Photo Weather Transfer: Comparing Neural Style and Deep Neural Style Transfer

Juliet Okwara
Department of Computer Science
Stanford University
jokwara@stanford.edu

Nicholas Seay
Department of Computer Science
Stanford University
nseay@stanford.edu

Abstract

The aim of this paper is to tackle the challenge of photorealistic style transfer, specifically in the context of transferring the weather, sky, and ambience elements of one photo into another. Existing methodologies such as Neural Style Transfer by Gatys et al. has been proven to be successful in generating images that merge the content of a content reference image and the style of a style reference image into a single image. However, these generated images tend to work best in an artistic context. The noisy image which the network trains on is often changed in ways reminiscent of a painting, possibly created by the artist who made the style image, rather than a photo. This paper shows the results of introducing to the original algorithm the ideas of semantic image segmentation and the Matting Laplacian. Since the goal is to transfer qualities mainly of the sky, image segmentation is used to target the sky such that the gradients in backpropagation only affect these areas. To reduce the "painterly" aspect of Neural Style Transfer, the Matting Laplacian is introduced to smooth the curvy, wavy lines that are often produced in the process. These ideas are inspired by Deep Photo Style Transfer by Luan et al. and this paper seeks to show the improvements that these changes make, as well as their shortcomings. Specifically, image segmentation was quite successful in preventing many unwelcome changes to the foreground, though sometimes the changes to the background were great enough that the difference in brightness or mood did not quite match the foreground. Meanwhile, the Matting Laplacian, while not unhelpful, often did not make as dramatic a change as expected.

Introduction

Realistic photographic style transfer has remained a long standing challenge in deep learning. Photographic style transfer seeks to transfer the style of a selected style image onto a provided content image. This allows a new generated photograph to appear to have been taken under different conditions--time of day, weather, ambiance, or location--while still producing a photographic result.

Familiar attempts at style transfer are limited in their ability to generate a realistic combination of style and content images, leaning too heavily into an artful interpretation of the content, one unfaithful to the original content photograph. Here, our approach builds upon the Neural Style Transfer of Gatys *et al.*, and attempts to recreate elements of Luan *et al.*'s Deep Photo Style Transfer by altering the original Neural Style Transfer to produce results that mimic photographs rather than the artistically inspired paintings of Neural Style Transfer [5,6].

To do so, we explore Semantic Image Segmentation of the content photograph with the inclusion of the Matting Laplacian.

We explore Semantic Image Segmentation in order to reduce the tendency of Neural Style Transfer to spill misplaced style onto the foreground of the content image rather than focusing solely on the background. This is especially relevant to weather and ambiance transfer as most transfers will be concerned with some transformation of the sky (e.g. presence of clouds, coloration, lighting).

We seek to preserve and create a locally affine colorspace with the addition of a term in the loss function that is dependent upon the Matting Laplacian. This new term hopes to reduce wavy lines, spirals, and content distortions that Neural Style Transfer often produces.

With these two approaches to the problem, given an input consisting of an RGB style and content image along with a black and white mask of the content image, we produce images that are both artistic and realistic and provide insights as to what extensions of Neural Style Transfer produce the most photorealistic results.

Related Work

Neural Style Transfer serves as the baseline model for this project. Neural Style Transfer is a well-known technique for generating an image while minimizing the difference between its content and that of a content reference image and the difference between its style and that of a style reference image.

Though this technique is effective in creating artistic interpretations of the style image, we desired to determine how well it fares with utilizing photographs as input for both content and style and producing a photorealistic output. More specifically, we were looking to test the limits of Neural Style in generating precise stylistic choices that create a new ambiance/weather effect while maintaining realism.

With a focus on photorealism, our goal is most similar to those papers more concerned with photocentric aspects of transfer.

We relate to Shih *et al.*'s time of day hallucination in our desire to realistically transfer, or hallucinate, photo ambiance like time of day [10]. Shih *et al.*'s time of day is specified by semantic time labels, and their method transfers the color appearance from videos with a similar scene to the input photo onto the input photo. After finding a frame that correctly matches the distribution of color in the input image, a target frame that represents the desired time of day is selected. Shih *et al.* then find dense correspondents between the input image and the matching frame using a Markov Random Field. The match and target are then warped to align with the input. The output is then hallucinated by transferring the color of the warped frames to the input.

Our approach differs however as Shih *et al.* utilizes a database generated from time-lapse videos in order to infer the transformation needed for a specific time of day rather than relying on only style and content images.

Another form of weather/ambiance based image transformation comes in the form of Laffont *et al.*'s Transient Attributes for Outdoor Scenes [7]**.** Here, each image is characterized by transient scene attributes. This allows each image in their photo database to be represented as a point in attribute space. Each dimension of this attribute space maps to a scene property that aligns to aspects like time, weather, lighting conditions, etc. Utilizing their Transient Attribute Database on a set of attributes to be altered in the scene, they find example images from the database that both correspond to the given scene and exhibit those desired attribute changes. The input image is then edited with an analogy-based appearance transfer approach.

This approach is rather different than the approach taken in our paper as we look to perform a weather/ambiance transfer without a large database of style images that are closely representative of the content images. We look to find a much more generic approach that can handle a broad range of style images.

Reinhard *et al.*'s color transfer returns to the concept of a style and content image [3]. But they use statistical analysis in order to impose the style image's color characteristics onto the content image. Our approach for this paper, however, makes use of deep learning.

Our work is largely a reproduction of Luan *et al.*'s Deep Photo Style Transfer which seeks to limit the transformation that occurs between the input content image and the output image while allowing the output to remain locally affine in colorspace. Luan *et al.* are able to successfully limit the amount of distortion that occurs while producing rather photorealistic transfers. To reduce the effects of style "spillover" onto undesirable areas for the created image, semantic segmentation of the input and style images is performed with manually created segmentation masks. This method resembles Champandard's *et al.*'s *Semantic Style Transfer and Turning Two-Bit Doodles into Fine Artwork* [1].

Their changes allowed for the transfer of time of day, weather, season, and other artful edits. The methodology explored in this paper are inspired by the approaches of both Gatys *et al.*'s Neural Style Transfer as well as Luan's *et al.*'s Deep Photo Style Transfer with the intent of replicating similar results but from a weather/ambiance transfer perspective rather than the more artistic perspective of Gatys *et al.* or the more broadly focused Luan *et al.*. It is Luan *et al.'s* work that this project seeks to both replicate and open more of a discussion on as it specifically applies to weather and ambiance transfer. In particular, we explore the growth from Neural Style Transfer to Deep Photo Style Transfer.

Dataset and Features:

Because we employ the pre-trained VGG-19 as our feature extractor, our dataset remains small, consisting of 7 style images and 2 content images. Content images consist of outdoor scenes of both nature and urban settings. The images contain a large background so that style transfer would be more apparent. Style images selected contain some interesting element of weather or ambiance (e.g. sunsets, colored sky, dark clouds, bright sunlit background, etc). These style and content images are then paired off to produce artful and interesting background transfer results. Images are pulled from Flickr as well as Luan *et al.'s* github repository.

Accompanying each content image is a semantic segmentation mask. The foreground pixels are converted to zeros while the background pixels are converted to ones.

Figure 1: content image of cityscape along (shown left) along with its semantic segmentation mask (shown right)

Methods

The Baseline Model: Neural Style Transfer

Because this project seeks to apply the weather and ambience-related elements of the background of a style image onto the background of a given content image, Neural Style Transfer was a reasonable baseline to set as the overall arrangement and foreground of the style image is perceptually recognizable while the colors and structures that pertain to the scenery are provided by the style image.

We minimize the following loss function created by Gatys *et al.* and summarized below by Luan *et al.* to transfer a style image S onto the content image I in order to produce an output image O:

$$\mathcal{L}_{\text{total}} = \sum_{\ell=1}^{L} \alpha_\ell \mathcal{L}_c^\ell + \Gamma \sum_{\ell=1}^{L} \beta_\ell \mathcal{L}_s^\ell$$

$$\text{with: } \mathcal{L}_c^\ell = \frac{1}{2N_\ell D_\ell} \sum_{ij} (F_\ell[O] - F_\ell[I])_{ij}^2$$
$$\mathcal{L}_s^\ell = \frac{1}{2N_\ell^2} \sum_{ij} (G_\ell[O] - G_\ell[S])_{ij}^2$$

Keeping with Luan's notation of the loss function, L is the total number of convolutional layers, and ℓ represents the ℓ-th convolutional layer. Each layer has N_ℓ filters with a D_ℓ sized vectorized feature map. $F_\ell[\cdot] \in R^{N_\ell \times D_\ell}$ is the feature matrix. The Gram matrix $G_\ell[\cdot] = F_\ell[\cdot]F_\ell[\cdot]^T \in R^{N_\ell \times N_\ell}$. Layer preference configuration is controlled by α_ℓ and β_ℓ respectively while Γ balances between content loss, L_c^ℓ, and style loss, L_s^ℓ.

Employing a tensorflow implementation of feed-forward Neural Style Transfer [8], we perform 1000 iterations with a style layer weight of 0.2. We opt to use average pooling rather than max pooling for a better gradient flow and because we found it to obtain more appealing results.

Extensions to Neural Style Transfer

Semantic Image Segmentation

One of the most prominent issues with Neural Style Transfer is the spilling over of the style image onto the foreground of the content image, decreasing the photographic quality of the image and muddling the original image context.

To solve this problem, semantic image segmentation masks are created for each of the content images through Adobe Photoshop (Fig. 1). With pixel values of zeros in the foreground region and pixel values of 1 in the background region, we limit gradient application to the regions where we desire the style image to affect, thus curtailing spillover.

Content loss is computed the same as the baseline Neural Style Transfer. However, style loss, L_s^ℓ, is altered to accommodate the inclusion of the mask. Recall,

$$\mathcal{L}_s^\ell = \frac{1}{2N_\ell^2} \sum_{ij} (G_\ell[O] - G_\ell[S])_{ij}^2$$

And that the Gram matrix $G_\ell[\cdot] = F_\ell[\cdot]F_\ell[\cdot]^T \in R^{N_\ell \times N_\ell}$. With the inclusion of the masks, utilizing the same notation as Luan *et al.*, we redefine style loss to be:

$$\mathcal{L}_{s+}^\ell = \sum_{c=1}^{C} \frac{1}{2N_{\ell,c}^2} \sum_{ij} (G_{\ell,c}[O] - G_{\ell,c}[S])_{ij}^2$$

where $F_{\ell,c}[O] = F_\ell[O]M_{\ell,c}[I] \quad F_{\ell,c}[S] = F_\ell[S]M_{\ell,c}[S]$

and C is the number of channels in the mask, $M_{\ell,c}[\cdot]$ represents c-th channel of the segmentation mask in layer ℓ, and the Gram matrix $G_{\ell,c}[\cdot]$ corresponds to $F_{\ell,c}[\cdot]$. At each layer of the convolutional neural network, we resize the mask appropriately to match the feature map.

Matting Laplacian

As detailed by Luan *et al.*, a strategy for generating an image that is photorealistic is to leverage the fact that the original content and style images are photorealistic. Therefore, the photorealistic quality of the image must simply be retained, not necessarily produced. The tactic used for this is the Matting Laplacian. This is a transform function which is affine in colorspace. Namely the RGB values are mapped to others in such a way that any edges within a large set of variations remain. Furthermore, the Matting Laplacian works on patches of the image, allowing for different affine functions in each patch, which frees up the image enough to still allow for faithful transfer of the desired style elements of the style reference image. As a result of this addition, the following is added to the overall loss function with a coefficient lambda as a hyperparameter to penalize images which are not well described by the transform:

$$\mathcal{L}_m = \sum_{c=1}^{3} V_c[O]^T \mathcal{M}_I V_c[O]$$

Here, if the input image I has N pixels, then matrix M_I is N x N. $V_c[O]$ is the vectorized version (N x 1) of the output image O in channel c.

Experiment/Results/Discussion

Comparison to Neural Style Transfer

To see the effect that each addition to neural style had towards the improvement of weather transfer, we took two content images, each with a style image, and ran the Neural Style Transfer (NST) algorithm on its own, then with the addition of the segmentation mask, and finally with the secondary addition of the Matting Laplacian. After doing so, we also ran these content image examples with all additions using different style images to see how the method worked with other scenes.

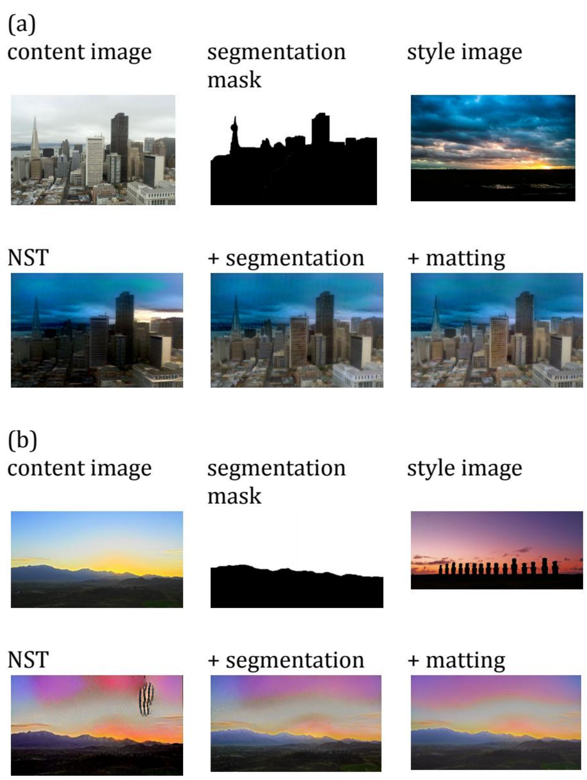

Figure 2: In these examples, the results of each step in the transfer can be seen. Neural Style Transfer gives a base to work with. Adding semantic segmentation had the large effect of eliminating unwanted artifacts and transfer of the foreground (seen in b). However, the Matting Laplacian did not appear to provide significant improvements.

The additions seem to reduce the amount of mistakes that the original Neural Style Transfer Algorithm makes. Notice the over-darkening of the city-scape as well as the very rectangular section of sunlight that neural style chose to keep which look out of place in (a) of Figure 2. Similarly, there is an artifact, potentially from the striped pattern of the statues, in the top-left corner of the mountain scene of Neural Style Transfer found in (b) of Figure 2.

While there were improvements, we did not expect to see as much distortion as we produced in the buildings (windows), and this is something that can be improved upon. Similarly, the small variances in the sky of the content image had a larger effect than expected, as it prevented the full propagation of the purples and reds throughout the sky of the generated images.

Extension Pitfalls

Figure 3: Adding segmentation masks results in more targeted transfer of the sky. This can lead to peculiar differences in tone, so choices must be carefully made in order to render a more convincing image. In addition, the current implementation suffers from occasional incomplete transfer, such as that in the purple cityscape sky, or poor transfer, such as that in the blue and red mountain sky

Conclusion/Future Work

Our additions of Semantic Image Segmentation and Matting Laplacian improved the photorealistic qualities from Neural Style Transfer but still had their shortcomings in regard to image distortions, even style application, and potential poor transfer results. The foreground remains largely untouched however we would like to observe a more dramaticized background transformation.

Considering the shortcomings seen in the examples, there are several possibilities for improvement.

The first involves our style of mask. As mentioned, currently wise choices must be made in order for the foreground and new weather to contain the same ambiance. If the new sky is too dark, the foreground will be too light, and there will be mismatch. To counter this, experiments could be made with the 0 and 1 nature of the mask. If these were changed into specific coefficients between 0 and 1, with the sky being close to 1 and the foreground being close to 0, perhaps some of the ambience of the lighting may also be moved to the foreground without actually changing the content of the image.

A second addition that might similarly be helpful could be the addition of a ternary mask rather than a binary one. This mask might include a foreground, background, and midground region, and each of these may have different coefficients as mentioned above, such that there are layers of the image which should have a different intensity of transfer. Additionally, perhaps these regions might have a gradient going from each to the other to allow for smoother transition between the regions. Other possibilities in this regard are to move back towards multiple classifications of the image, such as sky, water, city, etc, as is done by Luan *et al.*, but targeting classes which specifically have to do with weather such as clouds, sky, sun, city, etc. This, plus the addition of a segmentation mask for the style image as well could yield significant improvements to the current implementation.

Contributions

Juliet and Nicholas worked together on project idea and possible implementations. Juliet completed the addition of the segmentation mask to the original Neural Style Transfer Algorithm, while Nicholas headed the addition of Matting, and completed it with Juliet's help. Both worked together on the paper, but Juliet focused on the Intro, Related Work, Dataset and Features, and the first two parts of Methods, while Nicholas focused on the Abstract, the last part of Methods (Matting), Experiment/Results/ Discussion, the Figures, Conclusion/Future Work, and References.

Our code can be found in the following GitHub Repo:
julietokwara/Photo-Weather-Transfer-Comparing-Neural-Style-and-Deep-Neural-Style-Transfer

References

[1] A. J., "Semantic Style Transfer and Turning Two-Bit Doodles into Fine Artworks," *[astro-ph/0005112] A Determination of the Hubble Constant from Cepheid Distances and a Model of the Local Peculiar Velocity Field*, 05-Mar-2016. [Online]. Available: https://arxiv.org/abs/1603.01768. [Accessed: 16-Dec-2018].

[2] Anish Athalye. Neural Style. https://github.com/anishathalye/neural-style. 2015.

[3] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Computer Graphics and Applications*, vol. 21, no. 4, pp. 34–41, 2001.

[4] F. Luan, S. Paris, E. Shechtman, and K. Bala, "Deep Photo Style Transfer," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[5] L. A. Gatys, A. S. Ecker, and M. Bethge, "A Neural Algorithm of Artistic Style," *Cornell University Library*, 02-Sep-2015. [Online]. Available: https://arxiv.org/abs/1508.06576. [Accessed: 13-Dec-2018].

[6] Luan, Fujun and Paris, Sylvain and Shechtman, Eli and Bala, Kavita. Deep Photo Style Transfer. https://github.com/luanfujun/deep-photo-styletransfer. 2017.

[7] P.-Y. Laffont, Z. Ren, X. Tao, C. Qian, and J. Hays, "Transient attributes for high-level understanding and editing of outdoor scenes," *ACM Transactions on Graphics*, vol. 33, no. 4, pp. 1–11, 2014.

[8] Smith, Cameron. neural-style-tf. https://github.com/cysmith/neural-style-tf. 2016.

[9] Yang Liu. deep-photo-style-transfer-tf. https://github.com/LouieYang/deep-photo-styletransfer-tf. 2017.

[10] Y. Shih, S. Paris, F. Durand, and W. T. Freeman, "Data-driven hallucination of different times of day from a single outdoor photo," *ACM Transactions on Graphics*, vol. 32, no. 6, pp. 1–11, Jan. 2013.