

Local highlight and shadow adaptively repairing GAN for illumination-robust makeup transfer

Zhihao Song

School of Computer Science (National Pilot Software Engineering School), Beijing University of Posts and Telecommunications, Beijing, China

2354470847@qq.com

Abstract. Recently, makeup transfer task has been widely explored with the development of deep learning. However, existing methods have shortcomings in more complex lighting situations in the real world because they do not consider the interference of lighting factors on facial features. To solve the above problem, we propose a local highlight and shadow adaptively repairing GAN for illumination-robust makeup transfer. We first map the 2D face images to UV representations and perform makeup transfer in the UV texture space, which explicitly removes the spatial misalignment to achieve pose and expression invariant makeup transfer. Furthermore, we take advantage of the face symmetry in the UV texture space to design an illumination repair module. It can adaptively repair the features affected by asymmetric local highlight and shadow based on a process of flipping and multi-layer attention fusion. In addition, the multi-layer attention maps are obtained by a pre-trained illumination classification network and hence have the ability to indicate local highlight and shadow areas. Comprehensive experiment results demonstrate the consistent effectiveness and clear advantages of our method, which significantly improve the robustness against local light effects and generate natural transfer results.

Keywords: facial makeup transfer, deep learning, generative adversarial network

1. Introduction

Recently, facial makeup transfer has received more and more attention due to its broad application prospects in e-commerce and live broadcast industries, which allows users to easily find the most suitable cosmetics for themselves and intuitively see the makeup effects. Given a source image and a reference image, the fundamental goal of makeup transfer is to transfer the makeup style of the reference image to the source image while preserving the facial identity of the source image.

Benefiting from the development of deep learning, especially generative adversarial network (GAN) [1], recent makeup transfer works represented by BeautyGAN [7] adopt the adversarial training framework of CycleGAN [6] to generate high-quality transferred image. To further solve the spatial misalignment problem caused by variation in pose and expression, PSGAN [10] and SSAT [14] leveraged feature-level spatial attention to align makeup features with source features at the corresponding semantic positions. However, none of the previous works considers the interference of lighting factors in real-world environment, which will affect the final makeup transfer effect. Essentially, the unnormalized lighting distribution on the face, such as local highlights and shadows, could lead to blur or even loss of the extracted facial features, then the subsequent attention calculation and feature alignment may also be biased.

To address the above problems, we propose a local highlight and shadow adaptively repairing GAN for illumination-robust makeup transfer. First, 3D face technique is used to decouple the input 2D images into face shape (i.e. geometry structure) and texture. Specifically, we obtain the texture map in UV space, where each pixel corresponds to a fixed semantic point on the face and hence eliminate the spatial misalignment. Then we pass the texture map to our proposed network to perform makeup transfer in the UV space. Inspired by [16], we further exploit the underlying face symmetry in the UV texture space to design an illumination repair module, which can adaptively repair the facial feature areas affected by the asymmetric local highlights and shadows through flipping and attentional merging process. Moreover, this attentional merging process is driven by the illumination-inspired attention maps extracted from multiple layers of a pre-trained illumination classification network. After obtaining the repaired texture, we pass it to the makeup transfer module to generate the transferred texture and then render it back to the standard image representation.

Finally, we conduct both qualitative and quantitative comparative experiment to show the advantages of our method to improve the makeup transfer effect under unnormal lighting conditions, in particular, asymmetric local highlights and shadows. We also analyse the effectiveness of our key network components by ablation study. In addition, our method naturally achieves pose and expression invariant makeup transfer since the face shape are explicitly normalized in the UV space.

2. Related work

Makeup transfer can be roughly divided into three categories in terms of methods: traditional methods [2, 3], methods based on CNN [4, 5], and methods based on GAN [7-15]. Recently, makeup transfer methods based on GAN have become the main research direction since GAN can significantly improve the quality of generated images. Therefore, we focus on GAN-based related works in this section.

BeautyGAN [7] and PairedCycleGAN [8] were the earliest representative works to introduce GAN. They learned from the unsupervised learning framework and the cycle consistency loss of CycleGAN [6], and hence did not require paired training data. Specifically, BeautyGAN [7] realized makeup transfer and makeup removal simultaneously with a dual input/output generator, and proposed a histogram matching loss to constraint the color distribution of component-wise makeup style. PairedCycleGAN [8] designed three local generators and discriminators for the eye, lip, skin regions, and blended these three regions using Poisson blending. Similar to PairedCycleGAN [8], LADN [9] designed multiple local discriminators for overlapping image regions to better transfer local details. However, the above works are only applicable to frontal, aligned face images, and fall short when spatial misalignment is caused by large pose and expression variations between input images. To address the spatial misalignment problem, PSGAN [10] proposed attentive makeup morphing module which calculated spatial attention and aligned the extracted makeup features to the source features. In a different way, SCGAN [11] extracted the style code of each makeup component from the reference image, then integrated identity and style code. Furthermore, to solve the problem of contaminated makeup caused by occlusion and shadow on the reference face, SOGAN [12] utilized the symmetry of the face texture in the UV space to design a flip attention module which repaired the contaminated makeup features and finally improved the robustness under occlusion situation. For extreme makeup styles such as stickers and patterns, CPM [13] proposed a dual-branch network architecture including a color branch and a pattern branch which realized the transfer of makeup color and pattern parallelly. Emphasized the importance of semantic correspondence, SSAT [14] introduced extra encoders to extract semantic features from the face parsing map, and established dense semantic correspondences by self-attention mechanism to improve the transferred makeup quality and accuracy. To prevent the loss of makeup details, EleGANt [15] designed a dual-branch network architecture, where high-resolution and low-resolution feature maps can be transferred respectively through their own branch. Therefore, high-frequency details were preserved through high-resolution feature maps and a shifted overlapped windowing attention module was proposed to reduce the computational overhead. In this work, we solve the makeup transfer task in the UV texture space and improve the performance under more challenging lighting conditions by designing an illumination repair module.

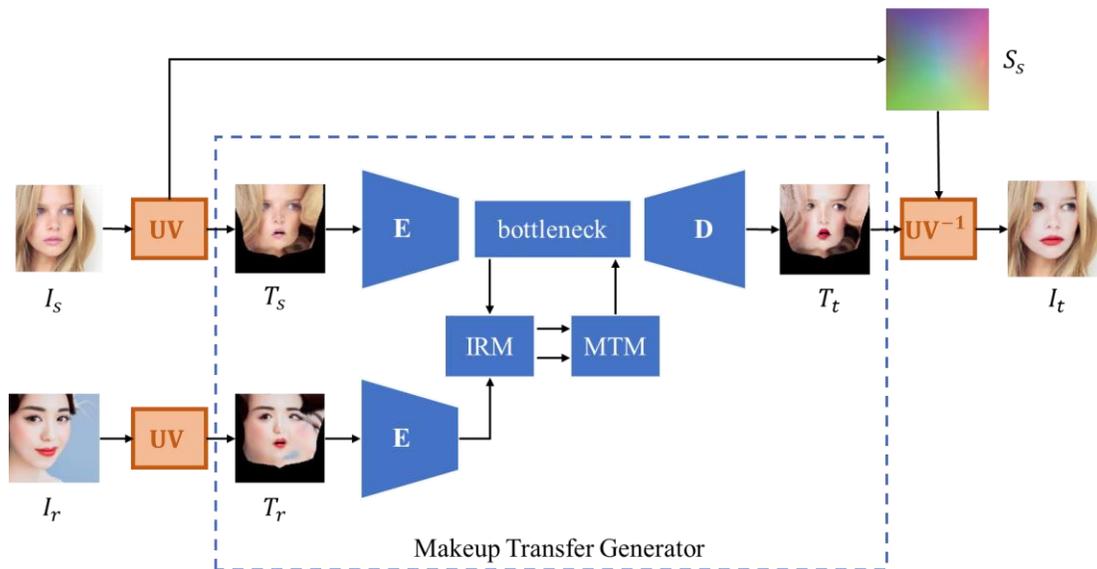


Figure 1. Overview of the pipeline of our method

3. Method

3.1. Overall pipeline

Different from previous 2D-to-2D methods, we perform the makeup transfer task in UV texture space based on the following two insights. First, the source face and the reference face are aligned in the UV texture space, and therefore avoid the influence of face structure, pose, and expression, which has been proven effective in CPM [13]. Second, we take advantage of the face symmetry in UV space to further alleviate the interference of asymmetric local highlights and shadows (detailed in subsection 3.2). In overview, our method consists of the following three main steps, as shown in Figure 1.

(1) UV mapping. Given a source image I_s and a reference image I_r , we first convert them to 3D representations including UV position maps S_s, S_r and UV texture maps T_s, T_r respectively, as in equation (1). In this work, we use PRnet [17] (denoted as UV) to obtain the disentangled UV position maps and UV texture maps. The UV position map only contains geometric information such as face shape, pose, expression while the UV texture map only contains color information related to makeup transfer. Therefore, the UV position map remains unchanged and does not participate in the subsequent makeup transfer process.

$$S_s, T_s = UV(I_s) \text{ and } S_r, T_r = UV(I_r) \quad (1)$$

(2) Makeup transfer in UV texture space. We pass the source texture map T_s and reference texture map T_r into a generator network for makeup transfer, and output the transferred texture map T_t , as in equation (2). Following previous works, the encoder-bottleneck-decoder architecture is employed as our network backbone, where the encoder applies two down-sampling layers to extract features from the input textures. In order to solve the problem of local highlights and shadows on the input face, in the bottleneck part, we design an illumination repair module (IRM) based on the face symmetry in the UV texture space to repair the illumination affected features. Then the repaired features are passed through the makeup transfer module (MTM) and the decoder to generate the transferred texture map T_t .

$$T_t = G(T_s, T_r) \quad (2)$$

(3) UV reverse mapping. The final output image I_t can be recovered from the source position map S_s and the transferred texture map T_t by a rendering function UV^{-1} , as in equation (3).

$$I_t = UV^{-1}(S_s, T_t) \quad (3)$$

3.2. Key module design

Inspired by [16], we exploit the horizontal symmetry of facial textures in UV space to eliminate the negative effects of local highlights and shadows. We assume that the illumination distribution on the input faces is asymmetric, which is a common case in real-world images with various illumination angles. Based on the above assumptions, when there are local highlights and shadows on one side of the face, we have a high probability of extracting features that are not affected by those light effects on the other symmetrical side of the face. For example, assuming a light from left direction is projected on the face and forms a local shadow on the right side of the nose, which causes corresponding feature details to be blurred or lost. However, the symmetrical area on the left side of the nose is normal lighted, so we can use it to repair the shadow affected area to get the refined feature details.

Furthermore, we need to focus on the local shadow and highlight areas and hence introduce the visual attention mechanism into our architecture. Generally, the goal of visual attention or visual saliency is to make something stand out from its surroundings. Since human faces have similar shape and similar surface reflectance, illumination classification is mainly depending on the changes in face appearance caused by illumination variation. Therefore, we use a pretrained illumination classification network to obtain the attention map, which can reflect the illumination distribution i.e. giving higher weight to local highlights and shadow areas. Along with the attention map, we design a flip and attentional fusion process in our illumination repair module, as shown in Figure 2.

Specifically, we first need to convert the original high-dimensional features extracted from the intermediate layers of the illumination classification network into a single-channel attention mask by channel-level aggregation and sigmoid normalization. Since it is not clear which layer of illumination attention map contributes the most to our task, we determine to adopt a multi-layer attention strategy. Based on the horizontal symmetry of the face in UV texture space, the illumination repair module extracts unaffected high-quality features from the original face features F and the horizontally flipped face features F' according to the 1-layer attention mask M_l and fused them into higher-quality features F_l , as in equation (4). Then multi-layer F_l are concatenated to obtain the final repaired feature \hat{F} , as in equation (5).

$$F_l = (1 - M_l) \otimes F \oplus M_l \otimes F' \quad (4)$$

$$\hat{F} = F_1; F_2; \dots; F_l \quad (5)$$

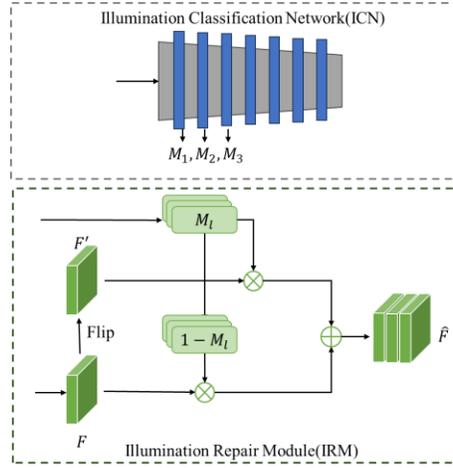


Figure 2. Illustration of our key network modules. The bottom figure shows detailed process of the illumination repair module and the top figure shows the illumination classification network, which consists of multiple convolutional blocks.

where \otimes denotes element-wise multiplication, \oplus denotes element-wise addition, ; denotes channel-wise concatenation and l denotes the layer number. Note that M_l at different scales should be up-sampled or down-sampled before the attentional fusion step so that it has the same width and height as the feature map F . In our implementation, l is set to $\{1,2,3\}$, which means the first three layers are used.

3.3. Loss function

We first use the histogram matching loss proposed by BeautyGAN [7]. The histogram matching is calculated between the generated image I_t and the reference image I_r to constrain them to have similar makeup styles at the corresponding local areas. Then the local histogram loss L_{item} is formulated as equation (6):

$$L_{item} = \|HM(I_t \otimes M_{item}, I_r \otimes M_{item}) - I_t \otimes M_{item}\|_2 \quad (6)$$

where $HM(I_t \otimes M_{item}, I_r \otimes M_{item})$ denotes the histogram matching. Specifically, M_{item} is the facial mask obtained by the face parsing model [18], where $item \in \{lips, eyes, face\}$. Therefore, the final histogram matching loss L_{hm} consists of three local histogram losses for lips, eyes and face, as in equation (7):

$$L_{hm} = \lambda_1 L_{lips} + \lambda_2 L_{eyes} + \lambda_3 L_{face} \quad (7)$$

In addition, we adopt perceptual loss in texture space to measure the high-dimensional feature distance between the generated texture T_t and the source texture T_s . Specifically, we use the VGG-16 model [19] pre-trained on Imagenet [20], whose *relu-4-1* layer output is taken as the high-dimensional feature. Then the perceptual loss L_{per} is formulated as equation (8):

$$L_{per} = \|VGG(T_t) - VGG(T_s)\|_2 \quad (8)$$

Since no paired training data is provided for supervision, we follow previous works to introduce the cycle consistency loss proposed by CycleGAN [6], which is defined as the pixel-level loss between the reconstructed texture and the input texture, as in equation (9):

$$L_{cyc} = \|G(G(T_s, T_r), T_s) - T_s\|_1 + \|G(G(T_r, T_s), T_r) - T_r\|_1 \quad (9)$$

The adversarial loss is a fundamental loss of the GAN-based makeup transfer network, aiming to improve the realism of the generated results. We introduce a discriminator D_{tex} on texture space and a discriminator D_{img} on image space to distinguish real and generated textures/images. Similar to BeautyGAN [7], the original adversarial loss is replaced by the least squares loss in LSGAN [21] to stabilize the training procedure:

$$L_{adv}^{D_{tex}} = \mathbb{E}[(D_{tex}(T_r) - 1)^2] + \mathbb{E}[(D_{tex}(T_t))^2] \quad (10)$$

$$L_{adv}^{G_{tex}} = \mathbb{E}[(D_{tex}(T_t) - 1)^2] \quad (11)$$

$$L_{adv}^{D_{img}} = \mathbb{E}[(D_{img}(I_r) - 1)^2] + \mathbb{E}[(D_{img}(I_t))^2] \quad (12)$$

$$L_{adv}^{G_{img}} = \mathbb{E}[(D_{img}(I_t) - 1)^2] \quad (13)$$

Finally, the total loss can be obtained by summing up all the above losses. The total loss for the generator and the discriminator is formulated as equation (14) and equation (15), respectively:

$$L_{total}^G = L_{adv}^{G_{tex}} + L_{adv}^{G_{img}} + \lambda_{cyc} L_{cyc} + \lambda_{per} L_{per} + \lambda_{hm} L_{hm} \quad (14)$$

$$L_{total}^D = L_{adv}^{D_{tex}} + L_{adv}^{D_{img}} \quad (15)$$

4. Experiment

4.1. Experiment settings

We use the MT dataset [7] to train our model. MT dataset contains 1115 non-makeup images and 2719 makeup images, from which 100 non-makeup images and 250 makeup images are randomly selected for testing, and the remaining images are divided into training set and validation set. Before fed into our network, all the face images are converted into UV texture representations by PRnet [17]. During training, the learning rate is set to 0.0001, and the batch size is set to 1.

When setting up comparative experiments, we choose several state-of-the-art methods that have released codes and pretrained model for a fair comparison, including BeautyGAN [7], PSGAN [10], SCGAN [11], and SSAT [14]. As shown in Table 1, we summarize the functionalities of different makeup transfer methods, including whether they can handle some unusual illumination situations, whether they can handle spatial misalignment under large pose and expression, and whether their makeup synthesis is shade-controllable. Our method not only covers the functionalities of existing methods, but is also the only one capable of handling illumination.

Table 1. Functionality analysis of makeup transfer methods

Methods	Illumination	Misalignment	Controlibility
BeautyGAN			
PSGAN		✓	✓
SCGAN		✓	✓
SSAT		✓	✓
Ours	✓	✓	✓

4.2. Comparative experiment

To qualitatively evaluate our method, we directly generate the makeup transfer results of other state-of-the-art methods for visual comparison, as shown in Figure 3. In the first row: the reference image has large area of shadow on the right side of the face, which makes existing methods suffer from severe artifacts, including erroneous shading at the corners of the mouth and eyes (appeared in almost all methods except ours), and obvious changes in irrelevant background contents (appeared in SCGAN). In the second row: the source image has a local shadow on the left cheek, which also causes non-negligible problems on existing methods, such as lipstick bleeding (appeared in BeautyGAN), unnatural eye shadow (appeared in PSGAN, SCGAN), and even missing pixels on the forehead (appeared in SSAT). As can be seen in Figure 3, our method not only avoids the artifacts in other methods, but also outperforms other methods in terms of overall visual effects and reconstruction of makeup styles such as lipstick and eye shadow.



Figure 3. Visual comparison with state-of-the-art methods

Following previous works, we also conduct a user survey to quantitatively compare and analyse our method with other state-of-the-art methods. First, we randomly sample 10 source-reference image pairs with frontal normal light from the MT dataset, and then obtain the makeup transfer results for each method, named "Normal lighting set". To further evaluate the ability to resist light interference, we additionally sampled another 10 source-reference image pairs with local highlights and shadows from the MT data set, and then obtain the makeup transfer results for each method, named "Unnormal lighting set". A total of 20 volunteers

participated in the user survey. They evaluated the makeup transfer results of different methods in terms of overall visual effect and similarity to the reference makeup, and then selected the best one in their own perspective. Table 2 shows the results of our user survey, where the data represents the best selection rate (%), that is, the proportion of the current method being selected as the best or ranked first. Our method achieved the highest best selection ratio, especially on the "Unnormal lighting set", thus demonstrating the significant advantage of our method under more challenging lighting conditions.

Table 2. User survey on two evaluation sets: "Normal lighting set" and "Unnormal lighting set"

	BeautyGAN	PSGAN	SCGAN	SSAT	Ours
Normal light set	18.5	14.0	14.5	19.5	33.5
Unnormal light set	13.0	6.5	8.0	11.0	61.5

4.3. Ablation study

In this work, an illumination repair module with multi-layer attention is proposed to adaptively repair the facial feature affected by local highlights and shadows to obtain better transfer results. So we conduct an ablation study to verify their effectiveness, as shown in Figure 4. Specifically, "w/o IRM" and "w/o MLA" denote different ablation configurations. "w/o IRM" means that the Illumination Repair Module (IRM for short) is skipped and other network structures remain unchanged. "w/o MLA" means using a variant illumination repair module without the Multi-Layer Attention strategy (MLA for short), that is, set $l=2$ in Equation (4) and omitting the concatenation operation of Equation (5).

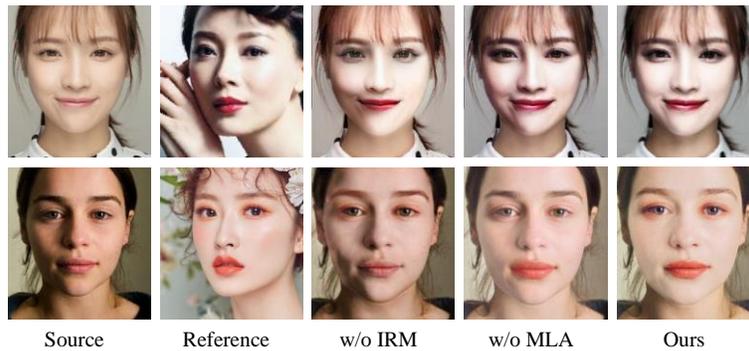


Figure 4. Ablation study of key designs in our method

As can be seen in the first row of Figure 4, without the illumination repair module, the generated image incorrectly transfers the shadows of the left cheek and eye corners of the reference image. Without multi-layer attention, this problem improves but there is still some shadow on the left face edge. With complete designs, our method can generate natural and real makeup transfer effect. In the second row, due to the large area of shadow on the left side of the face in the source image, the generated image without the illumination repair module produces unnatural artifacts. While the difference between the results without multi-layer attention and with complete designs is not obvious, thanks to the fact that it still uses the illumination repair module to handle negative light effects. Therefore, the ablation experimental results prove the effectiveness of the illumination repair module with multi-layer attention design of our method. Especially, the illumination repair module plays a key role in makeup transfer task under harsh lighting conditions, which can significantly improve the transfer quality.

4.4. More results under large pose and expression

To further test the robustness of our method under large pose and expression, we tested it on the Makeup-Wild dataset [10], which contains 403 makeup images and 369 non-makeup images with various poses, expressions and complex backgrounds. Figure 5 shows our test results on the Makeup-Wild dataset, where the first column is the input source image, and the first row is the input reference image. It can be observed that even if any pair of source and reference images have different poses and expressions, as well as different makeup styles, our method still achieves good effect on all transferred images since we perform makeup transfer in the spatially-aligned UV texture space.

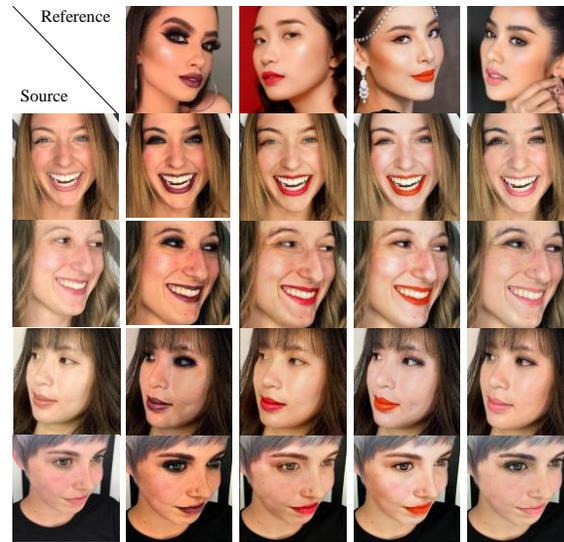


Figure 5. Test results on the Makeup-Wild dataset

5. Conclusion

In this work, we focus on the makeup transfer task under unnormal lighting conditions, which is still challenging to existing method. Therefore, we propose a local highlight and shadow adaptively repairing GAN for illumination-robust makeup transfer. First, we solve the spatial misalignment problem by performing makeup transfer in UV texture space, where the face shapes are explicitly normalized. Furthermore, since human face has horizontal symmetry in UV texture space, we design a process of flipping operation and attentional fusion to repair the features affected by asymmetric local highlight and shadow, thus propose the illumination repair module. In addition, an illumination classification network is introduced to form the multi-layer attention maps that indicates highlight and shadow areas. Thanks to the design of the illumination repair module, our network gains robustness to handle negative light effects, achieving promising results especially in the cases such as local highlight and shadow. Extensive experiments demonstrate the superiority of our method compared to other state-of-the-art methods.

References

- [1] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27, 2672–2680.
- [2] Tong, W. S., Tang, C. K., Brown, M. S., & Xu, Y. Q. (2007). Example-based cosmetic transfer. In *15th Pacific Conference on Computer Graphics and Applications* (pp. 211–218).
- [3] Guo, D., & Sim, T. (2009). Digital face makeup by example. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 73–79).
- [4] Liu, S., Ou, X., Qian, R., Wang, W., & Cao, X. (2016). Makeup like a superstar: Deep localized makeup transfer network. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence* (pp. 2568–2575).
- [5] Wang, S., & Fu, Y. (2016). Face behind makeup. In *Proceedings of the AAAI Conference on Artificial Intelligence* (pp. 58–64).
- [6] Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 2223–2232).
- [7] Li, T., Qian, R., Dong, C., Liu, S., Yan, Q., Zhu, W., & Lin, L. (2018). Beautygan: Instance-level facial makeup transfer with deep generative adversarial network. In *Proceedings of the 26th ACM International Conference on Multimedia* (pp. 645–653).
- [8] Chang, H., Lu, J., Yu, F., & Finkelstein, A. (2018). Pairedcyclegan: Asymmetric style transfer for applying and removing makeup. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 40–48).
- [9] Gu, Q., Wang, G., Chiu, M. T., Tai, Y. W., & Tang, C. K. (2019). Ladn: Local adversarial disentangling network for facial makeup and de-makeup. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 10481–10490).
- [10] Jiang, W., Liu, S., Gao, C., Cao, J., He, R., Feng, J., & Yan, S. (2020). Psgan: Pose and expression robust spatial-aware gan for customizable makeup transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5194–5202).
- [11] Deng, H., Han, C., Cai, H., Han, G., & He, S. (2021). Spatially-invariant style-codes controlled makeup transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 6549–6557).

- [12] Lyu, Y., Dong, J., Peng, B., Wang, W., & Tan, T. (2021). SOGAN: 3D-Aware Shadow and Occlusion Robust GAN for Makeup Transfer. In *Proceedings of the 29th ACM International Conference on Multimedia* (pp. 3601–3609).
- [13] Nguyen, T., Tran, A. T., & Hoai, M. (2021). Lipstick ain't enough: Beyond color matching for in-the-wild makeup transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 13305–13314).
- [14] Sun, Z., Chen, Y., & Xiong, S. (2022). Ssat: A symmetric semantic-aware transformer network for makeup transfer and removal. In *Proceedings of the AAAI Conference on Artificial Intelligence* (pp. 2325–2334).
- [15] Yang, C., He, W., Xu, Y., & Gao, Y. (2022). Elegant: Exquisite and locally editable gan for makeup transfer. In *Proceedings of European Conference on Computer Vision* (pp. 737–754).
- [16] Wu, S., Rupprecht, C., & Vedaldi, A. (2020). Unsupervised learning of probably symmetric deformable 3d objects from images in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 1–10).
- [17] Feng, Y., Wu, F., Shao, X., Wang, Y., & Zhou, X. (2018). Joint 3d face reconstruction and dense alignment with position map regression network. In *Proceedings of European Conference on Computer Vision* (pp. 534–551).
- [18] Yu, C., Wang, J., Peng, C., Gao, C., Yu, G., & Sang, N. (2018). Bisenet: Bilateral segmentation network for real-time semantic segmentation. In *Proceedings of European Conference on Computer Vision* (pp. 325–341).
- [19] Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations* (pp. 1–14).
- [20] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 248–255).
- [21] Mao, X., Li, Q., Xie, H., Lau, R. Y., Wang, Z., & Paul Smolley, S. (2017). Least squares generative adversarial networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 2794–2802).