Application of Style Transfer Using Neural Networks on Daily Photographs

Yi Li
Harvard University
Cambridge, MA
yli23@hsph.harvard.edu

Yiyun Zhang
Massachusetts Institute of Technology
Cambridge, MA
yiyunz@mit.edu

Manny Favela
Massachusetts Institute of Technology
Cambridge, MA
mfavela@mit.edu

I. Introduction

Convolutional neural networks (CNN) have been widely used in image tasks, such as image recognition, image classification and image segmentation. Research has shown that CNN trained with sufficient labeled data on object recognition tasks are able to extract high-level image contents in generic feature representations that generalize across datasets (Donahue et al., 2014). In lower layers, only the reconstruction of the pixels of the original image is produced. In higher layers, CNN develop a feature representation of an image that contains increasingly explicit contents of the original image along the hierarchy (Gatys, Ecker, & Bethge, 2015b).

Gatys et al. proposed neural style transfer (NST), where they developed an artificial system based on a deep neural network that creates artistic images of high perceptual quality that combine the content of a daily photograph with the appearance (more specifically, texture style) of well-known artworks paintings. (Gatys, Ecker, & Bethge, 2015a, 2016). The key finding of Gatys et al.'s work is that the representations of content and style are separable, which enables us to manipulate both representations individually and simultaneously to generate a synthesized picture.

In this paper, we explore the application of NST, transfering the world-famous artworks' style to our daily photographs. We use VGG-19, a pre-trained neural network on ImageNet, as proposed in Gatys et al.'s paper. The output image is created by findinging an image that preserves the content of the daily photograph we choose and the style of the artwork painting.

II. Methods

We feed a daily photograph (content image), and a painting (style image) to the CNN. To learn the features of the content image, network activations are sampled at a late convolution layer of the VGG-19, which is composed of 16 convolutional and 5 pooling layers. To learn the features of the artwork, network activations are sampled at early to middle layers, which are encoded into a Gram matrix representation. An ideal output image should maintain the content in the photograph we choose, but also incorporate the artistic style from the painting.

In the training phase, we first generate a noisy image by adding some random noise to the content image, and then set the input image to be a weighted average of the content image and the noisy image. We then reshape and normalize both the content image and the style image. In content (style) reconstructions, we perform gradient descent on a white noise image to find another image that matches the feature (style) representation of the original image. The total loss is

$$L_{total}(p, a, x) = \alpha \cdot L_{content}(p, x) + \beta \cdot L_{style}(a, x),$$

where $L_{content}$ is the squared-loss between two feature representations from the original image and from the generated image; L_{style} is the mean-squared distance between the entries of the Gram matrix from the original style image and from the generated image; p, a, x are the photograph, the artwork and the generated image, respectively; α and β are the weighting factors for content and style reconstruction, respectively. We use backpropagation to update the weights in each layer. Since the two loss functions are separate, we can generate a synthesized picture focused on either content or style by changing the weights.

III. Preliminary results

We use an MIT photograph (Figure 1) as the input content image, and Vincent van Gogh's *starry night* (Figure 2) as the input style image. The generated output Figure 3 is after 200 iterations. We see that the synthesized output picture preserves the content of the MIT photo, but also successfully transfers the artistic style.

Fig. 1: An MIT photo (Dizikes, 2020)

Fig. 2: *The Starry Night* (van Gogh, 1889)

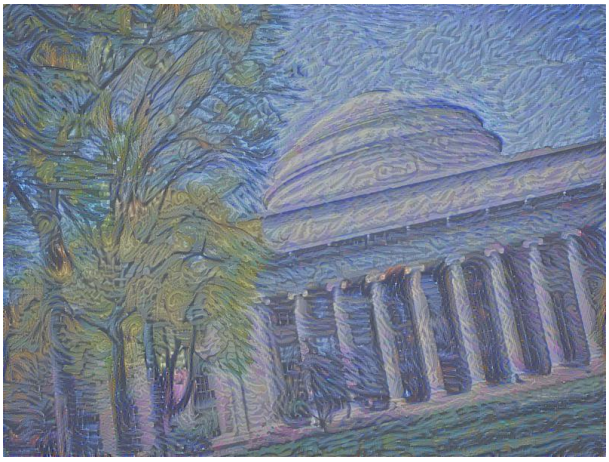

Fig. 3: Generated Output

The input, output images and python code can be found at: `https://github.com/einsley1993/cnn-style-transfer`.

IV. Next steps

Our preliminary results show that NST gives a promising result in creating a synthezied picture that combines the content of the phorograph and the style of the artwork. In the main project, we plan to test more painting styles, including Claude Monet, Pablo Picasso, and even contemporary artists, and evaluate the performance of NST. It has been found that NST works well with most impressionist paintings, and so we would like to see it in other types of artwork as well. Although NST is mainly focused on texture transfer, we also plan to explore color transfer. Moreover, we will experiment with feed-forward networks developed by Johnson et al. They use a VGG-16 model pretrained on ImageNet (Johnson, Alahi, & Fei-Fei, 2016). We will compare its performance to that of Gatys et al.'s VGG-19.

References

Dizikes, P. (2020). *Mit announces plans for fall 2020 semester.* (`https://news.mit.edu/2020/plans-fall-2020-semester-0707`)

Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., & Darrell, T. (2014). Decaf: A deep convolutional activation feature for generic visual recognition. In *International conference on machine learning* (pp. 647–655).

Gatys, L. A., Ecker, A. S., & Bethge, M. (2015a). A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576*.

Gatys, L. A., Ecker, A. S., & Bethge, M. (2015b). Texture synthesis using convolutional neural networks. *arXiv preprint arXiv:1505.07376*.

Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 2414–2423).

Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision* (pp. 694–711).

van Gogh, V. (1889). *The starry night.* (`https://en.wikipedia.org/wiki/The\_Starry\_Night`)