

Article

# The Intra-Class and Inter-Class Relationships in Style Transfer

Xin Cui, Meng Qi \* , Yi Niu and Bingxin Li

School of Information Science and Engineering, Shandong Normal University, Jinan 250014, China; twentycui@gmail.com (X.C.); niuyi@sdu.edu.cn (Y.N.); lbxsdnu@163.com (B.L.)

\* Correspondence: qimeng0914@gmail.com

Received: 13 August 2018; Accepted: 13 September 2018; Published: 17 September 2018



**Abstract:** Neural style transfer, which has attracted great attention in both academic research and industrial engineering and demonstrated very exciting and remarkable results, is the technique of migrating the semantic content of one image to different artistic styles by using convolutional neural network (CNN). Recently, the Gram matrices used in the original and follow-up studies for style transfer were theoretically shown to be equivalent to minimizing a specific Maximum Mean Discrepancy (MMD). Since the Gram matrices are not a must for style transfer, how to design the proper process for aligning the neural activation between images to perform style transfer is an important problem. After careful analysis of some different algorithms for style loss construction, we discovered that some algorithms consider the relationships between different feature maps of a layer obtained from the CNN (inter-class relationships), while some do not (intra-class relationships). Surprisingly, the latter often show more details and finer strokes in the results. To further support our standpoint, we propose two new methods to perform style transfer: one takes inter-class relationships into account and the other does not, and conduct comparative experiments with existing methods. The experimental results verified our observation. Our proposed methods can achieve comparable perceptual quality yet with a lower complexity. We believe that our interpretation provides an effective design basis for designing style loss function for style transfer methods with different visual effects.

**Keywords:** style transfer; loss function; convolutional neural network

## 1. Introduction

Transferring the look and feel (style) of one image to another image is an interesting but difficult problem. Many researchers are committed to the study of the automatic style transfer, and have developed many efficient and excellent methods [1–3]. Recently, Gatys et al. proposed a pioneering work which captures the artistic style of the images and transfers it to other images by using the convolutional neural network (CNN) [4]. This work formulated the style transfer problem as trying to create an image which matches both the content and style statistics for a given pair of images which provide content and style information separately based on the neural activations of each layer in CNN. Although this work has attracted lots of attention and demonstrated interesting and remarkable visual results in both research laboratory and industry, the key technology of the work, especially the reason that *Gram matrix* can represent artistic style, remains a mystery.

Recently, Li et al. demystified the Neural Style Transfer algorithm by theoretically showing that matching Gram matrices in neural style transfer can be reformulated as minimizing Maximum Mean Discrepancy (MMD) with the second order polynomial kernel [5]. The authors argued that the *Gram matrix* is not a necessary reference for style transfer, and the neural style transfer is actually a process of distribution alignment of the neural activations between images. Even though this work unlocks

the secret of the *Gram matrix* to a certain extent, no one has studied the relationships between feature maps obtained from the CNN for the statistical distributions of feature maps.

Here, we define the feature map for the same layer obtained from the CNN as a feature map *class*. The relationship between the feature map and itself (or its variants) is called *intra-class relationship*. All computation performed on a particular feature map without the effect of other feature maps are called *intra-class operation*. For example, for a particular feature map  $x$ , computing the mean and standard deviation of  $x$ , or multiplying  $x$  with itself, are all intra-class operations. Similarly, the *inter-class relationship* is defined for the relationship between different feature maps in the same layer, and the *inter-class operation* is defined as the computations performed on different feature maps.

We analyzed the methods proposed in [4,5], and discovered that in the former method, both intra-class and inter-class relationships are considered, while, in the latter, only the intra-class relationships are considered. According to our experimental results, we argue that the latter produces finer strokes and keeps more content details in the generated images. Furthermore, in this paper, we propose two new methods, called *Cov-Matrix* and *Cov-MDE-Matrix* algorithms, to perform style transfer. The former considers both intra-class and inter-class relationships, while the latter only considers the intra-class relationships. We compared our algorithms with the algorithms mentioned above, achieved appealing results and observed a similar trend as mentioned above. We believe that the properties we observed can help researchers to design style transfer methods for different visual demands.

The rest of this paper is organized as follows: Section 2 introduces some background and related work of style transfer. In Section 3, we analyze the intra-class and inter-class relationships in [4,5], and propose our argument. Furthermore, two new methods of style transfer are proposed to verify the argument we have presented. The experimental results are shown in Section 4, which verify our argument proposed in Section 3, and show our methods can generate comparable visual effect yet with a lower complexity. Section 5 summarizes our work.

## 2. Background and Related Work

### 2.1. Convolutional Neural Network

In the past few years, deep learning methods have proven to be superior to previous state-of-the-art machine learning techniques in several areas and computer vision is one of the most prominent cases. The recent success of CNNs in computer vision have surpassed human performance in image classification, image segmentation and image captioning [6–8]. The primary reason CNNs can boost these advances is that CNNs can construct spatially-preserved feature representations of the content of the image hierarchically. When a filter slides over an image, an activation map or feature map is generated, and each layer in a CNN comprises of a number of filters. By tacking many of these layers on top of one another, CNNs can develop abstract and high-level representations of the image content in the deeper layers of the network.

### 2.2. Style Transfer

Style transfer is an interesting and popular topic in both academic research and industrial engineering.

Many studies explore how to transfer the style from one image to another one automatically. In these studies, the progress of non-photorealistic rendering (NPR) is promotive and can be widely used in various fields [9–11]. However, the NPR algorithms are usually highly dependent on specific artistic styles (e.g., oil paintings, Watercolor, etc.), which means an oil painting-oriented NPR algorithm can only generate an oil painting style digital result, and cannot be easily extended to create other artistic style results.

Recently, inspired by the great success of CNNs in the field of computer vision, Gatys et al. first studied how to use CNN to reproduce famous painting styles on natural images [4,12]. In their work, the content and style of the image are represented separately: they used the Gram matrices of the

neural activations from different layers of a CNN to represent the artistic style of an image, and the feature maps from different layers of a CNN to represent the content of an image. For a given pair of content and style images, the authors used a white noise image as a start image, and an iterative optimization method to correct the result image by matching the neural activations with the content image and the Gram matrices with the style image.

This novel algorithm has attracted many follow-up works, and these follow-up works have improved the original algorithm on different aspects. Some recent works try to speed up the iterative optimization process by training a feed-forward generative network which provides up to three orders of magnitude speed-up over the original approach [13,14]. Furthermore, Ruder et al. incorporated temporal consistency terms by penalizing deviations between frames for video style transfer [15]. Now, the style transfer is realistic for real-time video applications [16,17].

In addition to improving the efficiency of style transfer, recent research works propose various schemes to explore high perceptual quality of the generated images. These schemes include spatial constraints method [18], semantic guidance method [19], Markov Random Field (MRF) method [20], patch-based transfer method [21], and so on. Recently, some extensions to the original neural style transfer method have significantly improved the perceptual quality of generated images by introducing a number of new features, such as multiple style transfer [22,23], color-preserving style transfer [24], and content-aware style transfer [25].

### 2.3. Demystifying Neural Style Transfer

Although all the methods mentioned above have improved over the original neural style transfer significantly, they all use the same *Gram matrix* as the original algorithm to represent the artistic style. As for why the *Gram matrix* can represent the style of an image in neural style transfer, this fundamental question has never been answered by anyone until, recently, the authors of [5] started to demystify the original neural style transfer algorithm.

In [5], the authors investigated the original neural style transfer algorithm and argued that the essence of neural style transfer is to match the feature distributions between the style images and the generated images. Furthermore, the authors proved that matching the *Gram matrix* is equivalent to minimizing maximum mean discrepancy (MMD) with the second order polynomial kernel [26], which means the *Gram matrix* is not a must for neural style transfer. To verify their theory, the authors also used other kernel functions, such as the linear and Gaussian kernels for the MMD to reproduce the style of the style image onto the content image. Besides the MMD, the authors also tried to use the statistics of Batch Normalization (BN) of a certain layer to represent the style of the image. The experimental results show that, compared to the methods which use the *Gram Matrix*, the methods which use the MMD with the second order polynomial kernel, or the BN to represent the style of the image, can produce stylized images with comparable visual effect.

In the following sections, the algorithm mentioned in [5] is termed as BN method, while the original neural style transfer algorithm [4] is termed as *Gram-Matrix* method. Before proposing our style transfer algorithms, the loss functions employed by the *Gram-Matrix* and BN methods are analyzed in the following section.

## 3. Method

In this section, we analyze the intra-class and inter-class relationships in the *Gram-Matrix* method and BN method, and present our argument. Then, we propose two methods of style transfer based on our argument.

### 3.1. Analysis for *Gram-Matrix* Method

For a given content image  $x_c$  and a style image  $x_s$ , the goal of style transfer is to make the generated image  $x^*$  both have content image's content and style image's style. When  $x_c$ ,  $x_s$ ,  $x^*$  pass the layer  $l$  of CNN, their feature maps are denoted by  $P^l \in R^{N_l \times M_l}$ ,  $S^l \in R^{N_l \times M_l}$ , and  $F^l \in R^{N_l \times M_l}$ ,

where  $N_l$  is the number of the feature maps in the layer  $l$ , and  $M_l$  is the width times the height of the feature map. *Gram-Matrix* method generates an image  $x^*$  which depicts the content of image  $x_c$  in the style of image  $x_s$  by minimizing the following loss function:

$$\mathcal{L}_{total} = \alpha \mathcal{L}_{content} + \beta \mathcal{L}_{style} \tag{1}$$

where  $\alpha$  and  $\beta$  are the weights for content and style losses. The content loss  $\mathcal{L}_{content}$  is the mean-squared distance between the feature maps of  $x_c$  and  $x^*$  at a specific layer  $l$ , while the style loss  $\mathcal{L}_{style}$  measures the distributional difference of the feature maps of  $x_c$  and  $x^*$  at several specified style layers. The  $\mathcal{L}_{content}$  and  $\mathcal{L}_{style}$  can be formalized as:

$$\mathcal{L}_{content} = \frac{1}{2} \sum_{i=1}^{N_l} \sum_{j=1}^{M_l} (F_{ij}^l - P_{ij}^l)^2 \tag{2}$$

$$\mathcal{L}_{style} = \sum_l w_l \mathcal{L}_{style}^l \tag{3}$$

where  $w_l$  is the weight of layer  $l$ , and  $\mathcal{L}_{style}^l$  is the mean-squared distance between the features correlations expressed by Gram matrices of  $x^*$  and  $x_s$ :

$$\mathcal{L}_{style}^l = \frac{1}{4N_l^2 M_l^2} \sum_{i=1}^{N_l} \sum_{j=1}^{N_l} (G_{ij}^l - A_{ij}^l)^2 \tag{4}$$

Here, the *Gram matrix*  $G_l \in R^{N_l \times N_l}$  is the inner product between the vectorized feature maps of  $x^*$  in layer  $l$ :

$$G_{ij}^l = \sum_{k=1}^{M_l} F_{ik}^l F_{jk}^l \tag{5}$$

Similarly,  $A_l$  is the *Gram matrix* corresponding to  $S_l$ .

According to our definitions for inter-class and intra-class operations, we observed that, for a particular feature map  $x$ , the *Gram matrix* contains both kinds of operations; therefore, the statistical distributions of feature maps in this method belong to both intra-class and inter-class relationships.

Furthermore, we found that all the intra-class relationships appear on the position of main diagonal elements of the matrix, while all the inter-class relationships appear on the off-diagonal positions. In other words, the main diagonal elements of the *Gram matrix* contains the intra-class relationships of the statistical distributions of feature maps. If we only keep the main diagonal elements of the *Gram matrix* and delete other elements, a method called *Gram-MDE-Matrix* can be generated. In Section 4, we conduct experiments to compare this method with others.

### 3.2. Analysis for BN Method

In the BN methods, the authors constructed the style loss function by aligning the BN statistics (mean and standard deviation) of two feature maps of the two images:

$$\mathcal{L}_{style}^l = \frac{1}{N_l} \sum_{i=1}^{N_l} \left( (\mu_{F^l}^i - \mu_{S^l}^i) + (\sigma_{F^l}^i - \sigma_{S^l}^i) \right)^2 \tag{6}$$

where  $\mu_{F^l}$  and  $\sigma_{F^l}$  are the mean and standard deviation of the  $i$ th feature channel among all the positions of the feature map in the layer  $l$  for image  $x^*$ :

$$\mu_{F^l}^i = \frac{1}{M_l} \sum_{j=1}^{M_l} F_{ij}^l \tag{7}$$

$$\sigma_{F^l}^i{}^2 = \frac{1}{M_l} \sum_{j=1}^{M_l} (F_{ij}^l - \mu_{F^l}^i)^2 \tag{8}$$

In the above formula, the BN statistics matching measures the statistical difference between the mean and standard deviation of each feature map of a style image in a layer and the average and standard deviation of the corresponding feature map in the generated image. In an image (a style image or the generated image), the feature map does not have a relationship with other feature maps, so all the relationships in this method are all intra-class relationships.

In Section 4, we compare this method with the *Gram-Matrix* method.

### 3.3. Cov-Matrix Method

In [4], they used *Gram matrix* to represent the style of an image. Here, we use covariance matrix, which can be applied to many fields such as principle component analysis (PCA) and spectral analysis [27,28], to represent the style of an image and can achieve good results (see Section 4). This method is termed as *Cov-Matrix* method in this paper.

In probability statistics, the variance describes the average distance from the sample point to the mean for each sample set. When faced with two-dimensional data, the covariance is often used to measure the relationship between two random variables (see Equation (9)). Here,  $n$  is the number of samples. When  $cov(X, Y) > 0$ , this shows that  $X$  and  $Y$  are positively related. When  $cov(X, Y) < 0$ , this shows that  $X$  and  $Y$  negatively related. When  $cov(X, Y) = 0$ , this shows that  $X$  and  $Y$  are independent of each other.

$$cov(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n - 1} \tag{9}$$

Covariance can only handle two-dimensional problems; when faced with multi-dimensional data, we usually need to calculate multiple covariances. We can use covariance matrix to represent these data (see Equation (10)).

$$C_{N_l \times N_l}^l = (c_{i,j}, c_{i,j} = cov(Dim_i, Dim_j)) \tag{10}$$

Here,  $N_l$  is the number of feature maps of layer  $l$ , which is also called the number of dimensions of layer  $l$ .  $Dim_i$  and  $Dim_j$  represent the  $i$ th feature map and the  $j$ th feature map of layer  $l$ , respectively. The value of  $cov(Dim_i, Dim_j)$  can be computed by:

$$\begin{aligned} & cov(Dim_i, Dim_j) \\ &= \frac{\sum_{k=1}^M (Dim_{ik} - \bar{Dim}_i)(Dim_{jk} - \bar{Dim}_j)}{M - 1} \end{aligned} \tag{11}$$

where  $M$  is the width times the height of the feature map and  $Dim_{ik}$  is the  $k$ th element of the  $i$ th feature map. The style loss  $\mathcal{L}_{style}^l$  can be computed by:

$$\mathcal{L}_{style}^l = \frac{1}{4N_l^2 M_l^2} \sum_{i=1}^{N_l} \sum_{j=1}^{N_l} (C_{ij}^l - H_{ij}^l)^2 \tag{12}$$

where  $C_l \in R^{N_l \times N_l}$  is the covariance matrix between the vectorized feature maps of  $x^*$  in layer  $l$ , i.e.,  $C_{ij}^l = cov(F_i^l, F_j^l)$ . Similarly,  $H_l$  is the covariance matrix corresponding to  $S_l$ . Apparently, this method contains both intra-class and inter-class relationships of the statistical distributions of feature maps.

### 3.4. Cov-MDE-Matrix Method

In covariance matrix, the main diagonal elements represent the variance of each dimension of the data. In style transfer, the main diagonal elements of the covariance matrix represent the variance

of each feature map of a specific layer. In this subsection, we use the main diagonal elements of the covariance matrix to represent the style of an image.

$$\mathcal{L}_{style}^l = \frac{1}{4N_l^2 M_l^2} \sum_{i=1}^{N_l} (C_i^l - H_i^l)^2 \quad (13)$$

where  $C_l \in R^{N_l \times 1}$  is the main diagonal elements of the covariance matrix between the vectorized feature maps of  $x^*$  in layer  $l$ , i.e.,  $C_i^l = cov(F_i^l, F_i^l)$ . Similarly,  $H_l$  is the main diagonal elements of the covariance matrix corresponding to  $S_l$ .

The algorithm proposed above is termed as *Cov-MDE-Matrix* method. Clearly, this method only contains the intra-class relationships of the statistical distributions of feature maps. Compared to the *Cov-Matrix* method, the *Cov-MDE-Matrix* method only computes the main diagonal elements of the covariance matrix, so it has lower complexity and can run faster. In Section 4, we compare the two methods, *Cov-Matrix* and *Cov-MDE-Matrix*, proposed in this paper.

## 4. Results

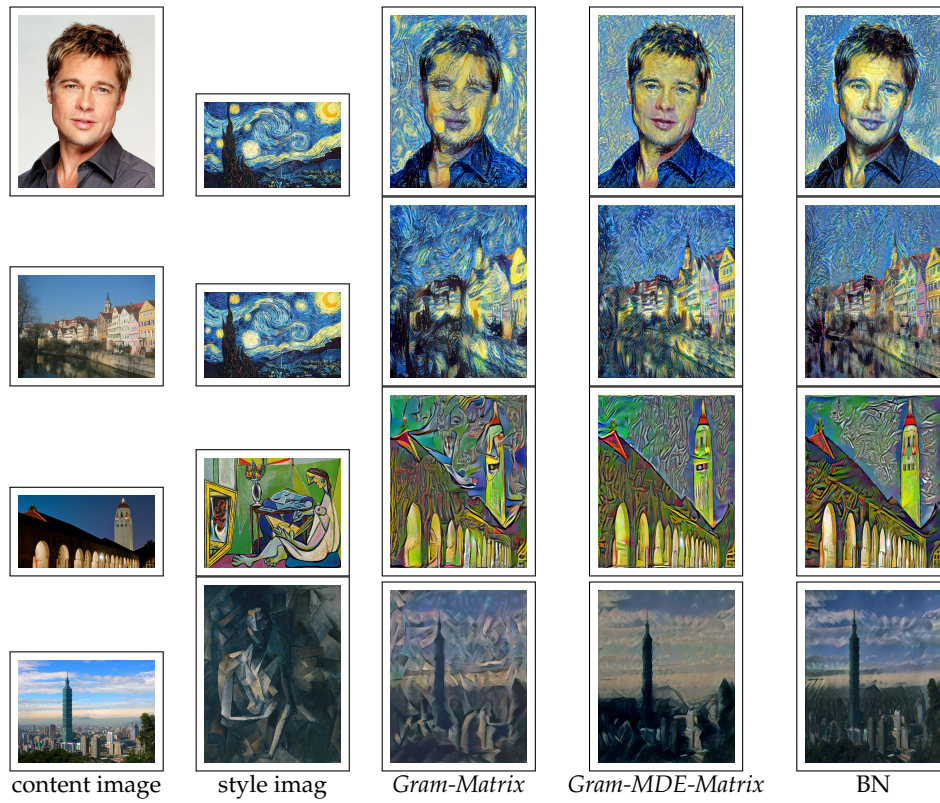
### 4.1. Implementation Details

We used VGG-19 (Oxford Visual Geometry Group) network [29] and adopted the *conv4\_2* layer for the content loss, and *conv1\_1*, *conv2\_1*, *conv3\_1*, *conv4\_1*, and *conv5\_1* ( $w_l = 0$  in all layers) for the style loss. The maximum number of iterations was set to 4000. When producing an image that combines the content of one image with the style of another, there is a trade-off between content and style matching (a linear combination of  $\alpha$  and  $\beta$ ). For a specific pair of content and style images, we can adjust the ratio of  $\alpha/\beta$  to generate visually appealing images. In this paper, since we wanted to conduct the comparative experiments with the *Gram matrix* algorithm, we set the value of ratio  $\alpha/\beta$  to be the same as in the *Gram matrix* algorithm.

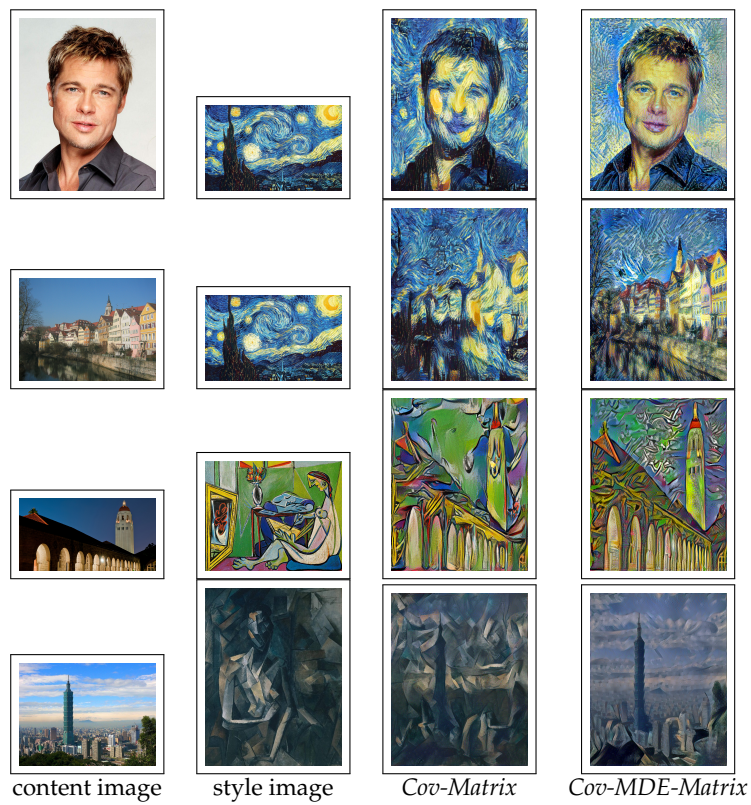
### 4.2. Result Comparisons

Figure 1 shows the results of three transfer methods: *Gram-Matrix*, *Gram-MDE-Matrix*, and BN method. As analyzed above, the *Gram-Matrix* method contains both intra-class and inter-class relationships of the statistical distributions of feature maps, while the *Gram-MDE-Matrix* and BN methods contain only intra-class relationships. As shown in the figure, the methods that only consider the intra-class relationship produce finer strokes and keep more content details compared to the method that does not.

Figure 2 shows the results of the two style transfer methods proposed in this paper: *Cov-Matrix* and *Cov-MDE-Matrix* method. The *Cov-Matrix* method contains both intra-class and inter-class relationships of the statistical distributions of feature maps, while the *Cov-MDE-Matrix* method contains only intra-class relationships. Again, the method that only considers the intra-class relationship produces finer strokes and keeps more content details compared to the method that does not. Compared with the results in Figure 1, our proposed method can generate comparable visual effect results.



**Figure 1.** Results of *Gram-Matrix*, *Gram-MDE-Matrix* and Batch Normalization (BN) method. The ratio of  $\alpha/\beta$  is  $5 \times 10^{-3}$ .



**Figure 2.** Results of *Cov-Matrix* and *Cov-MDE-Matrix* method. The ratio of  $\alpha/\beta$  is  $5 \times 10^{-3}$ .

According to our analysis and experimental results, we observe that algorithms that consider the inter-class relationships may keep more style information, while algorithms that only consider the intra-class relationships may generate finer strokes and keep more content details. When doing the style transfer, if we want to keep more content details (e.g., head portraits, as shown in Figures 1 and 2), the algorithms considering intra-class relationships are recommended. If we want to keep more style information (e.g., the building picture shown in Figures 1 and 2), algorithms considering both intra-class and inter-class relationships are recommended.

## 5. Conclusions

The Neural Style Transfer algorithm proposed by Gatys et al. produces fantastic stylized images with the appearance of a given famous artwork. Before the work of Li et al. [5], no one could theoretically explain why the Gram matrices used as the key technique in the original and many follow-up studies of the style transfer algorithms could represent the artistic style. The work in [5] unlocks the secret of the Gram matrices to some extent.

Since *Gram matrix* is not a must for the style transfer, how to design proper process for aligning the neural activation between images becomes an important problem. In addition to the *Gram matrix*, there are many kinds of matrices. Are these existing matrices all suitable to design the style loss algorithm? The answer should be obviously not. In this paper, after careful analysis of different matrices which are adopted successfully by some style transfer algorithms for style loss design, we discover that some algorithms take pairwise relationships between different feature maps into account, while some do not. Based on this observation, we define inter-class and intra-class relationships for statistical distributions of feature maps of a layer obtained from the CNN. The corresponding operations performed on the inter-class and intra-class relationships are defined as inter-class and intra-class operations, respectively. As a result, we discover that intra-class operations performed on the intra-class relationships are helpful to generate finer strokes and keep more content details during the style transfer, while the inter-class operations performed on the inter-class relationships are helpful to keep more style information. To further support our standpoint, we propose two new methods to perform style transfer; one takes inter-class relationships into account while the other does not. The experimental results verified our observation.

Recall all the existing work; people usually use some parameters to adjust the fusion effect of content and style information during style transfer. However, this process may depend on many experimental experiences. Our observation and conclusion proposed in this paper may provide a different direction for designing style loss function. In the future, we will focus on designing style loss functions which can adaptively adjust the inter-class and intra-class relationships in local image spaces, i.e., new algorithms could consider more intra-class relationships to generate finer strokes when more content details should be kept, while considering more inter-class relationships to keep more style information.

**Author Contributions:** Conceptualization, M.Q. and X.C.; Methodology, X.C. and M.Q.; Software, X.C. and B.L.; Validation, X.C., Y.N. and M.Q.; Formal Analysis, M.Q.; Supervision, M.Q. and Y.N.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Hertzmann, A.; Jacobs, C.E.; Oliver, N.; Curless, B.; Salesin, D.H. Image analogies. In Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, Los Angeles, CA, USA, 12–17 August 2001; ACM: New York, NY, USA, 2001; pp. 327–340.
2. Efros, A.A.; Leung, T.K. Texture synthesis by non-parametric sampling. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999; Volume 2, pp. 1033–1038.



3. Efros, A.A.; Freeman, W.T. Image quilting for texture synthesis and transfer. In Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, Los Angeles, CA, USA, 12–17 August 2001; ACM: New York, NY, USA, 2001; pp. 341–346.
4. Gatys, L.A.; Ecker, A.S.; Bethge, M. Image style transfer using convolutional neural networks. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2414–2423.
5. Li, Y.; Wang, N.; Liu, J.; Hou, X. Demystifying neural style transfer. *arXiv* **2017**, arXiv:1701.01036.
6. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS 2012), Lake Tahoe, NV, USA, 3–6 December 2012; Curran Associates Inc.: Red Hook, NY, USA, 2012; Volume 1, pp. 1097–1105.
7. Karpathy, A.; Fei-Fei, L. Deep Visual-Semantic Alignments for Generating Image Descriptions. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 664–676. [[CrossRef](#)] [[PubMed](#)]
8. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [[CrossRef](#)]
9. Gooch, B.; Gooch, A. *Non-Photorealistic Rendering*; AK Peters/CRC Press: Boca Raton, FL, USA, 2001.
10. Rosin, P.; Collomosse, J. *Image and Video-Based Artistic Stylisation*; Springer Science & Business Media: Berlin, Germany, 2012; Volume 42.
11. Strothotte, T.; Schlechtweg, S. *Non-Photorealistic Computer Graphics: Modeling, Rendering, and Animation*; Morgan Kaufmann: San Francisco, CA, USA, 2002.
12. Gatys, L.A.; Ecker, A.S.; Bethge, M. A Neural Algorithm of Artistic Style. *arXiv* **2015**, arXiv:1508.06576.
13. Johnson, J.; Alahi, A.; Fei-Fei, L. *Perceptual Losses for Real-Time Style Transfer and Super-Resolution*; European Conference on Computer Vision; Springer: Cham, Switzerland, 2016; pp. 694–711.
14. Ulyanov, D.; Lebedev, V.; Vedaldi, A.; Lempitsky, V.S. Texture Networks: Feed-forward Synthesis of Textures and Stylized Images. In Proceedings of the 33rd International Conference on International Conference on Machine Learning (ICML), New York, NY, USA, 19–24 June 2016; Volume 48, pp. 1349–1357.
15. Ruder, M.; Dosovitskiy, A.; Brox, T. *Artistic Style Transfer for Videos*; German Conference on Pattern Recognition; Springer: Cham, Switzerland, 2016; pp. 26–36.
16. Huang, H.; Wang, H.; Luo, W.; Ma, L.; Jiang, W.; Zhu, X.; Li, Z.; Liu, W. Real-Time Neural Style Transfer for Videos. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 7044–7052.
17. Ulyanov, D.; Vedaldi, A.; Lempitsky, V.S. Instance Normalization: The Missing Ingredient for Fast Stylization. *arXiv* **2016**, arXiv:1607.08022.
18. Selim, A.; Elgharib, M.; Doyle, L. Painting style transfer for head portraits using convolutional neural networks. *ACM Trans. Gr.* **2016**, *35*, 1–18. [[CrossRef](#)]
19. Champandard, A.J. Semantic style transfer and turning two-bit doodles into fine artworks. *arXiv* **2016**, arXiv:1603.01768.
20. Li, C.; Wand, M. Combining markov random fields and convolutional neural networks for image synthesis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2479–2486.
21. Chen, T.Q.; Schmidt, M. Fast Patch-based Style Transfer of Arbitrary Style. *arXiv* **2016**, arXiv:1612.04337.
22. Chen, D.; Yuan, L.; Liao, J.; Yu, N.; Hua, G. StyleBank: An Explicit Representation for Neural Image Style Transfer. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2770–2779.
23. Liao, J.; Yao, Y.; Yuan, L.; Hua, G.; Kang, S.B. Visual Attribute Transfer Through Deep Image Analogy. *ACM Trans. Gr.* **2017**, *36*, 120:1–120:15. [[CrossRef](#)]
24. Gatys, L.A.; Bethge, M.; Hertzmann, A.; Shechtman, E. Preserving Color in Neural Artistic Style Transfer; Technical Report. *arXiv* **2016**, arXiv:1606.05897.
25. Gatys, L.A.; Ecker, A.S.; Bethge, M.; Hertzmann, A.; Shechtman, E. Controlling Perceptual Factors in Neural Style Transfer. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 3730–3738.

26. Gretton, A.; Borgwardt, K.M.; Rasch, M.J.; Schölkopf, B.; Smola, A. A kernel two-sample test. *J. Mach. Learn. Res.* **2012**, *13*, 723–773.
27. Zorzi, M. Multivariate Spectral Estimation Based on the Concept of Optimal Prediction. *IEEE Trans. Autom. Control* **2015**, *60*, 1647–1652. [[CrossRef](#)]
28. Zorzi, M. An interpretation of the dual problem of the THREE-like approaches. *Automatica* **2015**, *62*, 87–92. [[CrossRef](#)]
29. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).