HAIRGAN: SPATIAL-AWARE PALETTE GAN FOR HAIR COLOR TRANSFER

Yifan Xie[1], *Yinjia Huang*[2], *Ricong Huang*[1], *Zhengtao Wu*[1], *Mingzhi Mao*[*], *Guanbin Li*[*]

[1]School of Computer Science and Engineering, Sun Yat-sen University, China
[2]New York University, USA

ABSTRACT

Hair color transfer aims to transfer the hair color from a reference image to an original image while maintaining the hair structure of the original image. However, due to the complex hair structure and the misalignment of hair regions between the original image and the reference image, existing methods cannot complete this task well. To address these issues, we propose a hair color transfer GAN (HairGAN). It first utilizes off-the-shelf network to extract the hair region from the original image and transfer it into intermediary hair color. Then, Spatial-Aware Palette Alignment (SAPA) module is introduced to align the hair regions in the original image and the reference image. Furthermore, HairGAN employs cycle consistency reconstruction module to ensure the global consistency between original image and transferred image. We collect a dataset containing 2850 images and a few paired samples for hair color transfer. Extensive experiments show that HairGAN could generate high-quality transferred results.

Index Terms— Hair color transfer, generative adversarial networks

1. INTRODUCTION

Hairstyle is one of the salient features of human image and people usually make a great effort in managing their hairstyle. Among them, the color of the hair is the focus of attention. However, without many failed attempts, it is difficult to find a suitable hair color at the beginning. Currently, some image and video-sharing applications, such as FaceApp, TikTok, provide some functions for users to transfer hair color style. However, the available colors can only be chosen from the candidate colors whose number is limited. And the transferred image is not realistic enough. To overcome these drawbacks, this paper studies a reference image guided hair color transfer method, which aims to transfer the hair color of the reference image to that of the original image while maintaining the consistency of the original hair texture details.

[*]Corresponding authors are Mingzhi Mao and Guanbin Li. This work was supported in part by the Guangdong Basic and Applied Basic Research Foundation under Grant No.2020B1515020048, in part by the National Natural Science Foundation of China under Grant No.61976250 and No.U1811463, and in part by the Guangzhou Science and technology project under Grant No.202102020633.

Fig. 1. Samples from self-collected hair transfer dataset. The first row is the raw images while the second row is the images we edited with Photoshop. It is worth noting that, some of them are selected as label images.

However, compared to other parts of the human body, hair has a finer and more diverse structure. In different regions, the curvature and direction of the hair are different. Meanwhile, affected by the ambient light, each region of the hair also shows different brightness and color. These problems have brought great technical challenges to the analysis and editing of hair. In addition, the hair of the original image and reference image are quite different in position, shape, outline, and detailed texture, so it is difficult to make perfect hair alignment and direct pixel-level hair color transfer.

At present, there have been some advanced methods in the field of style transfer. SANet [1], AdaAttN [2], using the original image and reference image as input, can generate high-quality results with both the original image content and the reference image style. However, these methods focus on the global style transfer of the image, rather than the transfer for a specific instance, and they are prone to lose some structure details of the original image after transferring. In addition, these methods do not perform explicit semantic and detail structure alignment when performing style transfer. Therefore, these methods cannot be directly applied to hair color transfer which requires strict maintenance of texture structure. Besides, in the field of hair editing, LOHO [3] and MichiGAN [4] can edit hair by adjusting the latent code. Yet the generative method cannot achieve precise hair color transfer, and it is also difficult to maintain the structure detail.

To address the above issues, we propose a novel hair color transfer GAN (HairGAN), which takes original image and

2022 IEEE International Conference on Multimedia and Expo (ICME) | 978-1-6654-8563-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICME52920.2022.9859758

reference image as input and can transfer the reference hair color to the original hair while maintaining the structural consistency. The network consists of three parts: (1) hair segmentation and intermediary color transfer (2) conditional hair color alignment transfer (3) cycle consistency reconstruction. The first part is the pre-processing stage of input image, we use BASNet [5] and retrain it to perform hair region segmentation, because hair mask is crucial for subsequent processing. Then, in order to minimize the negative impact of source color difference on the transfer performance , we use the Pixel2PixelHD [6] network to transfer hair color of the original image into an intermediary color(e.g., transferring all hair color to brown). For the second part, we design a Spatial-Aware Palette Alignment (SAPA) module, which can generate a horizontal palette and a vertical palette of reference hair. Based on the palettes and hair masks, HairGAN can perform alignment when coloring the original hair to reference hair color. Finally, we employ the third part to maintain the global consistency between the transferred hair and original hair. With our network, combined with advanced loss function, we can achieve smooth and robust training, and ultimately generate superior hair color transferred results.

To sum up, the main contributions are three-folds:

- We design a novel hair color transfer network. Experiments present the effectiveness of the transfer strategy, and the generated results have higher quality than state-of-the-art methods.
- We design a Spatial-Aware Palette Alignment module. It can effectively and naturally deal with the misalignment of hair in the original image and reference image.
- We collect a hair-related dataset including 2850 portrait images (shown in Fig. 1) with high-quality hair details. To the best of our knowledge, we are the first to use synthetic hair style pairs as an auxiliary for training hair color transfer without real ground truth.

2. RELATED WORK

Style Transfer. Style transfer aims at applying style patterns of style image onto the content image while preserving the structure of the content image simultaneously. It has been widely used in computer-aid art generation. In the early research, Gatys et al. [7] proposed an optimization-based approach that can generate high-quality results after hundreds of gradient descent iterations. Then, Johnson et al. [8] proposed to train a neural network that can directly generate results. However, it needs to train a network for each new style. Recently, some research [1,9–12] have been proposed to achieve arbitrary style transfer in one model. Among them, the common idea of [9, 10] is to apply feature modification globally. On the contrary, [1, 11, 12] implements attention mechanism to model fine-grained correspondence among local features of style and content images. However, all the methods above are domain-level style transfer methods. If these methods are forced to perform instance-level transfer, the transferred results are not stable.

Hair Modeling & Editing. Hair is one of the key features of the human image. Although its structure is delicate and diverse, it is still of great interest to researchers. To process the hair, previous work [13–15] first generated a 3D hair model, and then used the model for hair editing. However, these 3D hair models cannot represent the characteristics of hair comprehensively, so it is difficult to guarantee the quality of editing. Recently, GAN-based hair editing methods have made great progress. Tan et al. [4] proposed MichiGAN, which disentangles hair into four orthogonal attributes, and users can modify these attributes to realize the editing of the hair. Similarly, LOHO [3] also decomposes hair into three attributes. But the difference is that LOHO achieves hair editing by adjusting latent space code. Unlike LOHO, Our method takes images as input, which could preserve more hair details and generate better transferred results.

3. METHOD

3.1. Overview

Hair color transfer aims to transfer the hair color from a reference image to an original image while maintaining the hair structure of the original image. Specifically, given an original image I_{org} and a reference image I_{ref}, the network is expected to generate a hair color transferred image, denoted as $I_{syn} = G(I_{org}, I_{ref})$, where G denotes the generative network. I_{syn} has the hair structure of the original image and the hair color of the reference image.

As illustrated in Fig. 2, to achieve this goal, we propose HairGAN, which consists of three important parts including (1) Hair Segmentation and Intermediary Color Transfer (2) Conditional Hair Color Alignment Transfer, and (3) Cycle Consistency Reconstruction. In the following subsections, we will introduce these three parts in detail.

3.2. Hair segmentation and intermediary color transfer

Although there have been some instance-level style transfer models [16, 17], the disturbance from the background region to the transfer still cannot be ignored. To make the model focus on the hair region, we train a BASNet [5] to generate the mask of I_{org}, annotated as M_{org}, and the mask of I_{ref}, annotated as M_{ref}. Based on this, we use the following formula to extract the hair region of the original image and reference image respectively for subsequent transfer.

$$I^{h^o}_{org} = I_{org} \odot M_{org}, \quad I^{h}_{ref} = I_{ref} \odot M_{ref}, \tag{1}$$

where $I^{h^o}_{org}$, I^{h}_{ref} denotes the hair region of I_{org} and I_{ref}, $\odot$ refers to an operation of element-wise multiplication.

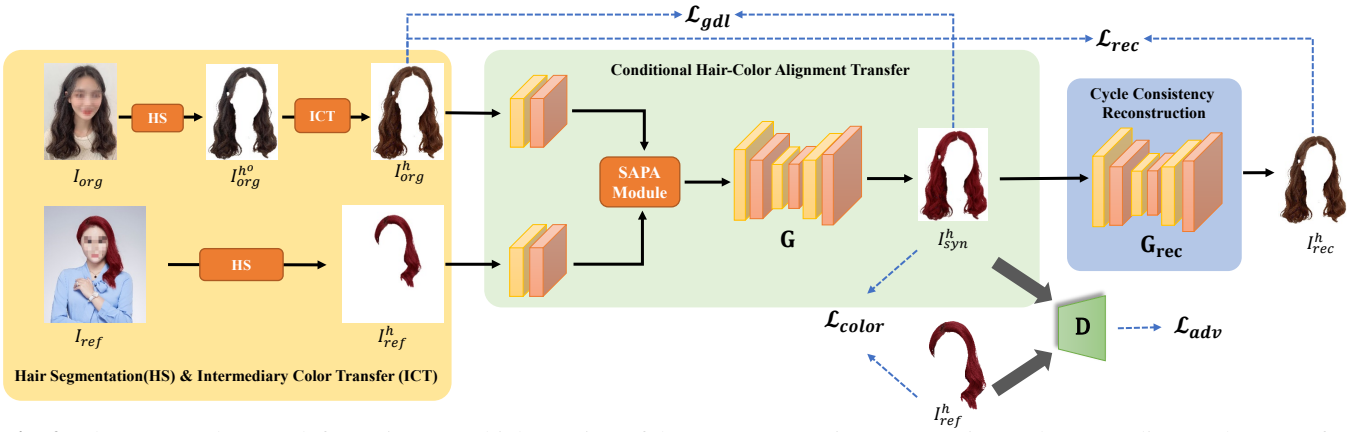

Fig. 2. The proposed network for HairGAN which consists of three parts (1) Hair Segmentation and Intermediary Color Transfer (2) Conditional Hair Color Alignment Transfer (3) Cycle Consistency Reconstruction.

In addition, due to the diversity of hair color in the original image and the reference image, it is hard for the model to converge while training. To address this issue, we train a pix2pixHD [6] to transfer the hair color of $I^{h^o}_{org}$ to a single intermediary color (brown), given by

$$I^h_{org} = \text{pix2pixHD}(I^{h^o}_{org}), \tag{2}$$

where I^h_{org} denotes the image with intermediary hair color. In the subsequent training, the hair color of I^h_{org} remains constant, and only the hair color of I^h_{ref} changes. Based on this, the training difficulty can be massively reduced and I^h_{org} could obey the same distribution.

3.3. Conditional hair color alignment transfer

When transferring hair color, the most critical issue is that the original hair cannot align with reference hair exactly in size, location, and shape, making it difficult to transfer the color of reference hair to original hair smoothly. To tackle this issue, in this part we propose Spatial-Aware Palette Alignment (SAPA) module as shown in Fig. 3, which could extract the color feature of the reference hair and align it with the original hair. We design the SAPA module in the following steps.

After getting the outputs (F_{org} and F_{ref}) of the two convolution branches, SAPA first generates their smallest bounding boxes containing hair region. Note that through the observation on hair images, we find out that hair characteristics may be similar in the same horizontal position or vertical position. Therefore, based on the bounding boxes, SAPA will further generate the horizontal palette and vertical palette of F_{ref}. In the horizontal palette, all the pixels whose horizontal position is y will be set to c_y, given by:

$$c_y = \frac{\sum_y (F_{ref} \odot M_{ref})}{\sum_y M_{ref}}, \tag{3}$$

where y denotes the horizontal position in both the horizontal palette and the smallest bounding box of F_{ref}.

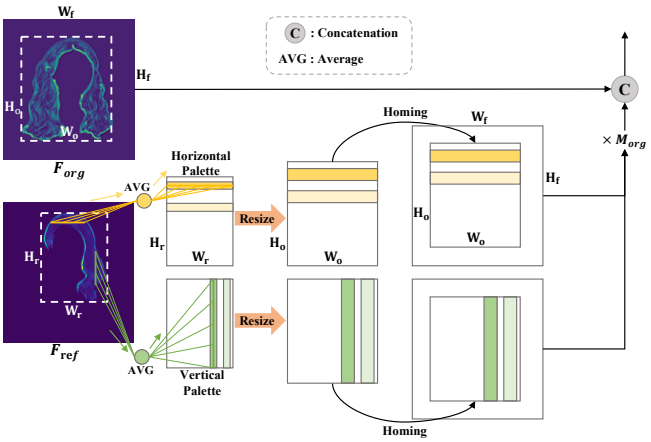

Fig. 3. The proposed Spatial-Aware Palette Alignment (SAPA) module. First, SAPA employs hair mask to generate the smallest bonding box containing hair region in the feature map. Then, based on it, combining F_{ref} and mask, SAPA get a horizontal palette and a vertical palette of same size. After that, SAPA will resize both palettes to fit the size of hair region in F_{org} and finally the two palettes will be homing to the corresponding location.

After obtaining the palettes, SAPA will resize the palettes so that they can cover the original hair exactly. Then, by referring to the position of the original hair in F_{org}, two palettes will be homing to the corresponding position of two full-zero feature maps to achieve alignment in position. Finally, SAPA will use M_{org} to match the shape of the original hair.

3.4. Cycle consistency reconstruction

Although HairGAN has generated a transferred hair image before this part, the global consistency between the transferred hair and the original hair cannot be ensured. For instance, the brightness of the hair color of the two images may be inconsistent. Therefore, we additionally introduce cycle consistency reconstruction, which aims to re-transfer the

color of transferred hair to the intermediary hair color, given by:

$$I^h_{rec} = G_{rec}(I^h_{syn}), \tag{4}$$

where I^h_{rec}, I^h_{syn} represent the reconstructed hair image and the generated hair image, respectively. G_{rec} is the reconstruction network of HariGAN. Based on this part, combined with the corresponding loss function, the global consistency can be maintained maximally.

3.5. Object function

Adversarial loss. We utilize a discriminator D to distinguish the generated hair image from the reference hair image to help generator G synthesize the realistic output. We use the following formulation to calculate their adversarial loss:

$$\begin{aligned}\mathcal{L}_{adv} = &\min_G \max_D \{\mathbb{E}_{I^h_{ref}}[log(D(I^h_{ref}))] \\ &+ \mathbb{E}_{I^h_{org}, I^h_{ref}}[log(1 - D(I^h_{syn}))]\}.\end{aligned} \tag{5}$$

Color matching loss. To provide coarse guidance for color transfer, we use the color histogram matching loss proposed by [16]. It first performs histogram matching on I^h_{ref} and I^h_{org} to obtain a pseudo label $HM(I^h_{ref}, I^h_{org})$, which has the same color distribution as I^h_{ref} but similar content information as I^h_{org}. Then we calculate the loss between the pseudo label image and generated hair image I^h_{syn}:

$$\mathcal{L}_{color} = ||I^h_{syn} - HM(I^h_{ref}, I^h_{org})||_2. \tag{6}$$

Gradient difference loss. Generated hair image usually loses some structure details. As proposed in [18], we directly penalize the difference on image gradient to maintain the hair structure. The Gradient Difference Loss (GDL) is given by:

$$\mathcal{L}_{gdl} = \mathcal{L}_{gdlx} + \mathcal{L}_{gdly}, \tag{7}$$

where $\mathcal{L}_{gdlx}$, $\mathcal{L}_{gdly}$ represent the horizontal and vertical gradients, respectively:

$$\begin{aligned}\mathcal{L}_{gdlx} = &\sum_{c,i,j} \big||I^h_{org}(i,j,c) - I^h_{org}(i-1,j,c)| \\ &-|I^h_{syn}(i,j,c) - I^h_{syn}(i-1,j,c)|\big|, \\ \mathcal{L}_{gdly} = &\sum_{c,i,} \big||I^h_{org}(i,j,c) - I^h_{org}(i,j-1,c)| \\ &-|I^h_{syn}(i,j,c) - I^h_{syn}(i,j-1,c)|\big|,\end{aligned} \tag{8}$$

where $I^h_{org}(i,j,c), I^h_{syn}(i,j,c)$ represent the value at the position (i,j) of the c^{th} channel in the original hair image and the generated image, respectively.

Cycle consistency reconstruction loss. Using the above loss, we can obtain a coarse generated image. However, the generated hair image is difficult to maintain global consistency with the original hair image. Therefore, we use cycle consistency reconstruction loss, which is able to couple color match loss and gradient difference loss to ensure that the generated image can maintain the global consistency.

$$\mathcal{L}_{rec} = ||I^h_{org} - I^h_{rec}||_2. \tag{9}$$

In summary, the overall training objective of generator and discriminator can be formulated as:

$$\mathcal{L}_{all} = \alpha\mathcal{L}_{adv} + \beta\mathcal{L}_{color} + \gamma\mathcal{L}_{gdl} + \delta\mathcal{L}_{rec}, \tag{10}$$

where α, β, γ and δ are the weights. Note that all four losses are only calculated on the hair region.

4. DATA COLLECTION

Since the hair region in the existing human-face datasets is not the focused region, the hair details are relatively blurred. We collect a new hair dataset consisting of 2850 images in total and refer to it as Hair-Transfer (HT) dataset. It includes a variety of hair colors, such as black, blue, red, orange and pink, and has various shapes, sizes, and structures. In addition, it contains 400 paired images whose hairs are dyed by Photoshop.

The raw dataset is collected from the Internet. To assist in training and analyze the experimental results quantitatively, we randomly select 400 images and use Photoshop to edit the hair color of them. Half of the images are set to blue and the other half to red. These additional images are used as label data. Examples are shown in Fig. 1. In the editing process, we carried out the professional operation. First, we used quick selection tool to segment the hair region from the human image to get the hair mask layer. Then, we set a hue-saturation layer on top of the mask layer. Based on it, we could adjust the hair color through hue and saturation. Finally, we created a curve layer to finetune the brightness of the hair.

Among 2850 images, for qualitative and quantitative comparison, we randomly select 100 original images, 50 reference images of blue hair, 50 reference images of red hair, and 100 label images matching the hair color of reference images to quantitatively calculate the transfer quality score. The remaining are separated into the training set and validation set.

5. EXPERIMENT

5.1. Implementation details

Network Structures. We illustrate the structure of HairGAN in Fig. 2. Specifically, the network contains two separate input branches with convolutional layers, which downsample the input by a factor of 2. In the middle, the SAPA will align and concatenate the outputs of these two branches, and feed them into the subsequent encoder module. After that, the output feature maps will pass through the transposed convolution module to generate the transferred image. Finally, another encoder-decoder module will reconstruct the transferred image to match the hair color of original image. Note that we use Instance Normalization [20] for the network.

Training strategy and parameters. In order to stabilize the training procedure and speed up convergence, we use a small amount of paired images to assist the training. To be specific, we use some original images with red-hair/blue-hair label im-

Fig. 4. Qualitative comparisons between ours and baselines. The first two columns are original images and reference images, respectively. Each row in the remaining columns shows the hair color transferred results of different methods.

ages during the training process, and we will select the reference image with the corresponding hair color. Based on this, we introduce an additional loss function to calculate the error between the generated image and its label image.

$$\mathcal{L}_{label} = ||I^h_{syn} - I^h_{label}||_2, \tag{11}$$

where I^h_{label} denotes the label image of the original image. It is worth noting that, the methods to be compared also add this loss and are trained by the same data.

In Eq.(10), α, β, γ and δ are set to 1, 1, 0.5 and 1. Adam [21] optimizer is used with batch size of 1. The learning rates for generator and discriminator are both set to 2e-4.

5.2. Comparison Against Baselines

Due to the lack of research on hair color transfer, we choose some relevant methods for comparison. BeautyGAN (BGAN) [16] and PSGAN (PGAN) [17] are makeup transfer networks, which aim to realize facial makeup transfer between a reference makeup image and a non-makeup image on instance-level. Sketch Image Colorization (SIC) [19] is a colorization work, which aims to use the reference image to color the original image. To adapt the hair color transfer task and fair comparison, we segment the hair region of the original image and the reference image as the input for above three works. In addition, for BGAN and PGAN, we transfer the hair color of the original image to the intermediary color, and for SIC, we transfer the original image to a grayscale image. Another method for comparison is MichiGAN (MGAN) [4], a hair editing work that includes hair color transfer.

Qualitative comparison. We show qualitative comparison results with baselines in Fig. 4. The results produced by MGAN have correct hair color but blurred hair structure since it accepts feature map as input, which will cause the loss of hair characteristics. On the contrary, SIC can maintain the structure of the hair, but it can only transfer hair into a specific color. The instance-level transfer methods outperform the above methods. However, BGAN does not align the original image with the reference image, which leads to artifacts and color deviations. We also use PGAN with an alignment mechanism, but it cannot model the correlation of hair features well, so the generated images may have some bright spots. Compared to these methods, our method is able to generate better images with the same hair color as the reference.

Quantitative comparison. To perform quantitative evaluation of HairGAN, we conduct a user study from 50 volunteers and use label images to calculate SSIM and PSNR of the results, respectively. It should be noted that we only calculate SSIM and PSNR of the hair region. As mentioned in section 4, 100 original images, reference images and label images are used for test. For the former evaluation method, in order to ensure the diversity of hair color, we randomly select 35 original images and divide them into 10 new original images and 25 reference images, which will generate 10×25 transferred images for each comparison method. Then volunteers are asked to choose the best transferred images by considering image realism and similarity with reference hair color. For the latter, we use the whole data for testing. Table 2 shows the human evaluation results. Our HairGAN outperforms other methods by a large margin. Table 1 shows the results of latter evalu-

Table 1. PSNR, SSIM of hair region for different methods. We randomly sampled 100 images with labels. (↑) indicates that higher is better.

Method	MGAN	SIC	BGAN	PGAN	Ours
SSIM (↑)	0.681	0.738	0.717	0.753	**0.783**
PSNR (↑)	21.553	20.101	25.044	24.850	**26.237**

Table 2. Ratio selected as best.

Method	MGAN	SIC	BGAN	PGAN	Ours
HT dataset	0.02	0.09	0.13	0.19	**0.57**

ation method, our HairGAN also outperforms other methods on both evaluation metrics. Compared with PGAN, our HairGAN achieves 4.0% and 5.6% higher of SSIM and PSNR, respectively. Overall, the outstanding performance demonstrates the superiority of our method.

5.3. Ablation Studies

To convincingly demonstrate the effectiveness of each component in our method, we conduct extensive ablation studies on HT dataset. As shown in Table 3, the HairGAN with all proposed modules removed does not perform well. SSIM and PSNR are only 0.714 and 24.129, respectively. After adding Intermediary Color Transfer (ICT), both evaluation metrics have been slightly improved. Further adding Cycle Consistency Reconstruction (REC), SSIM and PSNR rise to 0.761 and 24.789, respectively. Finally, when we add the SAPA module, both evaluation metrics' scores increase significantly and achieve the best performance. This shows the effectiveness of each component. Among them, SAPA contributes the most to the performance of model, which shows that this module can perform alignment and color transfer to the maximum.

6. CONCLUSION

In this work, we propose HairGAN which consists of three parts (1) hair segmentation and intermediary color transfer (2) conditional hair color alignment transfer (3) cycle consistency reconstruction. Based on this model, we can transfer the hair color from a reference image to an original image while maintaining hair structure of original image. We also collect a dataset containing 2850 images and 400 paired images with dyed hair. Extensive experiments show that HairGAN has the ability to generate high-quality transferred hair images.

Table 3. Result of Ablation Studies.

Method	SSIM(↑)	PSNR(↑)
HairGAN	0.714	24.129
HairGAN+ICT	0.730	24.419
HairGAN+ICT+REC	0.761	24.789
HairGAN+ICT+REC+SAPA	**0.783**	**26.237**

7. REFERENCES

[1] Dae Young Park and Kwang Hee Lee, "Arbitrary style transfer with style-attentional networks," in *CVPR*, 2019, pp. 5880–5888.

[2] Songhua Liu, Tianwei Lin, Dongliang He, Fu Li, Meiling Wang, Xin Li, Zhengxing Sun, Qian Li, and Errui Ding, "Adaattn: Revisit attention mechanism in arbitrary neural style transfer," in *ICCV*, 2021, pp. 6649–6658.

[3] Rohit Saha, Brendan Duke, Florian Shkurti, Graham W Taylor, and Parham Aarabi, "Loho: Latent optimization of hairstyles via orthogonalization," in *CVPR*, 2021, pp. 1984–1993.

[4] Zhentao Tan, Menglei Chai, Dongdong Chen, Jing Liao, Qi Chu, Lu Yuan, Sergey Tulyakov, and Nenghai Yu, "Michigan: multi-input-conditioned hair image generation for portrait editing," *TOG*, vol. 39, no. 4, pp. 95–1, 2020.

[5] Xuebin Qin, Zichen Zhang, Chenyang Huang, Chao Gao, Masood Dehghan, and Martin Jagersand, "Basnet: Boundary-aware salient object detection," in *CVPR*, 2019, pp. 7479–7489.

[6] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional gans," in *CVPR*, 2018, pp. 8798–8807.

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

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

[9] Yongcheng Jing, Xiao Liu, Yukang Ding, Xinchao Wang, Errui Ding, Mingli Song, and Shilei Wen, "Dynamic instance normalization for arbitrary style transfer," in *AAAI*, 2020, vol. 34, pp. 4369–4376.

[10] Xueting Li, Sifei Liu, Jan Kautz, and Ming-Hsuan Yang, "Learning linear transformations for fast arbitrary style transfer," *arXiv preprint arXiv:1808.04537*, 2018.

[11] Yuan Yao, Jianqiang Ren, Xuansong Xie, Weidong Liu, Yong-Jin Liu, and Jun Wang, "Attention-aware multi-stroke style transfer," in *CVPR*, 2019, pp. 1467–1475.

[12] Yingying Deng, Fan Tang, Weiming Dong, Wen Sun, Feiyue Huang, and Changsheng Xu, "Arbitrary style transfer via multi-adaptation network," in *ACMMM*, 2020, pp. 2719–2727.

[13] Menglei Chai, Lvdi Wang, Yanlin Weng, Xiaogang Jin, and Kun Zhou, "Dynamic hair manipulation in images and videos," *TOG*, vol. 32, no. 4, pp. 1–8, 2013.

[14] Menglei Chai, Linjie Luo, Kalyan Sunkavalli, Nathan Carr, Sunil Hadap, and Kun Zhou, "High-quality hair modeling from a single portrait photo," *TOG*, vol. 34, no. 6, pp. 1–10, 2015.

[15] Yi Zhou, Liwen Hu, Jun Xing, Weikai Chen, Han-Wei Kung, Xin Tong, and Hao Li, "Hairnet: Single-view hair reconstruction using convolutional neural networks," in *ECCV*, 2018, pp. 235–251.

[16] Tingting Li, Ruihe Qian, Chao Dong, Si Liu, Qiong Yan, Wenwu Zhu, and Liang Lin, "Beautygan: Instance-level facial makeup transfer with deep generative adversarial network," in *ACMMM*, 2018, pp. 645–653.

[17] Wentao Jiang, Si Liu, Chen Gao, Jie Cao, Ran He, Jiashi Feng, and Shuicheng Yan, "Psgan: Pose and expression robust spatial-aware gan for customizable makeup transfer," in *CVPR*, 2020, pp. 5194–5202.

[18] Michael Mathieu, Camille Couprie, and Yann LeCun, "Deep multi-scale video prediction beyond mean square error," *arXiv preprint arXiv:1511.05440*, 2015.

[19] Junsoo Lee, Eungyeup Kim, Yunsung Lee, Dongjun Kim, Jaehyuk Chang, and Jaegul Choo, "Reference-based sketch image colorization using augmented-self reference and dense semantic correspondence," in *CVPR*, 2020, pp. 5801–5810.

[20] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *arXiv preprint arXiv:1607.08022*, 2016.

[21] Diederik P. Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," in *ICLR*, 2015.