# BeautyGAN: Instance-level Facial Makeup Transfer with Deep Generative Adversarial Network


Tingting Li[1], Ruihe Qian[2], Chao Dong[2], Si Liu[4], Qiong Yan[3], Wenwu Zhu[1], Liang Lin[3, 5]

[1]Tsinghua University, [2]Chinese Academy of Science, [3]SenseTime Research, [4]Beihang University, [5]Sun Yat-sen University

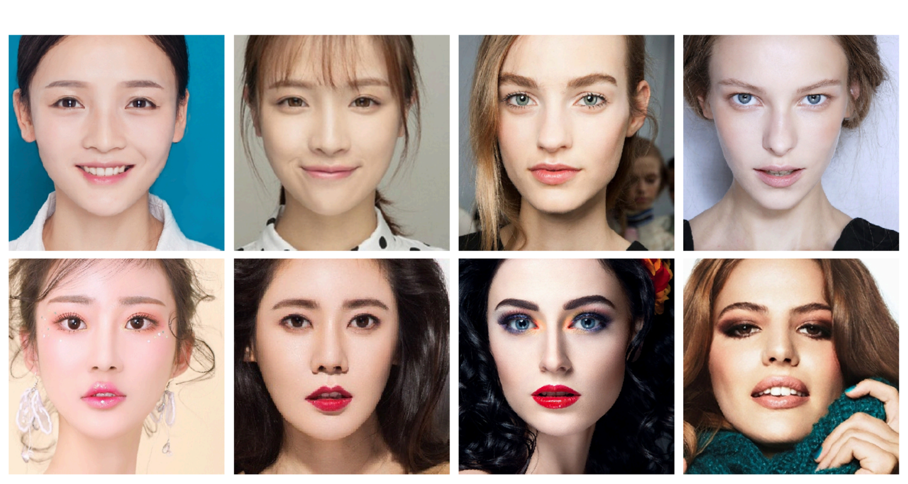

## Introduction

Facial makeup transfer aims to translate the makeup style from a given reference makeup face image to another non-makeup one while preserving face identity. Such an instance-level transfer problem is more challenging than conventional domain-level transfer tasks, especially when paired data is unavailable. Makeup style is also different from global styles (e.g., paintings) in that it consists of several local styles/cosmetics, including eye shadow, lipstick, foundation, and so on. Extracting and transferring such local and delicate makeup information is infeasible for existing style transfer methods.

In this paper, we address the issue by incorporating both global domain-level loss and local instance-level loss in an dual input/output Generative Adversarial Network, called BeautyGAN.

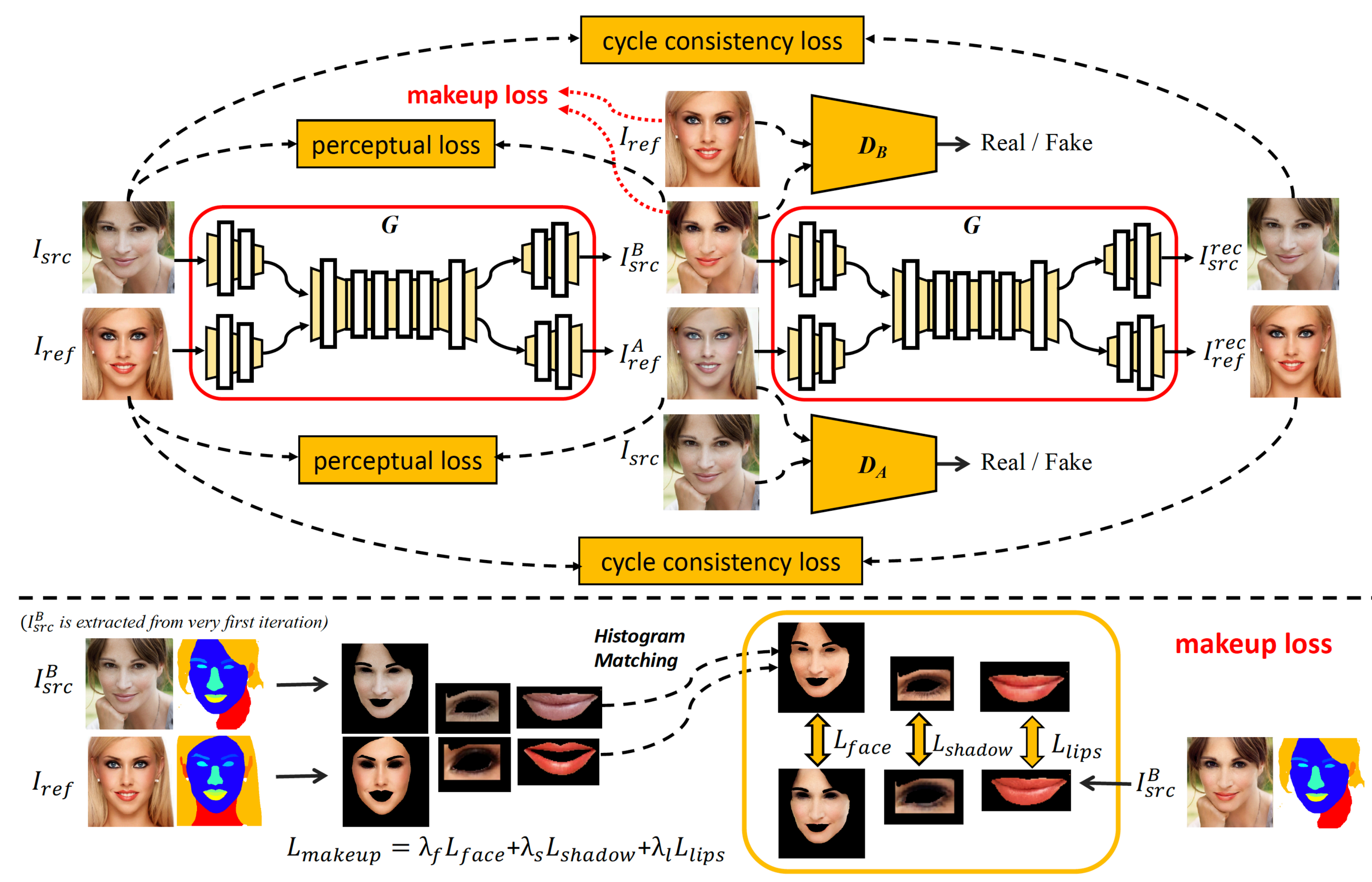
- The domain-level transfer is ensured by discriminators that distinguish generated images from domains' real samples.
- The instance-level loss is calculated by pixel-level histogram loss on separate local facial regions.
- We further introduce perceptual loss and cycle consistency loss to generate high quality faces and preserve identity.
- The overall objective function enables the network to learn translation on instance-level through unsupervised adversarial learning.

## Data Collection



We collect a new facial makeup dataset consisting of 3834 female images in total, with 1115 non-makeup images and 2719 makeup images. We refer to this dataset as the Makeup Transfer(MT) dataset. It includes some variations in race, pose, expression and background clutter. MT is the biggest makeup dataset comparing to other released makeup datasets. Existing makeup datasets mostly consist of no more than 1000 images.

## Methodology




- Consider two data collections, $A \subset \mathbb{R}^{H\times W\times 3}$ referring to non-makeup image domain and $B \subset \mathbb{R}^{H\times W\times 3}$ referring to makeup image domain. We simultaneously learn the mapping between two domains, denoted as $G: A \times B \rightarrow B \times A$, where '×' represents Cartesian product.
- The overall framework consists of one generator $G$ and two discriminators: $D_A$, $D_B$. In the formulation $\left(I_{src}^B, I_{ref}^A\right) = G(I_{src}, I_{ref})$, $G$ accepts two images, $I_{src} \in A$ and $I_{ref} \in B$, as inputs and generates two translated images as outputs, $I_{src}^B \in B$ and $I_{ref}^A \in A$.
- Objective Function

➢ $D_A$ and $D_B$ contain only adversarial losses. $D_A$ and $D_B$ distinguish the generated image $I_{ref}^A$ and $I_{src}^B$ from real samples in set $A$ and set $B$, respectively, given by:

$$\mathcal{L}_{D_A} = \mathbb{E}_{I_{src}}[log D_A(I_{src})] + \mathbb{E}_{I_{src},I_{ref}}[log(1 - D_A(I_{ref}^A))]$$
$$\mathcal{L}_{D_B} = \mathbb{E}_{I_{ref}}[log D_B(I_{ref})] + \mathbb{E}_{I_{src},I_{ref}}[log(1 - D_B(I_{src}^B))]$$

➢ The full objective function of generator $G$ contains four types of losses: adversarial loss, cycle consistency loss, perceptual loss and makeup constrain loss,

$$\mathcal{L}_G = \alpha\mathcal{L}_{adv} + \beta\mathcal{L}_{cyc} + \gamma\mathcal{L}_{per} + \mathcal{L}_{makeup}, \text{ where } \mathcal{L}_{adv} = \mathcal{L}_{D_A} + \mathcal{L}_{D_B}$$

➢ Perceptual loss calculates differences between high-level features extracted by deep convolutional networks. Cycle consistency loss formulated as the distances between the reconstruction results and the inputs because $(I_{src}, I_{ref}) \rightarrow G(I_{src}, I_{ref}) \rightarrow G(G(I_{src}, I_{ref})) \approx (I_{src}, I_{ref})$.

$$\mathcal{L}_{cyc} = \mathbb{E}_{I_{src},I_{ref}}\left[dist\,(I_{src}^{rec}, I_{src}) + dist(I_{ref}^{rec}, I_{ref})\right],$$
$$\text{where } \left(I_{src}^{rec}, I_{ref}^{rec}\right) = G(G(I_{src}, I_{ref}))$$

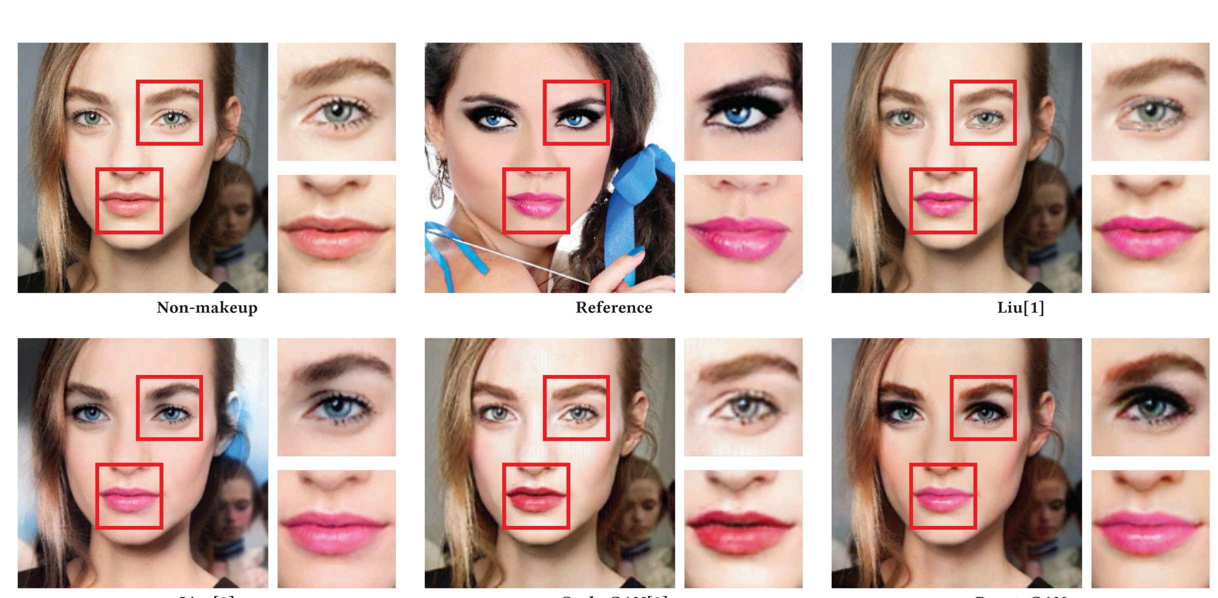
➢ Makeup loss are integrated by three local histogram losses acted on lips, eye shadows and face regions, respectively: $\mathcal{L}_{makeup} = \lambda_l\mathcal{L}_{lips} + \lambda_s\mathcal{L}_{shadow} + \lambda_f\mathcal{L}_{face}$

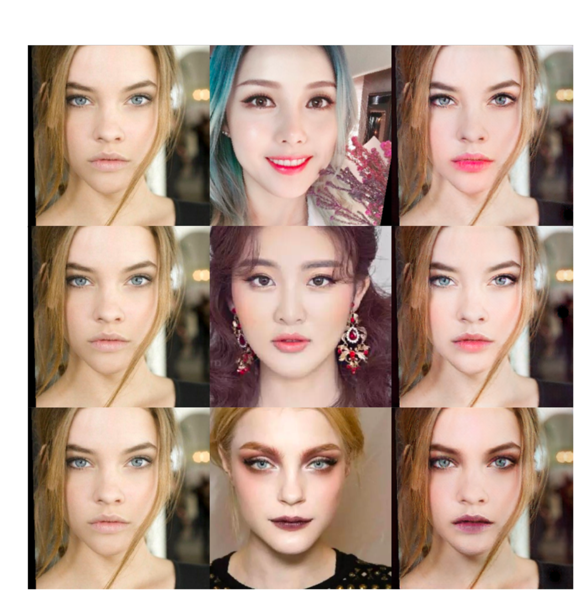
$$\mathcal{L}_{item} = \left\|I_{src}^B - HM(I_{src}^B \circ M_{item}^1, I_{ref} \circ M_{item}^2)\right\|_2$$
$$M^1 = FP(I_{src}^B),\ M^2 = FP(I_{ref})$$

## Experimental Results

- **Qualitative comparison**

The zoom in performance of eye makeup and lipsticks transfer are shown below on the right. It shows the comparisons among our BeautyGAN model and other baselines. We also listed one set of our generated results on the right. The last column shows the makeup transfer results.



## Results

The example results of our BeautyGAN model for makeup transfer are shown below. Three makeup styles on reference images (top row) are translated to three before-makeup images (left column). Nine generated images are shown in the middle.


- **Quantitative comparison**

For quantitative evaluation on BeautyGAN, we conduct a user study from 84 volunteers. We randomly choose 10 non-makeup test images and 20 makeup test images, which would obtain 10 x 20 after-makeup results for each makeup transfer method. Two representative baselines are in comparison: Liao et al. [1] and Liu et al. [2]. Each time, we present five images, including a non-makeup image, a makeup image as reference, and three randomly shuffled makeup transfer images generated from different methods. Participants are instructed to give a rank order of three generated images, based on quality, realism and makeup style similarity. Rank 1 represents the best makeup transfer performance while rank 3 represent the worst makeup transfer performance. Table below shows the results. For each method, we normalize the votes and obtain the percentages of three rank orders.

| Method | Rank 1 | Rank 2 | Rank 3 |
|---|---|---|---|
| Liu[1] | 4.25% | 10.78% | 84.97% |
| Liao[2] | 33.91% | 46.03% | 20.06% |
| Ours | 61.84% | 27.56% | 10.59% |


**Reference**

[1] Si Liu et al. Makeup like a superstar: deep localized makeup transfer network. AAAI 2016.
[2] Jing Liao et al. Visual attribute transfer through deep image analogy. ACM Transactions on Graphics (TOG) 36, 4 (2017), 120.
[3] Jun-Yan Zhu et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. CVPR 2017.


TBSI 清华-伯克利深圳学院 Tsinghua-Berkeley Shenzhen Institute
中国科学院 CHINESE ACADEMY OF SCIENCES
商汤 sensetime