Neural Weather Transfer Using Segmentation & Matting Laplacian

Juliet Okwara (jokwara@stanford.edu)
Nicholas Seay (nseay@stanford.edu)

Abstract and Motivation

- Photographic style transfer struggles between remaining faithful to realism and the content picture while balancing being a accurate representation of the style image.
- Existing techniques, like Neural Style Transfer by Gatys et al, largely distort the content image, creating a more artistic rather than contextually faithful result.
- The goal of this project is to create a model for transferring weather elements and general ambience of a given style photo into the content image with as little foreground distortion as possible.
- We used semantic image segmentation and matting laplacian to localize style transfer to the sky. This resulted in images with fewer distortions and artifacts appearing in the foreground and background.

Dataset and Features

Data

- We utilized a pretrained VGG19 network.
- 7 RGB style images were paired with 3 RGB content images.
- All images selected were photographs. In each style image, there is some observable weather/ambiance aspect to be transfered (storm clouds, blue skies, sunsets, etc). Content photos were selected to have a large background region for the style image to be transferred on to.

Features

- Raw input: RGB values of each pixel for both content and style images
- Segmentation Mask (black and white) for content image

semantic segmentation mask

Model

Base Model

The base of our model is the Gatys et al's **Neural Style Transfer Algorithm**.

- Input:
  - Content Image C
  - Style Image S
- Output:
  - Image with the content of C and artistic style of S
- Loss Functions:

$$\mathcal{L}_{total}(\vec{p},\vec{a},\vec{x}) = \alpha\mathcal{L}_{content}(\vec{p},\vec{x}) + \beta\mathcal{L}_{style}(\vec{a},\vec{x})$$

$$\mathcal{L}_{content}(\vec{p},\vec{x},l) = \frac{1}{2}\sum_{i,j}\left(F_{ij}^{l} - P_{ij}^{l}\right)^2$$

$$\mathcal{L}_{style}(\vec{a},\vec{x}) = \sum_{l=0}^{L} w_l E_l \qquad E_l = \frac{1}{4N_l^2 M_l^2}\sum_{i,j}\left(G_{ij}^{l} - A_{ij}^{l}\right)^2$$

Semantic Image Segmentation:

We incorporated an additional input:a semantic content segmentation mask. The content mask is of the same shape of the content image and contains ones in the pixel locations where the gradient ought to be applied (i.e. the sky) and zeros elsewhere (foreground).

- Loss Functions:
  - Content Loss: We compute the same content loss when conducting segmented style transfer.
  - Style Loss: Using the content input mask as a filter, we performed an element-wise multiplication between the mask and both the content and style image feature matrices.

Matting Laplacian:

We used Martin Benson's implementation for calculating the matting loss. It is used to smooth the image to keep straight lines straight and not curvy/wavy/squiggly.

$$\mathcal{L}_{\text{total}} = \sum_{l=1}^{L}\alpha_\ell\mathcal{L}_c^\ell + \Gamma\sum_{\ell=1}^{L}\beta_\ell\mathcal{L}_{s+}^\ell + \lambda\mathcal{L}_m \qquad \mathcal{L}_m = \sum_{c=1}^{3} V_c[O]^T\mathcal{M}_I V_c[O]$$

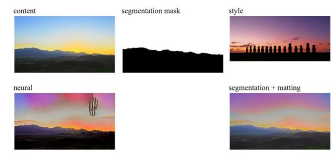
References

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

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

https://github.com/luanfujun/deep-photo-styletransfer
https://github.com/LouieYang/deep-photo-styletransfer-tf
https://github.com/cysmith/neural-style-tf
https://github.com/martinbenson/deep-photo-styletransfer/blob/master/deep_photo.py

Results



Discussions

Our method seems to reduce the amount of mistakes that the original Neural Style Transfer Algorithm (NS) 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. Similarly, there is a mysterious artifact in the top-left corner of the mountain scene of NS.
However, 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 larger than expected effects.

Future Work

We wish to explore more options with semantic image segmentation utilizing masks of the style image in a similarly binary way. We would also like to explore semantic labeling (coloring) similar regions in the style and content image masks to have a more intentional style transfer. Finally, we would like to work on incorporating finer style details as a whole.