Supplement: Style Transfer by Relaxed Optimal Transport and Self-Similarity

Nicholas Kolkin[1] Jason Salavon[2] Gregory Shakhnarovich[1]
[1]Toyota Technological Institute at Chicago [2]University of Chicago
nick.kolkin@ttic.edu, salavon@uchicago.edu, greg@ttic.edu

1 Extended Ablations

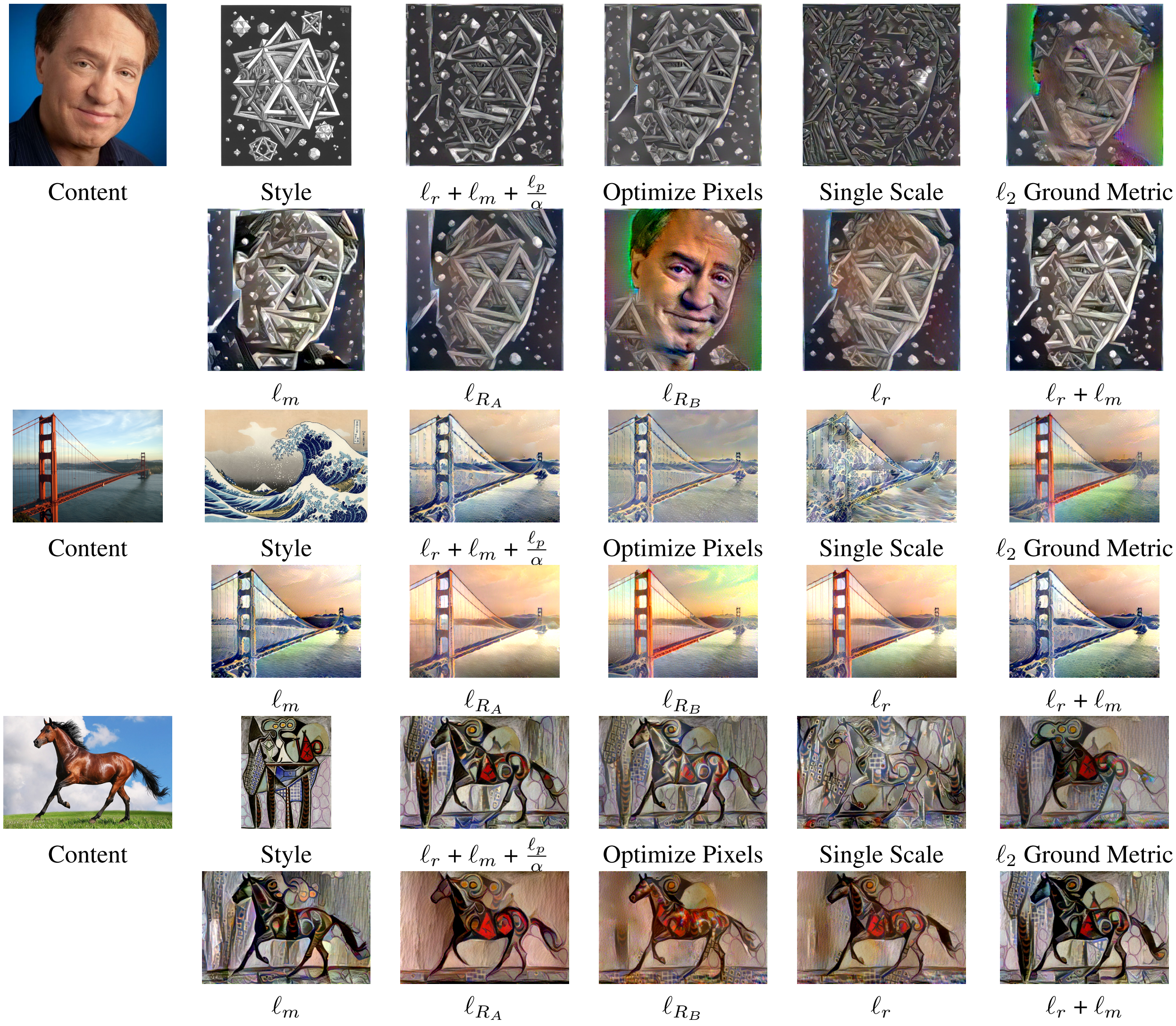
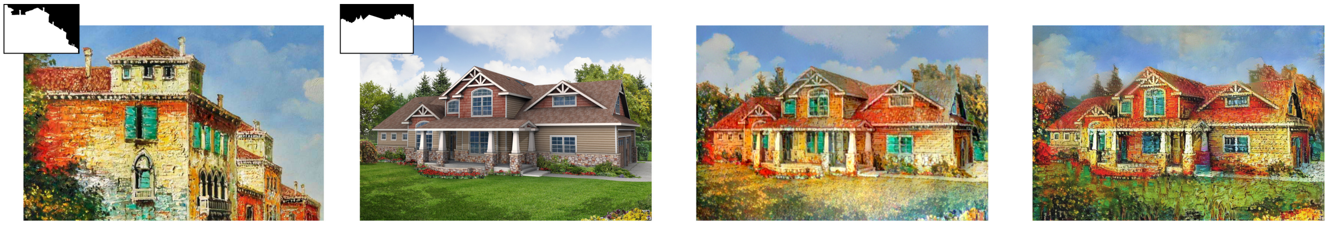


Figure 1: For each content image and style image we show the effect of different style losses or algorithmic decisions on our output. 'Optimize Pixels' refers to performing gradient descent on pixel values of the output directly, instead of the entries of a laplaccian pyramid (our default). In 'Single Scale' we perform 800 updates at the final resolution, instead of 200 updates at each of four increasing resolutions. In 'ℓ_2 Ground Metric' we replace the ground metric of the Relaxed EMD with euclidean distance (instead of our default, cosine distance). The other style loss ablations are explained in Section 4.2 of the main text.



Style & Mask

Content & Mask

Ours

Gatys

Figure 2: Qualitative comparison of the resulting output of our spatial guidance and that proposed in [2]

2 Spatial Guidance Comparison

demonstrate that our proposed method for spatial guidance gives the same level of user-control as those previously proposed we provide a qualitative comparison in Figure 2. For the same content and style with the same guidance masks we show the output of our method, and the output of the method proposed in [2] using one of the examples from their paper.

3 Random Qualitative Comparisons

In Figures 3 through 11 (on the following pages) we show 10 style/content pairs sampled uniformly at random (without replacement) from the images used in the AMT studies for the 'paired','unpaired', and 'texture' regimes (we sample 10 different pairs for each domain). In each figure we show the output of each algorithm for a certain hyper-parameter setting: low-content, high-content, or the defaults. Qualitatively we believe that the default hyper-parameter settings for each method tend to look best in the 'paired' and 'unpaired' regimes, but for 'texture' the best results are produced by the low-content parameter setting.

We also note that the output of the methods proposed by Gatys et al. [1] and Mechrez et al. [5] do not seem particularly sensitive to the range of content weights tested (0.25x to 2x the default), and the output varies only slightly. This is evident when comparing Figures 2 through 4, 5 through 7, or 8 through 10, and reflected in the similar scores different hyper-parameter settings of these methods received in our AMT study.

To facilitate future comparisons we will make the source of our AMT study available online, along with all outputs of our method and those of Gu et al. [3], Gatys et al. [1], Li et al.[4], and Mechrez et al. [5].

References

[1] L. A. Gatys, A. S. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2414–2423, 2016. 2, 3, 4, 5

[2] L. A. Gatys, A. S. Ecker, M. Bethge, A. Hertzmann, and E. Shechtman. Controlling perceptual factors in neural style transfer. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 2

[3] S. Gu, C. Chen, J. Liao, and L. Yuan. Arbitrary style transfer with deep feature reshuffle. 2, 3, 4, 5

[4] C. Li and M. Wand. Combining markov random fields and convolutional neural networks for image synthesis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2479–2486, 2016. 2, 3, 4, 5

[5] R. Mechrez, I. Talmi, and L. Zelnik-Manor. The contextual loss for image transformation with non-aligned data. *arXiv preprint arXiv:1803.02077*, 2018. 2, 3, 4, 5