Photorealistic Style Transfer via Adaptive Filtering and Channel Separation

Hong Ding[1,2]
[1]School of Big Data and Artificial Intelligence, Guangxi University of Finance and Economics
[2]School of Computer Science
Wuhan University
dhong20123@163.com

Fei Luo*
School of Computer Science
Wuhan University
luofei@whu.edu.cn

Caoqing Jiang
Guangxi Key Laboratory of Big Data in Finance and Economics, Guangxi University of Finance and Economics
jcqng@163.com

Gang Fu
School of Computer Science
Wuhan University
xyzgfu@gmail.com

Zipei Chen
School of Computer Science
Wuhan University
czpp19@whu.edu.cn

Shenghong Hu
Information Engineering School
Hubei University of Economics
wuhanhush@126.com

Chunxia Xiao*
School of Computer Science
Wuhan University
cxxiao@whu.edu.cn

(a) Reference (b) Content (c) An *et al.* [1] (d) Ours (e) Reference (f) Content (g) Li *et al.* [2] (h) Ours

Figure 1: Comparison with state-of-the-art methods. (a)-(d) is the facial image case, and (e)-(h) is the scene image case. Our model generates better results through adaptive filtering, channel separation and feature preservation methods.

ABSTRACT

The problem of color and texture distortion remains unsolved in the photorealistic style transfer task. It is mainly caused by the interference between color and texture during transferring. To address this problem, we propose a end-to-end network via adaptive filtering and channel separation. Given a pair of content image and reference image, we firstly decompose them into two structure layers through adaptive weighted least squares filter (AWLSF), which could better perceive the color structure and illumination. Then, we carry out RGB transfer in a channel separation way on the two generated structure layers. To deal with texture in a relatively independent manner, we use a module and a subtraction operation to get more complete and clear content features. Finally, we merge the color structure and texture detail into the ultimate result. We conduct solid quantitative experiments on four metrics NIQE, AG, SSIM, and PSNR, and make a user study. The experimental results demonstrate that our method is able to produce better results than previous state-of-the-art methods, and validate the effectiveness and superiority of our method.

CCS CONCEPTS

• **Computing methodologies** → **Computer graphics**.

KEYWORDS

Photorealistic style transfer, Adaptive filters, Channel separation, Feature synthesis

ACM Reference Format:
Hong Ding[1,2], Fei Luo*, Caoqing Jiang, Gang Fu, Zipei Chen, Shenghong Hu, and Chunxia Xiao*. 2022. Photorealistic Style Transfer via Adaptive Filtering and Channel Separation. In *Proceedings of Proceedings of the 30th ACM International Conference on Multimedia (MM '22).* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3503161.3548104

*Corresponding authors

MM '22, October 10–14, 2022, Lisboa, Portugal.

ACM ISBN 978-1-4503-9203-7/22/10…$15.00
https://doi.org/10.1145/3503161.3548104

1 INTRODUCTION

Image style editing is a fundamental task in the image processing domain. For example, the artistic style transfer [3–5] transfers the color and texture information between a photo and a painting image. The photorealistic style transfer [2, 6–12] mainly transfers the color of the reference image to the content image, while the content feature is unchanged.

In communities of artistic style transfer and photorealistic style transfer, there are two important representative methods. One is from Gatys *et al.* [3], and the other is from Li *et al.* [5]. Gatys *et al.* introduced an artistic image style transfer method by separating style and feature via gram and feature matrix. However, the model fails to realize complete individual texture preservation and color transfer because the loss function includes both the content and style. Subsequent methods [6, 7, 13] are further proposed to improve Gatys' work. Li *et al.* [5] proposed universal style transfer via feature transforms. It aims to transfer arbitrary visual styles to content images by a pair of feature transforms, whitening and coloring (WCT) are embedded in an image reconstruction network. Even though WCT can generate high-quality stylized images and has speed advantage, it fails to preserve features in content images well. There are many ways [1, 2, 8–10, 14] that improve the WCT to generate art or realistic images. Nevertheless, these methods perform color transfer on the feature extraction result and will lose texture details. For both two representative methods and their following ones, they would partially lose texture details (as shown in Fig. 1 (c) and (g)).

To address mutual interference between color and texture, this paper performs photorealistic style transfer via adaptive filter and channel separation (AFCS) framework. Briefly, we first adaptively extract both the structure and texture feature layers of the content and reference images. Then, we leverage RGB channel separation module on the structure layers to achieve clear and beautiful color transfer results. Simultaneously, texture feature extraction module extracts deep texture features with two ways to obtain full image feature. Finally, we merge the processed structure layer and the texture features to the ultimate image. Our results are shown in Fig. 1 (d) and (h). All modules are realized in one end-to-end network, whose diagram is illustrated in Fig. 2. The main novelty of our proposed method is to perform color transfer and feature editing respectively.

We evaluate our AFCS method on a variety of test images (including facial images and scene-level images) from the Internet and public datasets, such as the IMDB-WIKI dataset [15] and COCO dataset [16], to validate the superiority of our proposed method. The major contributions of this work can be summarized as follows:

(1) We introduce a novel method to independently deal with photorealistic style transfer and feature preservation, which could avoid color and feature interference.

(2) We propose an adaptive weighted least squares filter (AWLSF) and guide the filtering to obtain proper illumination transfer results.

(3) We design a feature-preserving feature extraction module.

2 RELATED WORK

Artistic style transfer. Gatys *et al.* [3] first used image representations derived from CNN optimised for artistic style transfer. However, their framework requires a slow iterative optimization process, which limits its practical application. Huang *et al.* [7] proposed a simple yet effective approach that firstly enables arbitrary style transfer in real-time. Yao *et al.* [4] introduced multiple feature maps reflecting different stroke patterns, which allows integrating multiple stroke patterns into different spatial regions of the output image harmoniously. However, these two methods fail to work well for photorealistic style transfer because of their limited feature preserving ability. Zhao *et al.* [17] proposed an automatic semantic style transfer using deep convolutional neural networks and soft masks. Their method removes some texture information of the content image and produces effects with more artistic style. Chen *et al.* [18] proposed image sentiment transfer using the filtered Visual Sentiment Ontology (VSO) dataset, and exhibited limited application because of the special datasets.

Photorealistic style transfer. To obtain photorealistic image style transfer, Luan *et al.* [6] combined content loss and style loss with semantic segmentation. Their parameters affect both content and style and interfere with texture preservation and color transfer. Li *et al.* [14] designed PhotoWCT to maximize the stylization effect. Although the strategies are valid, they fail to solve the texture loss problem, resulting in blurry artifacts. Yoo *et al.* [8] proposed WCT^2 with wavelet pooling and unpooling to correct transfer based on whitening and coloring transforms. However, the algorithm generates unnatural boundaries. Li *et al.* [2] presented data-driven fashion to transfer different levels of styles. An *et al.* [9] performed style transfer via photoNet and multiple style transfer modules. Later An *et al.* proposed ArtFlow [1] to prevent content leak during universal style transfer later. Hong *et al.* [10] introduced domain-aware style transfer networks. These style transfer methods often destroy the content detail of the input image and even blur the whole image. Makeup transfer only focuses transfering eye shadow, lipstick, the skin color, and not background [19–23]. These methods do not work well for image pair without strict dense corresponding. Further more, above methods have limited ability of keeping the content feature. Our work independently performs color-transfer and feature-preserving to solve this problem.

Edge-preserving image decomposition. The methods based on weighted average [24, 25] have been widely developed in past decades. Farbman *et al.* [26] proposed an edge-preserving smoothing operator based on the weighted least squares (WLS) optimization framework. This method does not adjust parameters for different images adaptively. Barron *et al.* [27] proposed a bilateral solver to accelerate the WLS filter smoothing. However, it is only capable of Gaussian guidance weights, and produce artifacts in the results. Liu *et al.* [28] proposed a global optimization-based method without additional information of a guidance image. Fanetal *et al.* [29, 30] performed images smoothing through convolutional neural networks. Most of the above-mentioned approaches are limited to a few applications because their inherent smoothing natures are usually fixed. Hence, we design an adaptive selection formula for the L parameter of [26] to guide the image smoothing for image photorealistic style transfer. Furthemore, Yim *et al.* [31] adjust the brightness, contrast and other aesthetic setting to produce filter effects. The work [31] and image smoothing have different meanings.

3 METHOD

We propose photorealistic color transfer via adaptive filter and channel separation (AFCS). As shown in Fig. 2, we utilize the adaptive weighted least squares filter (AWLSF) to smooth both the content (C) and reference images (R), perform color transformation in their structure layers, extract accurate textures via a feature extraction module, and finally obtain the photorealistic style transfer result O by the merging module.

3.1 Adaptive weighted least squares filter

Image style transfer models have two inputs: a content image (C) and a reference style image (R). Experiments show that the illumination differences between them have impacts on image filtering and style transfer results under different parameter $\mathscr{L}$ of WLS, as shown in Figs. 4 and 5. $\mathscr{L}(IN)$ is a source image for the affinity matrix that has the same dimensions as IN, IN is the input image. However, there are no methods filtering for a pair of images so far. To adjust the key parameter $\mathscr{L}(IN)$ of WLS adaptively for any input image pair to obtain the better image color transfer effect, we propose an adaptive weighted least squares filtering (AWLSF) for an image pair (C and R).

We utilize the WLS filter [26] to extract the structure layers of C and R and obtain S_C and S_R, respectively. The filter uses the parameter $\mathscr{L}(IN)$ to guide filtering, where the default value is $log(IN)$. When $\mathscr{L}(IN)$ is related to the image brightness, such as the L channel in the LAB color space, the filter focuses more on the illumination information of IN and produces the results preserving more changes in the color illuminated (see the boys's hand and face in Fig. 5). However, using the L channel also has some disadvantages, for example, it often blurs the boundaries in the result (see Fig. 4). In this paper, we propose an adaptive $\mathbb{L}$ combining $log(L(IN))$ and $log(IN)$. The WLS filter [26] has the following quadratic form:

$$(I + \lambda\mathscr{L}(IN))S = IN. \tag{1}$$

Both R and C are input into Eq. 1 as IN. S is the smoothing result. The default value $\mathscr{L}(IN)$ is $log(IN)$. I is the unit matrix, and λ balances the data term and the smoothness term. When IN is C, we set $\mathbb{L}(C)$ as the weighted sum of $L(C)$ (the luminance channel of C) and C:

$$\mathbb{L}(C) = \alpha \times log(L(C)) + (1 - \alpha) \times log(C), \tag{2}$$

where $\alpha = \beta \times (\frac{\Delta L}{\tau} - 1)$. $\Delta L = |mean(L(C)) - mean(L(R))|$. $mean(\cdot)$ means the mean value of all elements in a matrix. τ is a domain value controlling the choice between $log(L(C))$ and $log(C)$. β is a trigger whose value is 0 or 1. When $\Delta L \geq \tau, \beta = 1$, and when $\Delta L < \tau, \beta = 0$. Here, we set $\tau = 0.5$. When IN is R, we compute $L(R)$ like $L(C)$. We leverage $L(C)$ to guide the smoothing image, to preserve not only the image color, but also the brightness change information.

We show the influence of the L component in Eq. 2 for photorealistic style transfer in Figs. 3 and 4, the results with varying a in Fig. 5 and the ablation study for smoothing and channel separation photorealistic style transfer in Fig. 7.

3.2 RGB channel separation photorealistic style transfer

Encoder and decoder architecture. We use the same encoder and decoder in WCT^2 [8] for color transfer and feather extraction. The encoder applies the ImageNet-pretrained VGG-19 network and Haar wavelet pooling [8] from the conv1_1 layer to the conv4_1 layer. The decoder has the mirror structure of the encoder. We improve the decoder by adding structure optimization behind the convolution layer in each decoder layer to repair the unnatural boundary effect (as shown in Fig. 6).

Channel separation photorealistic style transfer. To avoid the mutual interference of the R, G, and B channels, we perform style transfer in the R, G, and B channels to achieve color transfer of the structure layer S'_C (see Fig. 2). Although the RGB channel separation design has been used in some color related tasks [32, 33], to our best knowledge we are the first to propose a deep learning network based on color channel separation in the field of photorealistic style transfer. It is inspired by the divide-and-grow idea. We leverage semantic segmentation for C and R to obtain their masks M_C and M_R. In each R, G, or B channel, we perform color transformation by inputting the R, G, and B components of S_C and S_R : S_C_R, S_C_G, S_C_B, S_R_R, S_R_G and S_R_B, and their masks into the encoder-decoder modules (see Fig. 2). The color transfer principle is based on WCT [3].

We extract the feature of S_C from the decoder of WCT:

$$\hat{f}_j^C = E_j^C (D_j^C)^{-1/2} (E_j^C)^T f_j^C, \tag{3}$$

where j denotes the R, G, or B channel. D_j^C is a diagonal matrix with the eigenvalues of the covariance matrix $\hat{f}_j^C, (\hat{f}_j^C)^T \in \mathcal{R}^{Ch \times Ch}$, and E_j^C is the corresponding orthogonal matrix of eigenvectors, satisfying $\hat{f}_j^C (\hat{f}_j^C)^T = E_j^C D_j^C (E_j^C)^T$. f_j^C is the vectorized VGG feature of S_C. Ch is the number of channels.

We transfer the color from S_R to S_C:

$$\hat{f}_j^{C'} = E_j^R (D_j^R)^{-1/2} (E_j^R)^T \hat{f}_j^C, \tag{4}$$

where $\hat{f}_j^{C'}$ is the jth channel of the color transfer result C. D_j^R is a diagonal matrix with the eigenvalues of the covariance matrix $\hat{f}_j^R (\hat{f}_j^R)^T \in \mathcal{R}^{Ch \times Ch}$, and E_j^R is the corresponding orthogonal matrix of eigenvectors. Then, $\hat{f}_j^{C'} = \hat{f}_j^{C'} + m_R$, and m_R is the mean vector of $\hat{f}_j^C$.

We invert $\hat{f}_j^{C'}$ to the decoder to obtain the color transformation result O'_j in the RGB channel.

Structure Optimization (SO). WCT^2 [8] has unnatural boundary problems ((b) and (d) in Fig. 8) caused by the imprecision of semantic segmentation. We repair the output image by replacing the texture layer with that of C. Firstly, we extract the structure layer O'_j of the output result using WLS filter, then extract C_j texture layer, finally we fuse O' and C_j texture layer to generate the style transfer result. This method can also repair the feature loss problem for other image style transfer methods. The structure optimization can be expressed as

$$O_j = O'_j + C_j - S_j^C, \tag{5}$$

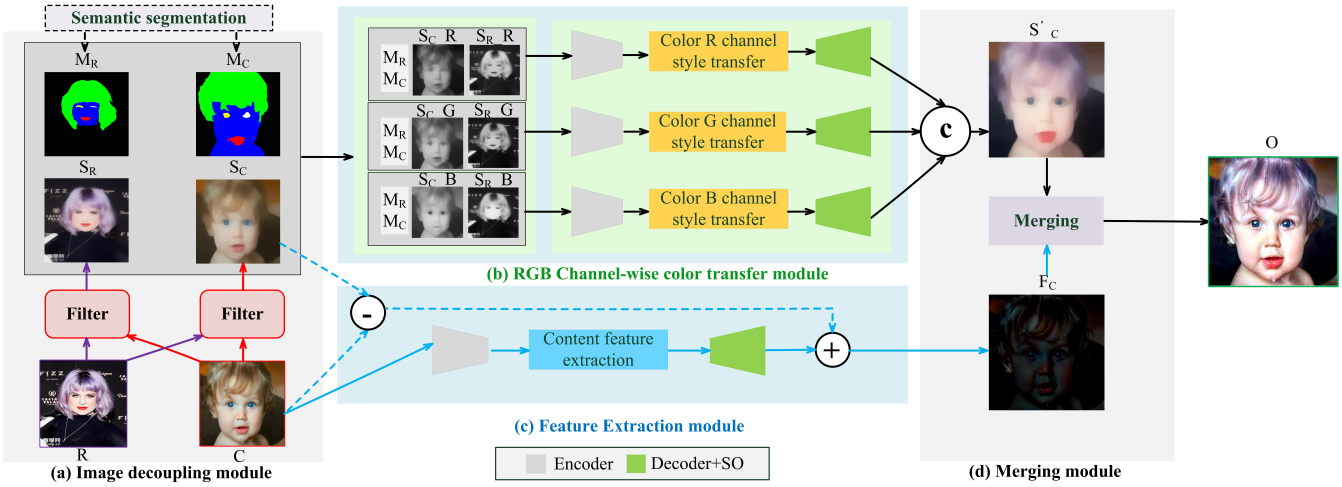


Figure 2: Overview of our proposed photorealistic style transfer network. SO is the structure optimization module. Our model takes the content image C, the reference image R, and their corresponding semantic segmentation maps M_C and M_R as input, and produces the photorealistic style transfer result in an end-to-end manner. In image filtering and semantic segmentation module (a), we leverage semantic segmentation for C and R. We utilize the modified filter to extract the structure layers S_C and S_R from C and R, and use RGB channel separation color transfer module (b) to perform color transformation to achieve the edited structure S'_C. We extract accurate texture detail features F_C by F_{C1} and F_{C2} in feature extraction module (c). Finally, we obtain the photorealistic color transfer result O by merging S'_C and F_C in merging module (d).

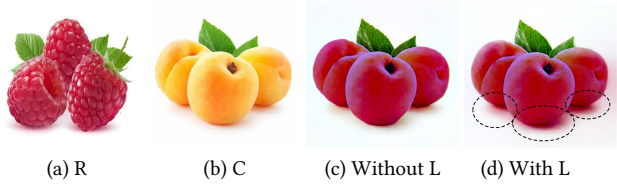
(a) Input (b) Without L (c) With L

Figure 3: Influence of the L channel in image smoothing. Smoothing the image with its L channel, we can obtain the changed shading information (see the shading of the fruits). However, without the L channel, the smoothed result has less shading information.

where O_j is the optimization result. O'_j is the structure layer of C after color transfer, and $C_j - S_j^C$ expresses the feature layer of C. O'_j is the output of the convolution layer in the decoder, S_j indicates the jth color channel of C, and S_j^C is the S_j smoothing result of C from the AWLSF. Fig. 8 (c) shows the visual results. With SO, the results have more natural texture details.

3.3 Feature Extraction (FE)

To obtain more accurate texture details in the final output, we adopt two methods for feature extraction and weighted fusion. We input C into the encoder and extract its texture feature:



(a) R (b) C (c) Without L (d) With L

Figure 4: Effect of $\delta L < 0.5$ in Eq. 2 to the result. The luminance L of C and that of R are similar. We obtain better contours without L than with L (see the fruits' shadows circled in black).

$$\hat{f}^C = E^C(D^C)^{-1/2}(E^C)^T f^C. \tag{6}$$

Then, we invert $\hat{f}^C$ to the decoder to achieve feature F_{C1} of C.

We utilize C and S_C to extract texture feature F_{C2}:

$$F_{C2} = k_1 \times C - k_2 \times S_C. \tag{7}$$

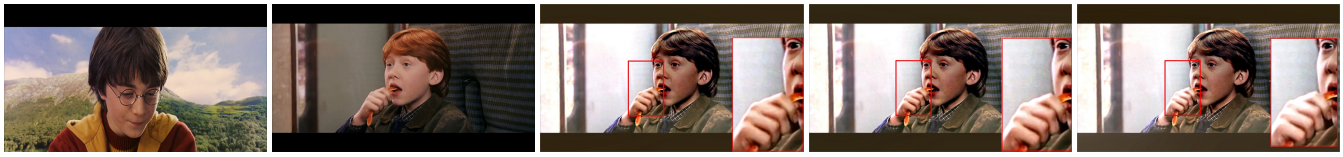
We sum F_{C1} and F_{C2} to obtain the final content feature F_C:

$$F_C = K_1 \times F_{C1} + K_2 \times F_{C2}. \tag{8}$$

where $K_1 + K_2 = 1$. Fig. 9 shows the ablation study for feature extraction. In Fig. 9, the feature map by WCT [5] looks unclean and has no clear edges, the feature map by WCT^2 is often with blurry edges, however, our feature map presents clear edges.

3.4 Merging structure and texture features

We combine all of O_j ($j \in R, G, B$ denotes a channel in RGB color space) in Eq. 5 to obtain color transfer result of the structure layer S'_C. Then we merge S'_C and F_C by



(a) Reference (b) Content (c) α=0 (d) α=0.6 (e) α=0.9

Figure 5: Style transfer results with varying parameter α when $\Delta L \geq 0.5$. The luminance of C and R is large. We obtain better luminance information with L than without L (see the boy's left face and hand marked by a red rectangle).

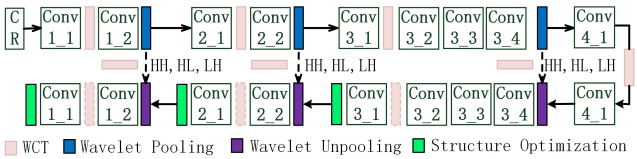


Figure 6: The structure of our used encoder-decoder network. We add a structure optimization operation after each convolution layer in the decoder [8].

$$O = \gamma_1 S'_C + \gamma_2 F_C, \tag{9}$$

where γ_1 and γ_2 are the weights of S'_C and F_C. Combined with Eqs. 7 and 8, we have

$$\begin{aligned} O &= \gamma_1 S'_C + \gamma_2 (K_1 F_{C1} + K_2 F_{C2}) \\ &= \gamma_1 S'_C + \gamma_2 K_1 F_{C1} + \gamma_2 K_2 (k_1 C - K_2 S_C) \\ &= \gamma_1 S'_C + \gamma_2 K_1 F_{C1} + \gamma_2 K_2 k_1 C - \gamma_2 K_2 k_2 S_C. \end{aligned} \tag{10}$$

According to Eq. 5, the sum of the weights of the two structural layers is 0, and the sum of the others is 1. Hence, we set

$$\begin{aligned} \gamma_2 K_1 + \gamma_2 K_2 k_1 &= 1, \\ \gamma_1 - \gamma_2 K_2 k_2 &= 0, \end{aligned} \tag{11}$$

where $\gamma_1 = k_2$, $\gamma_2 = K_2 = 1$, and $k_1 = 0.8, K_1 = 0.2$. γ_1 controls the weight of maintaining the reference style, and $-k_2$ is the weight of subtracting the style of C. The higher the parameters γ_1 and k_2, the more prominent the transfered reference style in the style transfer result, and the original style is removed more completely. We can adjust the γ_1 and k_2 values to achieve the different results shown in Fig. 10. We experimentally set $\gamma_1 = k_2 = 1$.

4 EXPERIMENTS

4.1 Comparison with state-of-the-art methods

Comparison methods. Our AFCS method can use semantic segmentation or not (see Figs. 11 and 12). We usually utilize semantic segmentation in facial images to obtain a more accurate photorealistic style transfer effect. We select four state-of-the-art photorealistic style transfer methods Luan *et al.* [6], Li *et al.* [14], Yoo *et al.* [8] and Li *et al.* [2] for vivid style transfer comparison of facial and scene images with semantic segmentation. Fig. We also compare five state-of-the-art photorealistic style transfer methods, Yoo *et al.* [8], Li *et al.* [2], An *et al.* [9], Hong *et al.* [10], and An *et al.* [1], for photorealistic style transfer without using semantic segmentation. For Hong's method [10], we choose the photorealistic style transfer results by running their code [1].

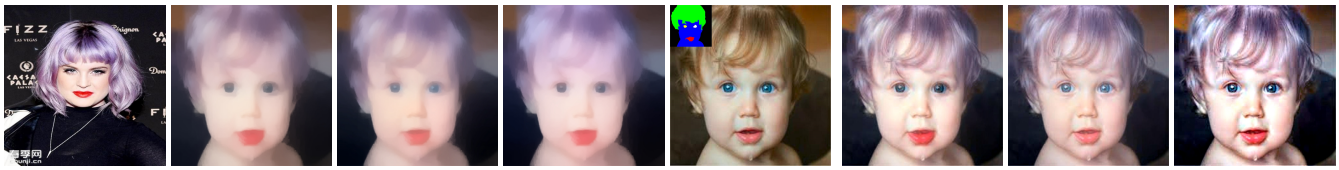
Visual comparison. Results of [8] fails to work well on the boundaries of some objects that have evident artificial traces. Results of [14], [6], [2], [10] and [1] have visible distortion in some local areas. Results of [14] have color overflowing and distortion. Results of [6] has visible distortion in some local areas. Some results of [2] fails to work well for the facial image. The method [10] does not work well enough for photorealistic images since this method focus on artistic style transfer. Results of [1] improved the method of [9]. However, they have not a good feature preservation ability. Our results have better photorealistic style transfer effects and are very similar to the reference color. Furthermore, our results have more clear content textures than the compared methods.

Quantitative evaluation. We use the natural image quality evaluator (NIQE) and average gradient (AG) to quantitatively evaluate the results. NIQE measures the difference in the multivariate distribution of an image. The distribution is constructed by extracting features from normal natural images. The AG refers to sharpness, it reflects the contrast of tiny details and texture changes in an image, as well as the sharpness of an image. Lower NIQE and higher AG values indicate better results. Tables 1 and 2 report the quantitative evaluation of Figs. 11 and 12 respectively. Our AG values are much higher than other methods. This means that our method has a stronger ability of feature preservation than other methods. Most of our NIQE values are smaller than those of the other methods. In Table 2, NIQE of [9] is lower than ours. However, their pattern on the glass is fuzzy. We also use SSIM and PSNR for quantitative comparison, and report the results in Tables 3 and 4. When the clarity of the content image is good, our advantage of SSIM and PSNR is not obvious. When the clarity of the content image is not very good, our SSIM and PSNR value is much less than other methods. However, we can find that our results are still better than others from the comprehensive evaluation, such as NIQE, AG.

Table 1: Comparisons of NIQE and AG evaluation. NIQE and AG respectively correspond to row 1 and row 2 of Fig. 11.

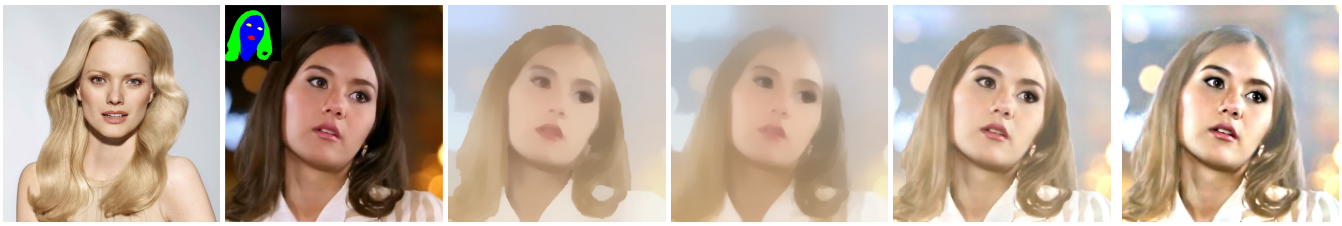
	[6]	[14]	[8]	[2]	**Ours**
NIQE	2.48	3.04	2.50	2.39	**2.38**
AG	7.72	3.14	4.84	6.12	**14.74**
NIQE	3.28	**3.26**	3.75	3.48	3.28
AG	5.82	4.86	8.66	12.81	**19.19**

[1] https://github.com/Kibeom-Hong/Domain-Aware-Style-Transfer



(a) Reference (b) (c) (d) (e) Content (f) (g) (h)

Figure 7: Ablation study for AWLSF filtering and RGB channel separation. (e) shows C and M_C (at the upper left corner). (b) and (f) show the final color and photorealistic style transfer results without AWLSF filtering. (c) and (g) show the final color and photorealistic style transfer results with AWLSF filtering without RGB channel separation. (d) and (h) show the final color and style transfer results with AWLSF filtering and RGB channel separation.



(a) Reference (b) C, M_C (c) Without SO (d) With SO (e) Without SO (f) With SO

Figure 8: Ablation study for structure optimization (SO). (b) is C, M_C (at the upper left corner). (c) and (d) are the style transfer results for the structure layers without and with SO respectively. (e) and (f) are are our final style transfer results without and with SO respectively.

(a) content (b) WCT (c) WCT^2 (d) Ours (e) Reference (f) Without FE (g) With FE

Figure 9: Feature extraction comparison and our style transfer results with and without FE. (b), (c) and (d) are the feature extraction results by WCT, WCT^2, and ours. (f) and (g) are the style transfer results produced by our method without and with FE respectively.

(a) Reference (b) Content (c) 0.8,-0.8 (d) 1.0,-1.0 (d) 1.2,-1.2

Figure 10: Influence of γ_1 and k_2 in Eq. 10. γ_1 and k_2 are shown under each result.

Artistic style comparison We select four recent state-of-the-art artistic style transfer methods including [4], [34], [35] and [36] for comparison. Results are shown in Fig. 13. The results of these methods are quite different from photorealistic style transfer. Therefore, they are more suitable for artistic style transfer than photorealistic style transfer.

4.2 User study

We performed a user study with 80 random volunteers to validate the effectiveness of the proposed method. We randomly show 80 groups of style transfer results obtained with our approach and other compared methods [6], [14], [2], [8] [9], [10] and [1] for each volunteer. Each volunteer browses the labeled images shown in

(a) Reference (b) C, M_C (c) Luan *et al.* [6] (d) Li *et al.* [14] (e) Yoo *et al.* [8] (f) Li *et al.* [2] (g) **Ours**

Figure 11: Comparison with semantic segmentation. (b) shows C and M_C (on the upper left corner). For (c) to (g), their NIQE values are 5.82, 7.98, 6.14, 5.66 and 5.20, and their their AG values are 2.90, 2.17, 2.68, 2.64 and 4.45, respectively.

(a) C, R (b)Yoo *et al.* [8] (c) Li *et al.* [2] (d) An *et al.* [9] (e) Hong *et al.* [10] (f) An *et al.* [1] (g) **Ours**

Figure 12: Comparison results without semantic segmentation. For (c) to (h), their NIQE values are 4.34, 4.56, 3.60, 5.07, 5.59 and 3.74, and their AG values are 4.33, 2.31, 4.43, 7.81, 2.36 and 8.28, respectively.

(a) R (b) C (c) Yao *et al.* [4] (d) Yu *et al.* [34] (e) Svoboda *et al.* [35] (f) Deng *et al.* [36] (g) **Ours**

Figure 13: Comparison results with artistic style methods.

Table 2: Comparisons of NIQE and AG evaluation. NIQE and AG respectively correspond to row 1 and row 2 of Fig. 12.

	[8]	[2]	[9]	[10]	[1]	**Ours**
NIQE	5.09	3.91	3.88	5.01	4.43	**3.67**
AG	6.03	8.83	6.13	7.00	7.36	**10.81**
NIQE	4.34	4.56	**3.60**	5.07	5.59	3.74
AG	4.33	2.31	4.43	7.81	2.36	**8.28**

Table 3: Comparisons of SSIM and PSNR. SSIM and PSNR respectively correspond to rows 1 and row 2 of Fig. 11.

	[6]	[14]	[8]	[2]	**Ours**
SSIM	0.90	0.68	0.87	0.80	**0.91**
PSNR	84.81	78.46	81.95	81.14	**84.83**
SSIM	**0.89**	0.88	0.81	0.59	0.62
PSNR	**83.98**	83.10	79.65	75.03	79.24

Table 4: Comparisons of SSIM and PSNR. SSIM and PSNR respectively correspond to rows 1 and row 2 of Fig. 12.

	[8]	[2]	[9]	[10]	[1]	**Ours**
SSIM	**0.93**	0.84	0.90	0.86	0.91	0.92
PSNR	**81.33**	77.89	80.24	76.83	81.25	81.29
SSIM	0.88	0.65	0.69	0.82	0.61	**0.89**
PSNR	81.07	76.03	76.63	81.06	75.27	**81.32**

Figs. 11 and 12 and the supplementary material. A survey was conducted to collect feedback on the following questions: (i) clear details and distinct contrast, (ii) natural and vivid color, (iii) no loss of detail, and (iv) well-preserved photorealism. For fairness, the results generated by the five methods are labeled as R_1, R_2, R_3, R_4, R_5, R_6 and R_7, while our results are labeled R_8. The final results are shown in Table 5.

Let V_{ij} denote the total votes of R_i on the jth question. To evaluate each method of the individual question, we compute the percentage of votes (PoV) obtained by R_i on the jth question by $PoV = (V_{ij}/8000) * 100\%$. To provide an overall evaluation of different methods, we further calculate the percentage of votes obtained by R_i on $\overline{PoV} = (\sum_{j=1}^{4} V_{ij})/32000 * 100\%$. In Table 5, we give the percentage of votes obtained by different methods, where Qu.x denotes the xth question. Table 5 shows that our method achieves the highest scores. This means that human evaluation prefers our results.

Table 5: Voting results for our method and compared methods. MD indicates the models here.

MD	Qu.1	Qu.2	Qu.3	Qu.4	Overall
Luan17	5.71%	7.39 %	6.24%	9.76%	7.28%
Li18	4.56%	5.09 %	5.26%	7.76%	5.67%
Yoo19	6.29%	8.35 %	9.84%	9.89%	8.59%
Li19	10.31%	10.24 %	9.41%	8.76%	9.68%
An20	8.21%	8.29 %	10.19%	12.66%	9.84%
Hong21	7.70%	10.19 %	9.65%	12.09%	9.91%
An21	6.98%	9.40 %	10.51%	12.76%	9.91%
Ours	**50.24%**	**41.06%**	**38.90%**	**26.31%**	**39.13%**

4.3 Discussion

Comparison with the LAB channel. What color space to be selected and how to make use of its components are very important for the color transfer related tasks. Color effect is determined by its constitutive components. The RGB color space uses three channels, R, G, and B, to represent all colors perceived by humans. The numerical representation range of RGB is [0, 255]. The LAB color space only uses two channels, A and B, and the range is [-128, 127] to represent the colors. However, this means that there are still some unseparated couplings in the A and B channels. Since different colors may interfere with each other during photorealistic style transfer, and this incomplete color separation of channels A and B will affect the photorealistic style transfer performance. Hence, we use the R, G, and B channels rather than the L, A, and B channels. Fig. 14 shows the comparison of the results.

(a) Reference (a) Content (c) With RGB (d) With LAB

Figure 14: Comparison of photorealistic style transfer with LAB. The LAB result tends to be yellow, while the RGB result is closer to the reference style.

Limitations. Our method also has some limitations. For instance, when the input content image is too dark, our output is not clear enough. When performing style transfer for facial images without semantic segmentation, we preserve feature well, however, we fails to have good photorealistic style transfer. Fig. 15 shows the results.

(a) Reference (b) Content (c) Result

(d) C, R, M_C (e) Without M_C (f) With M_C

Figure 15: Two failure cases. The first row is the case of scene images and the second row is the case of facial images.

5 CONCLUSION AND FUTURE WORK

We have proposed a novel pipeline of image photorealistic style transfer called AFCS method in this paper. We introduce an adaptive image smoothing method via AWLSF filter for a pair of images (C and R), RGB channel separation module, feature extraction module and image merging module. We evaluate the AFCS method on various natural and facial images to show its superiority over state-of-the-art methods. We will extend our method to handing photorealistic style transfer for videos in the future. Moreover, to produce more stable and vivid style transfer results, we will introduce lighting and shadow processing techniques [37–40] into our method in future work.

ACKNOWLEDGMENTS

We would like to thank the ACM MM reviewers for their feedback, thank Dr. Qing Zhang for his invaluable help in suggestion. The main work of this paper is done in Wuhan University. This work is partially supported by NSFC (No. 61972298), Bingtuan Science and Technology Program (No. 2019BC008), Guangxi First-class Discipline Statistics Construction Project Fund, Guangxi Key Laboratory of Big Data in Finance and Economics and the Humanities and Social Science Project of Ministry of Education of China (No. 18YJCZH050).

REFERENCES

[1] Jie An, Siyu Huang, Yibing Song, Dejing Dou, Wei Liu, and Jiebo Luo. Artflow: Unbiased image style transfer via reversible neural flows. In *CVPR*, 2021.

[2] Xueting Li, Sifei Liu, Jan Kautz, and Ming-Hsuan Yang. Learning linear transformations for fast image and video style transfer. In *CVPR*, pages 3809–3817, 2019.

[3] Leon A. Gatys and Ecker. Image style transfer using convolutional neural networks. In *CVPR*, pages 2414–2423, 2016.

[4] Yuan Yao, Jianqiang Ren, Xuansong Xie, Weidong Liu, Yongjin Liu, and Jun Wang. Attention-aware multi-stroke style transfer. In *CVPR*, pages 1467–1475, 2019.

[5] Yijun Li, Chen Fang, Jimei Yang, Zhaowen Wang, Xin Lu, and Ming-Hsuan Yang. Universal style transfer via feature transforms. *arXiv preprint arXiv:1705.08086*, 2017.

[6] Fujun Luan, Sylvain Paris, Eli Shechtman, and Kavita Bala. Deep photo style transfer. In *CVPR*, pages 6997–7005, 2017.

[7] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *ICCV*, pages 1501–1510, 2017.

[8] Jaejun Yoo, Youngjung Uh, Sanghyuk Chun, Byeongkyu Kang, and Jung-Woo Ha. Photorealistic style transfer via wavelet transforms. In *ICCV*, pages 9036–9045, 2019.

[9] Jie An, Haoyi Xiong, Jun Huan, and Jiebo Luo. Ultrafast photorealistic style transfer via neural architecture search. In *AAAI*, pages 10443–10450, 2020.

[10] Kibeom Hong, Seogkyu Jeon, Huan Yang, Jianlong Fu, and Hyeran Byun. Domain-aware universal style transfer. In *ICCV*, pages 14609–14617, 2021.

[11] Qing Zhang, Yongwei Nie, Lei Zhu, Wei Shi Zheng, and Weishi Zheng. A blind color separation model for faithful palette-based image recoloring. *IEEE Transactions on Multimedia*, 24(2022):1545–1557, 2022.

[12] Hong Ding, Gang Fu, Qingan Yan, Caoqing Jing, Tuo Cao, Shenghong Hu, and Chunxia Xiao. Deep attentive style transfer for images with wavelet decomposition. *Information Sciences*, 587(2022):63–81, 2022.

[13] Jimei Yang Yijun Li, Chen Fang. Diversified texture synthesis with feed-forward networks. In *CVPR*, pages 2732–2738, 2017.

[14] Yijun Li, Ming Yu Liu, Xueting Li, Ming Hsuan Yang, and Jan Kautz. A closed-form solution to photorealistic image stylizations. In *ECCV*, pages 1–16, 2018.

[15] Rasmus Rothe, Radu Timofte, and Luc Van Gool. Dex: Deep expectation of apparent age from a single image. In *ICCVW*, pages 252–257, 2016.

[16] Michael Maire Tsung-Yi Lin and Serge Belongie. Microsoft coco: Common objects in context. In *ECCV*, page 740–755, 2014.

[17] Hui-Huang Zhao, Paul L Rosin, Yu-Kun Lai, and Yao-Nan Wang. Automatic semantic style transfer using deep convolutional neural networks and soft masks. *The Visual Computer*, 36(7):1307–1324, 2020.

[18] Tianlang Chen, Wei Xiong, Haitian Zheng, and Jiebo Luo. Image sentiment transfer. In *Proceedings of the 28th ACM*, pages 4407–4415, 2020.

[19] Taihong Xiao, Jiapeng Hong, and Jinwen Ma. Elegant: Exchanging latent encodings with gan for transferring multiple face attributes. In *ECCV*, pages 168–184, 2018.

[20] Shuyang Gu, Jianmin Bao, and Yang. Mask-guided portrait editing with conditional gans. In *CVPR*, pages 3436–3445, 2019.

[21] Honglun Zhang and Chen. Disentangled makeup transfer with generative adversarial network. *arXiv preprint arXiv:1907.01144*, 2019.

[22] Si Liu, Xinyu Ou, and Qian. Makeup like a superstar: Deep localized makeup transfer network. *arXiv preprint arXiv:1604.07102*, 2016.

[23] Tingting Li, Ruihe Qian, and Dong. Beautygan: Instance-level facial makeup transfer with deep generative adversarial network. In *ACM MM*, pages 645–653, 2018.

[24] Longquan Dai, Mengke Yuan, Feihu Zhang, and Xiaopeng Zhang. Fully connected guided image filtering. In *ICCV*, pages 352–360, 2015.

[25] Xiao Tan, Changming Sun, and Tuan D Pham. Multipoint filtering with local polynomial approximation and range guidance. In *CVPR*, pages 2941–2948, 2014.

[26] Zeev Farbman, Raanan Fattal, Dani Lischinski, and Richard Szeliski. Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Transactions on Graphics (TOG)*, 27(3):1–10, 2008.

[27] Jonathan T Barron and Ben Poole. The fast bilateral solver. In *ECCV*, pages 617–632, 2016.

[28] Wei Liu, Pingping Zhang, Xiaolin Huang, Jie Yang, Chunhua Shen, and Ian Reid. Real-time image smoothing via iterative least squares. *ACM Transactions on Graphics (TOG)*, 39(3):1–24, 2020.

[29] Qingnan Fan, Dongdong Chen, Lu Yuan, Gang Hua, Nenghai Yu, and Baoquan Chen. A general decoupled learning framework for parameterized image operators. *IEEE transactions on pattern analysis and machine intelligence*, 2019.

[30] Qingnan Fan, Jiaolong Yang, David Wipf, Baoquan Chen, and Xin Tong. Image smoothing via unsupervised learning. *ACM Transactions on Graphics (TOG)*, 37(6):1–14, 2018.

[31] Jisung Yoo Yim, Jonghwa. Filter style transfer between photos. In *ECCV*, pages 103–119, 2020.

[32] Yoav HaCohen and Eli Shechtman. Non-rigid dense correspondence with applications for image enhancement. *ACM Transactions on Graphics*, 30(4):1–10, 2011.

[33] E. Reinhard, M. Ashikhmin, B. Gooch, and P. Shirley. Color transfer between images. *IEEE Computer Graphics and Applications*, 21(5):34–41, 2002.

[34] Yilin Wang Yulun Zhang, Chen Fang. Multimodal style transfer via graph cuts. In *ICCV*, pages 5943–5951, 2019.

[35] J. Svoboda, A. Anoosheh, Ch. Osendorfer, and J. Masci. Two-stage peer-regularized feature recombination for arbitrary image style transfer. In *CVPR*, pages 2761–2776, 2020.

[36] Yingying Deng, Fan Tang, Weiming Dong, Wen Sun, Feiyue Huang, and Changsheng Xu. Arbitrary style transfer via multi-adaptation network. In *ACM MM*, pages 2719–2727, 2020.

[37] Zhongyun Bao, Chengjiang Long, Gang Fu, Yuanzhen Li, Jiaming Wu, Daquan Liu, and Chunxia Xiao. Deep image-based illumination harmonization. In *CVPR*, pages 18542–18551, 2022.

[38] Zhipei Chen, Chengjiang Long, Ling Zhang, and Chunxia Xiao. Canet: A context-aware network for shadow removal. In *ICCV*, pages 4743–4752, 2021.

[39] Gang Fu, Qing Zhang, Lei Zhu, Ping Li, and Chunxia Xiao. A multi-task network for joint specular highlight detection and removal. In *CVPR*, pages 7752–7761, 2021.

[40] Qing Zhang, Ganzhao Yuan, Chunxia Xiao, Lei Zhu, and Shiwei Zheng. High-quality exposure correction of underexposed photos. In *ACMMM*, 2018.