Using VGG19 for Neural Style Transfer Learning

Yuqing Chen
Department of Statistics
Columbia University
New York, NY 10027
yc3753@columbia.edu

Levi Lee
Department of Statistics
Columbia University
New York, NY 10027
ll3248@columbia.edu

Yue Liang
Department of Statistics
Columbia University
New York, NY 10027
yl4391@columbia.edu

Michael Petkun
Department of Statistics
Columbia University
New York, NY 10027
mjp2262@columbia.edu

Abstract

The abstract paragraph should be indented ½ inch (3 picas) on both the left- and right-hand margins. Use 10 point type, with a vertical spacing (leading) of 11 points. The word **Abstract** must be centered, bold, and in point size 12. Two line spaces precede the abstract. The abstract must be limited to one paragraph.

1 Introduction

Style transfer from one image to another is considered a difficult image processing task, due to lack of semantic information representation of images. However, the advance on studies of deep convolutional neural networks have provided some ideas on performing such a task. It is shown that convolutional neural networks trained for tasks like object recognition extract high-level image content in generic feature representations, which can be used for the style transfer task.

In this project, we implement a neural style transfer algorithm based on Gatys et al and various Tensorflow tutorials.

Gatys et al (2016) discussed how the VGG19 neural network can be used for neural transfer learning by specific layers used for the process and the hyperparameters that can be changed when rendering a target image. Several tensorflow tutorials and Kaggle guides have outlined the steps to set this process up using Python 3–see tensorflow's "Neural Style Transfer" and "Fast Style Transfer for Arbitrary Styles" as well as Basu's "Style Transfer Deep Learning Algorithm" on Kaggle for more details. We have taken elements and ideas from each of these guidelines to not only replicate the process as described by Gatys et al but also created a tool that allows us to easily switch between different options and hyperparameters, including the choice of seed, style loss, content loss, optimizer, and training steps.

In the spirit of Columbia, we have decided to use the Alma mater statue and Morningside campus as our content images. For our style images, we chose two contrasting images: one is a more traditional notion of art–Jackson Pollock's abstract work No. 5, 1948–and one a more modern-day cartoon–a portrait of Homer J. Simpson, from the popular television show The Simpsons. We feel that this contrast of color, linework, and style help to better understand the general performance of the VGG19 neural network. We also utilize other style images–such as Caravaggio's Supper at Emmaus and Dali's Persistence of Memory as examples that highlight the limitations of the VGG19 neural network.

Preprint. Under review.

A list of references as well as a list of image references are included at the end of this report.

1.1 Content Representation

A layer with N_l distinct filters has N_l feature maps each of size M_l, where M_l is the height times the width of the feature map. So the responses in a layer l can be stored in a matrix $F_l \in \mathcal{R}^{N_l \times M_l}$, where F^l_{ij} is the activation of the i^{th} filter at position j in layer l.

Let $\vec{p}$ and $\vec{x}$ be the original and generated images, and P^l and F^l be their respective feature representations in layer l. The content loss is defined as

$$\mathcal{L}_{\text{content}}(\vec{p}, \vec{x}, l) = \frac{1}{2} \sum_{i,j} (F^l_{ij} - P^l_{ij})^2$$

Convolutional neural networks extract more explicit information of the content of a image at higher layers when trained on object recognition. Therefore, a high-level feature representation is chosen for content representation.

1.2 Style representation

To obtain a representation of the style if an image, the Gram matrix is used to capture the texture information. The Gram matrix $G^l \in \mathcal{R}^{N_l \times M_l}$ consists of correlations between different filter responses, where G^l_{ij} is the inner product between the vectorized feature maps i and j in layer l:

$$G^l_{ij} = \sum_k F^l_{ik} F^l_{jk}$$

Let $\vec{a}$ and $\vec{x}$ be the original and generated images, and A^l and G^l be their respective style representations in layer l. The contribution of layer l to the style loss is defined as

$$E_l = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} (G^l_{ij} - A^l_{ij})^2$$

and the total style loss is

$$\mathcal{L}_{\text{style}}(\vec{a}, \vec{x}) = \sum_{l=0}^{L} w_l E_l$$

where w_l are weighting factors of the contribution of each layers.

2 Methods

3 Results

3.1 Main Takeaways

In this section, we show four images generated by the algorithm described above, combining the content images and the style images. The images in this section were all generated using the content image to initialize the target, running 100 training steps with alpha = 1 and beta = 1e-14, but we will vary these parameters in later sections.

The first image combines the content of the Alma Mater statue picture with the Homer Simpson style. As we can see below, the image now shows Alma Mater with the crisp black outlines and mostly white yellow color palette of the Homer Simpson image.

The second image combines the content of the Alma Mater statue picture with Jackson Pollock's style. Here, Alma Mater is rendered with the darker colors and disorderly (lack of) pattern for which Pollock is known.

The third image combines the content of the Columbia campus picture with the Homer Simpson style. Again, we see the crisp outlines and bright white yellow colors from this distinctive style.

The fourth image combines the content of the Columbia campus picture with Jackson Pollock's style. As we would expect, here the campus looks much darker and less orderly.

3.2 Varying the Number of Training Iterations

In this section, we explore the evolution of the learning process after different numbers of training iterations (10, 25, 50, 75, and 100 steps).

For example, in the Alma Mater + Homer Simpson combination, we can see a clear progression of the style seeping into the content (recall that we are still initializing the target image using the content image). After 10 steps, the color palette has barely begun to transform. After 25 steps, the black outlines are forming and some of the white spaces are getting whiter. This process continues, adding the distinctive yellow coloring as well, as we go through more training steps. Between the 75-step and 100-step images, the returns are diminishing, with the latter image having only marginally brighter yellows and clearer white spaces.

3.3 Varying the Target Image Initialization

In this section, we explore different ways of initializing the target image. We compare initializing with the content image to initializing with random white noise, and compare different random seeds for the white noise.

For the Alma Mater + Homer Simpson combination, initializing with the content image gives us a much clearer picture of the Alma Mater statue. While initializing with random white noise still gives us a recognizable pattern, the outline is not nearly as crisp. Using different random seeds gives us different-looking images, though unsurprisingly neither is observably "better" than the other.

3.4 Varying the Content vs. Style Loss Weights

In this section, we explore how the output images look when we use different combinations of content loss (alpha) and style loss (beta).

For the Alma Mater + Homer Simpson combination, there are clear differences between these three parameter choices.

With beta = 1e-13, many of the statue's features have been washed away, with only the basic outlines remaining. Since this is the example where the style loss receives the highest weight, it is unsurprising that this image looks the most "cartoony".

With beta = 1e-14, we see more of the statue's finer features, and the yellow white color palette is not as sharp.

With beta = 1e-15, we see very little of the Homer Simpson style. This weighting seems to tilt too heavily toward the content loss.

4 Discussion

4.1 Main Takeaways

This project is a visual example of the power and limitations that neural transfer learning can provide. Being able to successfully train and select the right hyperparameters for a convolutional neural network gives rise to a unique approach to art and design.

Graphic designers could potentially find a new rival or ally in neural networks. Social media giants such as Facebook and Snapchat can utilize this technology to expand their filter options. The world of digital art–or art in general–could find its next major movement.

4.2 Commentary on Other Style Images

Beyond the examples shown above, we explored many other artistic styles, some of which were more successful than others in blending with the content images. Based on this experimentation, we noticed an interesting pattern: The style images that worked best for these combinations tended to be those with distinctive color palettes (e.g. Homer Simpson's bright white and yellow, or Jackson Pollock's dreary browns) and recognizable characteristics (e.g. Homer Simpson's sharp borders, or Jackson Pollock's random splatter). On the other hand, the style images that did not work as well tended to be those with a greater variety of shades and more diversity of shapes within the composition.

In our paper, we will show and discuss some examples of style images that did not work well.

4.3 Further Research

Although our work is closely aligned with Gatys et al and various tensorflow tutorials but with distinct changes in implementation, we feel that there are many more changes worth considering in future work. One challenge we have seen–which will be discussed in detail in our paper– is how certain style images do not work well with our current implementation. There are many possible reasons for this: (1) we need to train for a greater number of steps that goes beyond the memory and computing power we currently have access to (2) we have yet to discover the right combination of current hyperparameters within the range of training steps we are currently capable of (3) the choice of layers of in the vgg19 neutral network should be different from our current choice, and (4) we need an entirely different neural network different from vgg19 altogether.

Additionally, we also find difficulty in quantifying exactly which rendered image is arguably better based on changing the hyperparameters alone. However, this may not even be necessary as the issue is more a matter of opinion and artistic preference. If this were to be used in application, the "best" result would often be dictated by one's intentions, the context in which the output is used, and the audience that is viewing the output.

Broader Impact

Acknowledgements and Disclosure of Funding

References

[1] Abadi, Martín, et al. "Tensorflow: Large-scale machine learning on heterogeneous distributed systems." arXiv preprint arXiv:1603.04467 (2016).

[2] Basu, Victor. "Style Transfer Deep Learning Algorithm." Kaggle, Kaggle, 26 Apr. 2019, www.kaggle.com/basu369victor/style-transfer-deep-learning-algorithm.

[3] "Fast Style Transfer for Arbitrary Styles." TensorFlow, www.tensorflow.org/hub/tutorials/tf2$_{a}rbitrary_{i}mage_{s}tylization.$

[4] Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge. "Image style transfer using convolutional neural networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[5] "Neural Style Transfer." TensorFlow, www.tensorflow.org/tutorials/generative/style$_{t}ransfer.$

[6] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

[7] Surma, Greg. "Style Transfer." Kaggle, Kaggle, 13 Jan. 2019, www.kaggle.com/greg115/style-transfer.

Image References

[1] French, Daniel Chester, and Beyond My Ken. "2014 Columbia University Alma Mater Closeup." Wikimedia Commons, Wikimedia Commons, 1 Jan. 2015, commons.wikimedia.org/wiki/File:2014$_{C}olumbia_{U}niversity_{A}lma_{M}ater_{c}loseup.jpg.$

[2] Caravaggio, Michelangelo Merisi da, and National Gallery, London. "1602-3 Caravaggio, Supper at Emmaus National Gallery, London." Wikimedia Commons, Wikimedia Commons, 28 Mar. 2018, commons.wikimedia.org/wiki/File:1602-3$_{C}aravaggio, Supper_{a}t_{E}mmaus_{N}ational_{G}allery,_{L}ondon.jpg.$

[3] “Columbia University Morningside Campus.” Columbia University, Columbia University, images.app.goo.gl/i5ZkryXru6Sjav1t6.

[4] Dali, Salvador, and Xennex. “La Persistencia De La Memoria.” WikiArt, WikiArt, 13 Aug. 2020, www.wikiart.org/en/salvador-dali/the-persistence-of-memory-1931.

[5] “Homer Simpson Portrait.” NY Daily News, NY Daily News, interactive.nydailynews.com/2016/05/simpsons-quiz/.

[6] Pollock, Jackson. “Number 5, 1948.” Jackson Pollock Biography, Paintings, and Quotes, Www.Jackson-Pollock.org, www.jackson-pollock.org/number-5.jsp.