Portrait Style Transfer using Deep Convolutional Neural Networks and Facial Segmentation

Hui-Huang Zhao and Jing-Hua Zheng

Hunan Provincial Key Laboratory of Intelligent Information Processing and Application, Hunan, China

College of Computer Science and Technology, Hengyang Normal University, Hengyang China

Yao-Nan Wang and Xiao-Fang Yuan

College of Electrical and Information Engineering, Hunan University, Changsha, China

Yu-Hua Li

School of Computer Science and Informatics, Cardiff University, Cardiff, UK

Abstract

When standard neural style transfer approaches are used in portrait style transfer, they often inappropriately apply textures and colours in different regions of the style portraits to the content portraits, leading to unsatisfied transfer results. This paper presents a portrait style transfer method to transfer the style of one image to another. It first proposes a combined segmentation method for the portrait parts, which segments both the style portrait and the content portrait into masks of seven parts automatically, including background, face, eyes, nose, eyebrows, mouth and foreground. These masks are extracted to capture elements of the styles for objects in the style image and to preserve the structure in the content portrait. This paper then proposes an augmented deep Convolutional Neural Network (CNN) framework for portrait style transfer. The masks of seven parts are added into a trained deep convolutional neural network as feature maps in certain selected layers in the augmented deep CNN model. An improved loss function is proposed for the training of the portrait style transfer. The masks are used to preserve the structure of the portraits while having the style transferred. Results on various images show that our method outperforms the state-of-the-art style transfer techniques.

Keywords: Deep Convolutional Neural Networks; Portrait; Style Transfer; Facial

1. Introduction

Image style transfer is a technique of recomposing an image in the style of another single image or images. It can create impressive results covering a wide variety of styles [1], and it has been applied to many successful industrial applications, such as Cartoon [2], makeup [3, 4], Internet of things [5, 6], videos [7]. Image style transfer has also been widely employed to solve problems such as inpainting [8], head portraits [9], super-resolution [10, 11, 12], and font [13, 14].

When standard neural style transfer approaches are used in portrait style transfer, the textures and colours in different regions of the style portraits are often applied inappropriately to the content portraits, leading to undesired results . Li and Wand [15] presented a method that combines generative Markov Random Field (MRF) models for image synthesis Li and Wand [15]. Unlike other MRF-based texture synthesis approaches, their combined system can both match and adapt local features with considerable variability. We therefore adapted the method in Li and Wand [15] to develop our new portrait style transfer method in this paper.

The rest of this paper is organized as follows. Section 2 provides a brief overview of related work. Section 3 introduces the improved deep CNN architecture. Section 4 presents the portrait style transfer algorithm and details of automatic semantic mask extraction. Section 5 provides experimental results and discussions. Finally conclusions are drawn in section 6.

2. Related Work

The success of deep CNNs (DCNNs) in image processing has also raised interest in image style transfer [16]. Luan et al. [17] proposed a new style transfer method which can constrain the transformation from the input to the output to be locally affined in color space for photographic style transfer. Shih et al. [18] proposed a new style transfer method for headshot portraits.Their new multiscale technique was based on deep networks to robustly transfer the local statistics of an example portrait onto a new one. Yao et al. [19] developed an attention-aware multi-stroke style transfer model, which produce multiple feature maps reflecting different stroke patterns. Li et al. [20] proposed a fast style transfer method which could be used in video style transfer. Recently, Generative Adversarial Networks (GAN) became a popular method for solving problems in computer vision and graphics. Zhang and Dana [21] built a Multi-style Generative Network (MSG-Net), which achieved real-time performance. Li and Wand [22] proposed a new model named MGANs (Markovian Generative Adversarial Networks) to train generative neural networks for efficient texture synthesis. Chen et al. proposed Cartoon GAN, which can use real scenery photos as source images to generate "cartoons" [2]. Azadi et al. proposed a method based on GAN for font style transfer [13].

This paper presents a combined segmentation method for the portrait parts. The portraits are segmented into seven part masks automatically. We propose to add masks in portrait style transfer, since they preserve more information and are more robust when image regions have similar chances of belonging to multiple parts categories. They are used to capture elements in the style image and to preserve the structure of the content image. For the human face, in particular, we proposed a new segmentation method which use a more detailed segmentation, in which different facial parts such as the nose, eyes, eyebrows, face, and mouth are also automatically segmented.

3. Architecture

Visual Geometry Group (VGG) [23] has shown especial characteristics for style transfer and it is used by most existing methods to obtain feature maps for style transfer. We also use an augmented deep CNN architecture which is based on VGG for portrait style transfer. The deep CNN architecture combines pooling and convolution layers l with 3×3 filters (for example, the first layer after second pooling is named $Conv3\_1$). Like common deep CNNs, intermediate post-activation results denoted as x^l for the layer l consist of N channels, which capture patterns from the source images for each region of the image.

As shown in Figure 1 our augmented network chooses to use certain special layers

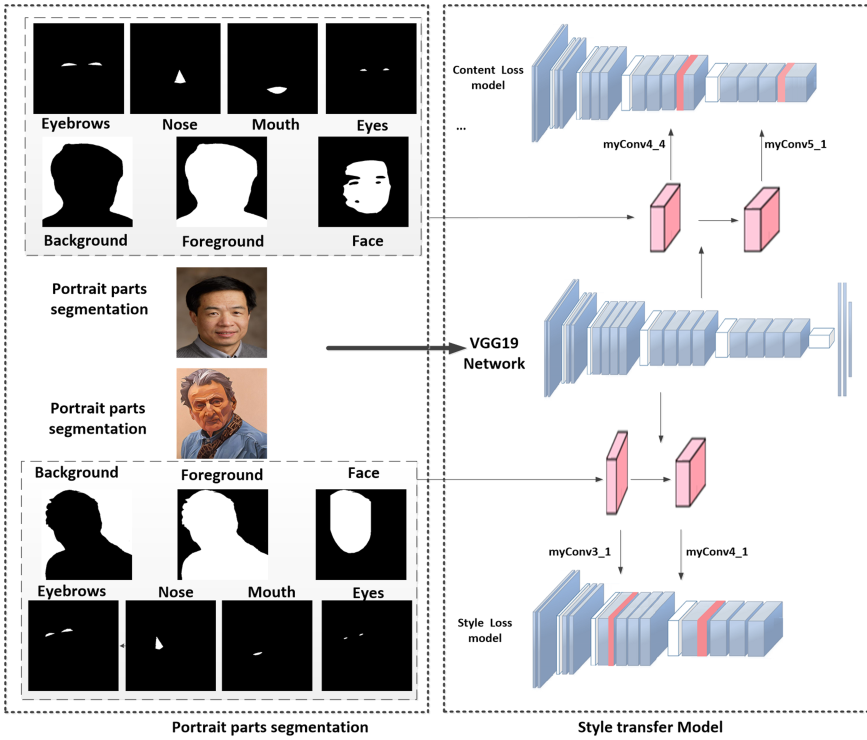


Figure 1: Portrait style transfer framework with deep neural networks.

to building content and style loss models, taking K masks as input. The K masks of content portrait and style portrait are edited as RGB images, then RGB images are down-sampled to produce semantic channels p^l at layer l with the same resolution as x^l. We concatenate them to form the new output with N + RGB channels, defined as d^l and labelled accordingly for each layer (e.g. $myConv4\_1$). In order to constrict the weight of masks, a parameter β is defined to balance their importance:

$$d^l = (x^l, \beta p^l). \tag{1}$$

Compared with other style transfer models, the main contributions of our augmented network are as follows:

1. We augment a trained deep convolutional neural network by concatenating semantic detection channels $n = 3$ (RGB) and channels of regular filters. Both the style and content images and their semantic detection results are input into the

augmented deep convolutional neural network. We set $\beta = 20$ which we found experimentally provides interesting results.

2. A novel loss function combining semantic detection is proposed. Unlike the common approaches, the new loss function consists of two parts: a new style loss which contains semantic detection map and a content loss. The semantic detection map is used to keep objects structures in the content image.

4. Portrait Style Transfer Optimization Function

For the convenience of describing our proposed method, all symbols used in the following text are listed below:

x_s, style image

x_c, content image

w_s, h_s, width and height in style image

w_c, h_c, width and height in content image

K, number of maps

m_s, semantic map of style image

m_c, semantic map of content image

L_s, style loss function

L_c, content loss function

$\Omega(x)$, feature maps of x

$\Omega(m_s)$, feature maps of m_s

$\Omega(m_c)$, feature maps of m_c

P, number of patches in the synthesised image

$\Gamma(\Omega(x))$, all the local patches from $\Omega(x)$

Given a style image $x_s \in \mathbb{R}^{3\times w_s \times h_s}$, a content image $x_c \in \mathbb{R}^{3\times w_c \times h_c}$, and semantic maps $m_{c_k} \in \mathbb{R}^{w_c \times h_c}$ and $m_{s_k} \in \mathbb{R}^{w_c \times h_c}$ associated with the content and style images, respectively ($k = 1, 2, \ldots, K$), our proposed portrait style transfer method introduces an augmented loss function based on a patch-based approach [15] for style transfer. It combines an MRF and a deep CNN model to minimise content reconstruction error L_c

and style remapping error L_s.

For simplicity, the semantic masks for the content and style images are also collectively represented as $m_c \in \mathbb{R}^{w_c \times h_c \times K}$ and $m_s \in \mathbb{R}^{w_s \times h_s \times K}$. The style transfer result image is denoted by $x \in \mathbb{R}^{3 \times w_c \times h_c}$. Since the synthesised image x is expected to have the same semantic layout as the content image, we treat m_c also as the semantic masks for the synthesised image. During our method, we make the high-level neural encoding of x similar to x_c and use the local patches similar to patches in x_s. As a result, the style of x_s is transferred onto the layout of x_c. Meanwhile, the semantic masks are also used to penalise patch matches. During our method, an energy function is defined as follows and seek x that minimises it:

$$L(x) = \mu_1 L_s(\Omega(x), \Omega(x_s), \Omega(m_c), \Omega(m_s)) + \mu_2 L_c(\Omega(x), \Omega(x_c)). \tag{2}$$

where $\Omega(m_c)$ and $\Omega(m_s)$ are the semantic masks of the content and style images down-sampled to the same resolution as $\Omega(x)$ and $\Omega(x_s)$. For our method, L_s aims to penalise inconsistencies in neural activations and/or semantic masks between x and x_s. L_c computes the squared distance between the feature map of the synthesised image and that of the content source image x_c.

The modified energy function L_s incorporates semantic masks as.

$$\begin{aligned} L_s(\Omega(x), \Omega(x_s), \Omega(m_c), \Omega(m_s)) = &\sum_{i=1}^{P} \| \Gamma_i^*(\Omega(x)) - \Gamma_{\phi(i)}^*(\Omega(x_s)) \|^2 \\ &+ \sum_{i=1}^{P} \sum_{k=1}^{K} \| \tilde{\Gamma}_i(\Omega(m_{c_k})) - \tilde{\Gamma}_{\phi(i)}(\Omega(m_{s_k})) \|^2, \end{aligned} \tag{3}$$

Each patch in $\Gamma_i(\Omega(x))$ has size $t \times t \times N$, and t is the width and height of the patch, and N is the number of channels. Similarly, $\tilde{\Gamma}(\Omega(m_{c_k}))$ and $\tilde{\Gamma}(\Omega(m_{s_k}))$ are the down-sampled semantic masks of extracted patches, each of size $t \times t$. For each patch $\Gamma_i(\Omega(x))$ with semantic masks $\tilde{\Gamma}_i(\Omega(m_{c_k}))$ we find its best matching patch $\Gamma_{\phi(i)}(\Omega(m_s))$ or $\Gamma_{\phi(i)}(\Omega(x_s))$ using normalised cross-correlation over all P_s example patches in $\Gamma^*(\Omega(x_s))$:

$$\phi(i) := \underset{j=1,\ldots,P_s}{argmax} \frac{\Gamma_i^*(\Omega(x)) \cdot \Gamma_j^*(\Omega(x_s))}{| \Gamma_i^*(\Omega(x)) | \cdot | \Gamma_j^*(\Omega(x_s)) |}, \tag{4}$$

Where $\Gamma_i^*(\Omega(x)) = \left(\Gamma_i(\Omega(x)), \beta\tilde{\Gamma}_i(\Omega(m_c))\right)$ is the concatenation of neural activation and semantic masks for the $i-$th patch of the synthesised image, and $\Gamma_j^*(\Omega(x_s)) =$

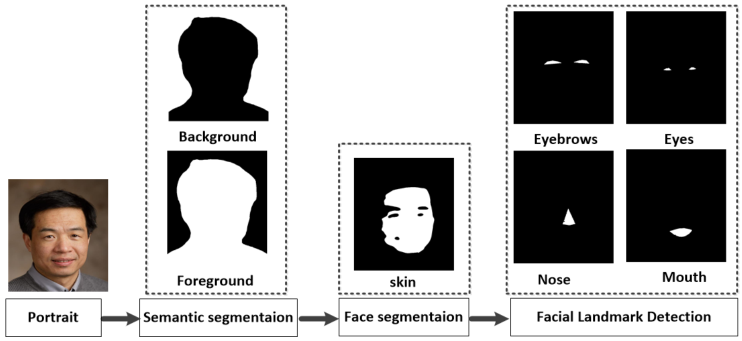


Figure 2: Portraits segmentation framework.

$\left(\Gamma_j(\Omega(x_s)), \beta\tilde{\Gamma}_j(\Omega(m_s))\right)$ is the concatenation of neural activation and semantic masks for the $j-$th patch of the style image. The nearest patch thus takes both style similarity and semantic consistency into account.

L_c is defined as the squared Euclidean distance between $\Omega(x)$ and $\Omega(x_c)$:

$$L_c(\Omega(x), \Omega(x_c)) = \|(\Omega(x) - \Omega(x_c)\|^2. \tag{5}$$

4.1. Facial Masks Extraction

In this paper we aim to automatically extract facial masks. Obviously, this would make mask-based style transfer more convenient for the user. Our automatic facial mask extraction step are shown in Figure 2.

In our method, facial component masks are automatically extracted using a combination of semantic segmentation, facial landmark detection, and face segmentation.

4.2. Semantic Image Segmentation

Zhou et al.[24] proposed a fusion model using Flexible Segmentation Graph (FSG) to explore multi-scale contexts for scene labelling problems. Zheng et al.[25] proposed a semantic segmentation method named Conditional Random Field Recurrent Neural Networks (CRF-RNN). CRF-RNN achieves a good result on the popular Pascal Visual Object Classes segmentation benchmark. This improvement can be attributed to the uniting of the strengths of CNNs and CRFs in a single deep network[26, 27] . In our work, we use CRF-RNN to produce semantic maps. These are treated as maps

(a) content (b) Semantic map for (a) (c) style (d) Semantic map for (c)

Figure 3: Semantic segmentations using CRF-RNN [25] for person.

predicting the probability of each pixel belonging to each object category. An example is shown in Figure 3.

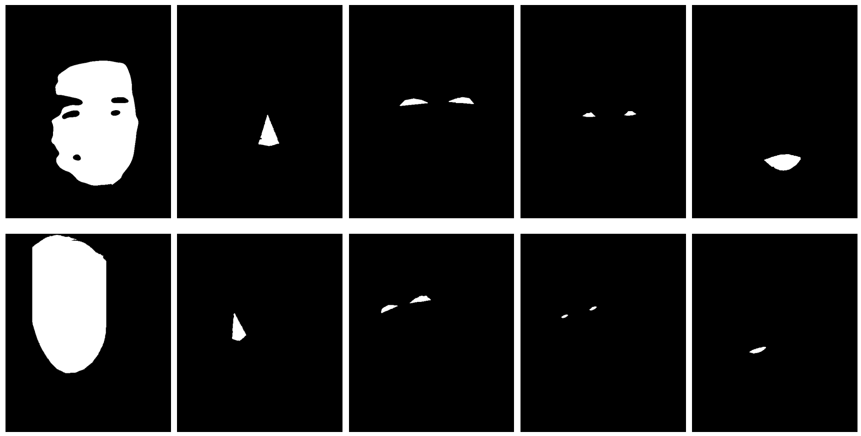
4.3. Face and Facial Part Segmentation

Skin detection is performed on the photographic images [28], using a rule-based analysis of pixels in YCbCr colour space. The skin mask is then intersected with the person mask provided by the CRF-RNN, so as to subdivide the person into skin and non-skin (e.g. hair, clothing). An example is shown in Figure 4.

Since skin detection is primarily colour based, it is generally not effective on artwork due to the typical colour shifts, as well as distortions caused by strong brush stroke textures. Therefore, for paintings, the facial region is detected using the face detector, rather than using skin detection.

Facial landmark detection is performed using OpenFace [29], which is based on Conditional Local Neural Fields, a version of the well-known Constrained Local Model approach. Sixty-eight facial landmarks are located, from which the eye, nose, inner and outer mouth regions are determined – see Figure 4.

Since the facial landmarks only cover the lower half of the face, the outline of the face is extended upwards, and intersected with the person mask provided by the semantic segmentation to produce a good approximation to the head region. This mask is used for artwork. For photographs the skin mask is used instead of the extended facial region as it is more accurate (although prone to noise).



(a) face (b) nose (c) eyes (d) eyes (e) mouth

Figure 4: Segmentation of facial components for the images in Figure 3.

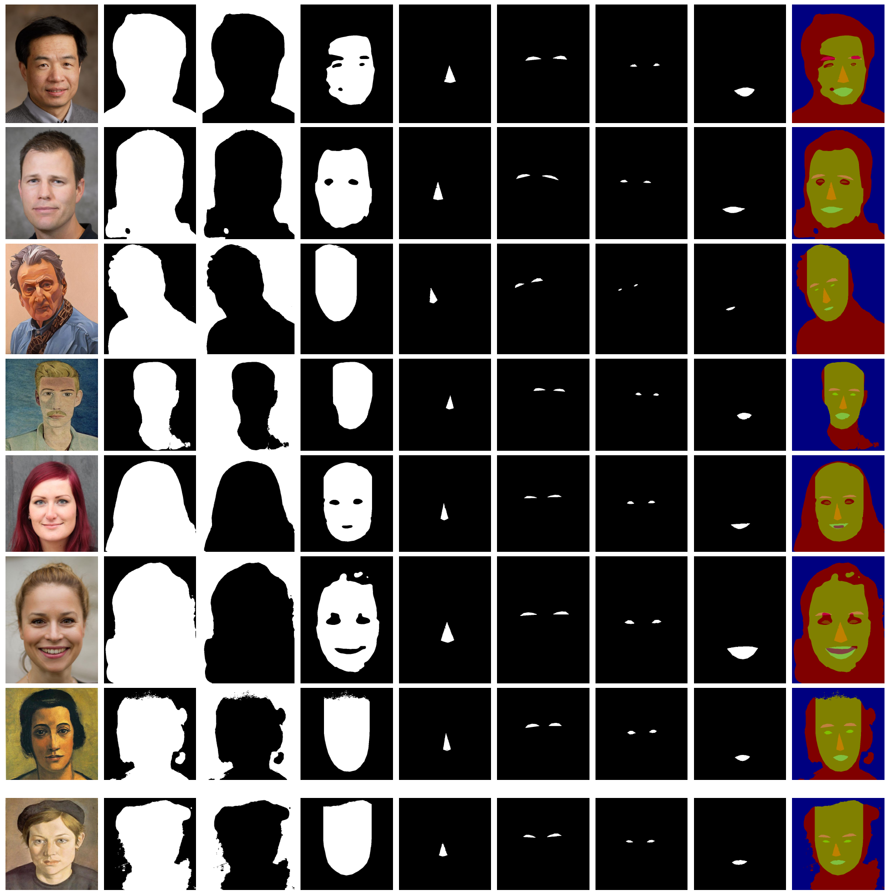
4.4. *Portrait parts segmentation results*

More portrait segmentation results are shown in Figure 5. One can see that the proposed portrait parts segmentation approach can achieve a good result in facial part segmentation.

Semantic segmentation is a difficult problem. For some images which are very abstract or contain high light, the CRF-RNN method cannot segment them very well. Some examples are shown in the following Figures 6, such as the faces in Figures 6 (b) and ïijĹdïijĽ, can not be segmented correctly.

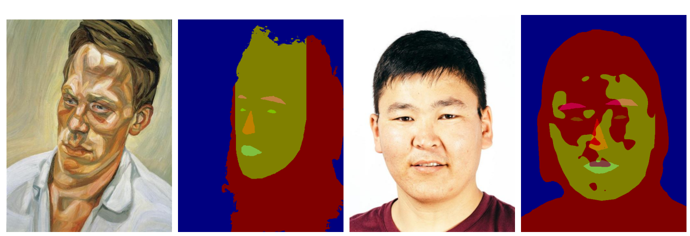
5. Results

We use the pre-trained 19-layer VGG-Network with the augmented style layers $myConv3\_1$ and $myConv4\_1$, and content layers $myConv5\_1$. For layers $myConv3\_1$, $myConv4\_1$, and $myConv5\_1$, we use 3×3 patches and set the stride to one. Following the patch-based approach of [15], we synthesise at multiple increasing resolutions, and randomly initialise the optimisation. On a GTX Titan with 12Gb of GPU RAM, synthesis takes from 5 to 30 minutes depending on the output quality and resolution. During our experiments, we compare the proposed method with several popular methods: Gatys



(a) Portrait (b) Back- ground (c) Fore- ground (d) Face (e) Nose (f) Eyebrows (g) Eyes (h) Mouth (i) RGB Masks

Figure 5: Content (rows 1,2,5,6) and style (rows 3,4,7,8) portraits segmentation.



(a) input (b) segmentation (c) input (d) segmentation

Figure 6: Failed examples of semantic segmentation.

et al. [1], Li and Wand [15], Ulyanov et al. [30], LFYW [31] and MGANs [22] which are representative global and local neural style transfer.

5.1. Comparison

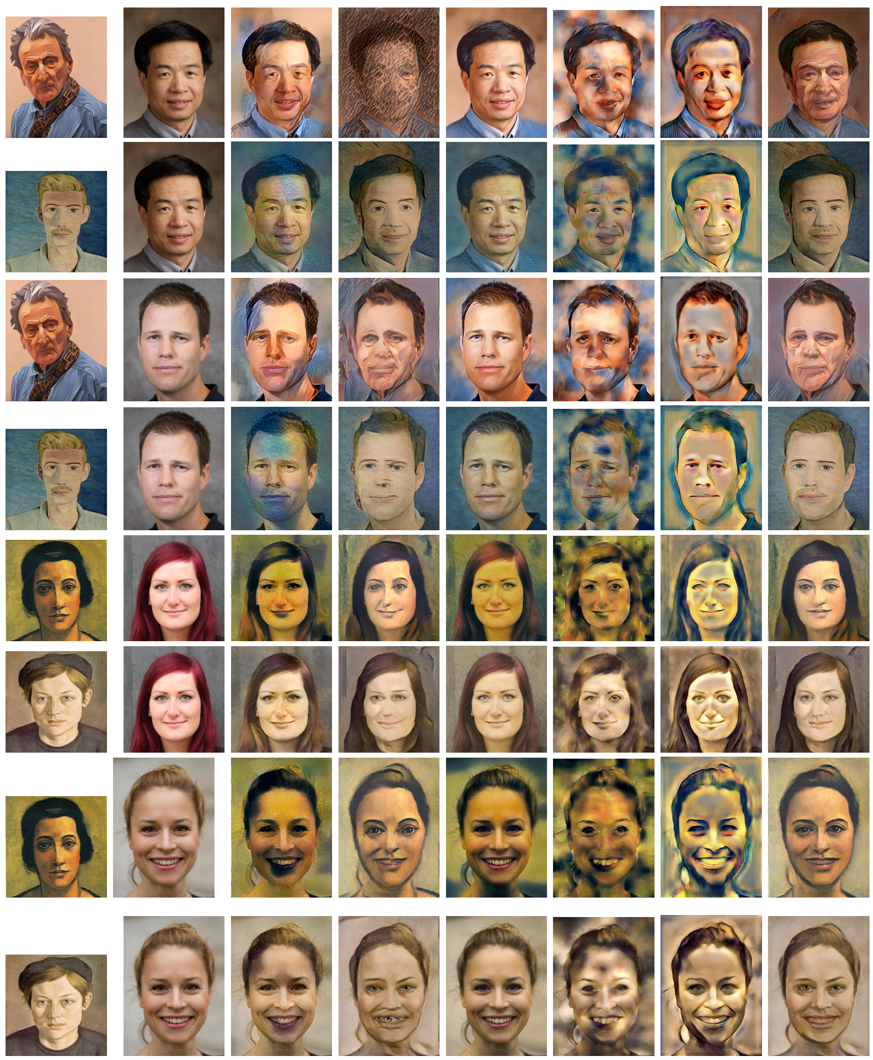
Given portrait content and style images in Figures 5, Figures 7 show style transfer applied separately to photographs of men and women. We transfer the style of each portrait style image to each content image. We can see from Figures 7 that our method can achieve better results than the CNNMRF method and avoid errors in applying style transfer to inappropriate parts. The style portrait images contain a range of simple and more complicated textures. In both cases, our method achieves effective results, and preserves the content of the portraits. the method of [15] can also achieve interesting results, but only for simple texture images. Further more, for some examples of Figures 5, our method can achieve better results in specific parts such as the mouth and eyes. For some examples of Figures 5(g),(h), our method can achieve better results in specific parts including eyebrows, mouth and eyes and mouth area.

5.2. Style transfer of other objects

More examples of style transfer for objects such as boat, and bus segmentation results are shown in Figure 8. Their corresponding style transfer results are shown in Figures 9. Our method achieves effective results, and preserves the content of the images. Li and Wand's method [15] can also achieve interesting results, but the results of their method still have many errors in which styles are misapplied. Such as in row 2 in Figures 9 (d), the boat style is transferred to sky. Gatys et al. [1] cannot transfer enough style to content images. Such as in row 1,2 in Figures 9 (c), very little style information is transferred. Ulyanov et al. [30] and LFYW [31] and MGANs [22] generate some imperfect results too. Such as in Figures 9 (f) and (g), there are many artifacts. Our method can achieve better results in specific parts in the background area.

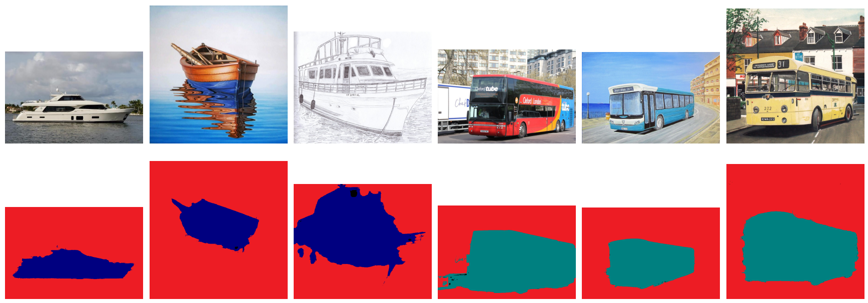
5.3. User evaluation.

In addition to visual inspection, we also performed a quantitative comparison with five existing methods. Since there is no standard automatic style transfer measure or



(a) style (b) content (c) Gatys et al. [1] (d) Li and Wand [15] (e) Ulyanov et al. [30] (f) LFYW [31] (g) MGANs [22] (h) Our

Figure 7: Portrait style transfer comparison.



(a) Content (b) style (c) style (d) Content (e) style (f) style

Figure 8: Content and style image (the first row) and segmentation results (the second row).

(a) style (b) content (c) Gatys et al. [1] (d) Li and Wand [15] (e) Ulyanov et al. [30] (f) LFYW [31] (g) MGANs [22] (h) Our

Figure 9: Style transfer comparison on other objects.

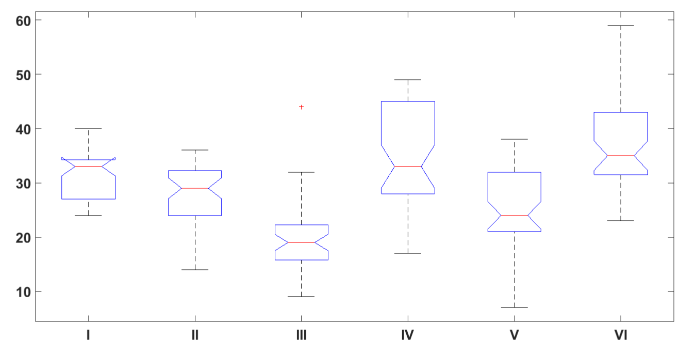


Figure 10: Boxplots of user preferences for six different style transfer methods in the task , showing the mean (red lines), quartiles (blues lines), and extremes (black lines) of the distributions.

test, we performed a user study in which the users were presented with a style image, a content image, and stylised output images from the following six methods: I [1], II [15], III [31], IV [30], V [22] and VI which is our method,.

The user study is designed using the 2AFC (Two-Alternative Forced Choice) paradigm, widely used in perceptual studies due to its simplicity and reliability. In each trial the user was asked to complete two tasks by answering the following question:

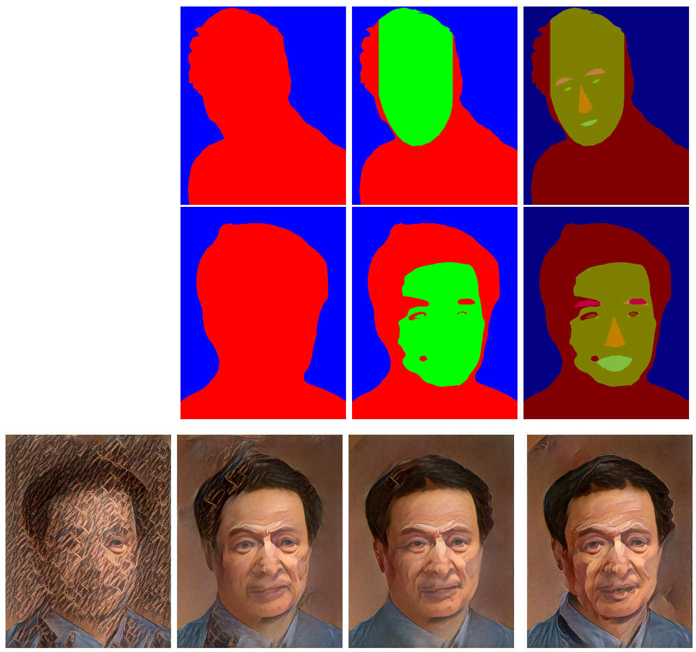
- Task : Given the two result images, which image do you prefer?

For the question, the user can choose either of the two result images. To make the comparison more meaningful while limiting the user effort to a reasonable level, we used the full set of results that were contained in Figures 7, and 9. The sets contain $12 \times 6 = 72$ test images (8 sets for human faces and 4 sets without human faces) and stylised results generated by the six methods, which total 96 images. During the task, 45 users (ages ranging from 19 to 40) participated in the user study where we randomised the order of image pairs. Altogether, the results of each method were compared against $12 \times 5 = 60$ results of alternative methods. We recorded the total number of user preferences (clicks) for each method, and treat these as random variables.

We performed the ANOVA test and the results are shown in Figures 10. The p-values comparing our method and alternative methods are shown in Table 1. They show

Table 1: The p-value of the ANOVA test of the proposed method against the other methods for both tasks

method	I [1]	II [15]	III [31]	IV [30]	V [22]
Task	0.0141	9.81808e-05	1.02749e-05	0.0219	4.45077e-06



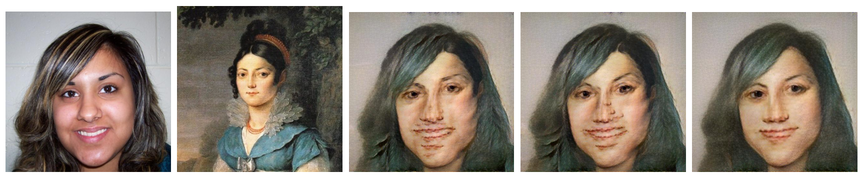
(a) no segmentation (b) 2 parts (c) 3 parts (d) 7 parts

Figure 11: Result showing the effects on style transfer with an increasing number of facial segmentation.

that the method proposed in this paper has the highest mean score and is preferred by the majority of the users. The difference between our method and alternative methods is statistically significant (at the level of 0.05).

5.4. Modifying the number of segmentation parts.

The semantic segmentation significantly affects the style transfer results. Figure 11 shows an experiment in which the number of labels in the facial segmentation is increased, and demonstrates the importance of separately labelling all the major components of the face.



(a) content (b) style (c) (d) (e)

Figure 12: Comparison of different scale parameters, (c) rotation=1, scale=1, (d) rotation=3, scale=1, (e) rotation=3, scale=4

5.5. Modify image scale parameters

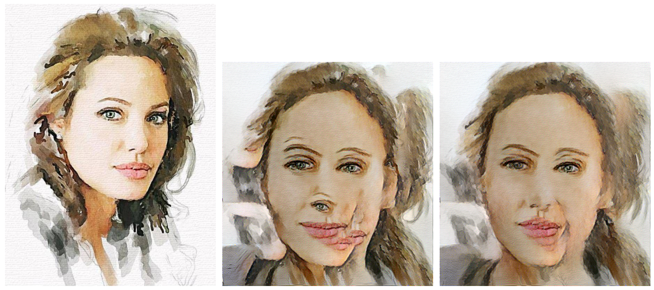
Normally, optimal results are obtained if objects in the source and style portraits have similar scales and orientations, otherwise suitable patches cannot be matched or transferred. In order to partially overcome the perspective and scale difference, we also define the scale parameter and rotation parameter to change the portrait size, and obtain more patches. If there are scale or orientation differences then this can be overcome by adding the scale and orientation parameters, but it also means the running time will increase too. Usually, we define rotation=N_1 and rotational copies are 0,$\pm\pi \setminus 24 * N_1$. We define scale=$N_2$, and scales are 0, $(1.0 \pm 0.05 * N_1)$ in Figure 12,rotation=1 and scale=1, we create three rotational copies $\{-\pi \setminus 24$,0, $\pi \setminus 24\}$ and for each rotational copy we generate three scales $\{0.95, 1, 1.05\}$. We modify rotation=3 , so the rotation copies are $\{-3\pi \setminus 24, -\pi \setminus 12, -\pi \setminus 24, 0, \pi \setminus 24, \pi \setminus 12, 3\pi \setminus 24\}$ and scale=4, so the sizes are $\{0.8, 0.85, 0.9, 0.95, 1, 1.05, 1.10, 1.15, 1.2\}$. The comparison results are shown in Figure 12.

We can see that, by modifying the scale parameter the results are improved. In the future we will do further experiments and find out their correlations.

5.6. Failure cases

Failures in the segmentation will cause some background texture to be embedded into the foreground elements in the synthesised image, thereby generating unsatisfactory results. One example is shown in Figure 13.

Due to failed semantic segmentation in the human area, the style of face in Figure 13(a) is transferred to the background in Figure 13(c), although compared with



(a) style image (b) Li and Wand [15] (c) Our

Figure 13: Result showing the effects on style transfer with faulty semantic segmentation.

Figure 13(b), our method still greatly improves the result greatly.

6. Conclusions

Our paper demonstrates the benefits of automatic semantic mask extraction by combining state-of-the-art methods for both semantic segmentation and facial features. The correctness and accuracy of the semantic masks are critical. Using masks helps mitigate this, but there is certainly scope to improve semantic segmentation, or to develop methods dedicated to generating semantic masks.

The segmentation result will affect the style transfer result. Better segmentation can generate more precise masks, which means that more appropriate patches will be matched and chosen, so that fewer error parts are transferred. In future, we will try to use different segmentation methods to obtain semantic maps.

Acknowledgements

This work was supported by National Natural Science Foundation of China [grant number 61733004, 61503128, 61602402,61573133], Science and Technology Plan Project of Hunan Province [grant number 2016TP1020], Scientific Research Fund of Hunan Provincial Education Department [grant number 18A333], Hengyang guided science and technology projects and Application-oriented Special Disciplines [grant number Hengkefa [2018]60-31], Double First-Class University Project of Hunan Province [grant number Xiangjiaotong [2018]469] and Scientific and Technological Development Project of Hengyang City [grant number (2018)No.60-31]. We would like to thank NVIDIA for the GPU donation.

References

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

[2] Y. Chen, Y.-K. Lai, Y.-J. Liu, Cartoongan: Generative adversarial networks for photo cartoonization, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 9465–9474.

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

[4] H. Lu, B. Li, J. Zhu, Y. Li, Y. Li, X. Xu, L. He, X. Li, J. Li, S. Serikawa, Wound intensity correction and segmentation with convolutional neural networks, Concurrency and computation: practice and experience 29 (6) (2017) e3927.

[5] S. Serikawa, H. Lu, Underwater image dehazing using joint trilateral filter, Computers & Electrical Engineering 40 (1) (2014) 41–50.

[6] H. Lu, Y. Li, M. Chen, H. Kim, S. Serikawa, Brain intelligence: go beyond artificial intelligence, Mobile Networks and Applications 23 (2) (2018) 368–375.

[7] M. Ruder, A. Dosovitskiy, T. Brox, Artistic style transfer for videos, in: German Conference on Pattern Recognition, Springer, 2016, pp. 26–36.

[8] A. Criminisi, P. Pérez, K. Toyama, Region filling and object removal by exemplar-based image inpainting, IEEE Transactions on Image Processing 13 (9) (2004) 1200–1212.

[9] A. Selim, M. Elgharib, L. Doyle, Painting style transfer for head portraits using convolutional neural networks, ACM Transactions on Graphics (TOG) 35 (4) (2016) 129–146.

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

[11] H. Lu, Y. Li, T. Uemura, H. Kim, S. Serikawa, Low illumination underwater light field images reconstruction using deep convolutional neural networks, Future Generation Computer Systems 82 (2018) 142–148.

[12] X. Deng, Enhancing image quality via style transfer for single image super-resolution, IEEE Signal Processing Letters 25 (4) (2018) 571–575.

[13] S. Azadi, M. Fisher, V. G. Kim, Z. Wang, E. Shechtman, T. Darrell, Multi-content gan for few-shot font style transfer, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7564–7573.

[14] H. Lu, Y. Li, S. Mu, D. Wang, H. Kim, S. Serikawa, Motor anomaly detection for unmanned aerial vehicles using reinforcement learning, IEEE internet of things journal 5 (4) (2017) 2315–2322.

[15] C. Li, M. Wand, Combining markov random fields and convolutional neural networks for image synthesis, in: Proc. Conf. Computer Vision and Pattern Recognition, 2016, pp. 2479–2486.

[16] F. Shen, S. Yan, G. Zeng, Neural style transfer via meta networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 8061–8069.

[17] F. Luan, S. Paris, E. Shechtman, K. Bala, Deep photo style transfer, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4990–4998.

[18] Y. Shih, S. Paris, C. Barnes, W. T. Freeman, F. Durand, Style transfer for headshot portraits, ACM Transactions on Graphics (TOG) 33 (4) (2014) 135–148.

[19] Y. Yao, J. Ren, X. Xie, W. Liu, Y.-J. Liu, J. Wang, Attention-aware multi-stroke style transfer, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 1467–1475.

[20] X. Li, S. Liu, J. Kautz, M.-H. Yang, Learning linear transformations for fast image and video style transfer, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 3809–3817.

[21] H. Zhang, K. Dana, Multi-style generative network for real-time transfer, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 1–8.

[22] C. Li, M. Wand, Precomputed real-time texture synthesis with markovian generative adversarial networks, in: European Conference on Computer Vision, Springer, 2016, pp. 702–716.

[23] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556.

[24] Q. Zhou, B. Zheng, W. Zhu, L. J. Latecki, Multi-scale context for scene labeling via flexible segmentation graph, Pattern Recognition (2016) 312–324.

[25] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, Du, Conditional random fields as recurrent neural networks, in: Proceedings of the IEEE international conference on Computer Vision, 2015, pp. 1529–1537.

[26] Q. Zhou, J. Cheng, H. Lu, Y. Fan, S. Zhang, X. Wu, B. Zheng, W. Ou, L. J. Lateckii, Learning adaptive contrast combinations for visual saliency detection, Multimedia Tools and Applications (2018) 1–29.

[27] Q. Zhou, W. Yang, G. Gao, W. Ou, H. Lu, J. Chen, L. J. Latecki, Multi-scale deep context convolutional neural networks for semantic segmentation, World Wide Web (2019) 555–570.

[28] N. Brancati, G. De Pietro, M. Frucci, L. Gallo, Human skin detection through correlation rules between the ycb and ycr subspaces based on dynamic color clustering, Computer Vision and Image Understanding 155 (2017) 33–42.

[29] T. Baltrušaitis, P. Robinson, L.-P. Morency, Openface: an open source facial behavior analysis toolkit, in: 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), 2016, pp. 1–10.

[30] D. Ulyanov, A. Vedaldi, V. Lempitsky, 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, 2017, pp. 6924–6932.

[31] Y. Li, C. Fang, J. Yang, Z. Wang, X. Lu, M.-H. Yang, Universal style transfer via feature transforms, in: Advances in Neural Information Processing Systems, 2017, pp. 386–396.

Authors

Hui-Huang Zhao received his Ph.D. degree in 2010 from XiDian University. He was a Sponsored Researcher in the School of Computer Science and Informatics, Cardiff University. Now he is an Associate Professor in the College of Computer Science and Technology, Hengyang Normal University. His main research interests include Solder Joint Inspection, Compressive Sensing, Machine Learning, and Image Processing.

Jing-Hua Zheng in-HuaZheng received a doctorate in control theory and control engineering from Central South University in 2000. He is a professor of information engineering at the Xiangtan University. His main research interests include evolutionary computation, multi-objective genetic algorithm, and machine learning.

Yao-Nan Wang Yao-Nan Wang received his Ph.D. degree in 1995 from Hunan University. Now he is a Professor, Ph.D. supervisor and the dean of College of Electrical and Information Engineering of Hunan University. His main research interests include Intelligent Control and Pattern Recognition.

Xiao-Fang Yuan received his Ph.D. degree in 2008 from Hunan University. Now he is a Professor, Ph.D. supervisor of College of Electrical and Information Engineering of Hunan University. His main research interests include Intelligent Control and Pattern Recognition.

Yu-Hua Li received the Ph.D. degree in general engineering from the University of Leicester, Leicester, U.K. He was a Senior Research Fellow with Manchester Metropolitan University, Manchester, U.K., and a Research Associate with the University of Manchester, Manchester, from 2000 to 2005. He is currently a in the School of Computer Science and Informatics, Cardiff University, Cardiff, U.K. His current research interests include Pattern Recognition, Machine Learning and Data Science.