Cognitive Robotics 1 (2021) 182–187

Contents lists available at ScienceDirect

Cognitive Robotics

journal homepage: http://www.keaipublishing.com/en/journals/cognitive-robotics/

Deep learning method for makeup style transfer: A survey

Xiaohan Ma[a,*], Fengquan Zhang[a], Huan Wei[a], Liuqing Xu[b]

[a] School of Information Science and Technology, North China University of Technology, Beijing 100144, China,
[b] Airborne remote sensing center, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China

ARTICLE INFO

Keywords:
Makeup transfer
Image style transfer
Reference makeup
Deep learning

ABSTRACT

Makeup transfer is one of the applications of image style transfer, which refers to transfer the reference makeup to the face without makeup, and maintaining the original appearance of the plain face and the makeup style of the reference face. In order to understand the research status of makeup transfer, this paper systematically sorts out makeup transfer technology. According to the development process of the method of makeup transfer, our paper first introduces and analyzes the traditional methods of makeup transfer. In particular, the methods of makeup transfer based on deep learning framework are summarized, covering both disadvantages and advantages. Finally, some key points in the current challenges and future development direction of makeup transfer technology are discussed.

1. Introduction

With the progress of society and the vigorous development of the beauty industry, how to help people quickly and accurately find their own beauty products has gradually become a research hotspot. Makeup transfer is a new application of Virtual Reality technology in images. How to quickly see virtual makeup effects on images is now a need for many young people. Therefore, makeup transfer technology has received more and more attention.

The ideal makeup transfer method needs to ensure the face appearance of the plain face image, only transferring the makeup style of the reference image, and the final output generated image automatically presents the perfect combination of the plain face portrait and the reference makeup. Because people's facial expressions, face shapes, lip shapes, eyebrow shapes, eye shapes and the distance between eyebrows and eyes are different, and the output generated image may not be well integrated, or even distorted. What's more, the style of makeup is changeable and irregular, and it is also affected by age and race. In order to solve these problems, the existing makeup transfer methods can be divided into two categories: traditional makeup transfer [1–3] and makeup transfer based on deep learning [4–8].

The first part of this paper introduces what is makeup transfer, the common problems in achieving makeup transfer, and the development process of makeup transfer technology; the second part is traditional methods of makeup transfer, which are classified and compared; the third part is a makeup transfer method based on deep learning, which classifies and introduces these methods and analyzes the advantages and disadvantages of various methods; the fourth part is challenges and prospects, discussing some of the challenges and future development directions of this technology; the fifth part is the result, summarizing the work of this paper and discussing the practical significance of this technology.

* Corresponding author.
E-mail address: m2016071025@163.com (X. Ma).

https://doi.org/10.1016/j.cogr.2021.09.001
Received 7 August 2021; Received in revised form 23 September 2021; Accepted 30 September 2021
Available online 8 October 2021

2. Traditional methods of makeup transfer

Most of the traditional makeup transfer methods have high requirements on the reference image and the target image. For example, the light and the posture of the characters in the two images are required to be the same. Therefore, the final result is not very ideal, which will seriously lead to the loss of some makeup or small parts of the face in the generated image. According to the different requirements of the data set, it can be divided into two types: supervised model [1] and unsupervised model [2]. Supervised models require paired data sets, while unsupervised models do not.

2.1. Algorithm based on paired data sets

Makeup transfer methods base on paired data sets requires the target image and the reference image pair before and after makeup, the three images should be obtained in the same posture and light. Firstly, calculate the changes in the image color and illumination before and after applying makeup. Then modify the skin texture and color difference between the reference face and the target face. Finally, quickly makeup transfer and adjustment of the makeup style for the given target face.

The transfer method of eye makeup is different from foundation and lip gloss, so we can use Bayesian matting [9] to remove eyebrows and eyelashes, and then use graph cutting texture synthesis algorithm [10] to fill the holes formed after removing eyebrows and eyelashes, and use the images around the holes as the texture patch. Freckles, moles and stains in the face should not be transfer, so skin patches can be selected from the faces of reference image pairs, and Independent Component Correlation Algorithm (ICA) [11] is applied to the patches to obtain the separation matrix of two kinds of pigments of some visible spots on human skin. The facial geometry can be distorted into a standard facial model by manually specifying the two-dimensional point correspondence between the input image and the standard face model, and we can reduce manual processing by using the active appearance model [12] for face training.

The mapping of the plain reference image to the make-up reference image can be expressed by the quotient of the pixels after and before make-up [1], and can also be found by the 3D deformation model of the face [13]. For the transfer of eye makeup, the extracted eyebrows and eyelashes mask can be used to superimpose the eyebrows and eyelash selections of the reference picture before makeup on the reference picture after makeup to realize the migration of eyebrows and eyelashes.

2.2. Algorithm based on unpaired data sets

Because it is difficult to obtain a reference image pair before and after makeup, a method that only requires the plain target image and the makeup reference image is proposed. Traditional makeup transfer methods based on unsupervised model mostly decompose two images into three layers.

Makeup transfer methods base on paired data sets needs to align the faces of the two images. We can use ASM [14] or the data acquisition method described in paper [15] to obtain the required control points. Because of the diversity of faces, we can add control points manually. Then use Thin Plate Spline (TPS) [16] to warp the reference image into the target image or use Gaussian Mixture Model (GMM) to adjust the skin area [17].

Convert the target image and reference image to CIELAB color space, decompose it into color layer and brightness layer, use WLS algorithm or bilateral filtering method [18] to perform edge-preserving and smoothing on the brightness layer to obtain the face structure layer. The face structure layer is subtracted from the brightness layer to obtain the skin detail layer.

Different layers use different transfer methods to achieve better results. For the skin detail layer, the weighted addition method can be used. For the color layer, when it belongs to the skin, the weights of the two are added. The effect of highlight and shadow is retained in the face structure layer, and the gradient algorithm [19] is used to maintain the light of the target graph, and the smooth effect is generated during transfer.

2.3. Comparison of traditional methods

In traditional methods, whether it is a makeup transfer method based on a supervised model or an unsupervised model, the posture and illumination requirements of the input image are relatively high. Comparing the methods of Gou et al. [2] and Tong et al. [1], the results are shown in Fig. 1. It can be seen that the method of Gou et al. preserves the face structure better. For the skin color, Gou et al. extracted the skin color from the reference image, and Tong et al. retained the original skin color.

3. Makeup transfer methods based on deep learning

With the rapid development of deep learning technology, the makeup transfer method has made breakthrough progress. In order to make makeup more integrated and more realistic, the use of deep learning methods for makeup migration is the current mainstream algorithm. At present, it mainly includes methods based on image pixel iteration [4,17] and methods based on model iteration [7,20–22].

3.1. Methods based on image pixel iteration

The goal of image pixel iteration is to minimize the total loss function, so that the generated image can simultaneously match the content of the target image with the makeup of the reference image. Gatys et al. [23] first proposed the use of image iteration

Fig. 1. Comparison of the methods of Tong et al. and Gou et al. The picture comes from the paper [2], the first column is the reference face, the second column is the target face, the third column is the transfer result of Tong et al., and the fourth column is the transfer result of Gou et al.

for style transfer. The method is to select a higher-level feature layer on the Convolutional Neural Network (CNN) [24] and use the Euclidean distance to calculate the content loss to preserve the content and position of the main body of the image. The lower-level feature layer represents the style texture information, and uses the Gram matrix to describe the texture of the image [25].

The focus of makeup migration with this method is to make the partial area of the plain face correspond to the partial area of the makeup. Full Convolutional Network (FCN) [26] can be used to semantically segment the reference image and the without makeup image, and establish the corresponding relationship between the two images [4,27]. It is also possible to select an image similar in appearance and posture to the input the without makeup image from the face database [28], and then adjust the posture, expression, lighting, color and other information of the selected reference image, and finally merge the two images [29][9].

3.2. Methods based on model iteration

Although the image generation based on image pixel iteration method has a good effect, it has the problem of low computational efficiency, and the makeup transfer method based on model iteration can effectively solve this problem. At present, the method is mainly based on GAN generation model [7,20,21]. In the past two years, some scholars have also studied the method of makeup transfer based on Glow generation model [22].

3.3. Methods based on GAN generative model

The core idea of Generative Adversarial Network (GAN) [30] is to continuously optimize the generator and discriminator to reduce the distribution difference between the real image and the generated image.

Inputting a random noise in the traditional GAN network will output a random image. Although the generated image looks good, when you zoom in on the details, you will find that it is very fuzzy. Pix2pix [31][1] draws on the idea of cGAN [32][2], the generator uses the U-Net structure [33][3], and the discriminator uses the conditional discriminator PatchGAN [31][1], but the model requires paired data sets. The CycleGAN [31][1] model can be regarded as the fusion of two GANs, which respectively realize image generation and discrimination from X domain to Y domain and image generation and discrimination from Y domain to X domain. The two networks form a cyclic process, and there is no need paired data sets. This model can realize makeup on without makeup face and remove makeup on reference face, but it can only realize general makeup transfer and the quality of the generated image is not very high. Pix2pixHD [34][4] uses a multi-scale cGAN [32][2] structure for picture-to-picture changes. First, it outputs a low-resolution image, then using this image as an input to another network, and finally generating a higher-resolution image. The StarGAN [35][5] model can realize the mapping between multiple domains with only a pair of generators and discriminators, and can effectively train images in each domain. Compared with the CycleGAN [31][1] and the cGAN [32][2], this model retains more facial features, generates better image quality, and produces better transfer results. The BeautyGAN [7] model uses a pair of discriminators to transfer the non-makeup face to the makeup domain. The discriminator distinguishes the generated image from the real samples of the domain. Based on the domain set transfer, it uses pixels based on different face regions. Level histogram loss is used to achieve instance-level migration. In order to maintain the consistency of the face and eliminate artifacts, perceptual loss and cyclic consistency loss are added to the overall objective function. The model can transfer makeup images and without makeup images at instance level. In order to solve the large difference between the reference face and without makeup face, Jiang et al. proposed the PSGAN [21] model. This model uses the encoder-bottleneck part of the generator structure in StarGAN [35][6] to extract facial features, and then uses the AMM [21] module to adaptively modify the makeup matrix by introducing an attention mechanism to solve the robust transfer of posture and expression. The FSGAN [36][6] model evaluates the occlusion area by combining face segmentation, which solves the

Table 1
Comparison of makeup transfer methods based on deep learning.

Methods	The quality of the generated image	Demand for hardware	Expansion	Training data	Speed of calculation
Using CNN [31]	good	ordinary	ordinary	unnecessary	slow
CycleGAN [31]	ordinary	ordinary	high	necessary	faster
BeautyGAN [31]	ordinary	ordinary	high	necessary	faster
BeautyGlow [31]	good	high	ordinary	necessary	quick

Fig. 2. Comparison of makeup transfer methods based on deep learning. The picture comes from the literature [20], the first column is the reference face, the second column is the target face, the third column is a generated graph based on the Cycle GAN model, the fourth column is a generated graph based on the Beauty GAN model, and the fifth column is a generated graph based on the Beauty Glow model.

occlusion problem in the makeup migration process to a certain extent. The SCGAN [37][7] model decomposes the makeup transfer problem into two steps of extraction and allocation. The part-specific style encoder extracts the features of each part and maps them into a disentangled style latent space, and the face identity encoder extracts the face identity features of the target image. The makeup fusion decoder fuses the style code with the facial identity features to generate final result. This design eliminates the problem of spatial misalignment. By editing the style code, you can easily achieve global/local tone transfer and removal and shadow control without additional calculation work. Thao Nguyen et al. [38] proposed to build a unified template that can align the 3D head pose, facial shape and facial expressions of the source image and the target image. Makeup migration is based on BeautyGAN [7], and proposes to use UV texture map instead of original image for makeup exchange.

3.4. Methods based on Flow generative model

The flow-based generative model was first proposed in NICE [39] and extended in Real NVP [40]. Compared with GAN and VAG, it has been ignored. Until the article Glow [41] appeared, it was considered that the flow-based generative model was pushed to the forefront of academics. Glow has achieved good results in the field of image generation, especially in the accurate inference of latent vectors obtained from image coding. In the direction of makeup transfer, compared with the methods based on the GAN generation model, the methods based on the Glow generation model does not need to train the two large networks of the discriminator and the generator, and the time to automatically synthesize the results is very fast.

The Glow model introduces 1×1 reversible convolution on the basis of RealNVP [40] and removes other components of RealNVP, thus simplifying the overall architecture and making it easier to understand and use. Chen et al. proposed the Beauty Glow model [22], which uses the Glow model to obtain the respective latent space of the input image (makeup reference image and without makeup target image), and decomposes the latent space into facial features and makeup features respectively. Then, the makeup features of the reference image and the facial features of the target image are added to get the latent space of the target image with makeup. Finally, the Glow model is used to reverse and transform it into the RGB target image with makeup.

3.5. Comparison based on deep learning methods

As shown in Table 1, each method of makeup transfer based on deep learning has advantages and disadvantages. On the MT data set, the three models of CycleGAN [31], BeautyGAN [7], and BeautyGlow [22] were compared for makeup transfer, and the results are shown in Fig. 2.

4. Challenges and prospects

4.1. Challenges of makeup transfer technology

- The quality of the generated image. As shown in Fig. 3, it is the result of the style transfer of the source image by several existing style transfer methods. From the results of the transfer of the face structure, the symmetry of the eyes, the corners of the eyes, the corners of the lips and other details need to be further increased in accuracy. From the results of makeup transfer, it should be both beautiful and real.
- Timeliness. The network structure can be improved to realize rapid makeup transfer and apply to mobile terminals. As shown in Table 2, Huang et al. [44]calculated the time consumption of 28 samples to complete the makeup results. From the table, we can

Fig. 3. Comparison of makeup transfer methods. The picture comes from the literature [37], showing the results of a variety of makeup transfer methods.

Table 2
The average generation time and average covariance of makeup transfer for 28 sample pairs by different methods.

	Gatys et al [23]	Luan et al [42]	Li et al [43]	Huang et al [44]
Time(s)	184.30	85.07	0.42	0.23
Log(Cov)	6.3153	6.3710	6.3100	6.3088

see that the makeup transfer method has been improving in the direction of timeliness, but now there are fewer methods applied to the mobile terminal, and the timeliness is not enough.

- Data set. There are few types of makeup styles available, and there may not be the most desired makeup for the target face in the reference makeup.
- Makeup recommendation algorithm. The algorithm can be improved to find the most suitable makeup for the target face from the data set.

4.2. Future development of makeup transfer technology

With the development of 5G technology, makeup transfer technology has also received more and more attention. In my opinion, the work can be carried out from the following four aspects in the future: (1) Realize the migration of special makeup, such as Beijing Opera makeup, animation character makeup, etc.; (2) Realize regional makeup transfer, only the lip gloss of one reference makeup can be transferred, and then the foundation of another reference makeup can be transferred to create the makeup that is most suitable for the target face; (3) Develop makeup transfer technology in video; (4) Combine expression transfer with makeup transfer.

5. Result

In recent years, makeup transfer technology has achieved good results, which are jointly promoted by social needs and scientific challenges. The traditional makeup transfer method is computationally complex, inefficient, and the transfer result is not real enough; while the method based on deep learning can separate the high-level features of the image, improve the calculation speed and the quality of the generated image.

In general, makeup transfer technology based on deep learning not only promotes the application of Virtual Reality technology in images, but also promotes the development of the beauty makeup industry. From the perspective of both academic and commercial applications, we should continue to explore and study this technology, further optimize the indicators of the makeup transfer model, and improve the generation speed and generalization degree while ensuring the quality of the generated image.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This paper is supported by Humanities and Social Sciences Foundation of Chinese Ministry of Education (No. 19YJC760150), (KM201810009006), National Natural Science Foundation (No. 61402016), the open funding project of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University (No. VRLAB2018A05).

References

[1] D. Tong, CK. Tang, MS. Brown, et al., Example-based cosmetic transfer" [C], IEEE Computer Society, 2007.
[2] G. Dong, T. Sim, "Digital face makeup by example" [C], Computer Vision and Pattern Recognition, CVPR 2009, IEEE, 2009.
[3] K. Scherbaum, T. Ritschel, M. Hullin, et al., Computer-Suggested Facial Makeup" [C], Computer Graphics Forum, (2011).
[4] L. Si, X. Ou, R. Qian, et al., Makeup like a superstar: deep localized makeup transfer network" [J], AAAI Press, 2016.
[5] J. Johnson, A. Alahi, L. Fei-Fei, Perceptual losses for real-time style transfer and super-resolution", [C], European Conference on Computer Vision, 2016.
[6] J. Liao, Y. Yao, L. Yuan, et al., Visual attribute transfer through deep image analogy", [J], ACM Trans. Graphics Papers 36 (4) (2017) 120.

[7] T. Li, R. Qian, C. Dong, et al., BeautyGAN: instance-level facial makeup transfer with deep generative adversarial network", [C], 2018 ACM Multimedia Conference, 2018.
[8] Kaur, P., Zhang, H., Dana, K J., "Photo-realistic facial texture transfer" [J], (2017).
[9] YY. Chuang, B. Curless, DH. Salesin, et al., A bayesian approach to digital matting", [C], IEEE Computer Society Conference on Computer Vision & Pattern Recognition, 2003.
[10] V. Kwata, Graphcut textures: image and video synthesis using graph", [J], ACM Trans. Graphics, Proc. SIGGRAPH2003 22 (3) (2003) 277–286.
[11] N. Tsumura, N. Ojima, K. Sato, et al., Image-based skin color and texture analysis/synthesis by extracting hemoglobin and melanin information in the skin", [J], ACM Trans. Graphics Papers 22 (3) (2003) 770–779.
[12] TF. Cootes, GJ. Edwards, CJ. Taylor, Active appearance models", in: [C], Proceedings of the 5th European Conference on Computer Vision-Volume II, II, 1998 - Volume.
[13] V. Blanz, T. Vetter, A. Rockwood, A morphable model for the synthesis of 3D Faces, [J], acm siggraph (2002) 187–194 Papers.
[14] S. Milborrow, F. Nicolls, Locating facial features with an extended active shape model", [C], European Conference on Computer Vision, 2008.
[15] C. Lee, MT. Schramm, M. Boutin, et al., An algorithm for automatic skin smoothing in digital portraits", [C], IEEE International Conference on Image Processing, 2010.
[16] FL. Bookstein, Principal warps: thin-plate splines and the decomposition of deformations", [J], IEEE Trans. Pattern Anal. Machine Intell. Papers 11 (6) (2002) 567–585.
[17] X. Lin, Y. Du, Y. Zhang, An automatic framework for example-based virtual makeup", [C], 2013 20th IEEE International Conference on Image Processing (ICIP), 2014.
[18] C. Tomasi, R. Manduchi, Bilateral filtering for gray and color images", [C], International Conference on Computer Vision, 2002.
[19] P. Perez, M. Gangnet, A. Blake, Poisson image editing", [J], ACM Trans. Graphics (TOG), Papers 22 (3) (2003) 313–318.
[20] H. Chang, J. Lu, et al., PairedCycleGAN: asymmetric style transfer for applying and removing makeup" [C], 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018.
[21] W. Jiang, S. Liu, C. Gao, et al., PSGAN: pose and expression robust spatial-aware GAN for Customizable Makeup Transfer", 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
[22] HJ. Chen, KM. Hui, SY. Wang, et al., BeautyGlow: on-demand makeup transfer framework with reversible generative network", [C], 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.
[23] LA. Gatys, AS. Ecker, M. Bethge, Image Style Transfer Using Convolutional Neural Networks", [C], 2016 IEEE conference on computer vision and pattern recognition (CVPR), 2016.
[24] A. Krizhevsky, I. Sutskever, G. Hinton, ImageNet classification with deep convolutional neural networks", [J], Adv. Neural Inf. Processing Syst., 25, 2012 Papers.
[25] LA. Gatys, AS. Ecker, M. Bethge, Texture Synthesis Using Convolutional Neural Networks" [J], MIT Press, 2015.
[26] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation", [J], IEEE Trans. Pattern Anal. Machine Intell. 39 (4) (2015) 640–651 Papers.
[27] Y. Nirkin, I. Masi, AT. Tran, et al., On face segmentation, face swapping, and face perception, IEEE Computer Society, 2017.
[28] A. Dantcheva, C. Chen, A. Ross, Can facial cosmetics affect the matching accuracy of face recognition systems?", [C], IEEE Fifth International Conference on Biometrics: Theory, 2012.
[29] D. Bitouk, N. Kumar, S. Dhillon, et al., Face swapping: automatically replacing faces in photographs", [J], ACM Trans. Graphics Papers 27 (3) (2008) 251–258.
[30] IJ. Goodfellow, J. Pouget-Abadie, M Mirza, et al., Generative adversarial networks, [J], Adv. Neural Inf. Process. Syst. 3 (2014) 2672–2680 Papers.
[31] JY. Zhu, T. Park, P. Isola, et al., Unpaired image-to-image translation using cycle-consistent adversarial networks, [J], IEEE (2017).
[32] M. Mirza, S. Osindero, Conditional generative adversarial nets, in: [J], Comput. Sci., Papers 2014, 2014, pp. 2672–2680.
[33] O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical Image segmentation" [J], Springer International Publishing, 2015.
[34] Wang, TC., Liu, MY., Zhu, JY., et al., "High-resolution image synthesis and semantic manipulation with conditional GANs" [J], (2017).
[35] Y. Choi, M. Choi, M. Kim, et al., StarGAN: unified generative adversarial networks for multi-domain image-to-image translation, 2018.
[36] Nirkin, Y., Keller, Y., Hassner, T., "FSGAN: Subject Agnostic Face Swapping and Reenactment" [J], (2019).
[37] H. Deng, Ch. Han, H. Cai, GQ. Han, ShF. He, spatially-invariant style-codes controlled makeup transfer, CVRP, 2021.
[38] Nguyen, T., Tran, A., Hoai, M., "Lipstick ain't enough: beyond color matching for in-the-wild makeup transfer." [J], (2021).
[39] L. Dinh, D. Krueger, Y. Bengio, NICE: non-linear independent components estimation, [J], Comput. Sci. (2014).
[40] Dinh, L., Sohl-Dickstein, J., Bengio, S., "Density estimation using Real NVP" [J], (2016).
[41] Kingma, DP., Dhariwal, P., "Glow: Generative Flow with Invertible 1×1 Convolutions" [J], (2018).
[42] F. Luan, S. Paris, E. Shechtman, Deep photo style transfer, [C], 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, 2017.
[43] Li, Y., Fang, C., Wang, Z., Lu, X., Yang, MH., "Universal Style Transfer via Feature Transforms" [J], (2017).
[44] Huang, Y., He, ZW., Zhang, WS., "Fast multi-way regional makeup transfer deep network", (2019).