Evaluation of GAN Makeup Transfering Models

Qinxi Yu, Yexiong Xu, Hanlai Chen
Department of Computer Science
University of Toronto
Github Link
27 King's College Cir, Toronto, ON M5S

Abstract

Facial makeup researches mainly focus on tasks like makeup removal, and makeup transfer. By using the decoder and remover framework, the algorithm would remove the makeup in any given face. Makeup transferring tries to apply the same makeup given the image of a reference person to the image of a target person. The conventional algorithms focus on dividing facial makeup into different parts and for each part, some unique methods are conducted. Unlike the previous makeup transfer algorithm, the two methods we will analyze in this paper use the Generative Adversarial Network(GAN) to accomplish their tasks. In this report, we will evaluate and compare the performance of 2 popular makeup transfer models, BeautyGAN, and PSGAN.

1 Introduction

Facial makeup transfer is a demanding technique as it is used in many portrait apps. Most conventional makeup transfer methods are not based on Generative Adversarial Networks(GAN). It is significant for us to investigate and learn how efficient and effective are those models. In the field of makeup transfer, the objective is to transfer the makeup of one person to the facial identity of another person. As a result, it is challenging to gather data set with paired makeup data and non-makeup data. Therefore, an existing method like BeautyGAN[1] and PSGAN[2] both adopt the framework CycleGAN[6] so that they can be trained on unpaired data. We will explore and compare the performance of these two models while performing some sensitivity analysis on different hyper-parameters. To compare the performance of makeup transfer require user studies on the visual effect of the images generated by the same data set with the two models. Besides, sensitivity analysis on different learning rates, activation function at generator and discriminator, and the dimension of convolution layers.

2 Related Works

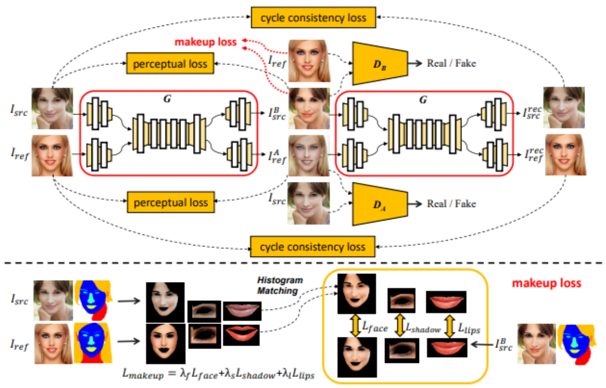
Recently, makeup transferring and facial analysis-related studies have aroused more and more attention. The Paper Face Behind Makeup [3] proposed a makeup detector and makeup remover algorithm based on locality-constrained dictionary learning. This can then be used as a part of makeup transferring applications. The Paper Anti-Makeup: Learning A Bi-level Adversarial Network for Makeup-Invariant Face Verification [4] studies how an adversarial network generates non-makeup images for makeup-invariant face verification. To extend those applications to makeup transfer, the main challenge would be that while transferring makeup style from the reference image, the identity of the source image is still preserved. The paper Digital Face Makeup by Example [5] explains an algorithm that decomposes images into three different layers and extracts makeup features layer by layer. Unlike some existed methods, the two models we exported could realize makeup transfer and do the makeup removal at the same time.

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

3 Algorithm

3.1 Beauty-GAN

Figure 1: BeautyGAN

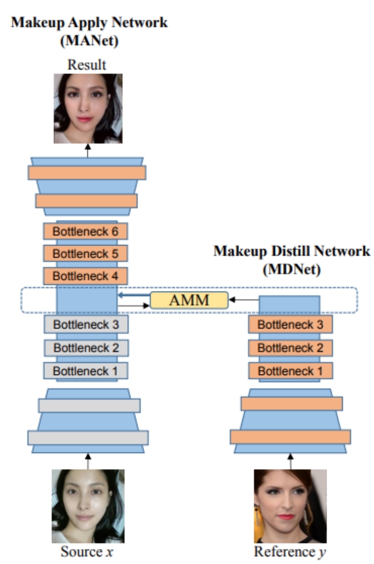


The overall architecture of Beauty-GAN is shown in Figure 1(image src). BeautyGAN uses an encoder-decoder generator G that can take a non-makeup image A and a makeup image B as input. Then it can simultaneously generate the makeup image A' with makeup style from B and an anti-makeup image B'. During training, BeautyGAN first maps A and B to A' and B'. Then it will try to ensure the face identity remains unchanged after mapping by using the VGG-16 model. The face identity of B and B' are expected to be the same. Also face identities of A and A' are expected to be the same. B' and A' then are fed back to G. The resulting images are expected to be the same as A and B. The cycle consistency loss is added to the overall objective function. Each time after the generator returns outputs, we will also add the adversarial loss given by the discriminator which tries to tell if a given image is real or fake. Besides, after generating A', the model will separate the region of faces, eye shadows, and lips from both images A' and B. It then performs a histogram matching on each region using the corresponding regions from the reference image. The result regions are used as ground truth. Then the loss given by the difference in makeup style is added as makeup loss. By minimizing the overall objective function the trained generator G can perform the task of makeup transfer.

3.2 PS-GAN

PSGAN also makes use of CycleGAN architecture like BeautyGAN. The main difference is the generator structure which is shown in Figure 2 (image src). The Makeup Distill Network(MDNet) will first extract the makeup style matrices γ and β from the reference image y. Considering the sources image x may have different poses or positions from y, γ, β, as well as x, are fed into the Attentive makeup morphing module(AMM). AMM aims to calculate how a pixel in x is related to the pixels in y by calculating an attention Matrix A. Attention matrix A is calculated by considering the similarities of both visual appearances and relative positions of 68 facial landmarks. The output of AMM, Γ' and B' are again fed into Makeup Apply Network(MANet). MANet follows the encoder-decoder architecture. The encoder part takes x as inputs and outputs a source image feature map V_x. Using the output of AMM, we calculate the transferred feature map $V_x' = \Gamma' V_x + B'$. Then $V'x$ is fed back into the decoder part of MANet. The resulting image would be the person from the source image with the makeup style from the reference image.

Figure 2: PSGAN



4 Experiments And Discussion

We performed sensitivity analysis on hyper-parameters: learning rates, dimension of the convolution layers, types of activate function. The two models are trained with the same makeup data-set with 2719 makeup images and 1115 non-makeup images, and we set the epochs to 4 at training. First, we studied the influence on the learning rates since we realized both of the models start to converge at the early stage of the training. According to $Table1$[1], the difference in learning rates would increase the average converge time, however, smaller learning rates tend to result in higher generator training loss in Beauty-GAN. This effect might be due to the limited amount of data-set and epochs, the models did not learn enough information from each image. Decreasing the learning rate in PS-GAN does not change the loss of the model, since this model obtains data from 3 sources: RGB images, masks, landmarks, and it gathers enough information including the orientations of the face features and makeup color features even with small learning rate and small epochs. However, from the loss graphs, we have noticed that the larger learning rate tends to result in spikier graphs. This might due to the models converge in the early stage of the training and a larger learning rate leads to a larger oscillation near the optimal point. According to $Table2$[2], for both of the models, similar behavior is recorded. The smaller convolution layer tends to result in higher training loss and lower resolution. This could be due to the model with a smaller convolution layer tends to ignore some of the facial information, thus the images generated tend to be in a lower resolution which meets our observation. According to $Table3$[3], a replacement of activation function at connections does not lead to great impacts on the result. In this experiment, we replaced all ReLU activation functions with LeakyReLU to prevents dead neurons. However, the result indicates that the two models do not suffer from the dying ReLU problem. The loss graphs for both of the models with different hyper-parameter settings are collected and can be viewed here.

Table 1: Effect of learning rate

Model	Learning Rate	Converge	Generator Loss
Beauty-GAN	2e-4	∼100 iterations	∼ 0.34
Beauty-GAN	2e-5	∼200 iterations	∼ 0.40
PS-GAN	2e-4	∼200 iterations	∼ 0.25
PS-GAN	2e-5	∼250 iterations	∼ 0.25

Table 2: Effect of convolution layer layers dimension of the decoder

Model	Convolution Layer Size	Converge	Generator Loss
Beauty-GAN	64x64	∼100 iterations	∼ 0.34
Beauty-GAN	32x32	∼100 iterations	∼ 0.43
PS-GAN	64x64	∼200 iterations	∼ 0.25
PS-GAN	32x32	∼200 iterations	∼ 0.30

Table 3: Effect of different activation functions

Model	Activation Function	Converge	Generator Loss
Beauty-GAN	ReLU	∼100 iterations	∼ 0.34
Beauty-GAN	LeakyReLU	∼100 iterations	∼ 0.35
PS-GAN	ReLU	∼200 iterations	∼ 0.25
PS-GAN	LeakyReLU	∼200 iterations	∼ 0.26

4.1 Quantative Comparison

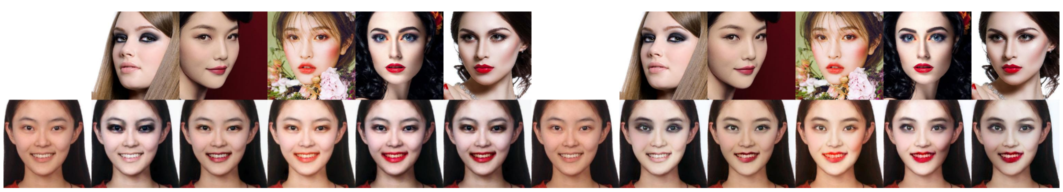
To compare the performance of BeautyGAN and PSGAN, we will conduct a user study. We will randomly select 5 non-make images and 10 makeup images as reference images. For each reference image, we will generate. Using both makeup transfer techniques, we will generate 200 result images. Then some volunteers are asked to select the better generated images. The comparison is based on image quality and the similarity to makeup style from the reference image. Then we will compare which method can have more generated images being preferred.

Method	BeautyGAN	PSGAN
Votes	25	22

From the comparison, we can see that the two methods have a similar vote. During the experiments, as suggested by the volunteers, BeautyGAN in general has a better performance when transferring the makeup color of the face and areas around the mouth, while PSGAN performs better when transferring the makeup color of the eyes. Also, we find out that PSGAN can sometimes consider the shadow on the face as makeup. Meanwhile, BeautyGAN seems to be not be affected by the shadows.

4.2 Qualitative Comparison

From the results of PSGAN and BeautyGAN [4], we can see most features like skin color, eye shadow and lipstick are transferred perfectly form the reference image if the pose of reference image is similar to the source image. However, there are some special cases in both model. For the case where the pose of reference image is very different from the source image, there is a chance to have the make-up at the wrong position for BeautyGAN model[5]. PSGAN model performs well in different poses and expressions, but the transferred image sometimes can be darker in one side[6]. The reason is that shadow features of side faces are captured by PSGAN and transferred to the source image.
Part of our Comparison is shown below, full qualitative comparison can be viewed here.



BeautyGAN PSGAN

Figure 3: Generated results by using the first row as references

5 Conclusion

In this report, we evaluate the makeup transfer performance of PAGAN and BeautyGAN. We sampled 5 non-make images and 10 makeup images to generate 50 results. We asked some volunteers to choose the better generated images. We also retrained the model using different model configurations to test the hyper-parameter sensitivity. We tried to change the learning rate, convolution layer dimension, and activation function of the model. The new configuration is trained on a data set of smaller size than the original data set.

To extend the research to other application, for example, a makeup remover can be made by modifying a part of our BeautyGAN model. We have noticed that when a makeup transferred image is generated, original and reference images are first decoded into its facial identity and its makeup. Then encode the original identity with the reference makeup. In the middle of this process, we can take the identity frame after decoding the original image as a makeup removed image. Although it requires more researches on how to make the model reliable and how to select a new data set, simply take the decoded identity as the result seems reasonable and reliable in our experiment.

References

[1] Tingting Li, Ruihe Qian, Chao Dong, Si Liu, Qiong Yan, Wenwu Zhu, Liang Lin. 2018, BeautyGAN: Instance-level Facial Makeup Transfer with Deep Generative Adversarial Network. http://liusi-group.com/pdf/BeautyGAN-camera-ready_2.pdf

[2] Wentao Jiang, Si Liu, Chen Gao, Jie Cao, Ran He, Jiashi Feng, Shuicheng Yan. 2019, PSGAN: Pose and Expression Robust Spatial-Aware GAN for Customizable Makeup Transfer. arXiv:1909.06956 https://arxiv.org/pdf/1909.06956.pdf

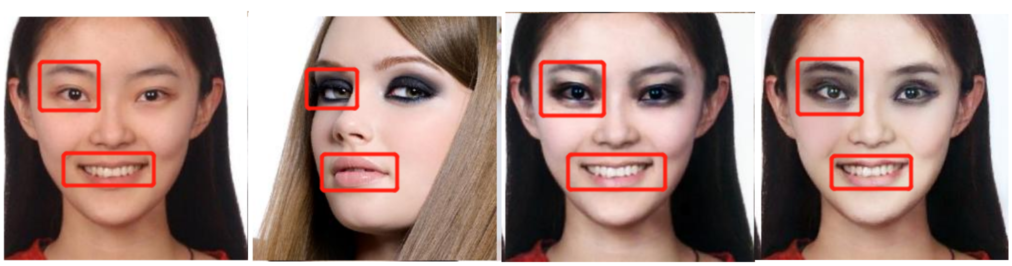
[3] Shuyang Wang and Yun Fu. 2016. Face Behind Makeup. In the Association for the Advance of Artificial Intelligence. 58–64.
https://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/download/12044/11564

[4] Yi Li, Lingxiao Song, Xiang Wu, Ran He, and Tieniu Tan. 2017,
https://arxiv.org/abs/1709.03654

[5] Dong Guo and Terence Sim. 2009. Digital face makeup by example. In Computer Vision and Pattern Recognition, 2009.
https://www.comp.nus.edu.sg/~tsim/documents/face_makeup_cvpr09_lowres.pdf

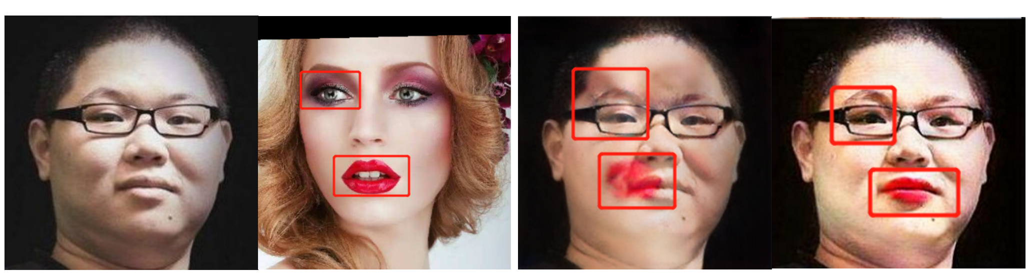
[6] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycleconsistent adversarial networks. In ICCV, 2017.

Appendix



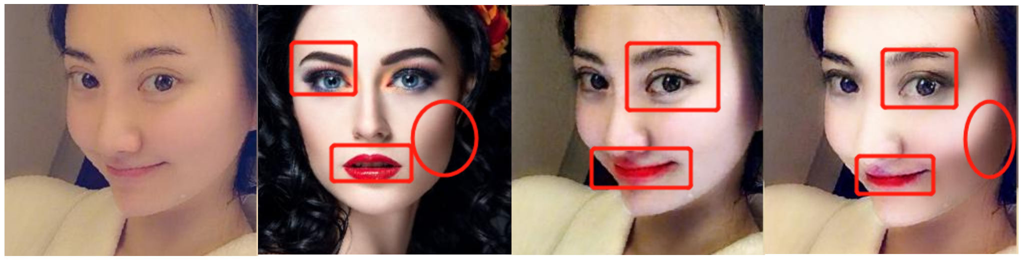
Source Reference BeatuyGan PSGan

Figure 4: most features like skin color, eye shadow and lipstick are transferred



Source Reference BeatuyGan PSGAN

Figure 5: make-up at the wrong position for BeautyGAN because of different poses and facial expressions



Source Reference BeatuyGan PSGan

Figure 6: transferred image sometimes is darker on the right part face for PSGAN model

Contribution

Hanlai Chen: Select subset of data. Tune the hyper parameters and train the models using each configuration;
Yexiong Xu: Analyze the resulting performance of each model. Design, conduct and conclude the experiments(user study);
Qinxi Yu: Understand and analyze models, explain the architecture and weakness of each model;