

Supplementary: Non-local Representation based Mutual Affine-Transfer Network for Photorealistic Stylization

Ying Qu¹, Zhenzhou Shao², Hairong Qi¹

¹The University of Tennessee, Knoxville, TN ² Capital Normal University, Beijing, China

yqu3@vols.utk.edu zshao@cnu.edu.cn hqi@utk.edu

1. Failure Cases

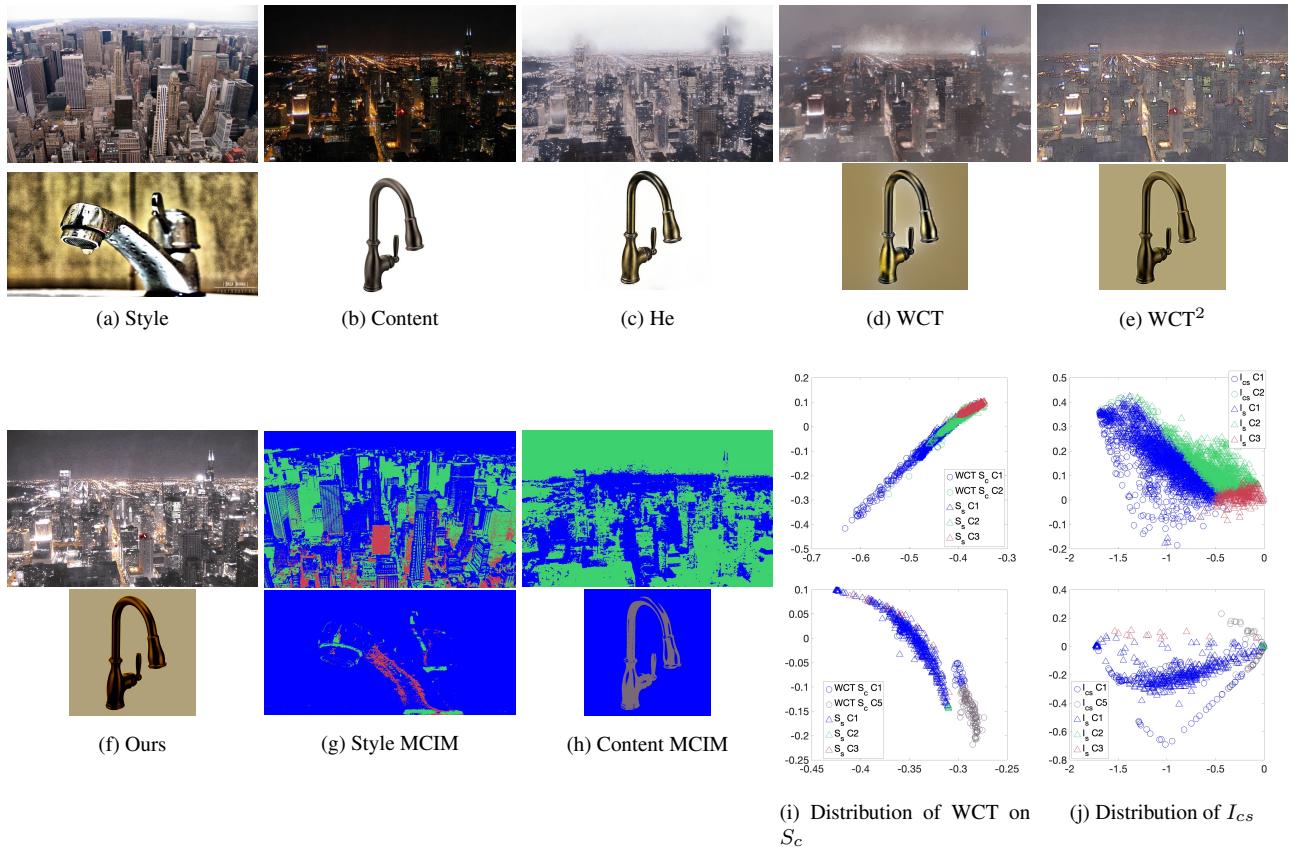


Figure 1: Failure case. The top row of each subfigure shows the results of the building, and the bottom row of each subfigure shows the results of the faucet.

In real photorealistic applications, the contexts of the content and style images may not be matched perfectly. Without any pre-trained segmentation models, the proposed scheme works well in most scenarios, even when there are some mismatches between the content and style images, as shown in Fig.14 in the manuscript. However, it may fail on some challenging image pairs where there exist large mismatches between the content and style images, as shown in Fig.1. In the top example, as shown in Figs.1g and 1h, the “white sky” in the content image is mismatched to the “light” in the style image, which happen to share similar distributions, as shown in Fig.1j. In the bottom example, the background of image pairs matches well.

However, since the faucet in the style image shares the same color basis as background in the style image, it is mismatched with the one in the content image. We can observe from Fig.1g to 1j that the proposed method could still transfer the distributions of the contexts in the content image to the matched ones in the style image. However, since the representations are mismatched, the color could not be transferred effectively. In our future work, we will exploit the usage of prior knowledge serving as additional physical constraints to regulate the learning process to further enforce semantic matching.

2. Downsampling Scale



Figure 2: The stylized results with different downsampling factors. Top row: downsampled content image; Middle row: downsampled style image. Bottom row: the stylized results. The resolution of the content image is 2560×1600 , and the resolution of the style image is 1920×1200 .

We conduct experiments to evaluate the downsampling factor and the results are shown in Fig. 2. For this image pair, the ‘light’ in the content image and style images should be matched and the ‘light’ is generally brighter in the content image than that in the style image. Thus, the stylized image should have dimmer light to match that in the style image. We can observe that the proposed method works well even when the downsampling factor is set to 32. When the downsampling factor is 64, the tower at the right part of the image starts to become difficult to recognize. This indicates that the light of the tower has not been transferred effectively. Thus, to transfer the tower effectively, we should set the downsampling factor below 32. We also observe that the sky and buildings seem not to be affected. Although some buildings in the content image may disappear, other buildings may still be there after downsampling. In this case, the stylization results will not be affected much.

3. Stylized results for the first dataset [3]



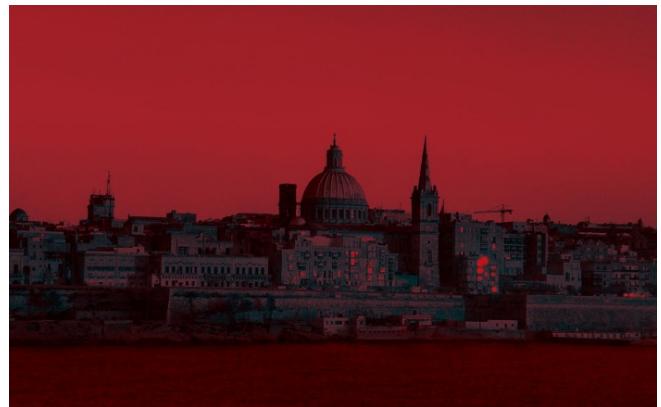
(a) Content



(b) Style



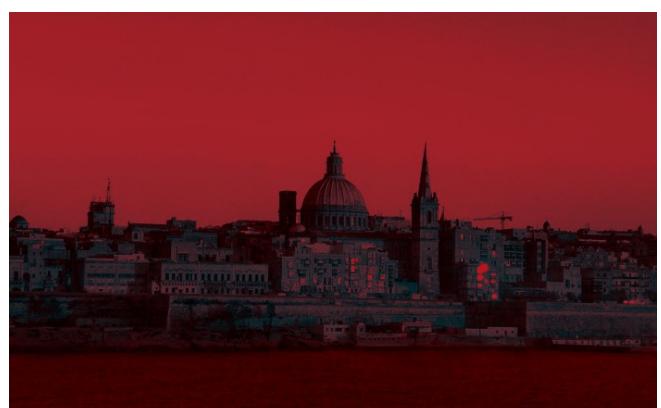
(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 3: Visual comparison for input 1 in the first dataset.



(a) Content



(b) Style



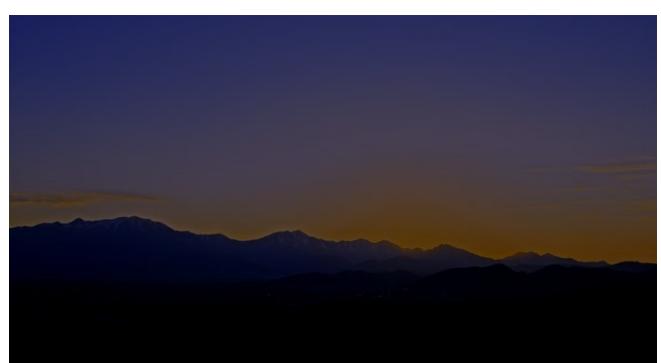
(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 4: Visual comparison for input 2 in the first dataset.



(a) Content



(b) Style



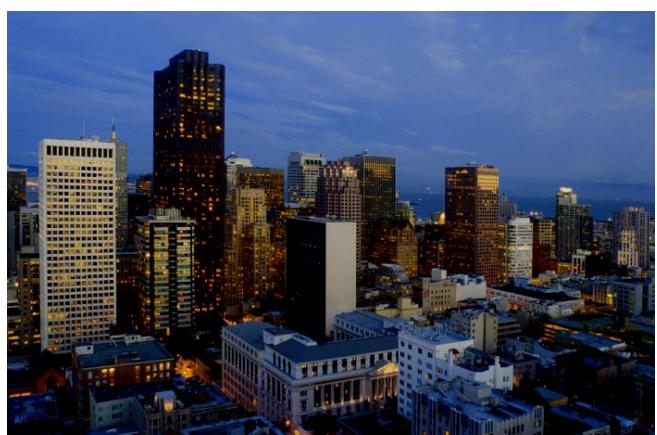
(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 5: Visual comparison for input 3 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

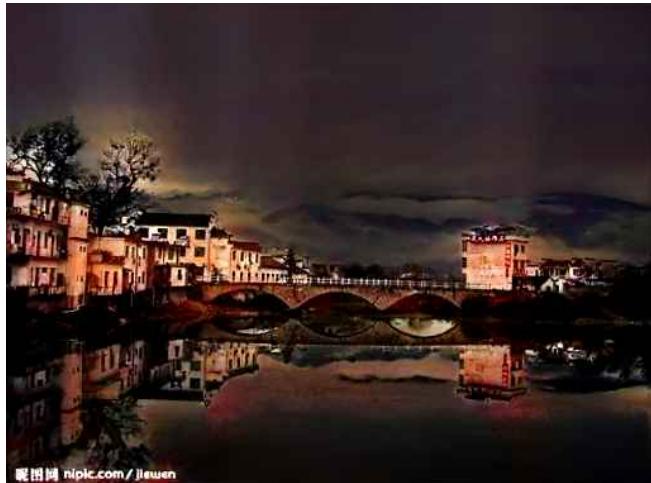
Figure 6: Visual comparison for input 4 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 7: Visual comparison for input 5 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 8: Visual comparison for input 6 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 9: Visual comparison for input 7 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 10: Visual comparison for input 8 in the first dataset.



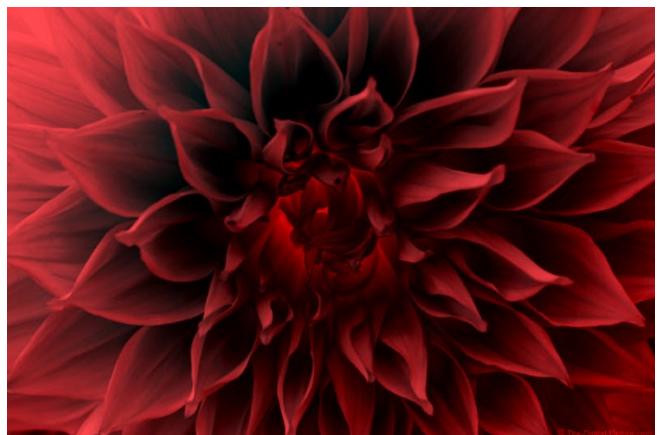
(a) Content



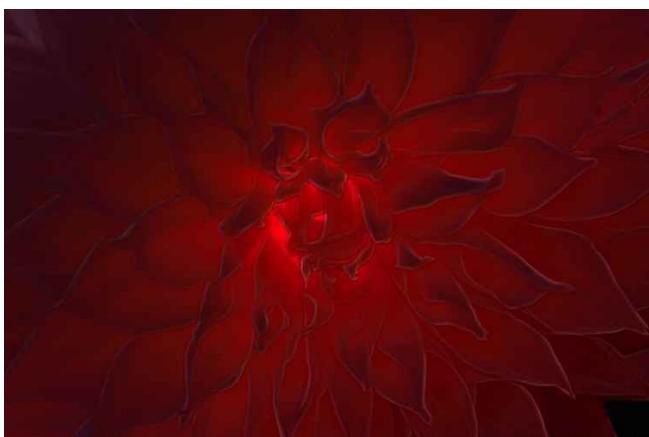
(b) Style



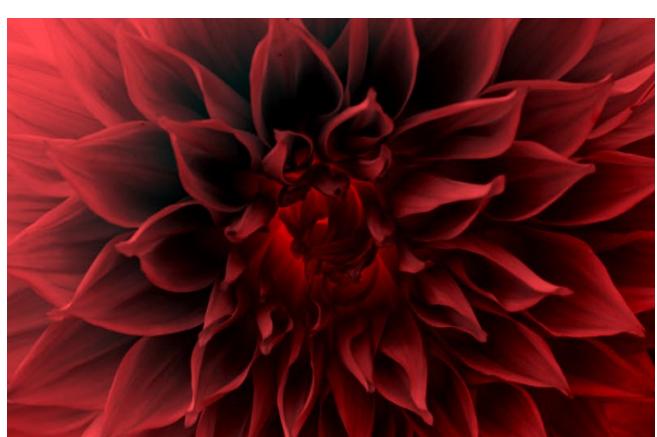
(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 11: Visual comparison for input 9 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 12: Visual comparison for input 10 in the first dataset.

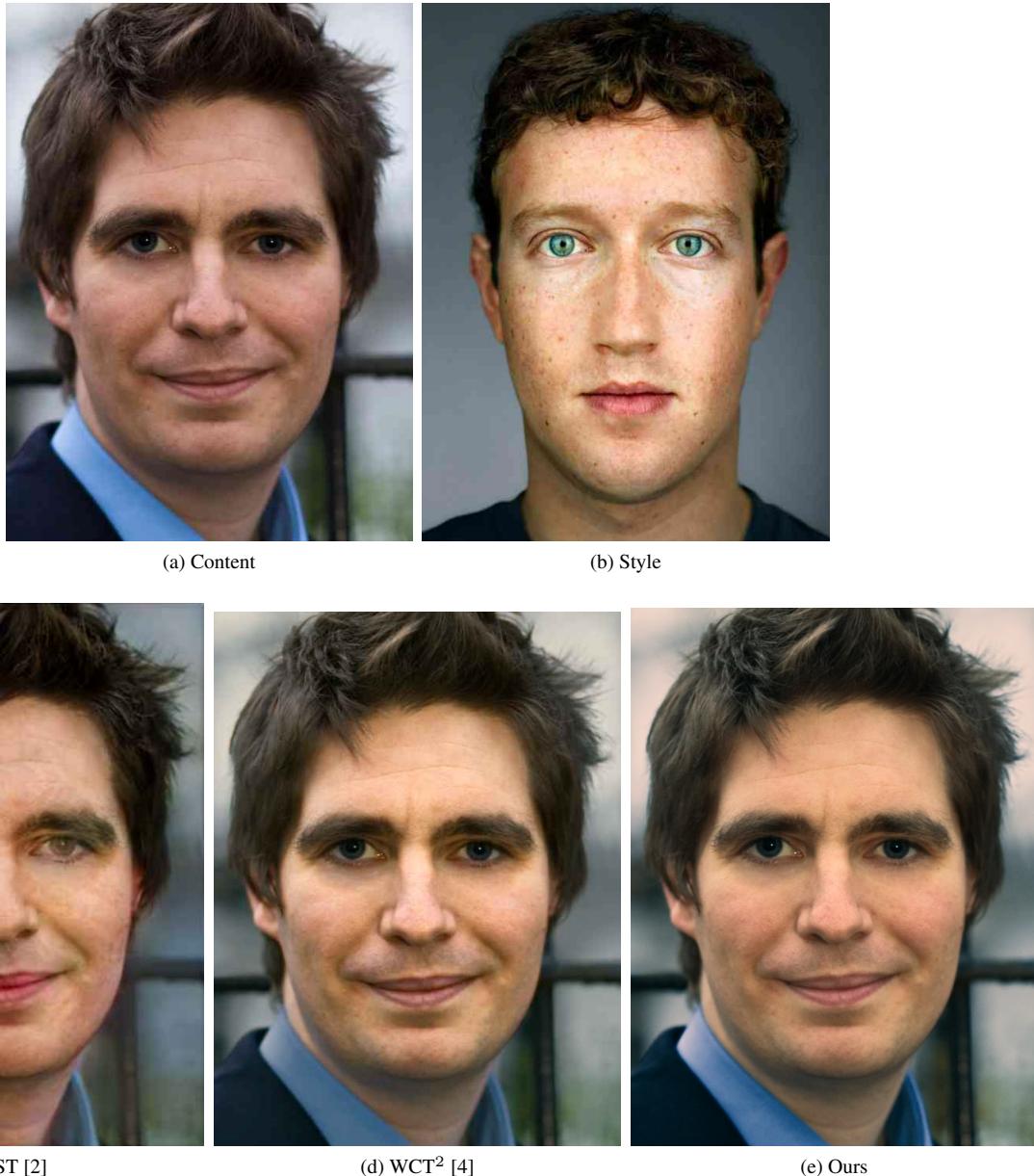


Figure 13: Visual comparison for input 11 in the first dataset.

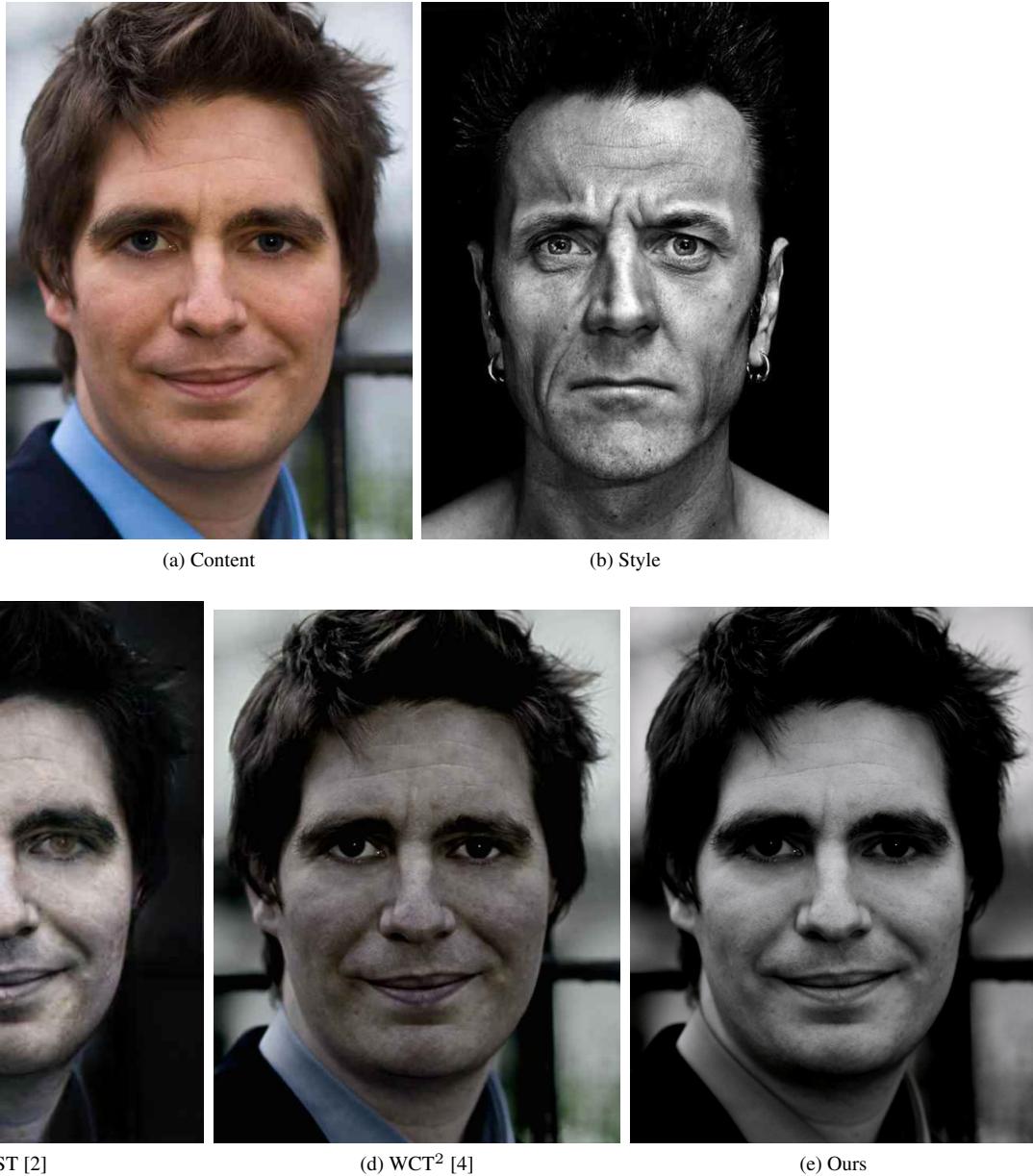


Figure 14: Visual comparison for input 12 in the first dataset.



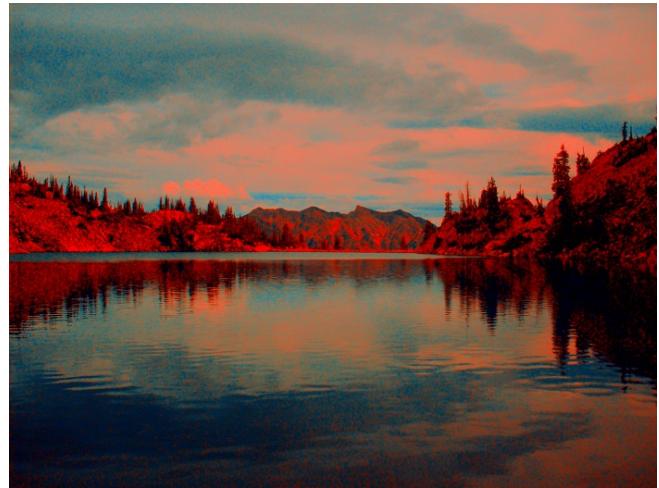
(a) Content



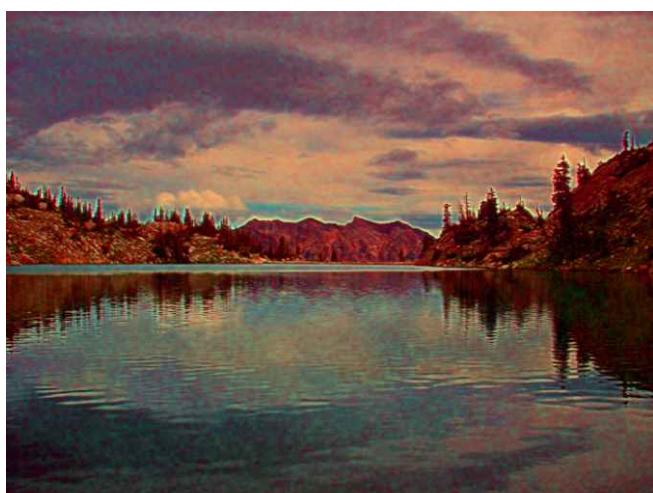
(b) Style



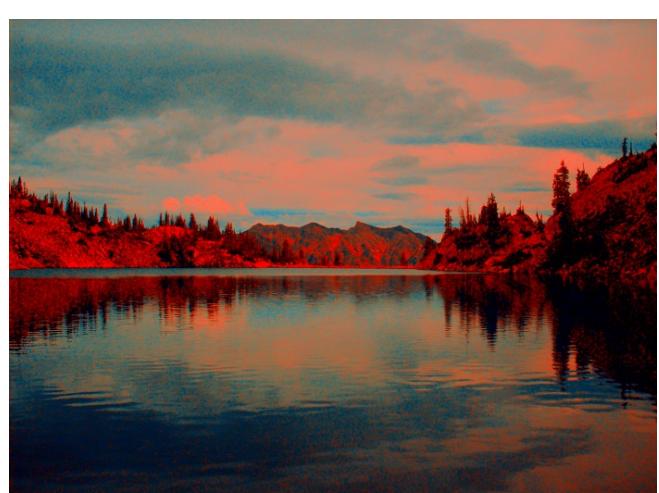
(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 15: Visual comparison for input 13 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 16: Visual comparison for input 14 in the first dataset.



(a) Content



(b) Style



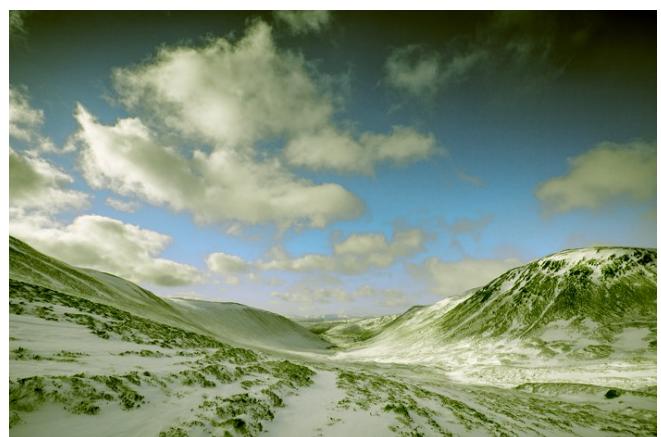
(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 17: Visual comparison for input 15 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 18: Visual comparison for input 16 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 19: Visual comparison for input 17 in the first dataset.



(a) Content



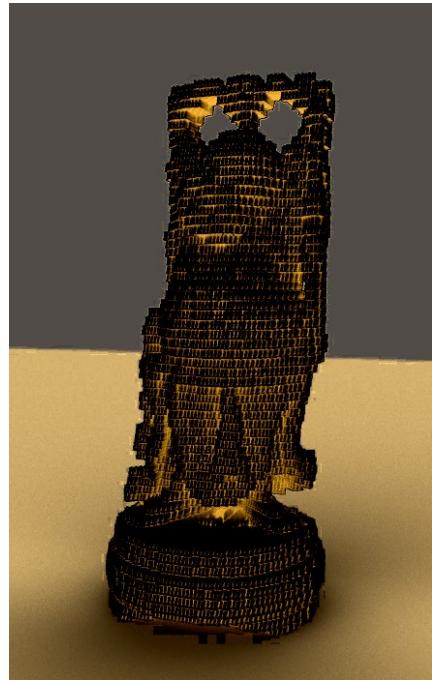
(b) Style



(c) LST [2]



(d) WCT² [4]



(e) Ours

Figure 20: Visual comparison for input 18 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 21: Visual comparison for input 19 in the first dataset.



(a) Content



(b) Style



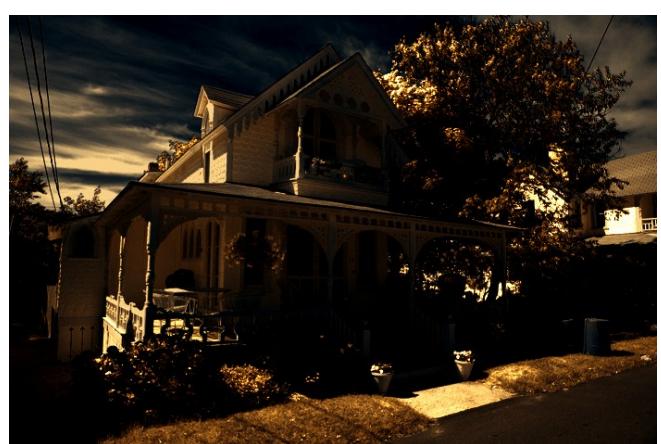
(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 22: Visual comparison for input 20 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 23: Visual comparison for input 21 in the first dataset.



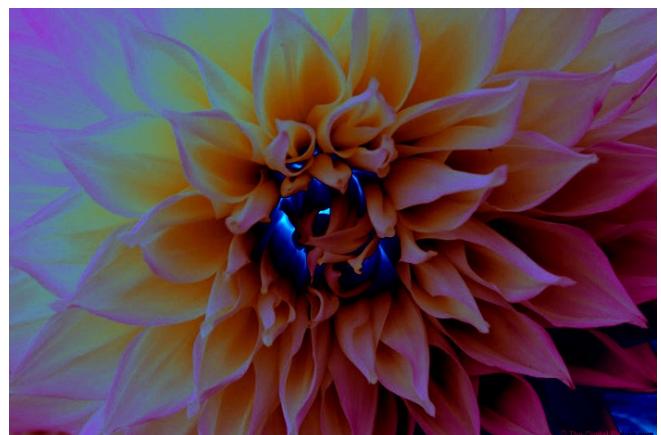
(a) Content



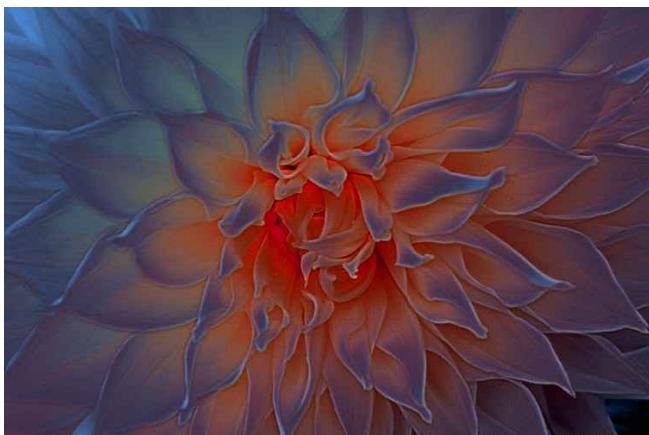
(b) Style



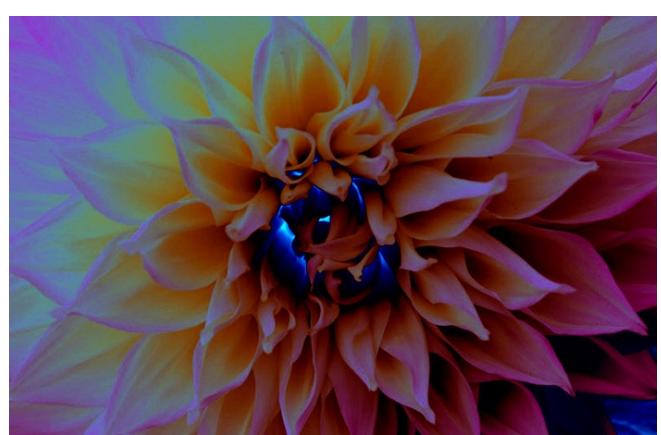
(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 24: Visual comparison for input 22 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]

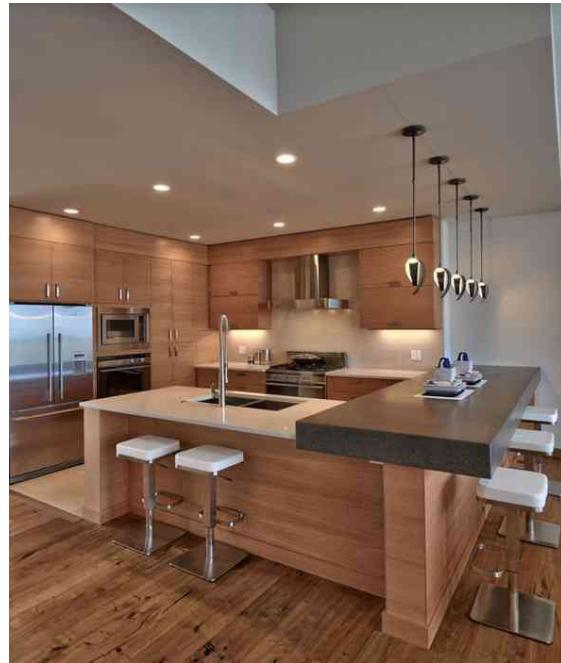


(f) Ours

Figure 25: Visual comparison for input 24 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 26: Visual comparison for input 25 in the first dataset.



(a) Content



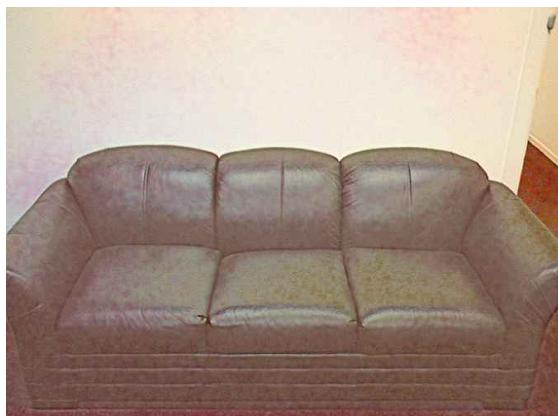
(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 27: Visual comparison for input 26 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 28: Visual comparison for input 27 in the first dataset.

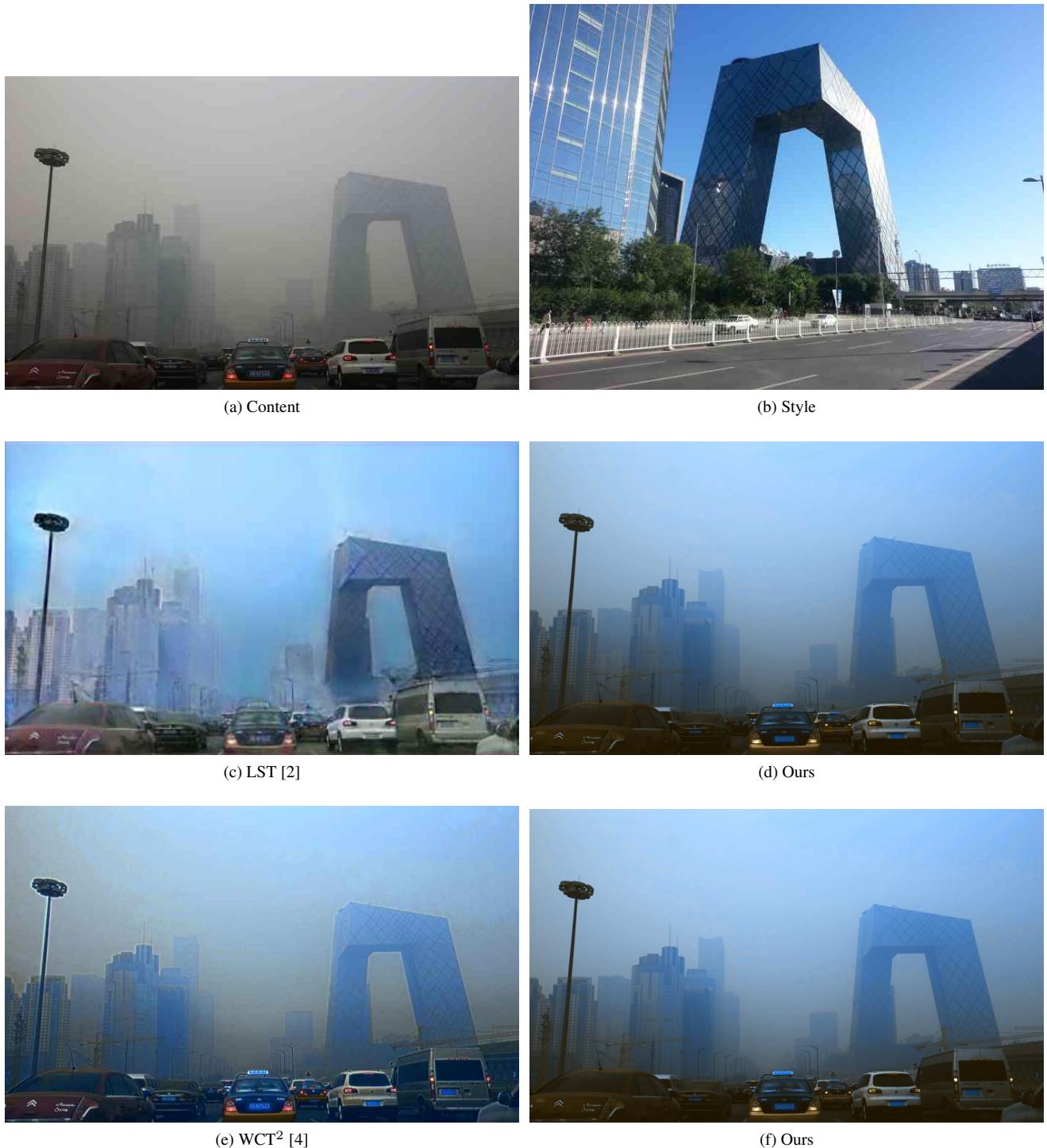


Figure 29: Visual comparison for input 28 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours

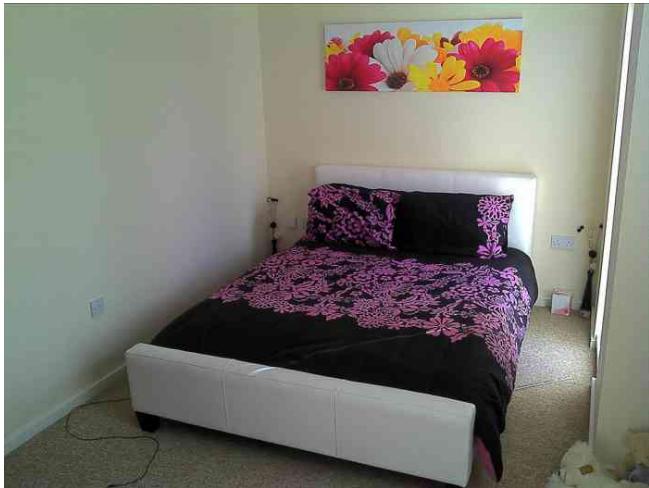


(e) WCT² [4]



(f) Ours

Figure 30: Visual comparison for input 29 in the first dataset.



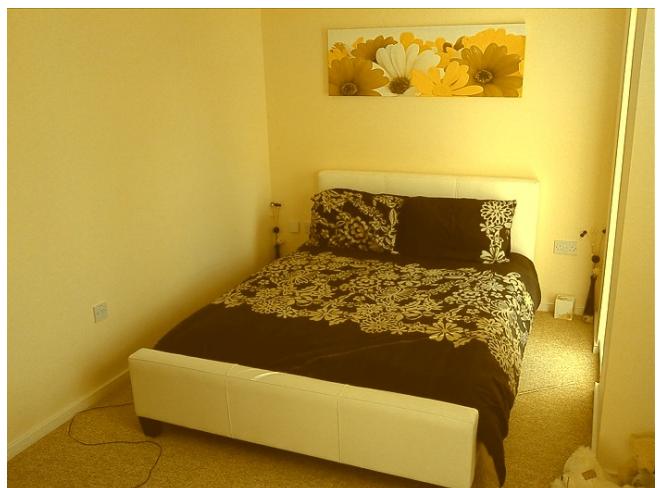
(a) Content



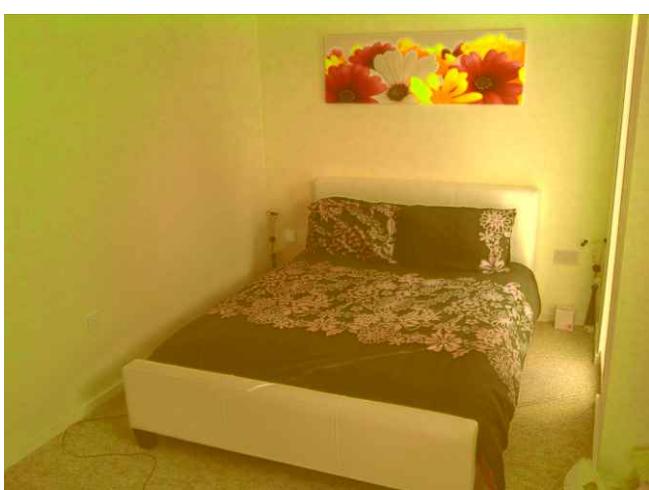
(b) Style



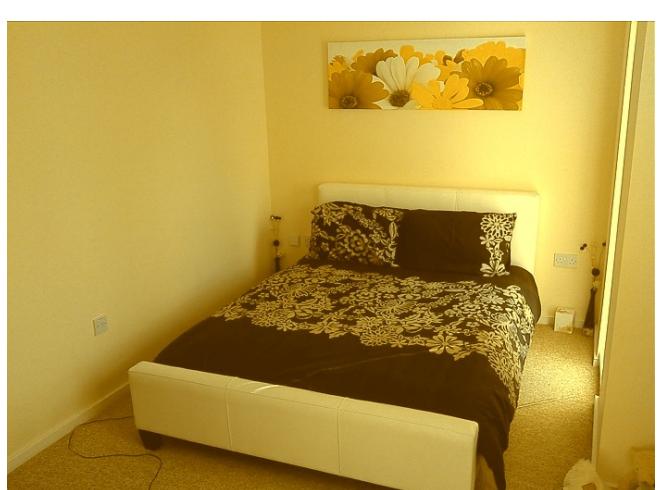
(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 31: Visual comparison for input 30 in the first dataset.



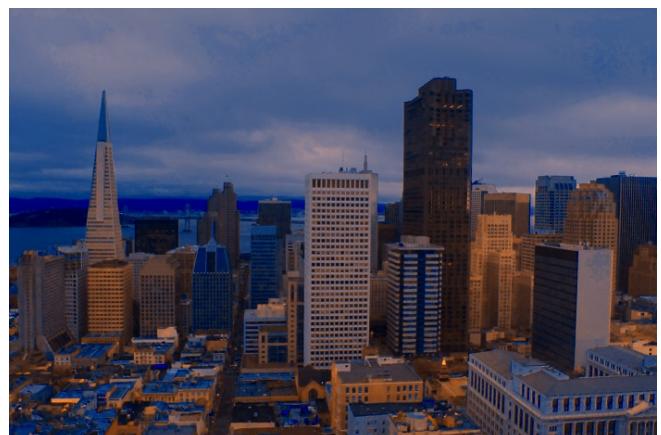
(a) Content



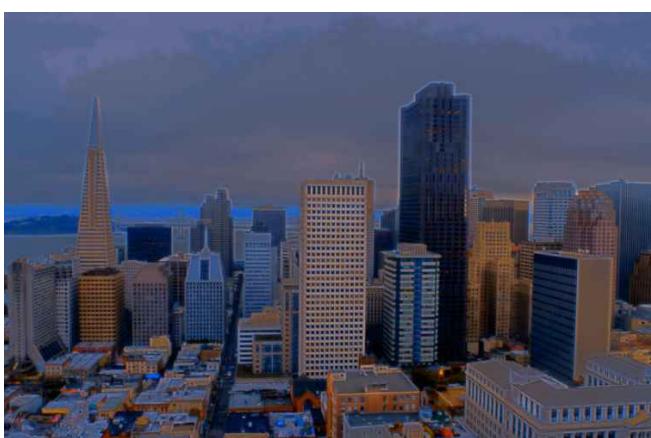
(b) Style



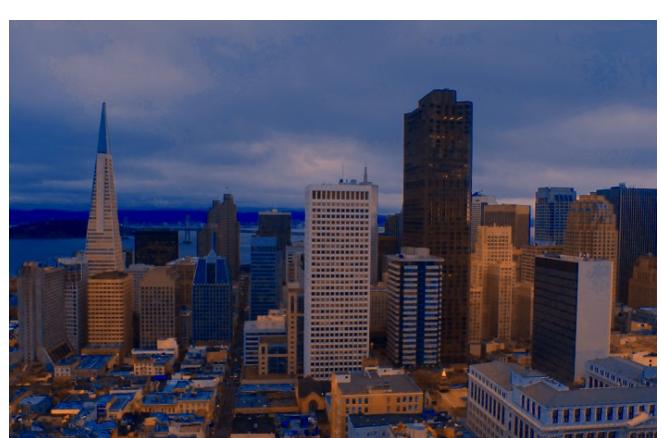
(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 32: Visual comparison for input 31 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

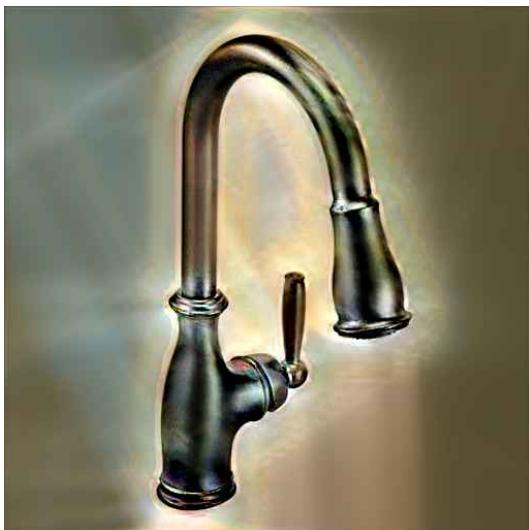
Figure 33: Visual comparison for input 32 in the first dataset.



(a) Content



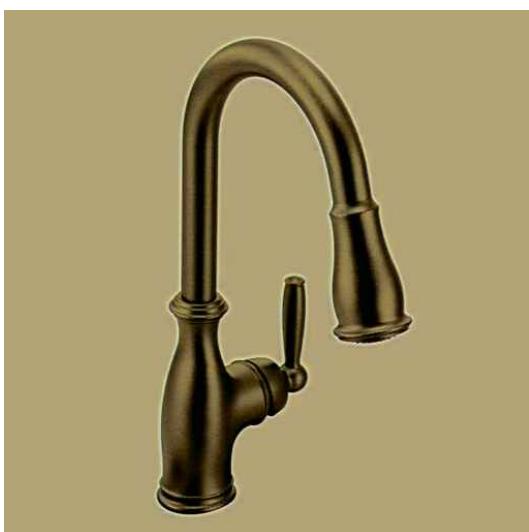
(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 34: Visual comparison for input 33 in the first dataset.



(a) Content



(b) Style



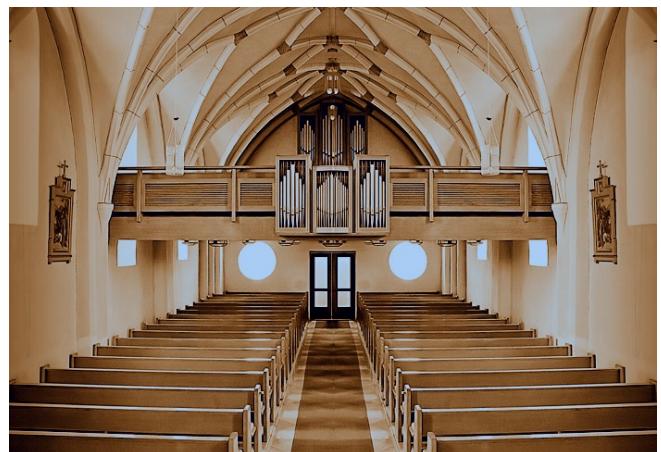
(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 35: Visual comparison for input 34 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 36: Visual comparison for input 35 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 37: Visual comparison for input 36 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 38: Visual comparison for input 37 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 39: Visual comparison for input 38 in the first dataset.



(a) Content



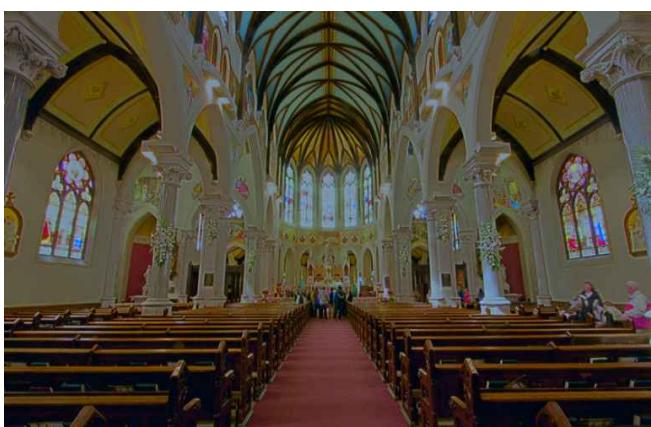
(b) Style



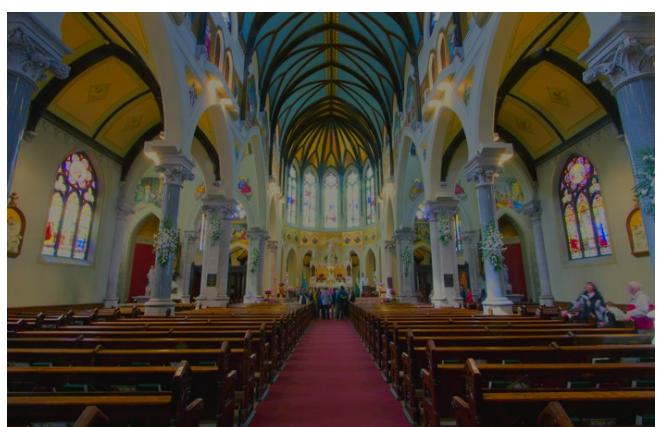
(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 40: Visual comparison for input 39 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 41: Visual comparison for input 40 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 42: Visual comparison for input 41 in the first dataset.



(a) Content



(b) Style



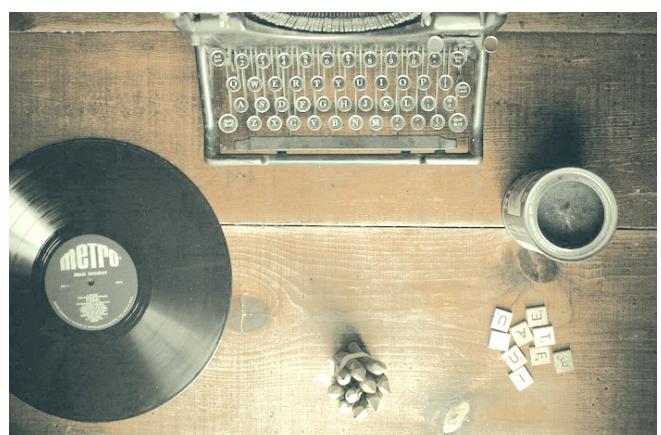
(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 43: Visual comparison for input 42 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 44: Visual comparison for input 43 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 45: Visual comparison for input 44 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 46: Visual comparison for input 45 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 47: Visual comparison for input 46 in the first dataset.



(a) Content



(b) Style



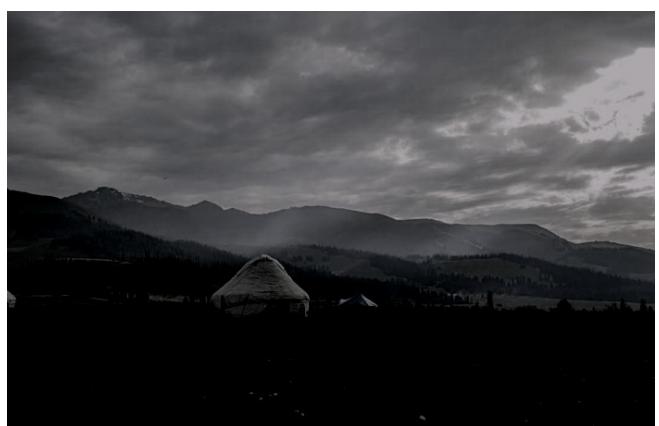
(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 48: Visual comparison for input 47 in the first dataset.



(a) Content



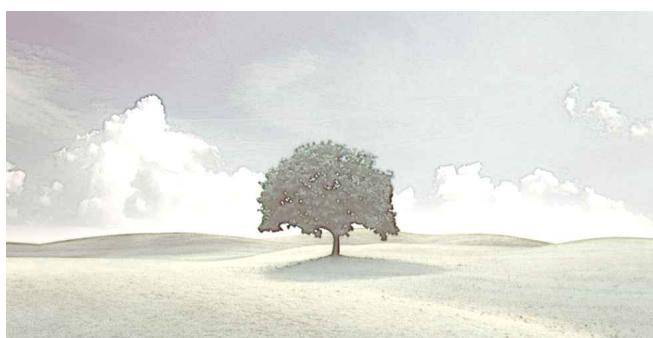
(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 49: Visual comparison for input 48 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 50: Visual comparison for input 49 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

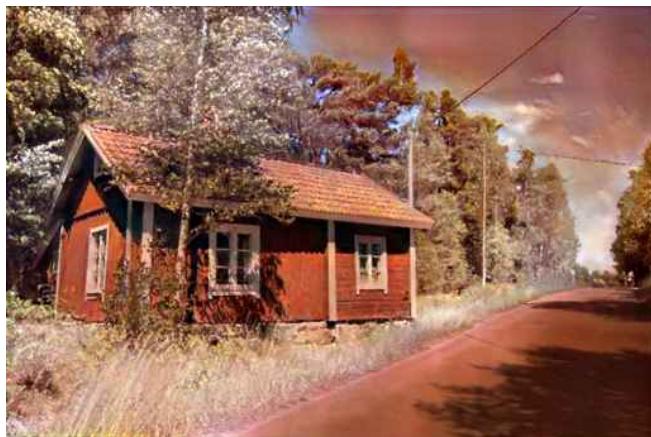
Figure 51: Visual comparison for input 50 in the first dataset.



(a) Content



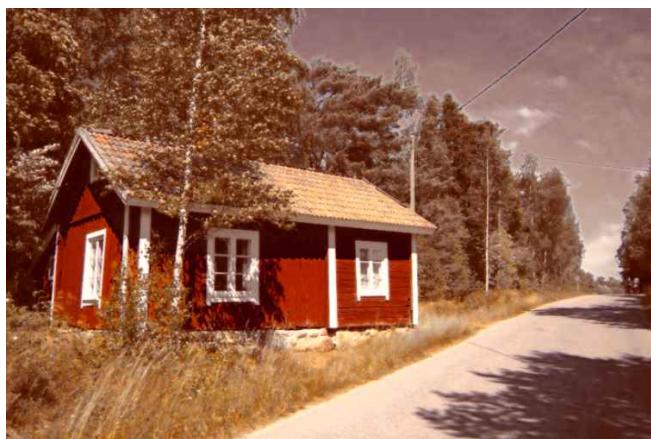
(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 52: Visual comparison for input 51 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]

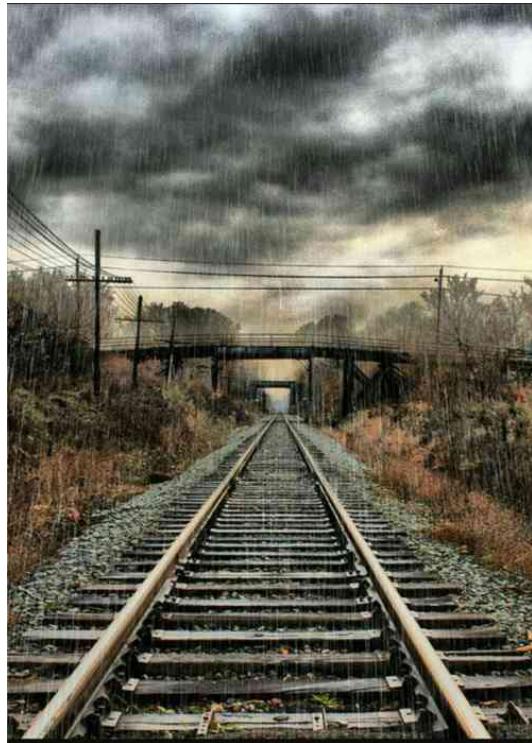


(f) Ours

Figure 53: Visual comparison for input 52 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 54: Visual comparison for input 53 in the first dataset.



(a) Content



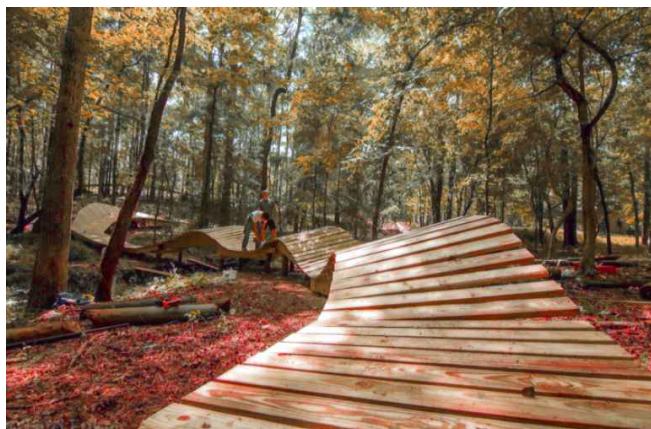
(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 55: Visual comparison for input 54 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 56: Visual comparison for input 55 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 57: Visual comparison for input 56 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 58: Visual comparison for input 57 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 59: Visual comparison for input 58 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 60: Visual comparison for input 59 in the first dataset.



(a) Content



(b) Style



(c) LST [2]



(d) Ours



(e) WCT² [4]



(f) Ours

Figure 61: Visual comparison for input 60 in the first dataset.

4. Stylized results for the second dataset [1]



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 62: Visual comparison for input 1 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 63: Visual comparison for input 2 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 64: Visual comparison for input 3 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 65: Visual comparison for input 4 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 66: Visual comparison for input 5 in the second dataset.



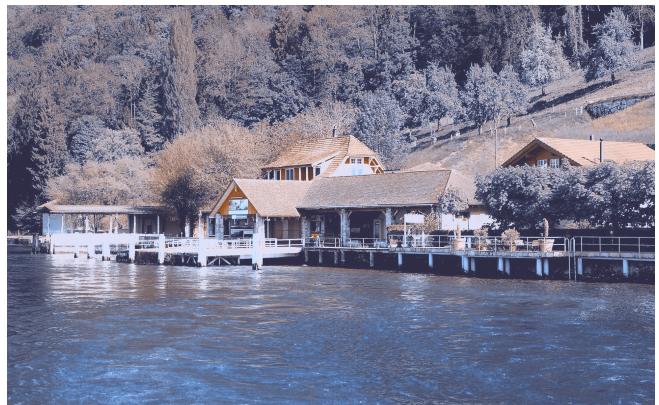
(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 67: Visual comparison for input 6 in the second dataset.



(a) Content



(b) Style

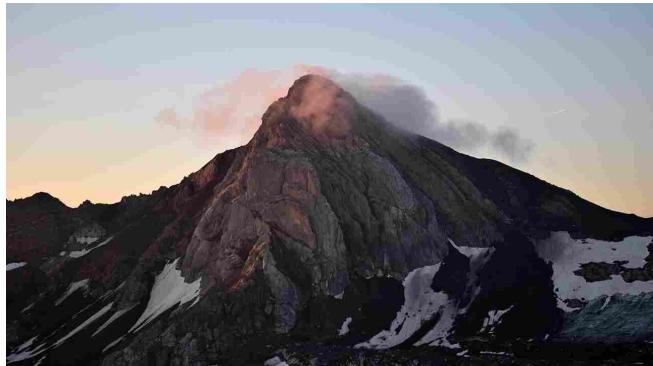


(c) PhotoNAS [2]



(d) Ours

Figure 68: Visual comparison for input 7 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 69: Visual comparison for input 8 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 70: Visual comparison for input 9 in the second dataset.



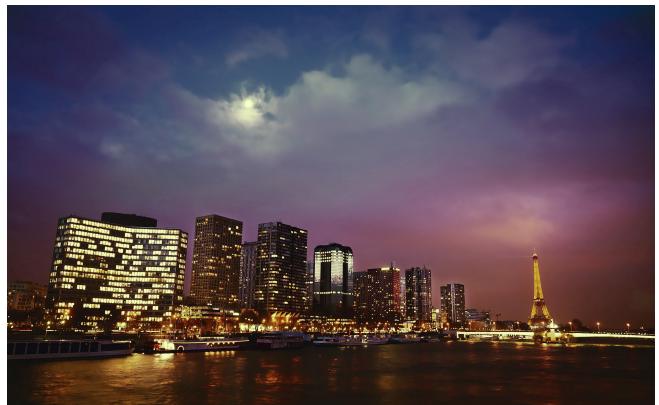
(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 71: Visual comparison for input 10 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 72: Visual comparison for input 11 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]

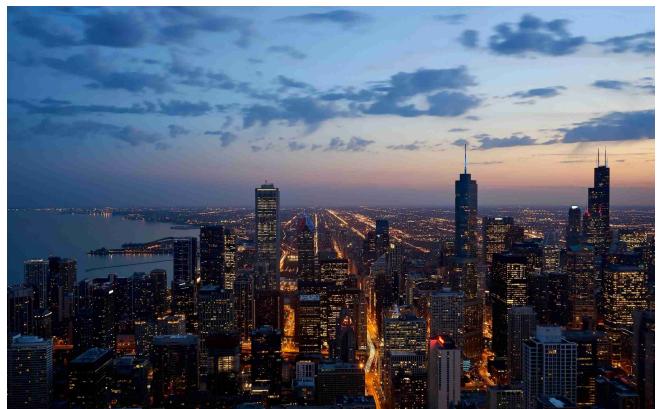


(d) Ours

Figure 73: Visual comparison for input 12 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 74: Visual comparison for input 13 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]

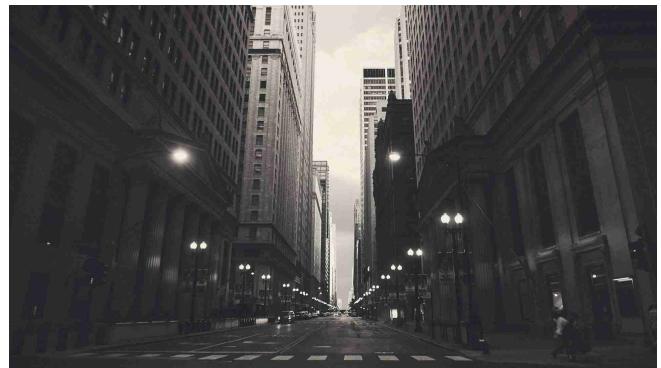


(d) Ours

Figure 75: Visual comparison for input 14 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 76: Visual comparison for input 15 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 77: Visual comparison for input 16 in the second dataset.



(a) Content



(b) Style

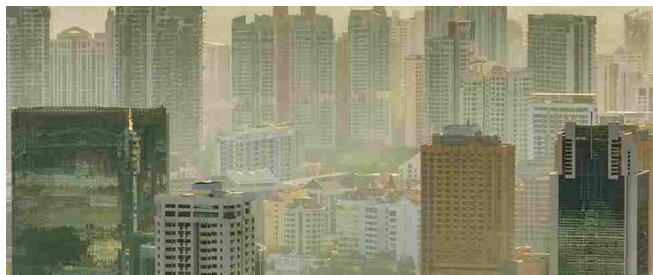


(c) PhotoNAS [2]



(d) Ours

Figure 78: Visual comparison for input 17 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 79: Visual comparison for input 18 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 80: Visual comparison for input 19 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 81: Visual comparison for input 20 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 82: Visual comparison for input 21 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 83: Visual comparison for input 22 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 84: Visual comparison for input 23 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 85: Visual comparison for input 24 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 86: Visual comparison for input 25 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 87: Visual comparison for input 26 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 88: Visual comparison for input 27 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

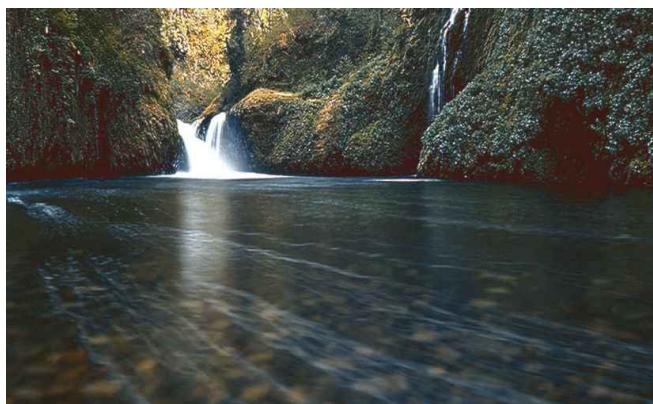
Figure 89: Visual comparison for input 28 in the second dataset.



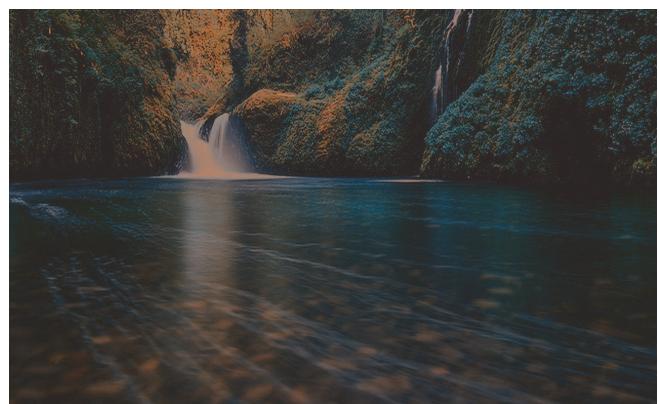
(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 90: Visual comparison for input 29 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 91: Visual comparison for input 30 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 92: Visual comparison for input 31 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 93: Visual comparison for input 32 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 94: Visual comparison for input 33 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 95: Visual comparison for input 34 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 96: Visual comparison for input 35 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 97: Visual comparison for input 36 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 98: Visual comparison for input 37 in the second dataset.



(a) Content



(b) Style

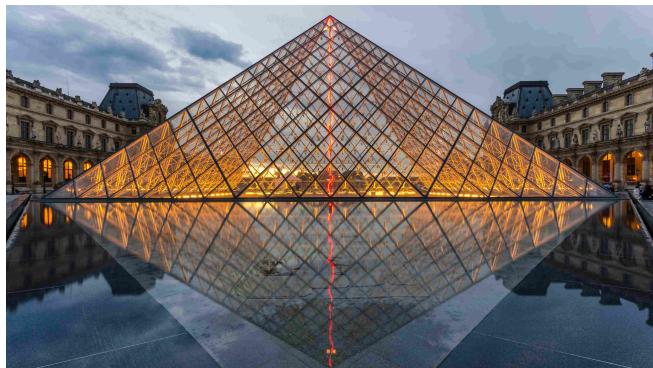


(c) PhotoNAS [2]

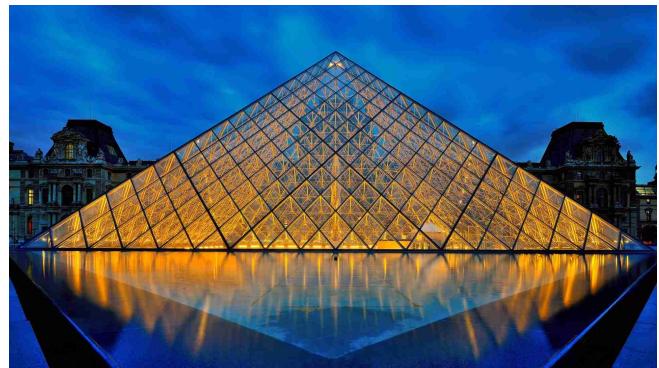


(d) Ours

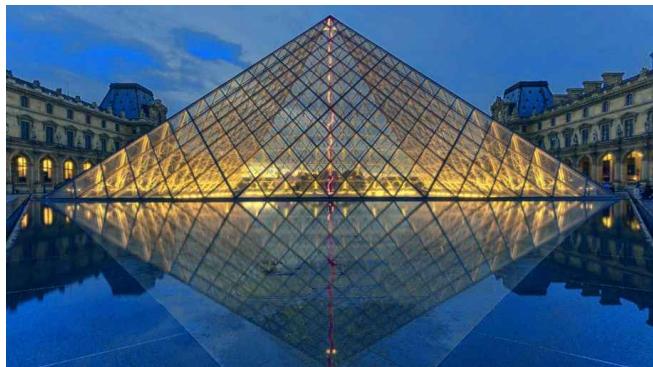
Figure 99: Visual comparison for input 38 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 100: Visual comparison for input 39 in the second dataset.



(a) Content



(b) Style



(c) PhotoNAS [2]



(d) Ours

Figure 101: Visual comparison for input 40 in the second dataset.

References

- [1] Jie An, Haoyi Xiong, Jun Huan, and Jiebo Luo. Ultrafast photorealistic style transfer via neural architecture search. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(07):10443–10450, 2020.
- [2] Xuetong Li, Sifei Liu, Jan Kautz, and Ming-Hsuan Yang. Learning linear transformations for fast arbitrary style transfer. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [3] Fujun Luan, Sylvain Paris, Eli Shechtman, and Kavita Bala. Deep photo style transfer. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4990–4998, 2017.
- [4] Jaejun Yoo, Youngjung Uh, Sanghyuk Chun, Byeongkyu Kang, and Jung-Woo Ha. Photorealistic style transfer via wavelet transforms. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 9036–9045, 2019.