IMAGE-TO-IMAGE FACIAL MAKEUP TRANSLATION

A PREPRINT

Furkan Coskun
Department of Computer Science
Sabancı University
Istanbul, TR
furkancoskun@sabanciuniv.edu

Anil Ozdemir
Department of Computer Science
Sabancı University
Istanbul, TR
aozdemir@sabanciuniv.edu

June 20, 2020

1 Introduction

Image-to-image translation is the task of converting one representation into another [1]. This can be achieved by learning the mapping between an input image and an output image using paired images. However, paired data-sets are not commonly available for many tasks, we will be investigating this translation with unpaired source and target domains. With the recent advances in generative adversarial networks (GAN) image generation has gained huge attention [2]. This unpaired image translation can be done using CycleGAN architecture [3].

Figure 1: Adding Facial Makeup [4]

Image to image translation studies started with the non-parametric texture models by Hertzmann et al.'s Image Analogies [5]. Then, the same study improved by using CNNs to learn parametric translation function [6] and finally "pix2pix" method has been proposed to learn a mapping from an input domain to an output domain by uses a conditional generative adversarial networks [2]. However, major restriction with the "pix2pix" model is the requirement of paired-images. Afterwards, Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks named as "CycleGAN" is published in 2018 and came up with a solution to the lack of paired-data problem by introducing cycle consistency loss [3]. CycleGAN presented a new methodology for learning to translate an image from a source domain X to a target domain Y in the absence of paired examples and results are presented on several tasks including object transfiguration, season transfer, collection style transfer, etc.

In this work, we looked for the solution to the facial make-up translation between images using CycleGAN architecture. Facial makeup translation is the task of translating a makeup style from one image to another image see figure 1. We consider this problem as a conventional domain-level transfer task just as in the CycleGAN. We aim to learn a mapping $G : X \rightarrow Y$ such that the distribution of images from G(X) is indistinctive from the distribution Y with benefiting from adversarial loss. CycleGAN, extend this formula with using inverse-mapping $F : Y \rightarrow X$ and introduce a cycle

consistency loss to enforce F(G(X)) ≈ X (and vice versa) [3]. In our case X and Y accounts for the distribution of the makeup images and non-make images respectively. We used Makeup Transfer dataset [2.2] and CelebA dataset [2.1] during perform our experiment.

2 Dataset

Two different datasets will be used. One can see the data descriptions below:

2.1 CelebFaces Attributes Dataset (CelebA)

Figure 2: Non-makeup and makeup data gathered from CelebA dataset. The non-makeup and makeup images are shown in upper row and lower row [7].

(CelebA) is comprehensive face attributes data which contain over 200 thousands of celebrity images with 40 attribute annotations. The dataset contains heavy_makeup and wearing_lipstick and male binary attributes which will be benefited for this task. We have collected the data from only female face images. Gathered makeup and non-make data according to the wearing_lipstick and heavy_makeup attributes.

For both domains, there were more than 50 thousands of images existed. However, because of GPU limitations, we have limited the data to 5000 images (2500 makeup, 2500 non-makeup).Training and validation parts account for 90% and 10% of data. Image shapes are 218x178x3 [7].

2.2 Makeup Transfer Dataset (MT)

Figure 3: Samples from MT dataset. The non-makeup and makeup images are shown in upper row and lower row [4].

Li and Quan collected 3834 female images with 1115 non-makeup and 2719 makeup images. The data is crawled from the web. All images are aligned and prepared specifically for makeup translation. Image shapes are 361x361x3 and

have a little bit of high quality when compared to the CelebA dataset. Thus, for the computational limitations, we have selected 2000 images (1000 makeup, 1000 non-makeup) for the experiment [4].Training and validation parts account for 90% and 10% of data we used.

3 Experiments

3.1 Formulation

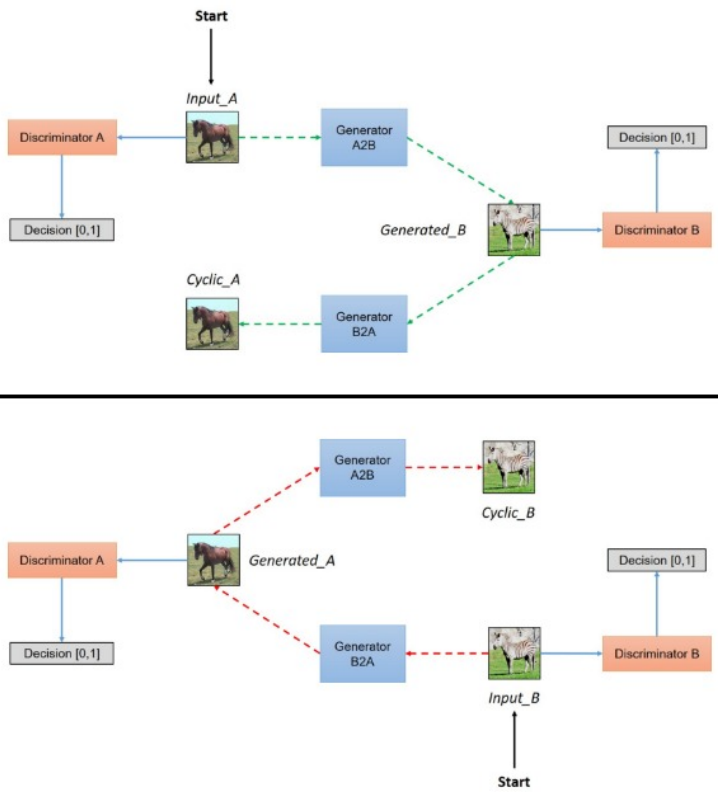

Figure 4: Network Architecture [3]

Our objective is the extract and the learn mapping function between makeup domain X and non-makeup domain Y given unpaired traning samples $\{x_i\}_{i=1}^{N}$ where $\{x_i\} \in X$ and $\{y_i\}_{J=1}^{K}$ where $\{y_j\} \in Y$. Model has two counter-mapping function G : X $\rightarrow$ Y and F : Y $\rightarrow$ X , two discriminator D_x and D_y, where D_x tries to seperate real makeup images $\{x\}$ and translated (reconstructed) real images $\{F(y)\}$ while D_y tries to distinguish non-makeup images $\{y\}$ and $\{G(y)\}$.Following is the two type of loss we applied as mentioned in CycleGAN [3]:

- Adversarial Loss (AL):
  Similar ALs have been applied to both mapping functions G : X $\rightarrow$ Y and F : Y $\rightarrow$ X. This type of loss comprehensively described in the generative adversarial networks (GAN) paper [2].
- Cycle Consistency Loss (CCL):
  CCL is fully described in the CycleGAN Paper [3]. The idea is the for particular sample image from domain X, the image translation cycle illustrated in 4 must be able to bring x back to the original image. This provides us with two concepts named as "Forward Cycle Consistency" and "Backward Cycle Consistency".Both together combined and form the cycle consistency loss.
  Forward Cycle Consistency: $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$.
  Backward Cycle Consistency: $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$.

Figure 5: Makeup transfer using CelebA dataset

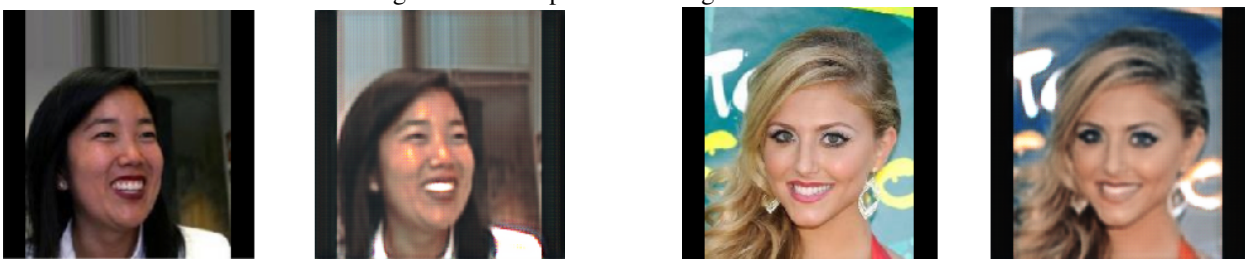

Figure 6: Adding makeup

Figure 7: Removing makeup

Figure 8: Makeup transfer using MT dataset

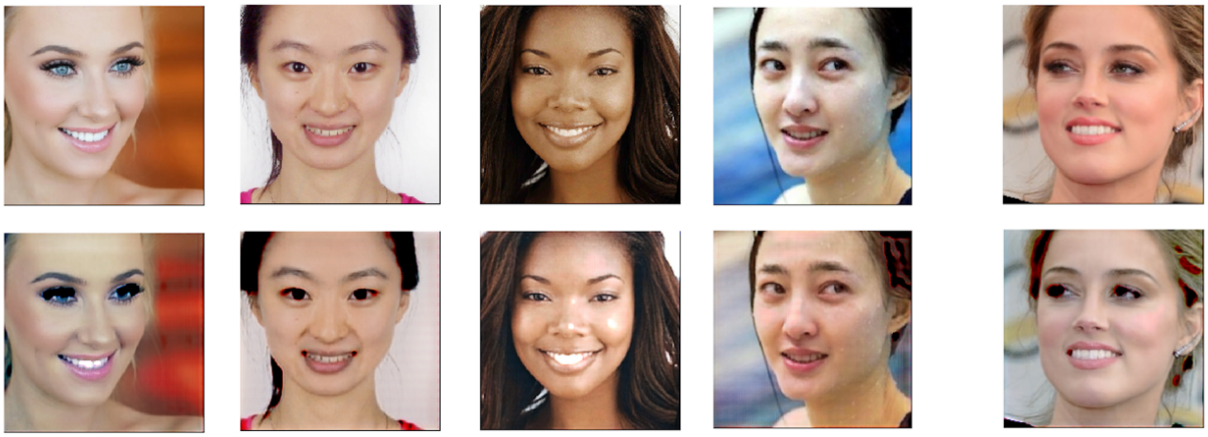

Figure 9: Adding makeup

Figure 10: Removal

To retrieve the best translation A to B (vice-versa) , full objective becomes the addition of adverserial losses and cycle consistency loss.

$L(G, F, D_x, D_y) = L_{GAN}(G, D_Y, X, Y) + L_{GAN}(F, D_X, Y, X) + \lambda L_{cyc}(G, F)$ where λ controls the relative importance of two objectives (see [3]).

3.2 Network Architecture

Underlying architecture [4] is the similar with the generative and discriminative networks from the CycleGAN [3] and implemented via Keras [8]. Generative networks contains 2 two-strided convolutions, 2 fractionally strided convolutions , several residual blocks [9]. We use 9 blocks for 256 256 and higher quality images (364 x 364) and Instance Normalization. For the discriminator network contains 70 70 PatchGANs [3] whose tries to classify whether 70 70 overlapping image patches are real or fake. This methodology helps to reduce number of parameter than full-image discriminator and does not affected by different-sized images.

4 Results

GANs have proven to be useful for high-quality image generation. However, to obtain high-quality images with GANs enormous amount of data and computational power is needed. In this work, we demonstrated that even with a limited amount of training (10-15 hours) and a small dataset, appreciable facial makeup translation results are obtained (see Figure 5 and 8).

We observed that CelebA dataset have provided poor quality translations. This is mainly due to the purpose of the CelebA dataset is not constructed for specifically makeup translation with lack of preciseness in annotations of CelebA.On the other hand, the MT dataset is constructed specifically for makeup translation and this is why outputs are

structurally smoother and visually more logical. Therefore, as it can be seen in the highlighted figures (see Figure 5 and 8), results from the MT dataset were a bit more satisfying then CelebA dataset.

5 Conclusion

Image to Image translation is hot-topic in current days. Recent works [10],[3],[1] has display highly successful results with and without lack of paired-data issue in the various tasks including object transfiguration, season transfer, collection style transfer, etc .In our case, we adopt the idea of an image to image translation from CycleGAN and apply it to the facial makeup translation task. Highlighted results illustrated in the figures 5 and 8 are the most successful ones. However, every test instance does not show satisfying results. This is due to the both nature of the data-sets and the limited equipment of us.

References

[1] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. *CoRR*, abs/1611.07004, 2016.

[2] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014.

[3] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *CoRR*, abs/1703.10593, 2017.

[4] 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. pages 645–653, 10 2018.

[5] Aaron Hertzmann, Charles E Jacobs, Nuria Oliver, Brian Curless, and David H Salesin. Image analogies. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 327–340. ACM, 2001.

[6] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[7] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, December 2015.

[8] François Chollet et al. Keras. `https://keras.io`, 2015.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2015.

[10] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. *CVPR*, 2017.