronghang Hu, Piotr Dollár, Kaiming He, Trevor Darrell, and Ross Girshick. Learning to Segment Every Thing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- [4] Hsin-Ying Lee, Hung-Yu Tseng, Jia-Bin Huang, Maneesh Kumar Singh, and Ming-Hsuan Yang. Diverse Image-to-Image Translation via Disentangled Representations. In *Proceedings of the European Conference on Computer Vision*, 2018.
- [5] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. Unsupervised image-to-image translation networks. In *Advances in Neural Information Processing Systems*, 2017.

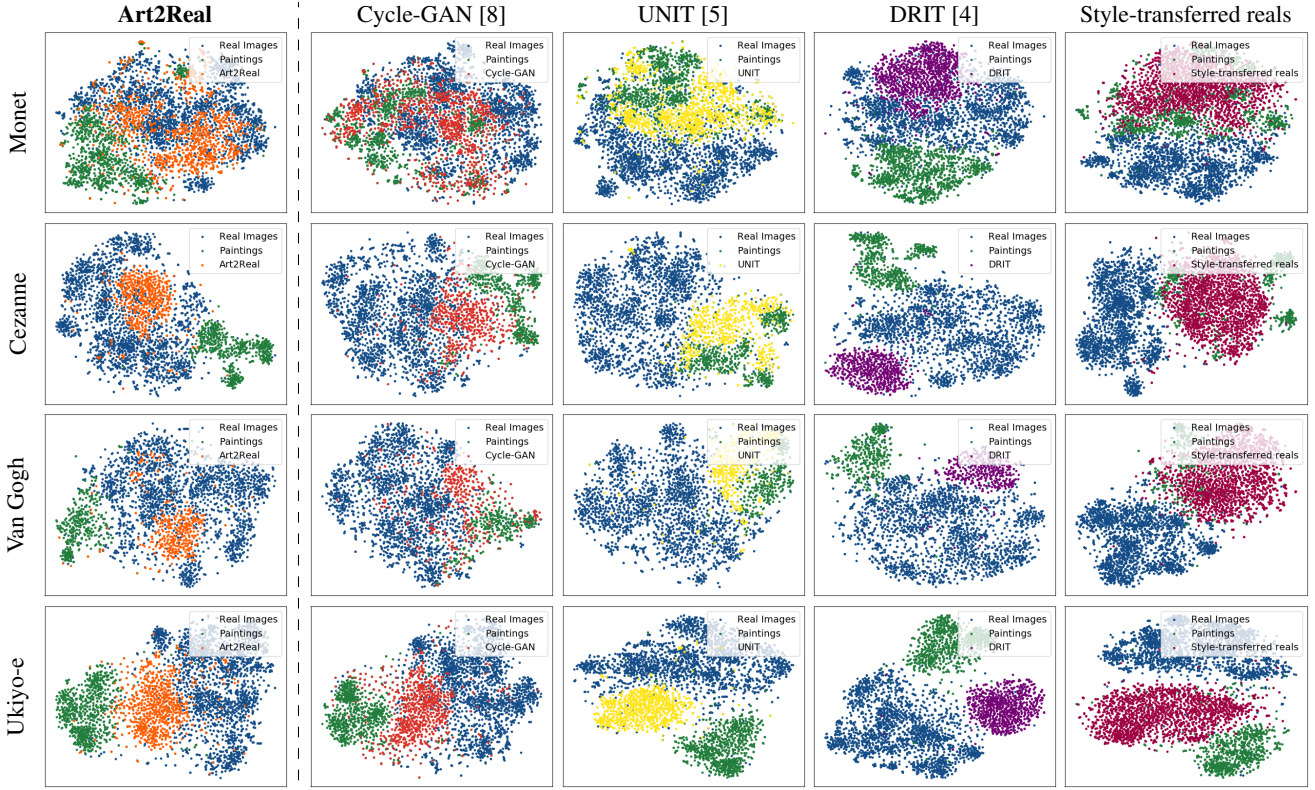


Figure 2: Distribution of ResNet-152 features extracted from Monet, Cezanne, Van Gogh and Ukiyo-e images. Each row shows the results of our method and competitors on a specific setting.



Figure 3: Sample failure cases.

- [6] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(Nov):2579–2605, 2008.
- [7] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.
- [8] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the International Conference on Computer Vision*, 2017.

Original Painting



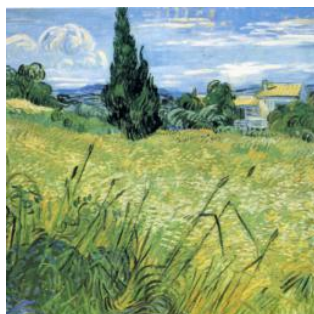
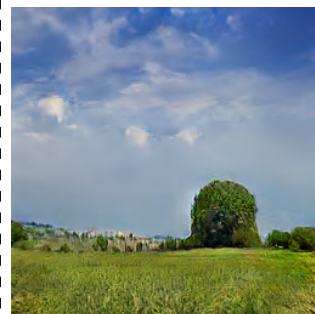
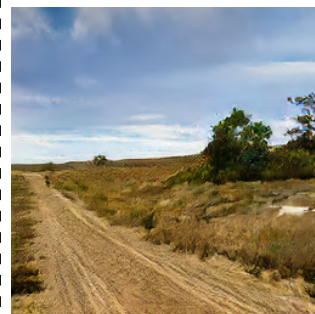
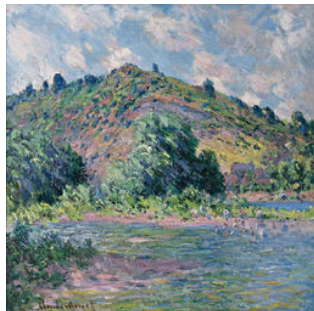
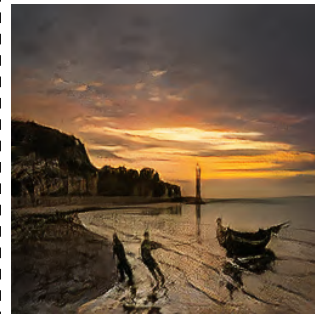
Art2Real



Original Painting



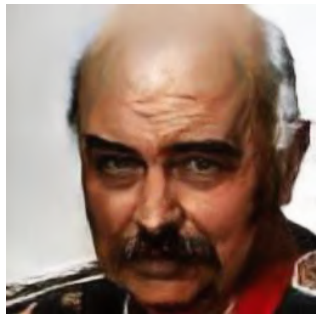
Art2Real



Original Painting



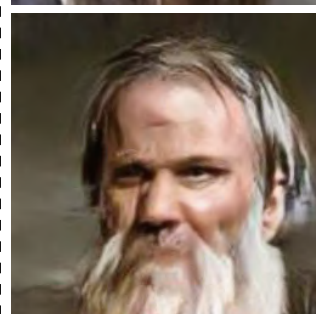
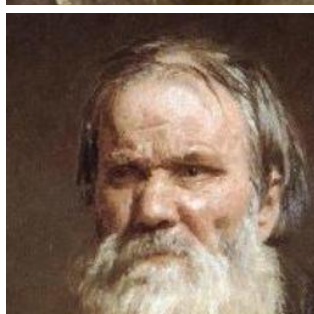
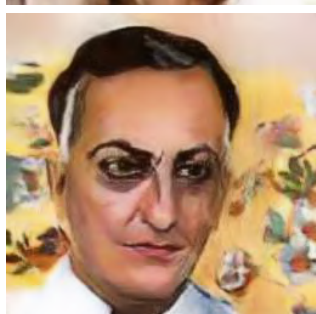
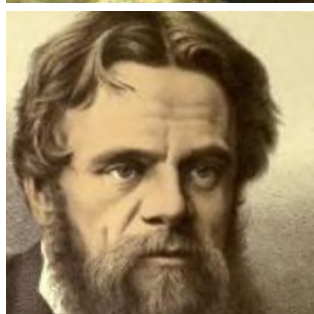
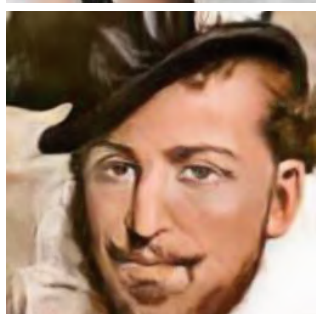
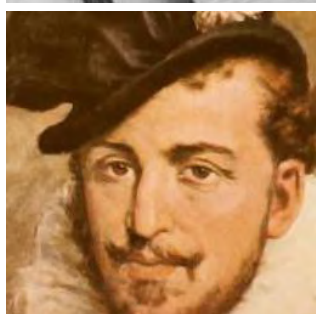
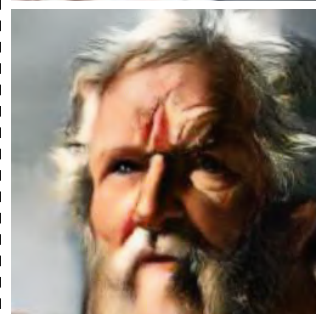
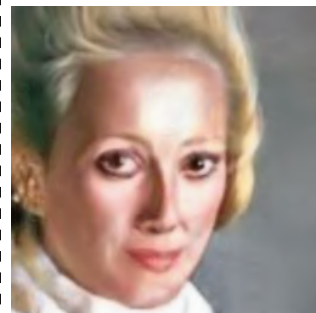
Art2Real

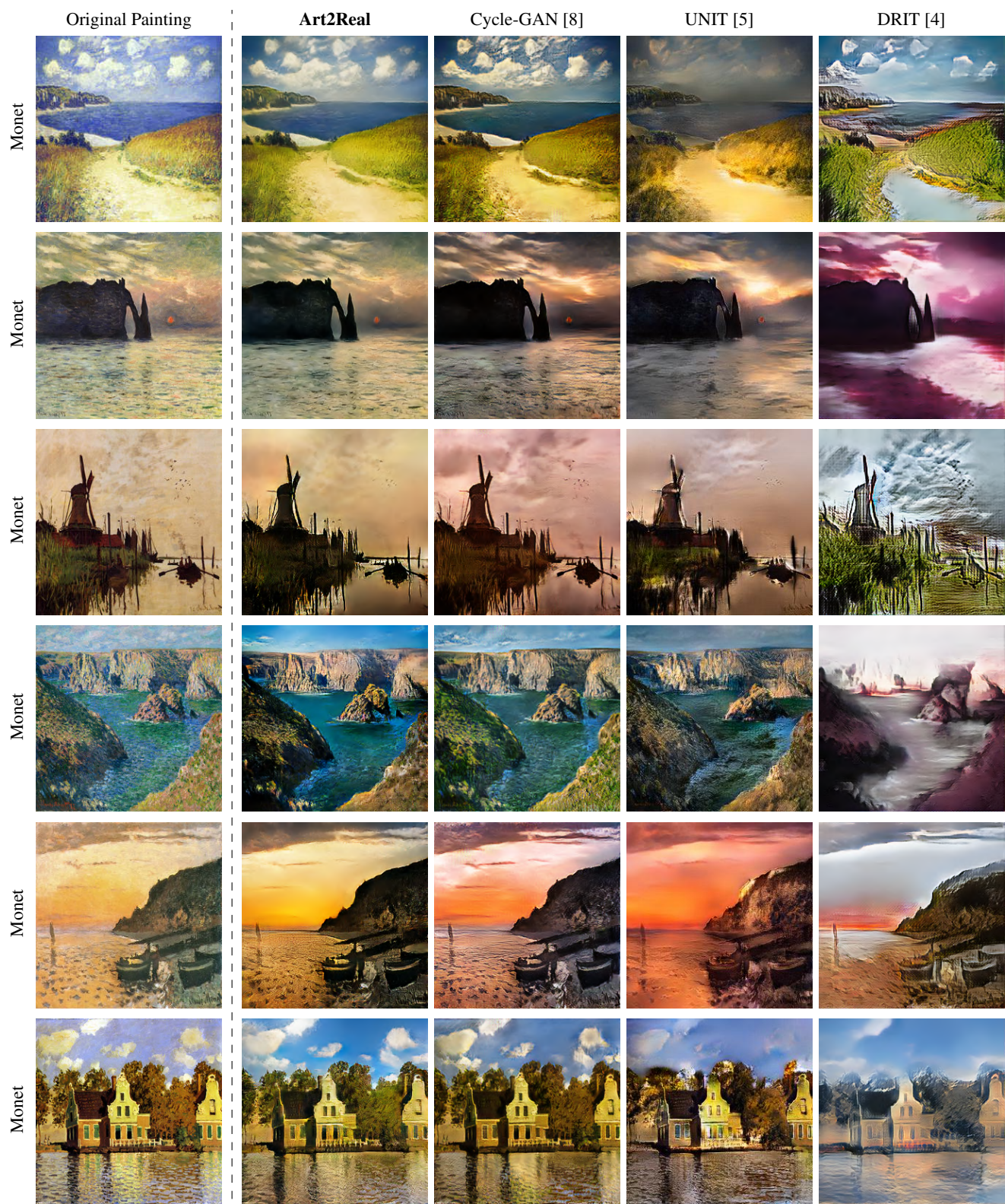


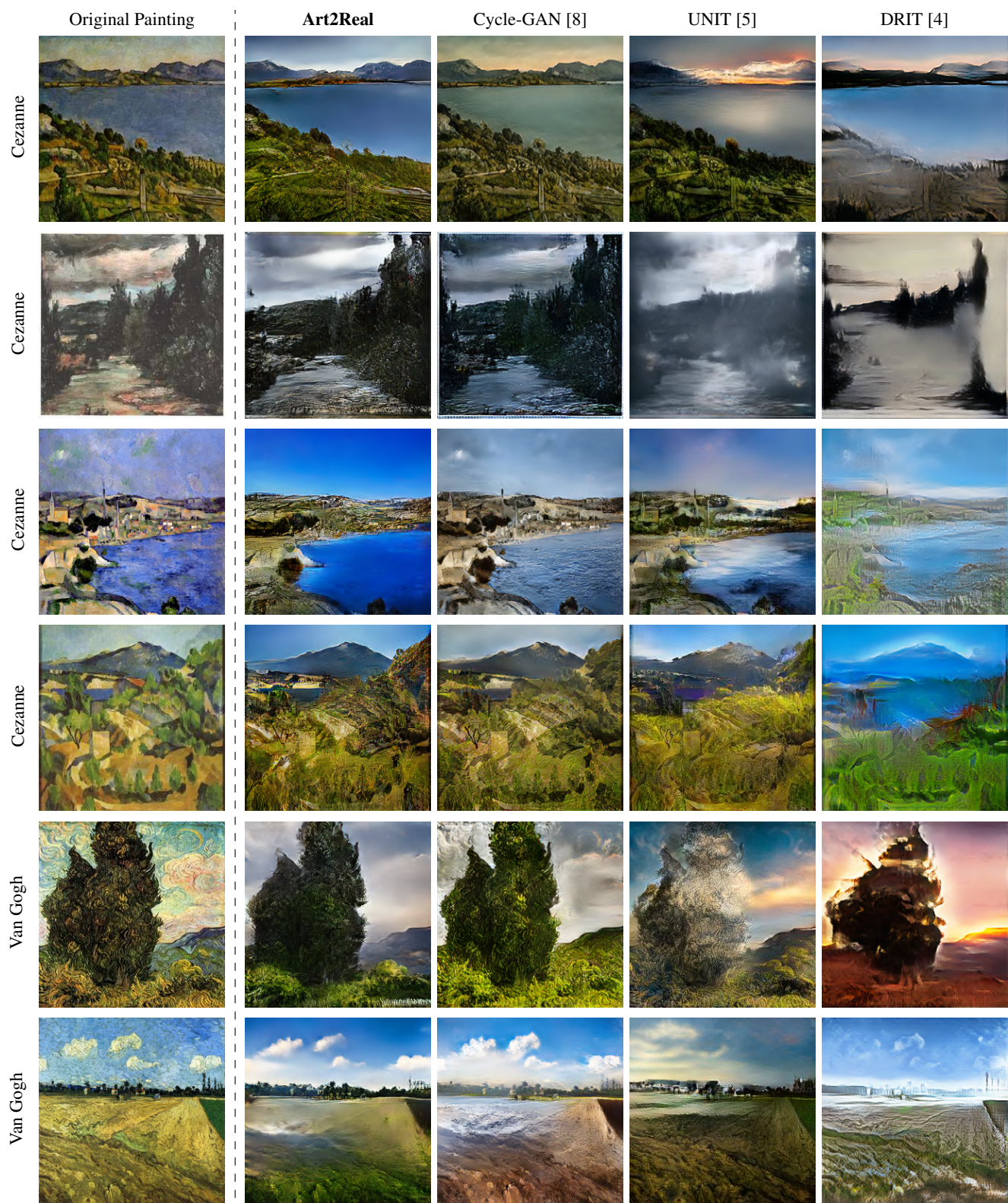
Original Painting

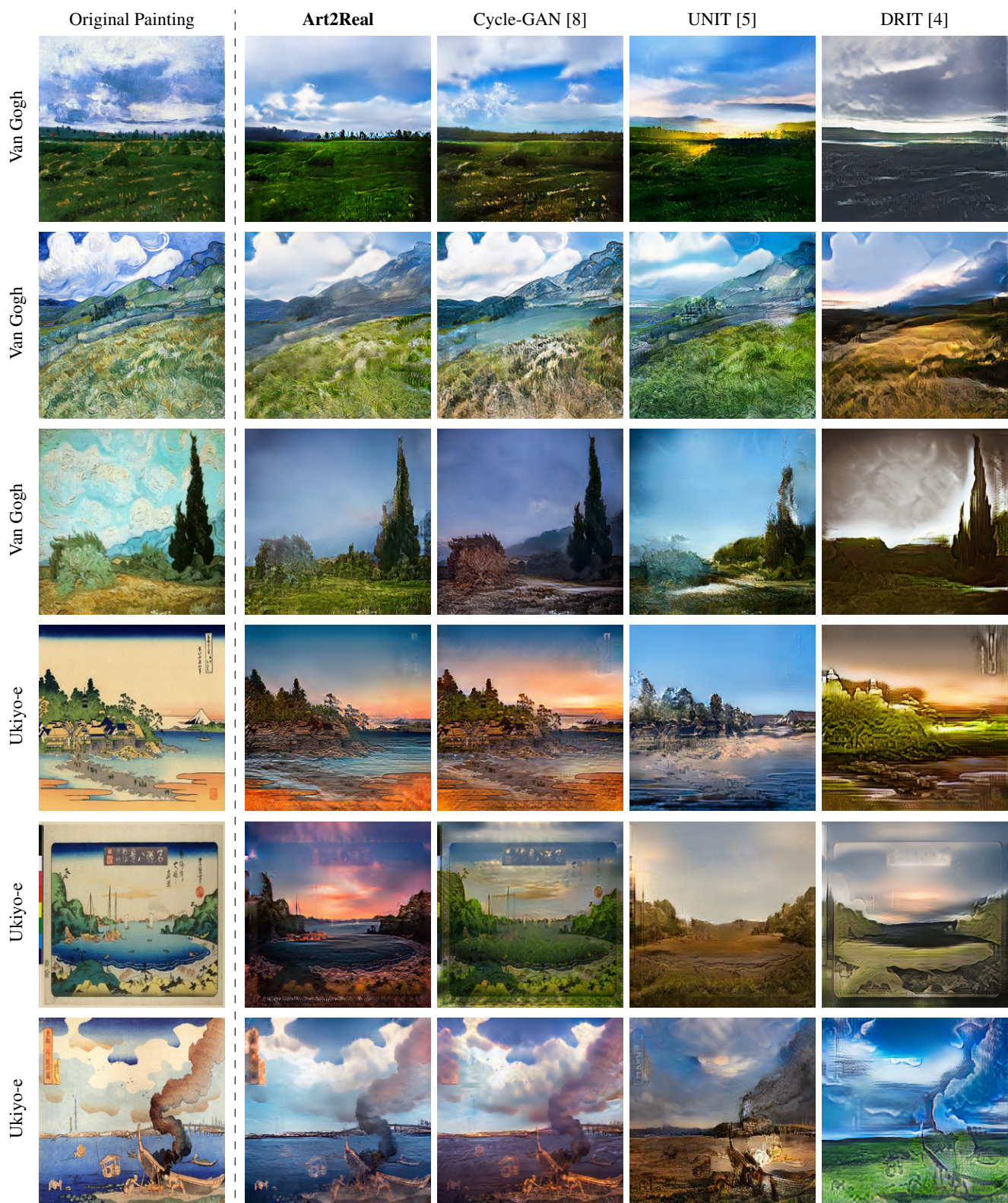


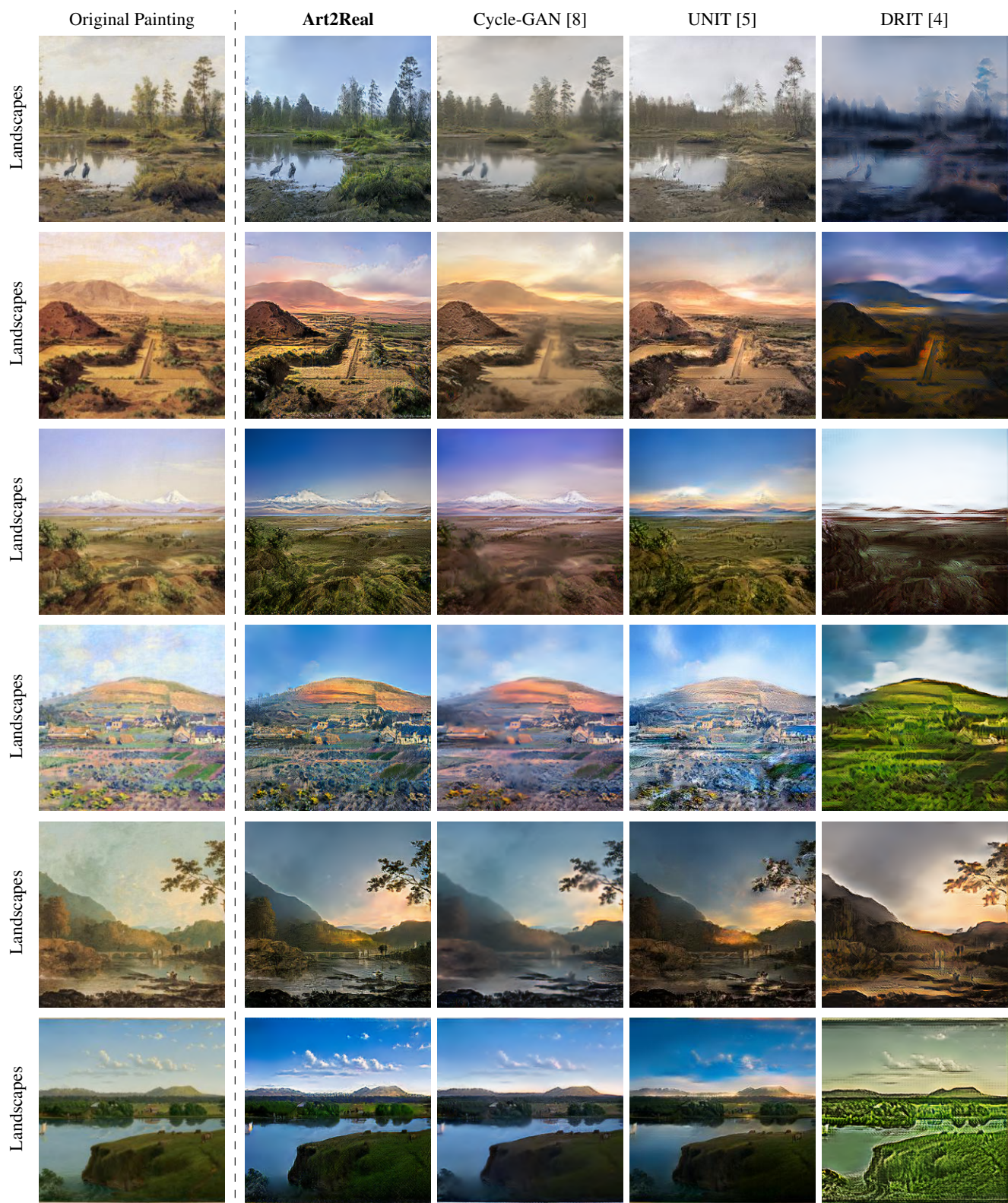
Art2Real

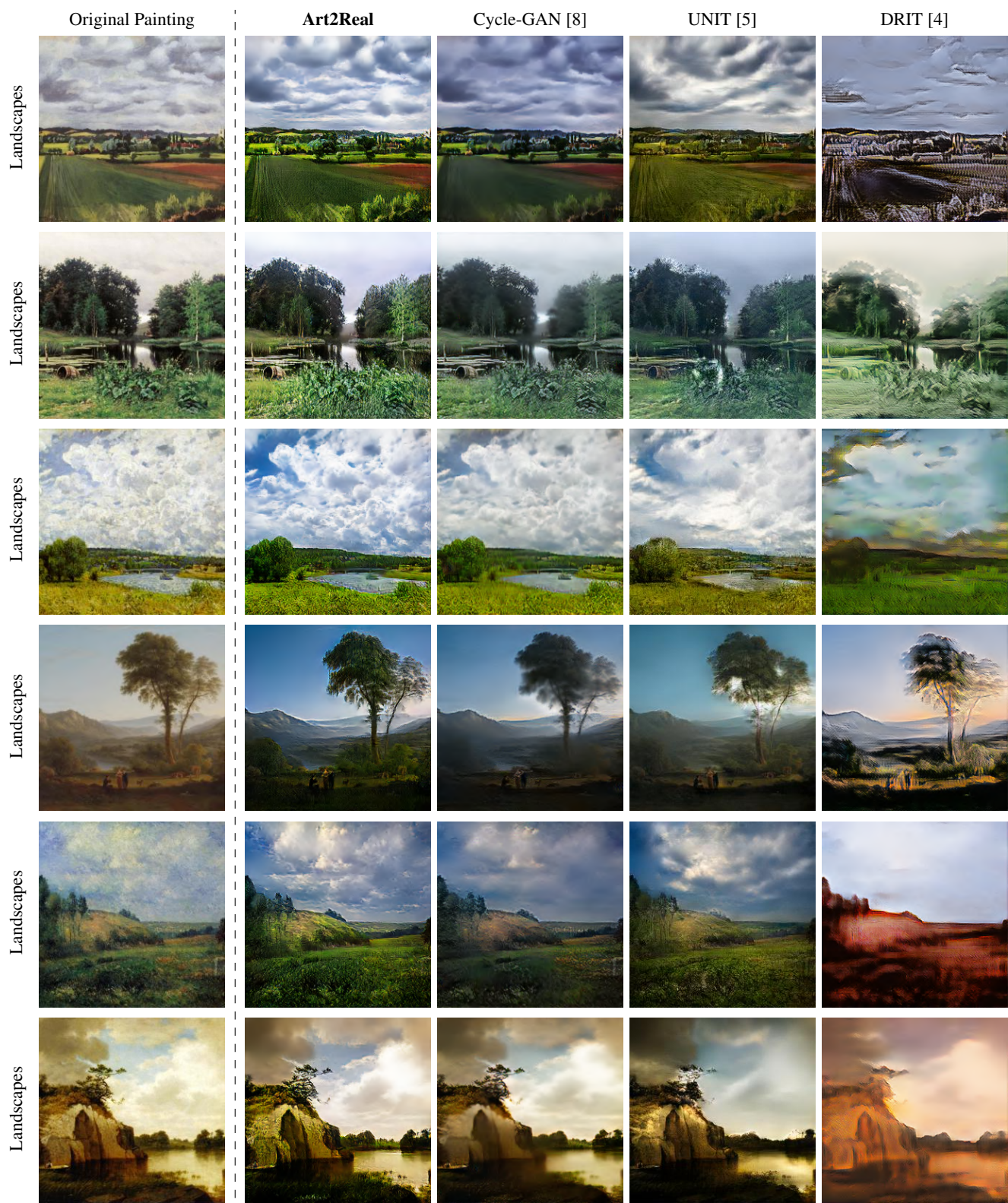


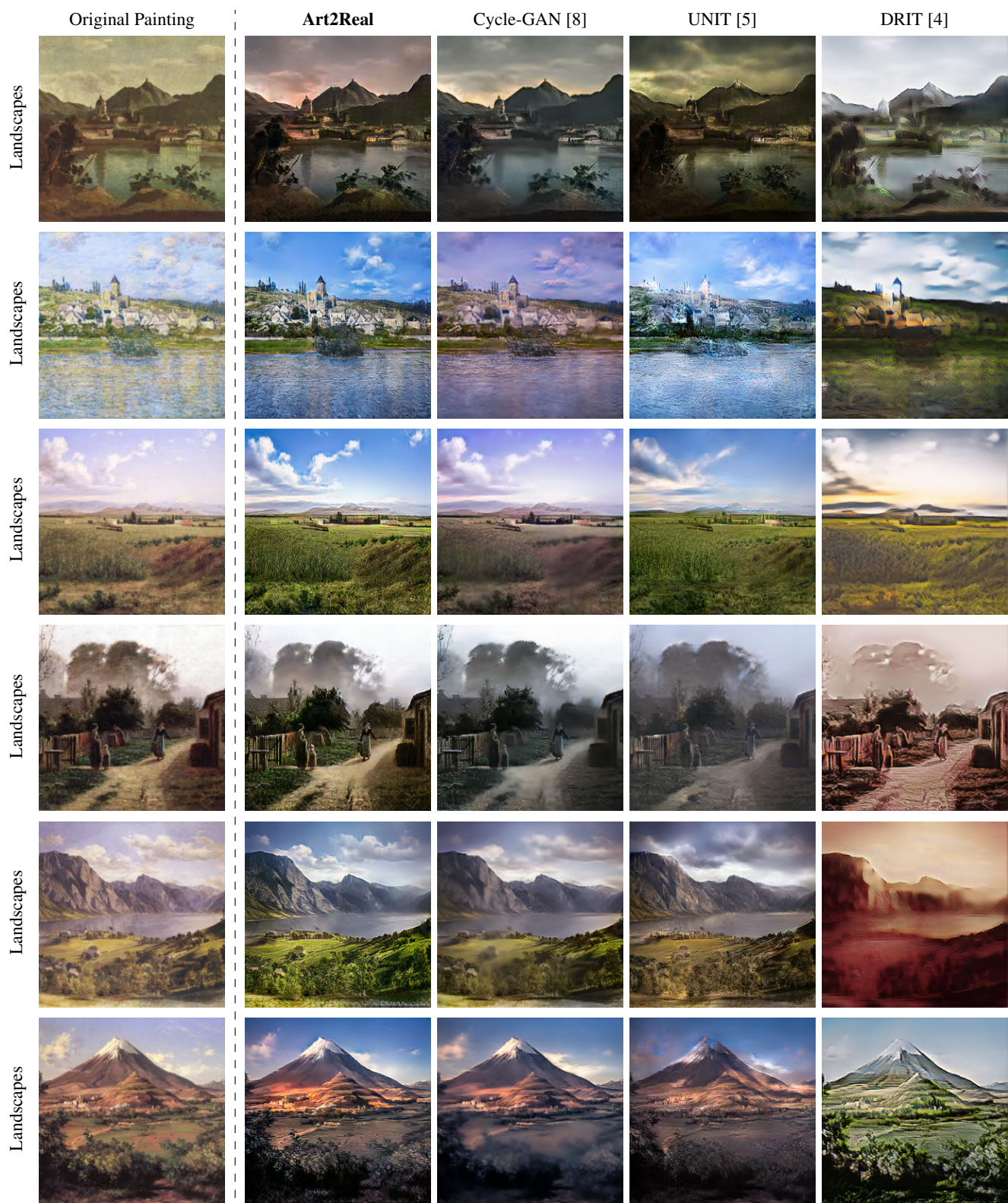




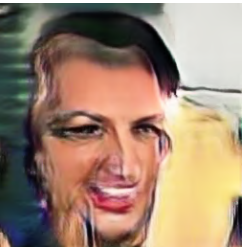

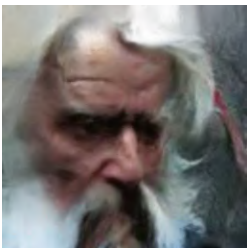






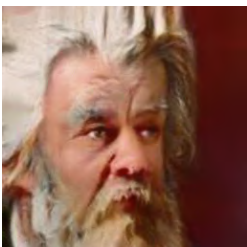
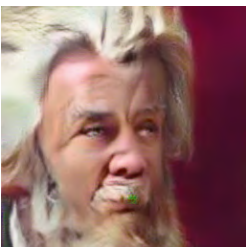

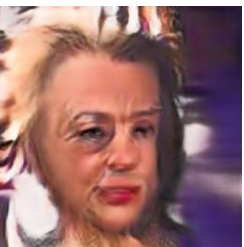

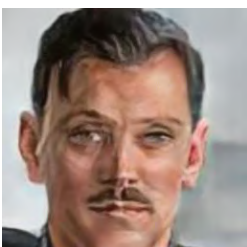
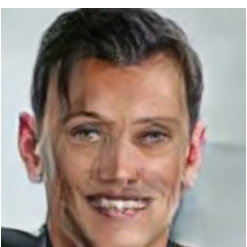

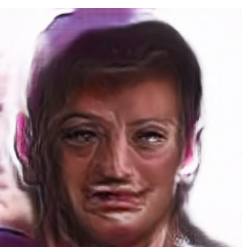


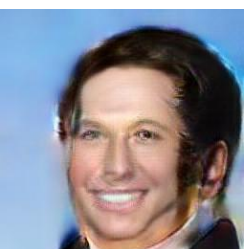
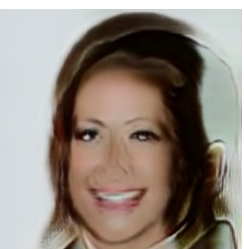










	Original Painting	Art2Real	Cycle-GAN [8]	UNIT [5]	DRIT [4]
Portraits					
Portraits					
Portraits					
Portraits					
Portraits					
Portraits					

Original Painting

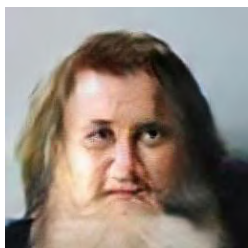
Art2Real

Cycle-GAN [8]

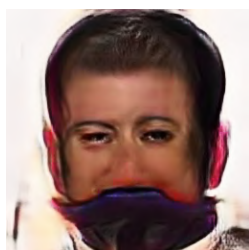
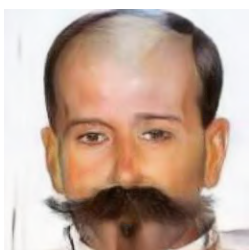
UNIT [5]

DRIT [4]

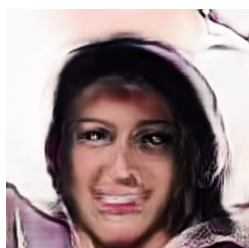
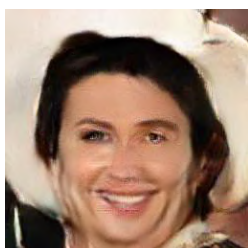
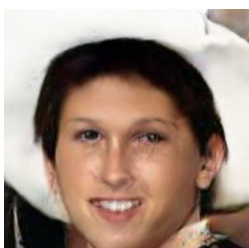
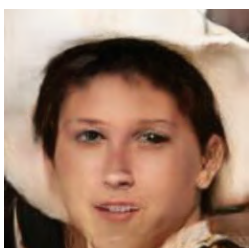
Portraits



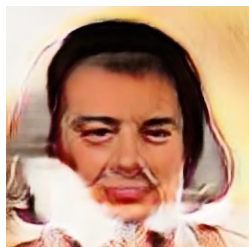
Portraits



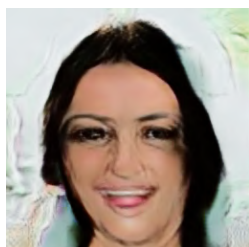
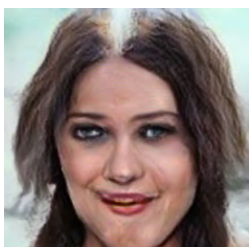
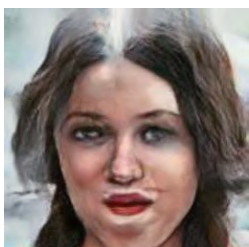
Portraits



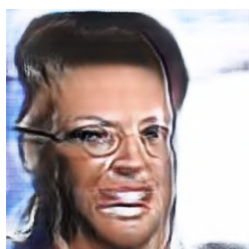
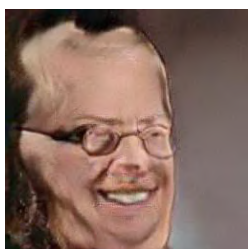
Portraits



Portraits



Portraits



Original Painting

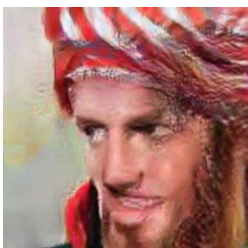
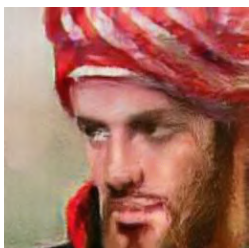
Art2Real

Cycle-GAN [8]

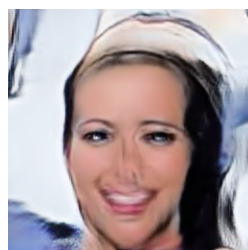
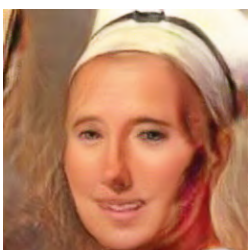
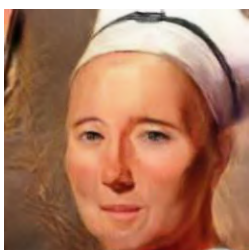
UNIT [5]

DRIT [4]

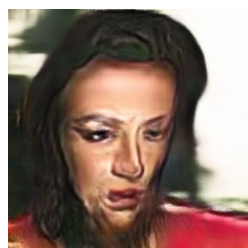
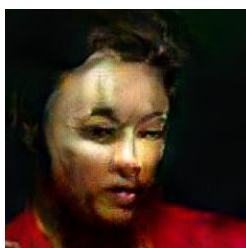
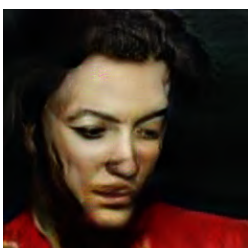
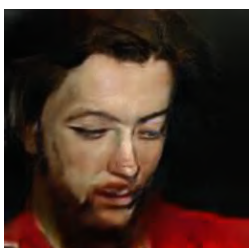
Portraits



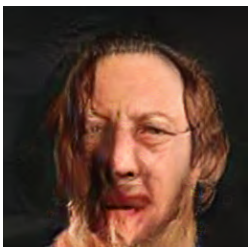
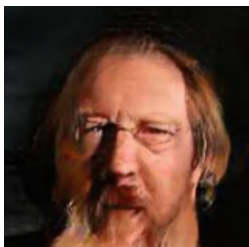
Portraits



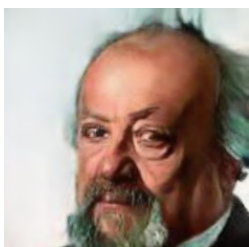
Portraits



Portraits



Portraits



Portraits

