Neural Style Transfer with Content Feature Segmentation

Xiaotian Fang*
School of Telecommunication, Xidian University, Xi'an, China
* Corresponding Author Email: xtfang@stu.xidian.edu.cn

Abstract. In our work we propose a new model by segmenting content features (i.e., segregating the feature map into several parts) of the content image in style transfer. We demonstrate that our novel way for style transfer with segmentation ties together the style of both global effect and local appearance that are used in majority of parametric and non-parametric techniques respectively. Our new model can successfully figure out the difficult problems with the current transfer techniques. For one thing, the new model in our research is able to avoid the distortion of local style maps, at the same time it also allows transfer in the level of semanteme compared to parametric means. For another thing, in contrast to other nonparametric approaches, it keeps the style image's overall appearance consistent and avoids washing-out images. We also present a feature optimization strategy based on this. Our trials have proven that our approach is suitable to a wide range of style genes and simultaneously yields higher-quality results over earlier papers.

Keywords: Neural Network, Neural Style Transfer, Feature Segmentation.

1. Introduction

Neural (or image) style transfer is a kind of sophisticated technology in computer vision, which stands for the conversion of images in two different domains, specifically by providing a style image and converting any image to this style, and at the same time, storing original image's content as much as possible. These days, academics, industry, and the arts are all interested in neural style transfer for its flexibility in a variety of applications. The methods to achieve neural style transfer are various. Many astonishing outcomes have recently been reported in various forms. However, picture structure representations should be stressed as key components in making a visually pleasing image.

Traditionally, as a texture creation problem, style transfer has been investigated. According to the classification of previous studies on style transfer, it might be divided into two kinds of types. Image-optimization techniques and model-optimization ways. The former approaches perform iterative optimization of stylized images through different networks. At first, Gatys and his researchers [1] pioneered to utilize convolutional neural networks (CNN) in the job of style transformation task and showed that pre-trained classification networks can extract correlations between features, capture style patterns well, and iteratively generate stylized images. The way Gatys et al. [1] applied is capable of combining the information and design of a given image, but it also has limitations in that the optimization process is too slow. Following works improve of [1], Li et al. [2] demonstrated that Gram matrix matching approach is conceptually equal to reducing Maximum Mean Discrepancy (MMD) in NST through the polynomial kernel. E. Risser et al. [3] used histogram losses to synthesize textures. Style transfer was also proposed by Nicholas et al. [4] using Relaxed Optimal Transport and Self-Similarity (STROTSS), which can both achieved a particular visual effect. Image-optimization approaches, on the other hand, suffer from a sluggish online optimization process, also their difficulties are essentially unsolved. Meanwhile, model-optimization approaches reconstitute stylized pictures using feedforward networks, which may be divided into three categories. (1) Per-Style-Per-Model approaches [5-8] use a single transfer network to train for each desired style. (2) Multi-Style-Per-Model approach [9-12] is to produce multiple styles of images by training a single transmission network, or even fusing more than one style together. (3) Arbitrary-Style-Per-Model techniques [13-17] can be used to accept a content alone with a style picture as the network's input and execute the transfer in a single feed-forward process. To put it another way, it can not only extract but also apply any style to a picture.

Traditional and neural image synthesis approaches can be roughly characterized as parametric and non-parametric. Parametric approaches, in particular, match global statistics of deep features for neural methods, such as Gram matrices and their approximations, means and variances, and histograms, while nonparametric methods look directly for neural patches that are similar to a given instance. However, to our knowledge, there is few current works that join these neural approaches to provide a supplementary answer.

Parametric models produce results that preserve the image content and the whole appearance of the picture. These strategies, however, may alter local style. These challenges are adequately addressed by neural nonparametric models, however their example matching use greedy optimization, as a result of which the richness of the style maps is reduced and the it also results in washed-out artifacts. This implies that neural nonparametric models based on neural parametric models should include global limitations.

In 2018, MSRA introduced a revolutionary style transfer technique which not only has the benefits from parametric methods but also maintain the advantages of non-parametric approaches. This significant assumption can successfully be testified via deep feature shuffling. This procedure involves changing the placement of patches spatially. For style transfer, they rearrange patches in style picture based on content image. For one thing, feature rearrangement changes the stylized patterns and ensures that they can be distributed consistently at the global level. In theory, rearranging some characteristics of a style picture equals optimizing the Gram matrices. Certain sorts of reshuffling features, in conjunction with non-parametric approaches, can, for another thing, help in the accomplishment of local semantic matching between images.

Figure 1. Some local details from the style image, such as distortions on the ridge and wrong colors on the face, are not preserved by metric style transfer.

Figure 2. Some results may occasionally be less like the style image and repeatedly use the same patches in non-parametric experiments, finally resulting in a bad performance.

Inspired by this method, we noticed that the Gram matrix is a global feature-based statistic, because it is difficult to take into account the details of style transfer in the global context, so it is naturally difficult to achieve successful local transfer using only parametric methods, but if the content image can be segmented into several parts, so that the parametric method no longer needs to take into account the global features of a whole image, but only needs to consider the features of a certain part of the image, and then combine the migrated images. That is, to achieve both global and local style transfer.

Our significant contributions are concluded briefly as follows:

1. Our work provide an unconventional segmentation picture framework for the research of style transfer that enhances the output images' quality via combining global and local structure loss.

2. We demonstrate that visual structure representation is crucial in style transfer, and if design an appropriate image structure representation, more visually appealing stylized results can be produced.

3. Our experiments demonstrate that our approach yields efficient and high-quality stylized results in which the structural patterns are efficiently synthesized, both globally and locally.

Arbitrary Style Transfer with Deep Feature Reshuffle (MSRA):

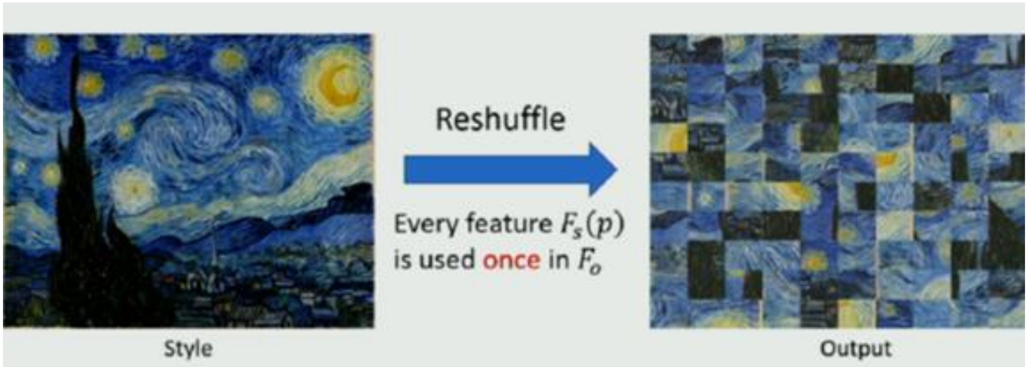

Figure 3. By reshuffling the style images, the output image is made to innately meet the requirements of the non-parametric model.

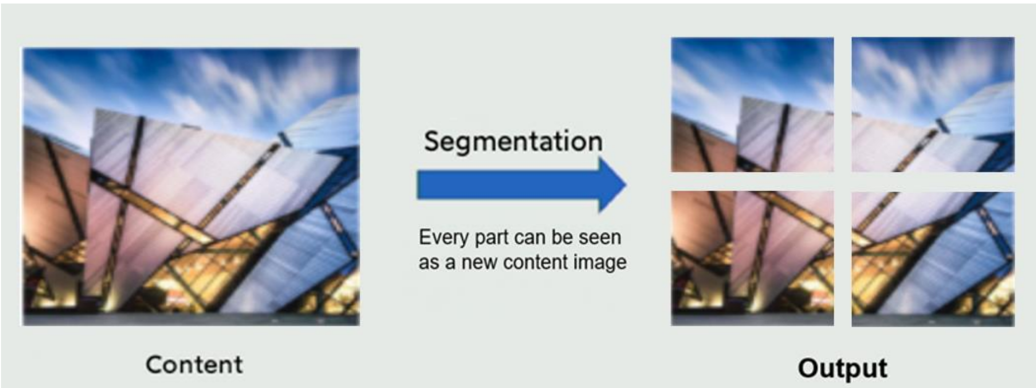

Figure 4. Segment the content image into several small pieces, each small piece individually as a content image for style migration, and finally integrated together.

2. Related work

Arbitrary Style Transfer. The primary and most important pursuit of this work is to utilize only one trained network to transmit pictures of any given style with the best efficiency and generation quality feasible. For style transfer, many prior researches have improved the compromise between the quality and efficiency of synthesized picture. Li and his members [2] have already verified that the batch normalization (BN) statistic representation may be utilized to express the style by simply change the variance and mean of feature maps in a certain layer. Ulyanov et al. [8] discovered that incorporating Instance Normalization (IN) into feedforward texture synthesis and styled networks can increase the quality and diversity of the outputs. Dumoulin and his team released the famous conditional instance normalization (CIN) approach. Its primary concept is trying to discover style patch in feature space that matches content patch, change the content patches for the style patches, and then utilize rapid image reconstruction technique to quickly recreate the feature map produced by the swap. However, due to the time overhead required for style swap, the effect of real-time is not yet achieved. Huang et al. then released the game-changing adaptive instance normalization (AdaIN) method, it successfully changes the variance or the mean in content picture to appropriately suit the mean and variance in the style image by transferring global feature data. His work is mainly inspired by the CIN layer of MSPM. The style and content characteristics collected by VGG are AdaIN's input; the data-driven technique enables AdaIN to immediately normalize the content in a picture by repeatedly training on a tremendous quantity of style images and content graphs. By changing the parameters of the affine transformation in the CIN layer, a different style can be obtained, thus, given an arbitrary style, we only need to know the parameters of the affine transform in its CIN layer. Along this line, Google Brain researchers designed and trained a style prediction network to specifically predict the parameters of the affine transform of each style. The drawback of this approach is also obvious: the data-driven approach’s output effect of style is very reliant on the gene and quantity of styles in the dataset. The methods and models for image style transfer have become more and more powerful, but the generated images still need some improvements and enhancements, among which image structure representation is a very important part. In our work, we offer a novel content segmentation for picture structure representation in arbitrary style transfer.

3. Method

We subsequently introduce a unique feature segmentation-based style transfer technique that blends global and local style loss across the objective function. We begin by proposing a novel style loss, segmentation loss, which will be paired with content loss. The optimization is then demonstrated in each single feature layer. Similar to the " forward and backward network iterations " strategy, such optimization may be performed in the picture domain. To speed up, we take two assumptions.

1) Optimization may be performed in feature domain without repeatedly returning back to the input image.

2) We improve the features in several layers step by step and do exhaustive patch matching in fine layers. It can be led by the coarse layer matching results.

3.1. Segmentation Loss Function

For the style loss, we create a new segmentation loss that only slightly affects the local style loss:

$$\text{Lseg} = \sum_{p} || \boldsymbol{\Psi}_{p}(\boldsymbol{F_o}) - \boldsymbol{\Psi}_{NNC(P)}(\boldsymbol{F_s})||_{F}^{2} \quad (1)$$

a new function replaces the initial closest neighbor (NN) search NN(p), which is NNC(p). Perform layer-by-layer optimization, and decode the image by decoder using image features at the second layer.

The architecture of the network in our proposed method:

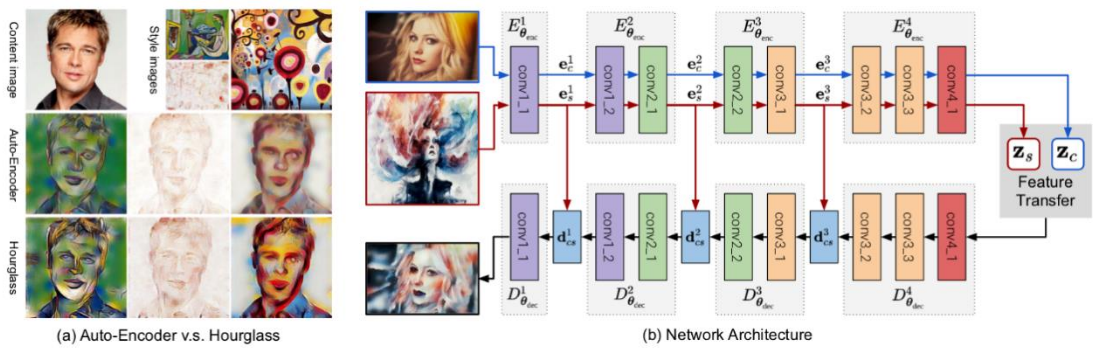

Figure 5. (a) Comparison between Autoencoder and style-augmented hourglass network in stylization. (b) The suggested method's network design.

3.2. Optimization of a Single Layer

The single layer objective function is defined as:

$$\boldsymbol{L_{total}} = \alpha \boldsymbol{L_{cont}} + (1\text{-}\alpha) \ \boldsymbol{L_{seq}} \tag{2}$$

The simplest way is to optimize in the picture domain. Initially, we assume that the output picture Io is either the content image or the random noise. These feature maps or feature images are then acquired through VGG-19 network (pre-trained in famous ImageNet). We optimize NNC by using the feature map, then compute the gradient of our energy function Ltotal and back propagate it to update Io. To converge by using the gradient decent approach, such processing always requires hundreds of iterations.

4. Experiment

Figure 6. Synthesized image comparison before and after segmentation.

As you can see, after segmentation, the synthesized picture's details are more prominent, the style image's features are also better preserved, and the style transfer is more successful, both globally and locally. We illustrate the suggested method's cutting-edge efficacy and efficiency in creating high-quality styled pictures through a range of successful applications, such as multi-style integration, video stylization, and so on.

4.1. Segmentation Loss Analysis

Our work investigates different experiment results of various style loss functions, such as global loss, local loss, and segmentation loss, while excluding common content loss. As a result, they are solely evaluated in texture synthesis. We gathered photos from pervious articles. Then began with random noise in order to minimize the three losses individually.

Figure 7. shows some of the results. We can observe that the outputs obtained via reducing global loss can preferably recreate the comprehensive sense of the input style image, whilst the outputs generated via reducing local loss can be more accurate to the local form of the sample texture. There are two advantages of using our segmentation loss: global resemblance and local plausibility. This can also be demonstrated by a quantitative comparison, just as shown in Figure 7. As we can see, our loss function results in a smaller global loss than the non-parametric technique; it can also achieve a smaller local loss than the parametric technique, and in all circumstances, the local loss is much lower.

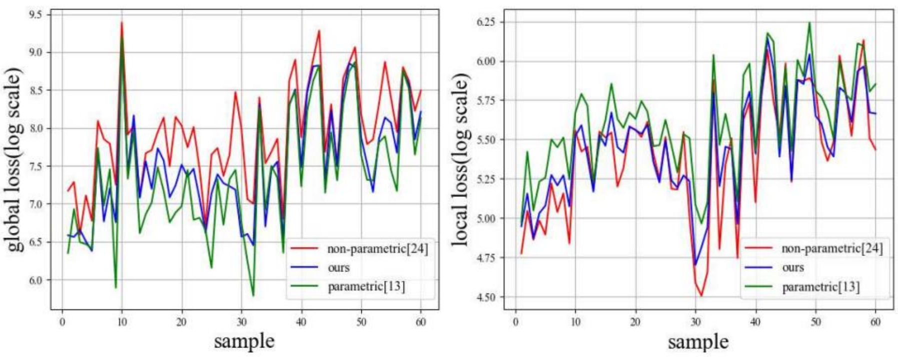

Figure 7. The comparison of the final convergence of different loss function.

4.2. Selection of Patch Size

The patch size is another technique hyperparameter. To some extent, enlarging the patch size may compensate for the loss of global style. In other circumstances, however, a bigger patch size is required to ensure spatial consistency. The bigger the patch size, the more of the style pattern's local structure is kept. Another consideration is that when it comes to some coarser layers of the network, these patches are likely to encompass greater reception fields, which will possibly make accurate matching much harder.

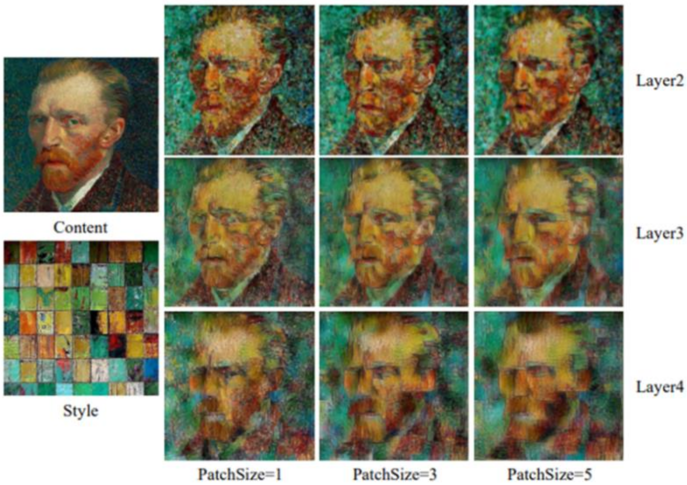

Figure 8. An instance of the transfer outputs from distinct layers when patch sizes are various.

4.3. Comparison results

We compare some of our results to those of existing neural style transfer approaches, both representative parametric and non-parametric. All of our results are created using set parameters for fair comparison. And we obtain them by executing author-released codes and programs with the default parameters.

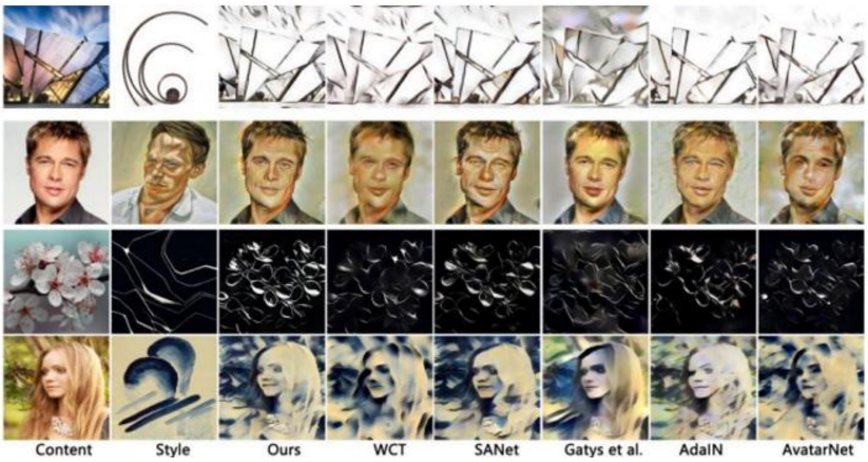

Figure 9. Comparison of transfer outcomes from several models.

5. Conclusion

We present a new feature-segmented picture framework in arbitrary style transfer driven by a mix of global and local structure loss in this paper. We effectively apply arbitrary creative effects in a trained network and at the same time, maintaining the image's structure. The results of our experiments reveal that our strategy is both successful and efficient. In comparison to other parametric methods, our model can preserve more content structural information while still creating outstanding aesthetic effects.

However, the method still has some limitations, especially when the color distribution of stylized images is not uniform. How to improve the model so that it can be more widely used will be an important practical issue to be explored in future work.

References

[1] Gatys, L.A., Ecker, A.S., Bethge, M.: Image style transfer using convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2414–2423 (2016)

[2] Li, Y., Wang, N., Liu, J., Hou, X.: Demystifying neural style transfer.arXiv preprint arXiv:1701.01036 (2017)

[3] Risser, E., Wilmot, P., Barnes, C.: Stable and controllable neural texture synthesis and style transfer using histogram losses.arXiv preprint arXiv:1701.08893 (2017)

[4] Kolkin, N., Salavon, J., Shakhnarovich, G.: Style transfer by relaxed optimal transport and self-similarity. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10051–10060 (2019)

[5] Johnson, J., Alahi, A., Fei-Fei, Li.: Perceptual losses for real-time style transfer and super-resolution. In: European Conference on Computer Vision, pp. 694–711 (2016). Springer

[6] Ulyanov, D., Lebedev, V., Vedaldi, A., Lempitsky, V.S.: Texture networks: Feed-forward synthesis of textures and stylized images. In: ICML, vol.1, p.4(2016)

[7] Ulyanov, D., Vedaldi, A., Lempitsky, V.: Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6924–6932 (2017)

[8] Li, C., Wand, M.: Precomputed real-time texture synthesis with Markovian generative adversarial networks. In: European Conference on Computer Vision, pp. 702–716 (2016). Springer

[9] Dumoulin, V., Shlens, J., Kudlur, M.: A learned representation for artistic style. arXiv preprint arXiv:1610.07629 (2016)

[10] Chen, D., Yuan, L., Liao, J., Yu, N., Hua, G.: Stylebank: An explicit representation for neural image style transfer. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.1897–1906 (2017)

[11] Li, Y., Fang, C., Yang, J., Wang, Z., Lu, X., Yang, M.-H.: Diversified texture synthesis with feed-forward networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.3920–3928 (2017)

[12] Zhang, H., Dana, K.: Multi-style generative network for real-time transfer. In: Proceedings of the European Conference on Computer Vision (ECCV) Workshops, pp. 0–0 (2018)

[13] Park, D.Y., Lee, K.H.: Arbitrary style transfer with style-attentional networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5880–5888 (2019)

[14] Huang, X., Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1501–1510 (2017)

[15] Li, Y., Fang, C., Yang, J., Wang, Z., Lu, X., Yang, M.-H.: Universal style transfer via feature transforms. Advances in neural information processing systems 30 (2017)

[16] Ghiasi, G., Lee, H., Kudlur, M., Dumoulin, V., Shlens, J.: Exploring the structure of a real-time, arbitrary neural artistic stylization network. arXiv preprint arXiv:1705.06830 (2017)

[17] Sheng, L., Lin, Z., Shao, J., Wang, X.: Avatar-net: multi-scale zero-shot style transfer by feature decoration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8242–8250 (2018)