Portrait-Aware Artistic Style Transfer

Conference Paper · October 2018

DOI: 10.1109/ICIP.2018.8451054

6 authors, including:

Xing Yeli
Tsinghua University
3 PUBLICATIONS **15** CITATIONS
SEE PROFILE

Jiawei Li
20 PUBLICATIONS **89** CITATIONS
SEE PROFILE

Tao Dai
Tsinghua University
73 PUBLICATIONS **431** CITATIONS
SEE PROFILE

Qingtao Tang
Tsinghua University
18 PUBLICATIONS **158** CITATIONS
SEE PROFILE

PORTRAIT-AWARE ARTISTIC STYLE TRANSFER

Yeli Xing⋆*, Jiawei Li*⋆*, Tao Dai*⋆*, Qingtao Tang*⋆*, Li Niu*†*, Shu-Tao Xia*⋆

⋆Department of Computer Science and Technology, Tsinghua University, China
† Electric and Computer Engineering department, Rice University, USA
Email: {xyl16, li-jw15, dait14, tqt15}@mails.tsinghua.edu.cn, ln7@rice.edu, xiast@sz.tsinghua.edu.cn

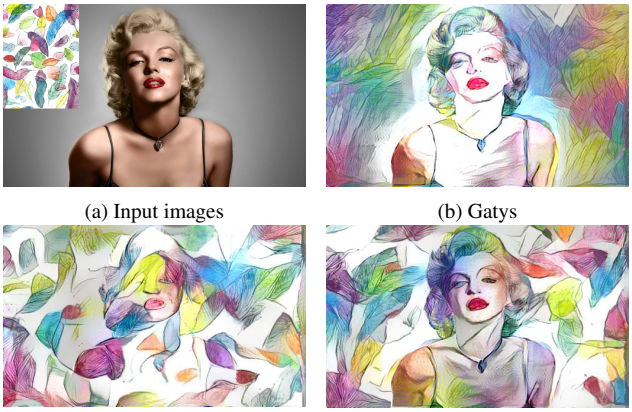
ABSTRACT

The goal of artistic style transfer is to transfer the style of artistic works into photos. However, the performances of existing style transfer algorithms on portraits are not very satisfactory, because the synthetic photo is either not sufficiently stylized or distorted severely in the portrait domain (*i.e.*, foreground), which limits the use of style transfer for portraits. In this paper, we propose a novel portrait-aware artistic style transfer algorithm, which treats foreground and background differently. Particularly, we separate the foreground from the background, and apply fine-grained style transfer to the background and coarse-grained style transfer to the entire image at the same time, so that the artistic style of entire image can be transferred with the details of the portrait well preserved. Extensive experiments demonstrate the effectiveness of our proposed method.

Index Terms— Artistic Style Transfer, Texture Synthesis, Convolution Neural Network (CNN), Semantic Segmentation

1. INTRODUCTION

Image style transfer is an image synthesis problem, which aims to transfer the style of one image (Style-image) to another (Content-image). Recently, Gatys [1, 2] used a pre-trained network (*i.e.*, pre-trained VGG-19 for image classification) as feature extractor to drive texture synthesis and style transfer, yielding impressive performance benefiting from the rich feature representation. Following [1, 2], Prisma Labs launched its application Prisma which draw ten million users in five weeks. Programms like Google Deep StyleOstagramand picsartalso attract a large number of users. People can upload their photos and transfer them with various filters. Thus, style transfer is becoming a new fashion and people show great enthusiasm for portrait style transfer.[1] [2] However, we observe that Gatys' algorithm [1] for portrait style transfer is subject to some limitations: the synthetic photo is either not sufficiently stylized or distorted severely in the portrait domain. In particular, as shown in Fig. 1(b), the synthetic artistic work is not stylized enough and appears quite different from the Style-image. To get a more stylized result, we enhance the weight of the Style-image by 20 times, leading to the result in Fig. 1(c). The new result matches the style of the Style-image but induces new problems: the portrait domain is seriously distorted by many small color chunks, which renders the subject unrecognizable. In practical applications, however, users prefer well-preserved structure details of the portrait domain, which will make the synthetic artistic work more visually pleasing. To tackle this issue, [3] proposes a patch-based style transfer algorithm, generating photorealistic results with recognizable portrait details. However, this method fails to match the photo with the artistic style as well as previous style transfer methods. Therefore, to the best of our knowledge, there exist no works on portrait style transfer, which can seek a perfect tradeoff between the Content-image and the Style-image, *i.e.*, it remains unclear how to make the synthetic image preserve the structure of the portrait while transferring the style of the Style-image as much as possible.



(a) Input images (b) Gatys (c) Gatys+ (d) Ours

Fig. 1: (a) shows the Content-image and the Style-image in the insets. (b) and (d) are the results of Gatys' algorithm [1] and ours respectively. (c) is also a result of [1], but we enhance its weight of the Style-image by 20 times. It is obvious that a more emphasis on the Style-image will introduce either a distortion or local color-chunks in the portrait domain.

In this paper, we propose a new method for portrait artistic style transfer that can capture the style of the Style-image while preserving the structure of the portrait as much as possible. Specifically, we employ different style trans-

[1]Tao Dai's contribution was made when visiting The Hong Kong Polytechnic University. This work is supported by the National Natural Science Foundation of China under grant Nos. 61771273.

[2]The authors thank the support of NVIDIA Corporation with the donation of the Titan Xp GPUs used for this project.

fer strategies on the foreground (*i.e.*, the portrait domain) and the background. Our method consists of two steps: 1) generate background mask for the Content-image to separate foreground from background with the aid of semantic segmentation; 2) apply fine-grained style transfer to the background, and coarse-grained style transfer to the entire image simultaneously. On one hand, the background mask prevents the style of the Style-image from spilling over into the portrait domain of the Content-image. On the other hand, the coarse-grained style transfer on the entire image tunes the global style of the synthetic image while preserving the structure details of portrait domain. Thus, our proposed portrait-aware style transfer algorithm can obtain a proper trade-off between preserving the structure of the portrait and style-transferring the Style-image.

2. RELATED WORK

Representation of style Basically, there are two kinds of ways to define an image's style in neural style transfer. One (Gram-based algorithms) uses a CNN as the feature exactor and defines the style of an image as the correlations of feature maps [1, 4]. The other (patch-based algorithms) [3, 5] uses the cosine distance to measure the similarity of two feature maps, which helps find the best-match patches in activation space and construct the patch-based loss function.

Generally, patch-based algorithms produce more impressive results in photorealistic style transfer task, while Gram-based ones work better in non-photorealistic style transfer.

Guided style transfer Many works [6, 7] add various constraints to achieve precise control for style transfer. Algorithm that appends spatial control to style transfer is called guided style transfer. This kind of method can prevent the style of the Style-image spilling over into mismatching domains of the Content-image.

For example, Castillo [8] used a Markov random field (MRF) criterion to combine style transfer and semantic segmentation, which leading to an accurate targeted style transfer. There is another kind of algorithm, which combines style transfer with image semantic segmentation, belonging to guided style transfer. These algorithms segment the Content-image and the Style-image into slice pairs, and style transfer each content slice with their corresponding style slice. Specifically, Alex and Champandard [9] and Luan [10] combined the image semantic segmentation with patch-based and Gram-based algorithms respectively. Although transferring styles between two images' corresponding slices can drive very meaningful style transfer, these methods cannot be directly applied to portrait artistic style transfer, since not all of the artistic images can be semantically segmented. Besides, these methods cannot capture the global style of the Style-image, which is of great significance for artistic style transfer.

3. NEURAL STYLE TRANSFER

We first briefly review the original neural style transfer algorithm [1]. Given a Content-image $\boldsymbol{x}_c$ and Style-image $\boldsymbol{x}_s$, the goal is to synthesize an image $\hat{\boldsymbol{x}}$ which simultaneously shares the content representation of $\boldsymbol{x}_c$ and the style representation of $\boldsymbol{x}_s$. For any input image $\boldsymbol{x}$, we vectorize its CNN representation in layer l as $F_l(\boldsymbol{x})$, thus $F_l(\boldsymbol{x}) \in R^{D_l(\boldsymbol{x}) \times N_l}$, where $D_l(\boldsymbol{x})$ is the height times the width of the feature map in layer l and N_l is the number of feature maps in layer l. Then neural style transfer is the process of using these representations to synthesize a new artistic work by minimizing the following loss function:

$$E_{total} = \alpha E_c + \beta E_s, \tag{1}$$

where E_c and E_s are the loss terms for content and the style reconstruction respectively, with their corresponding weighting factors α and β.

The content loss on the layer l can be defined as the squared Euclidean distance between feature representations of Style-image $\boldsymbol{x}_c$ and the synthetic image $\hat{\boldsymbol{x}}$:

$$E_c = \frac{1}{N_l D_l(\boldsymbol{x}_c)} \sum_{i,j} \left(F_l(\hat{\boldsymbol{x}}) - F_l(\boldsymbol{x}_c)\right)^2_{i,j}. \tag{2}$$

Similarly, the style loss E_s on the layer l can be defined as the squared Euclidean distance between $G_l(\hat{\boldsymbol{x}})$ and $G_l(\boldsymbol{x}_s)$:

$$E_s^l = \frac{1}{4N_l^2} \sum_{i,j} \left(G_l(\hat{\boldsymbol{x}}) - G_l(\boldsymbol{x}_s)\right)^2_{i,j}, \tag{3}$$

where the Gram Matrix G_l is the correlations of feature maps, that is, $G_l(\boldsymbol{x}) = \frac{1}{D_l(\boldsymbol{x})} F_l(\boldsymbol{x})^T F_l(\boldsymbol{x})$.

Different from the content loss on one single layer, the style loss E_s is employed on multiple layers:$E_s = \sum_l E_s^l$. Specifically, Gatys [1] uses conv1_1, conv2_1, conv3_1, conv4_1 and conv5_1 to represent E_s, and applys conv4_2 for content representation.

As mentioned above, using this method for portrait style transfer is subject to some limitations: a stronger emphasis on the style will introduce either a distortion or local color-chunks in the portrait domain. Though various methods [11, 12, 13, 14, 15, 3, 5, 4] are proposed afterwards, Gram-based algorithms represented by [1] are still the best algorithms at present [16].

4. OUR METHOD

In this paper, we introduce a new method for portrait artistic style transfer, and the pipeline of our algorithm is shown in Fig. 2, from which it can be seen that our algorithm is a two-stage method: portrait segmentation and portrait-aware style transfer. In portrait segmentation stage, we separate the portrait domain from the background and get a masked image. Based on the masked image, we apply a fine-grained style transfer on the background and a coarse-grained style transfer on the entire image simultaneously.

4.1. Portrait Segmentation

Unlike Gatys' algorithm [1], in which only one style loss function is used to guide style transfer, our method style-transfers the portrait and background separately by using d-

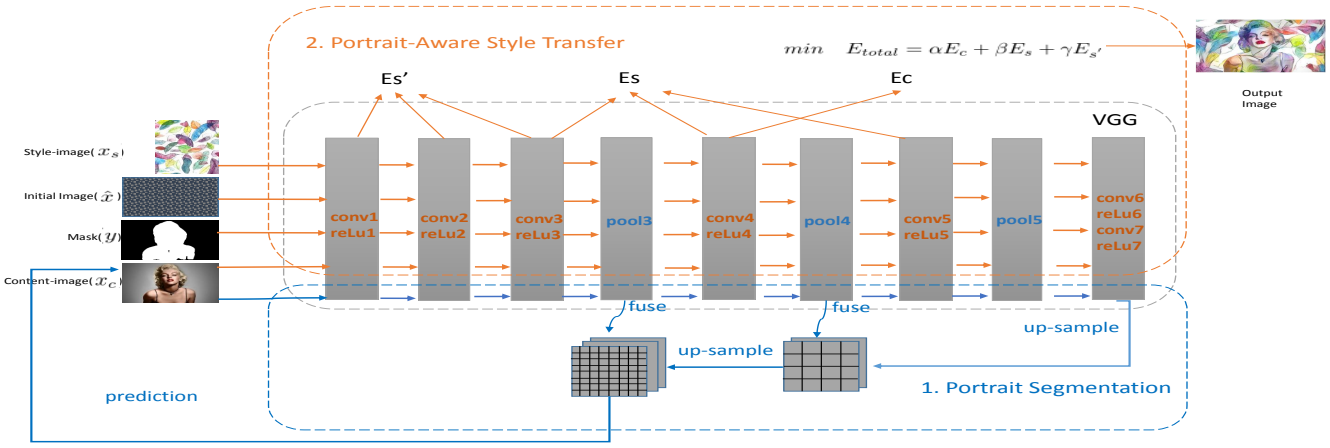


Fig. 2: An overview of our portrait-aware artistic style transfer algorithm.

ifferent style loss functions for the portrait domain and background, which requires portrait segmentation to separate the foreground from the background.

To be specific, we use semantic segmentation to separate the foreground from the background. Numerous algorithms have been developed for segmentation [17, 18, 19]. Here, we use FCN-8s [19] to identify different slices for the Content-image $\boldsymbol{x}_c$. By masking the portrait slice, we separate the portrait from the background and get a masked image $\boldsymbol{y}$.

4.2. Portrait-Aware Style Transfer

The goal of style transfer is to produce a synthetic image $\hat{\boldsymbol{x}}$ which simultaneously shares the content representation of $\boldsymbol{x}_c$ and the style representation of $\boldsymbol{x}_s$. To this end, we first use white noise to initialize the image $\hat{\boldsymbol{x}}$, which is of the same size as $\boldsymbol{x}_c$. Meanwhile, we apply the masked image $\boldsymbol{y}$ which is an image map of values in $\{0, 1\}$ to label the background domain of $\hat{\boldsymbol{x}}$. This background domain is defined as $\hat{\boldsymbol{x}}' = \hat{\boldsymbol{x}} \circ \boldsymbol{y}$, where $\circ$ denotes element-wise multiplication. Together with the Content-image $\boldsymbol{x}_c$ and the Style-image $\boldsymbol{x}_s$, these four images $(\hat{\boldsymbol{x}}, \boldsymbol{x}_c, \boldsymbol{x}_s, \boldsymbol{y})$ are fed into the CNN, as shown in Fig. 2.

In our method, the style loss is split into two parts: the E_s focusing on the entire image and $E_{s'}$ focusing on the background. Thus our new loss function is defined as:

$$E_{total} = \alpha E_c + \beta E_s + \gamma E_{s'}, \tag{4}$$

where E_c and E_s are the same as mentioned in Section 3. However, we only keep the upper layers (layers selection is detailed in the experimental part) to represent E_s. $E_{s'}$ is the style loss on some lower layers between $\boldsymbol{x}_s$ and $\hat{\boldsymbol{x}}'$, defined to guide the process of style transfer in background domain:

$$E_{s'} = \sum_l E^l_{s'}, \tag{5}$$

$$E^l_{s'} = \frac{1}{4N_l^2} \sum_{i,j} \left(G_l(\hat{\boldsymbol{x}}') - G_l(\boldsymbol{x}_s)\right)^2_{i,j}. \tag{6}$$

As illustrated in [1], lower layers in CNN tend to capture local styles, while the upper layers for style representation will lead to a smoother and more continuous visual experience. Inspired by such observations, we choose some lower layers for $E_{s'}$ to achieve fine-grained style transfer in background, and apply upper layers to the entire image to capture the global style of the Style-image. In this way, the upper layers for E_s can not only contribute to handling the local color-chunks problem effectively, but also making the synthetic image's style be more consistent with the style image as a whole. Thus, these two loss terms can guide the background and the whole image to style-transfer simultaneously with brushes of different strengthes and sizes.

Finally, we also add a total variation loss [4, 20], which is widely used in image generation methods, to E_{total}, in order to improve the output image's smoothness.

5. EXPERIMENTAL RESULTS

Settings of Portrait Segmentation In our method, we use FCN-8s [19] for image semantic segmentation. This model is trained on VGG-16 with PASCAL VOC 2011, using the same parameters as Long's [19]. After the image semantic segmentation, we get a sliced image, which can separate the image into up to 21 slices. By masking the portrait slice, we separate the portrait domain from the background and get the masked image.

Settings of Portrait-Aware Style Transfer As for the style transfer, our implementation is based on the VGG-19 network [21], which is pre-trained on the ImageNet dataset for image classification. The synthesis image $\hat{\boldsymbol{x}}$ is initialized as random white noise, and iteratively updated by minimizing (4) using back-propagation. Similar to [1], we set the parameters empirically. Specifically, (α, β, γ) in (4) are set as (5e0, 2e2, 2e4) respectively. For optimization, we minimize (4) by L-BFGS [22] and stop L-BFGS until 1000 iterations. As for the layer

Table 1: The structural similarity in foreground of different methods. The best result is shown in boldface.

methods	SSIM	FSIMc	GMSD
Gatys	0.878	0.930	0.196
Gatys+	0.908	0.952	0.157
Ours	**0.932**	**0.966**	**0.123**

selection, we choose one single layer relu4_2 as the content representation, but use relu3_1, relu4_1 and relu5_1 to represent the style loss E_s, and relu1_1, relu2_1 and relu3_1 for the mask loss $E_{s'}$.

Subjective and Qualitative Evaluation To verify the effectiveness of the proposed portrait-aware method, we have performed various experiments on different portrait images with different Style-images. Fig. 3 shows zoom-in results of different methods for portrait image, from which we can observe that the original Gatys' method [1] preserves the structure of portrait but produces less stylized results (see the background area), while the enhanced Gatys' algorithm generates more stylized results (Gatys+) but distorts the portrait severely (see the eye area). Only our method preserves the portrait well while producing more stylized results. This is mainly because the original Gatys' algorithm implicitly assumes that the foreground (portrait domain) and background have the same characteristics by treating the foreground and background with the same parameters, which is not true in practice. Our method, however, considers the characteristics' differences between the foreground and background by treating the foreground and background differently. More experimental results are shown in Fig. 4, and the results in Fig. 3 and Fig. 4 demonstrate that the proposed method can obtain a better trade-off between preserving the structure of the portrait and style-transferring the Style-image.

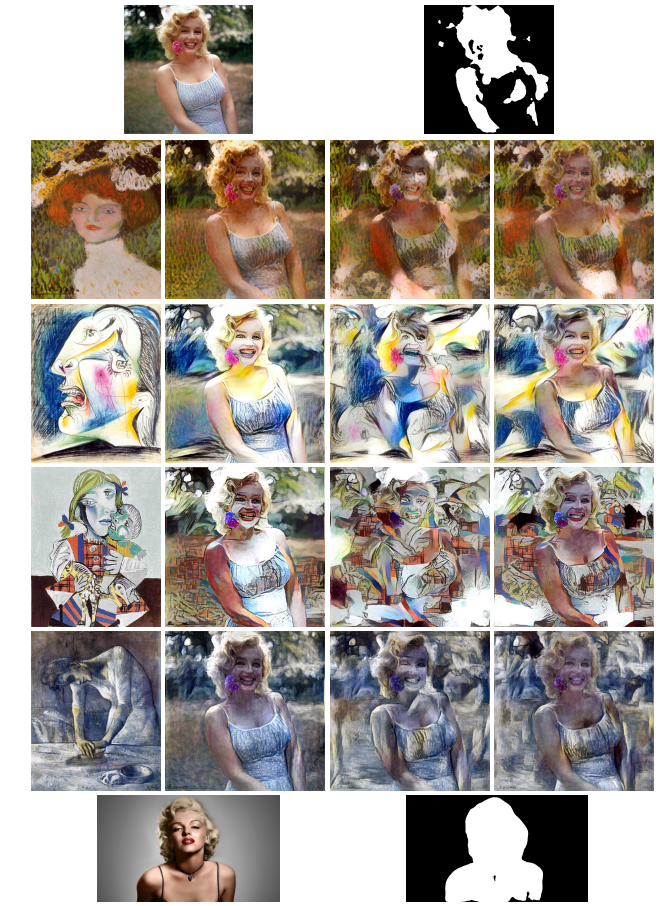
To further validate our method's effectiveness, we choose the commonly used full-reference metrics, including SSIM [23],FSIMc [24] and GMSD [25], to evaluate the foreground regions' structural similarity between the stylized images and the Content-image. For SSIM and FSIMc, the higher values means better structural similarity between two images, while for GMSD, the lower values represents better performance. As shown in Table 1, our method performs best on these three metrics, that is, our method can preserve the image's structure better in foreground.

(a) Gatys (b) Gatys+ (c) Ours

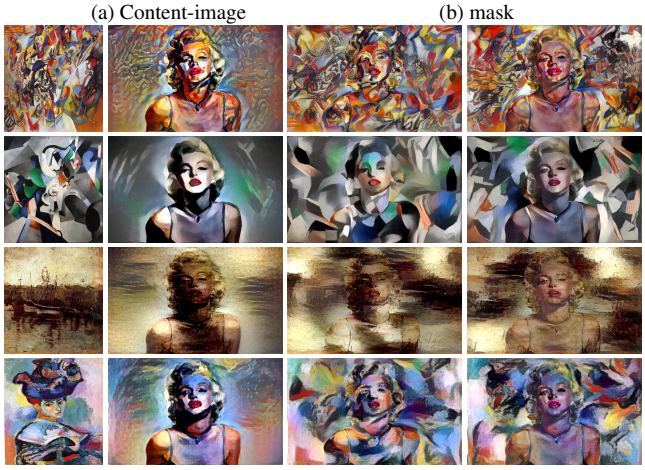
Fig. 3: Some details of the corresponding results.

6. CONCLUSION

In this paper, we tailor the original artistic style transfer to portrait artistic style transfer by combining style transfer with image semantic segmentation. Unlike the original style transfer, whose synthetic image is the result of a tradeoff between



(a) Content-image (b) mask



(c) Style-image (d) Gatys (e) Gatys+ (f) Ours

Fig. 4: (a)(c) are the Content-image and Style-image respectively. (b) is the mask image generated in Section 4.1. (d)(f)(e) are results of Gatys[1], ours and Gatys+ (with the style weight of [1] enhanced by 20 times) respectively.

the Content-image and the Style-image [26], we apply fine-grained style transfer to the background, and coarse-grained style transfer to the entire image simultaneously. By this strategy, we get a more visually pleasing result, which can be generalized to the industry for portrait artistic style transfer.

7. REFERENCES

[1] Leon A Gatys, Alexander S Ecker, and Matthias Bethge, "A neural algorithm of artistic style," *arXiv preprint arXiv:1508.06576*, 2015.

[2] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge, "Texture synthesis using convolutional neural networks," *Febs Letters*, vol. 70, no. 1, pp. 51–55, 2015.

[3] Chuan Li and Michael Wand, "Combining markov random fields and convolutional neural networks for image synthesis," in *CVPR*, 2016, pp. 2479–2486.

[4] Justin Johnson, Alexandre Alahi, and Li Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *ECCV*. Springer, 2016, pp. 694–711.

[5] Tian Qi Chen and Mark Schmidt, "Fast patch-based style transfer of arbitrary style," *arXiv preprint arXiv:1612.04337*, 2016.

[6] Pierre Wilmot, Eric Risser, and Connelly Barnes, "Stable and controllable neural texture synthesis and style transfer using histogram losses," *arXiv preprint arXiv:1701.08893*, 2017.

[7] Leon A Gatys, Alexander S Ecker, Matthias Bethge, Aaron Hertzmann, and Eli Shechtman, "Controlling perceptual factors in neural style transfer," in *CVPR*, 2017.

[8] Carlos Castillo, Soham De, Xintong Han, Bharat Singh, Abhay Kumar Yadav, and Tom Goldstein, "Son of zorn's lemma: Targeted style transfer using instance-aware semantic segmentation," in *ICASSP*, 2017, pp. 1348–1352.

[9] Alex J Champandard, "Semantic style transfer and turning two-bit doodles into fine artworks," *arXiv preprint arXiv:1603.01768*, 2016.

[10] Fujun Luan, Sylvain Paris, Eli Shechtman, and Kavita Bala, "Deep photo style transfer," *CoRR, abs/1703.07511*, 2017.

[11] Roman Novak and Yaroslav Nikulin, "Improving the neural algorithm of artistic style," *arXiv preprint arXiv:1605.04603*, 2016.

[12] Ahmed Selim, Mohamed Elgharib, and Linda Doyle, "Painting style transfer for head portraits using convolutional neural networks," *ACM Transactions on Graphics*, vol. 35, no. 4, pp. 129, 2016.

[13] Chuan Li and Michael Wand, "Precomputed real-time texture synthesis with markovian generative adversarial networks," in *ECCV*. Springer, 2016, pp. 702–716.

[14] Dmitry Ulyanov, Vadim Lebedev, Andrea Vedaldi, and Victor S Lempitsky, "Texture networks: Feed-forward synthesis of textures and stylized images.," in *ICML*, 2016, pp. 1349–1357.

[15] Vincent Dumoulin, Jonathon Shlens, and Manjunath Kudlur, "A learned representation for artistic style," *CoRR*, vol. abs/1610.07629, 2016.

[16] Yongcheng Jing, Yezhou Yang, Zunlei Feng, Jingwen Ye, and Mingli Song, "Neural style transfer: A review," *arXiv preprint arXiv:1705.04058*, 2017.

[17] Jifeng Dai, Kaiming He, and Jian Sun, "Instance-aware semantic segmentation via multi-task network cascades," in *CVPR*, 2016, pp. 3150–3158.

[18] Pedro O Pinheiro, Ronan Collobert, and Piotr Dollár, "Learning to segment object candidates," in *NIPS*, 2015, pp. 1990–1998.

[19] Jonathan Long, Evan Shelhamer, and Trevor Darrell, "Fully convolutional networks for semantic segmentation," in *CVPR*, June 2015.

[20] Aravindh Mahendran and Andrea Vedaldi, "Understanding deep image representations by inverting them," in *CVPR*, 2015, pp. 5188–5196.

[21] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *Computer Science*, 2014.

[22] Galen Andrew and Jianfeng Gao, "Scalable training of l 1-regularized log-linear models," in *ICML*, 2007, pp. 33–40.

[23] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[24] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang, "Fsim: A feature similarity index for image quality assessment," *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, vol. 20, no. 8, pp. 2378, 2011.

[25] Wufeng Xue, Lei Zhang, Xuanqin Mou, and Alan C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 684–95, 2014.

[26] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge, "Image style transfer using convolutional neural networks," in *CVPR*, 2016, pp. 2414–2423.