Image Style Transfer using Deep Neural Network

Jayabrata Chowdhury
Robert Bosch Centre for Cyber Physical Systems
Indian Institute of Science
Bangalore, India
jayabratac@iisc.ac.in

Abstract

The style transfer is a process of modifying the style of an image while still preserving its content.A deep learning based approach to artistic and photo realistic style transfer handles a large variety of image content while faithfully transferring the reference style. In literature a Neural Algorithm of Artistic Style transfer[**1**] is introduced that can separate and recombine image content and style of artistic image. But this neural style transfer does not work well for photo realistic style transfer.To address this in Deep Photo Style Transfer[**2**], transformation from input to output has been made locally affine in colorspace. This approach successfully suppresses distortion and gives satisfying photorealistic style transfers in a broad variety of scenarios.

1 Introduction

Transferring style is a problem of image texture transfer in image. In texture transfer the goal is to synthesise a texture from a input image while constraining the texture synthesis in order to preserve the semantic content of a target image. Previous texture transfer algorithms rely on non parametric techniques for synthesis of texture while using different ways preserve structure of target image.For example,some techniques use image analogies to transfer the texture from an already stylised image onto a target image**[3]**. All previous algorithms suffers from low level image features of target image. Deep convolutional neural networks are better in extracting high level features from images. In **[1]**, a Neural algorithm for artistic style transfer is proposed. Actually this algorithm transfer the texture of style image to content image like an art.
Now there is a problem with artistic style transfer algorithm. It does not work very well for photo realistic style transfer. There can be distortions in target image which are not acceptable for realistic style transfer. This problem is addressed in **[2]**. In **[2]**, they have prevented painting like distortions by transfer operation to be happened only in color space. Also they have shown the use semantic segmentation to address the issue of difference in content between reference and input photos.

2 Methods

2.1 Neural Image Style Transfer

This algorithm uses VGG-19 network **[4]** for image content and style representation. The VGG network is normalized by scaling the weights such that mean activation of each convolutional filter over images and positions is equal to one.

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

2.1.1 Content representation of neural style transfer

Let **p** and **x** are original image and image that is generated. P^l and F^l are their respective feature representation in layer l. A squared-error loss is defined between the two feature representations:

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

The gradient of this loss with respect to activation is used then for standard back propagation. The feature responses in higher layers of CNN is referred to content representation of image since higher layers are sensitive to actual content.

2.1.2 Style representation of neural style transfer

Feature space is used to find representation of image. Feature space consists of correlation between the different filter responses. Gram matrix, where G^l_{ij} is the inner product between vectorised feature maps i and j in layer l gives correlation.

$$G_{ij}^{l} = \sum_{k} F_{ik}^{l} F_{jk}^{l} \tag{2}$$

Let **a** and **x** be the original image and the image that is generated, and A^l and G^l their respective style representation in layer l. The contribution of layer l to the total loss is then

$$E_l = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} (G_{i,j}^{l} - A_{i,j}^{l})^2 \tag{3}$$

and total style loss is

$$\mathcal{L}_{style}(\mathbf{a}, \mathbf{x}) = \sum_{l=0}^{L} w_l E_l \tag{4}$$

where N_l is no of feature maps and M_l is the height times the width of feature map and w_l is weight of contribution in loss.

2.1.3 Total loss of neural style transfer

For total style transfer, loss function is

$$\mathcal{L}_{total}(\mathbf{p}, \mathbf{a}, \mathbf{x}) = \alpha \mathcal{L}_{content}(\mathbf{p}, \mathbf{x}) + \beta \mathcal{L}_{style}(\mathbf{a}, \mathbf{x}) \tag{5}$$

where α and β are weighting factors of content and style representation respectively. Gradient of total loss is used for optimization.

2.2 Deep Photo Style Transfer

This method introduces two new ideas for photo realism.

- A regularization term is introduced in loss function during optimization. It is constrained to prevent image distortions.
- It introduces semantic mapping to transfer technique for avoiding content mismatch problem.

2.2.1 Regularization for photorealism

It tries to find an image transform that is locally affine in color space. It uses **M**atting Laplacian of Levin **[5]**. It describe a least-squares penalty function that can be minimized with a standard linear system represented by a matrix M_I that only depends on the input image I. M_I is N X N where N is no of pixels in input image I. $V_c[O]$ is vectorized version of output image O in channel c. The regularization term is

$$\mathcal{L}_m = \sum_{c=1}^{3} V_c[O]^T M_I V_c[O] \tag{6}$$

2.2.2 Augmented style loss with semantic segmentation

Gram matrix computes exact distribution of neural response. This can cause inability to adapt to variations of semantic context. Neural Doodle [**6**] and semantic segmentation [**7**] is used. Segmentation masks are added to input image as additional channels.So new style loss is

$$\mathcal{L}_{s+}^{l} = \sum_{c=1}^{C} \frac{1}{2N_{l,c}^{2}} \sum_{i,j} (G_{l,c}[O] - G_{l,c}[S])_{ij}^{2} \tag{7}$$

where C is number of semantic segmentation masks. $M_{l,c}[.]$ is the channel c of the segmentation mask in layer l, and $G_{l,c}[\cdot]$ is Gram matrix corresponding to $F_{l,c}[\cdot]$.

2.2.3 Total loss

Total loss is combination of all 3 components.

$$\mathcal{L}_{total} = \sum_{l=1}^{L} \alpha_l \mathcal{L}_c^l + \Gamma \sum_{l=1}^{L} \beta_l \mathcal{L}_{s+}^{L} + \lambda \mathcal{L}_m \tag{8}$$

where L is no of convolutional layers. Γ controls the style loss. αl and β_l are the weights to configure layer preferences. λ is a weight that controls the photorealism regularization.

3 Implementation

3.1 Neural Image Style Transfer

In this algorithm, content is represented by conv4-2 layer and style is represented by conv1-1, conv2-1, conv3-1, conv4-1, conv5-1 layers of VGG19 network. In the paper, authors initialized gradient descent with white noise. But in my implementation I initialized gradient descent with content image.

3.2 Deep Photo Style Transfer

Here same layers are used for content and style representation as Neural style transfer. Γ=10^2 and λ=10^4 is used as in the paper. Due to unavailability of CUDA, I used CPU for implementation. It took approximately 2 hours to find result for a single image. Matting Laplacian is also implemented in python.

4 Result

4.1 Neural Style Transfer

I have tried different weights to content and style image. For α/β=2 and α/β=100 here are images.Left one is content image. Middle one is style image. Right one is output image.

Figure 1: Neural style transfer for α/β=2

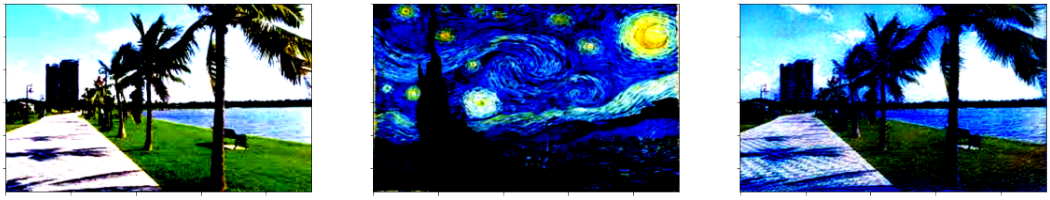

Figure 2: Neural style transfer for α/β=100

Figure 3: Neural style transfer for α/β=100

4.2 Deep Photo Style Transfer

In deep photo style transfer I have tried with the regularization term. Images are shown bellow:

Figure 4: Input image

Figure 5: Style image

Figure 6: Transformed image

5 Discussion

These two techniques are very good. Deep photo style transfer[**2**] improves on the problems of neural style transfer[**1**]. But in deep photo style transfer, some style images result in failure cases. In Visual attribute transfer through deep image analogy[**8**], this uses semantic style transfer. Also temporarily coherent style transfer is used for video style transfer. The main object is semantically meaningful results for style transfer. It is useful with input image pairs that may look completely different visually but have some semantic components that are similar. One common application is regular photo to style transfer. Another interesting application is color transfer between photographs which will allow creating time lapse video. Universal Style Transfer via Feature Transforms[**9**] is much more robust to any kind of style transfer. This new algorithm does not need to be trained on a set of style images.Here an autoencoder is trained for image reconstruction. The encoder part finds highly compressed representation of an image. Then reconstruction of full image occurs from compressed representation using decoder network. In paper, they split autoencoder in half and use encoder part on both input image and style image. Also some mattes can be created for target image and different artistic styles to different part. These are the further ways to improve transformed image quality and work is on the way.

References

[1] L. A. Gatys, A. S. Ecker and M. Bethge, "Image Style Transfer Using Convolutional Neural Networks," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 2414-2423, doi: 10.1109/CVPR.2016.265.

[2] F. Luan, S. Paris, E. Shechtman and K. Bala, "Deep Photo Style Transfer," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 6997-7005, doi: 10.1109/CVPR.2017.740.

[3] A. Hertzmann, C. E. Jacobs, N. Oliver, B. Curless, and D. H. Salesin. Image analogies. In Proceedings of the 28th annual conference on Computer graphics and interactive techniques, pages 327–340. ACM, 2001.

[4] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition." International Conference on Learning Representations, 2015.

[5] A. Levin, D. Lischinski and Y. Weiss, "A Closed-Form Solution to Natural Image Matting," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 2, pp. 228-242, Feb. 2008, doi: 10.1109/TPAMI.2007.1177.

[6] S. Bae, S. Paris, and F. Durand. Two-scale tone management for photographic look. In ACM Transactions on Graphics (TOG), volume 25, pages 637–645. ACM, 2006.

[7] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. arXiv preprint arXiv:1606.00915, 2016.

[8] J.Lao, Y.Yao, L.Yuan, G.Hua, S.B.Kang, Visual Attribute Transfer through Deep Image Analogy. *arXiv preprint arXiv: 1705.010882v2*

[9] Y.Li, C.Fang, J.Yang, Z.Wang, X.Lu, M.H.Yang. Universal Style Transfer via Feature Transforms. 31st International Conference on Neural Information Processing Systems, December, 2017.