



OPEN ACCESS

EDITED BY

Matteo Ferrara,
University of Bologna, Italy

REVIEWED BY

Hongfu Liu,
Brandeis University, United States
Sudha Velusamy,
Samsung, India

*CORRESPONDENCE

Xin Li
xin.li@mail.wvu.edu

SPECIALTY SECTION

This article was submitted to
Computer Vision,
a section of the journal
Frontiers in Computer Science

RECEIVED 01 April 2022

ACCEPTED 29 September 2022

PUBLISHED 28 September 2022

CITATION

Liu X, Wang R, Peng H, Yin M,
Chen C-F and Li X (2022) Face
beautification: Beyond makeup
transfer. *Front. Comput. Sci.* 4:910233.
doi: 10.3389/fcomp.2022.910233

COPYRIGHT

© 2022 Liu, Wang, Peng, Yin, Chen
and Li. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Face beautification: Beyond makeup transfer

Xudong Liu^{1,2}, Ruizhe Wang², Hao Peng², Minglei Yin¹,
Chih-Fan Chen² and Xin Li^{1*}

¹Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV, United States, ²ObEN, Inc., Pasadena, CA, United States

Facial appearance plays an important role in our social lives. Subjective perception of women's beauty depends on various face-related (e.g., skin, shape, hair) and environmental (e.g., makeup, lighting, angle) factors. Similarly to cosmetic surgery in the physical world, virtual face beautification is an emerging field with many open issues to be addressed. Inspired by the latest advances in style-based synthesis and face beauty prediction, we propose a novel framework for face beautification. For a given reference face with a high beauty score, our GAN-based architecture is capable of translating an inquiry face into a *sequence* of beautified face images with the referenced beauty style and the target beauty score values. To achieve this objective, we propose to integrate both style-based beauty representation (extracted from the reference face) and beauty score prediction (trained on the SCUT-FBP database) into the beautification process. Unlike makeup transfer, our approach targets many-to-many (instead of one-to-one) translation, where multiple outputs can be defined by different references with various beauty scores. Extensive experimental results are reported to demonstrate the effectiveness and flexibility of the proposed face beautification framework. To support reproducible research, the source codes accompanying this work will be made publicly available on GitHub.

KEYWORDS

face beautification, GAN, beauty representation, prediction, translation

1. Introduction

Facial appearance plays an important role in our social lives (Bull and Rumsey, 2012). People with attractive faces have many advantages in their social activities, such as dating and voting (Little et al., 2011). Attractive people have been found to have higher chances of dating (Riggio and Woll, 1984), and their partners are more likely to gain satisfaction compared to less attractive people (Berscheid et al., 1971). Faces have also been found to affect hiring decisions and influence voting behavior (Little et al., 2011). Overwhelmed by the social fascination with beauty, women with unattractive faces can suffer from social isolation, depression, and even psychological disorders (Macgregor, 1989; Phillips et al., 1993; Bradbury, 1994; Rankin et al., 1998; Bull and Rumsey, 2012). Consequently, there is strong demand for face beautification both in the physical world (e.g., facial makeup and cosmetic surgeries) and in the virtual space (e.g., beautification cameras and filters). To our knowledge, there is no existing work on face beautification that can achieve fine-granularity control of beauty scores.

The problem of face beautification has been extensively studied by philosophers, psychologists, and plastic surgeons. Rapid advances in imaging technology and social networks greatly expedited the popularity of digital photos, especially selfies, in our daily lives. Most recently, virtual face beautification based on the idea of applying or transferring makeup has been developed in computer vision communities, such as PairedCycleGAN (Chang et al., 2018), BeautyGAN (Li et al., 2018), BeautyGlow (Chen et al., 2019). Although these existing works have achieved impressive results, we argue that makeup transfer-based face beautification has fundamental limitations. Without changing important facial attributes (e.g., shape and lentigo), makeup application, abstracted by image-to-image translation (Zhu et al., 2017; Huang et al., 2018; Lee et al., 2018)—can only improve the beauty score to some extent.

A more flexible and promising framework is to formalize the face beautification process by translating *one-to-many* translation, where the destination can be defined in many ways (Figure 1). The motivation behind this work is two-fold. On the one hand, we can target producing a sequence of output images with monotonically increasing beauty scores by gradually transferring the style-based beauty representation learned from a given reference (with a high beauty score). On the other hand, we can also produce a variety of personalized beautification results by learning from a sequence of references (e.g., celebrities with different beauty styles). In this framework, face beautification can be made more flexible; for example, we can transfer the beauty style from a reference image to reach a specified beauty score, which is beyond the reach of makeup transfer (Li et al., 2018; Chen et al., 2019).

To achieve this objective, we propose a novel architecture based on generative adversarial networks (GANs). Inspired by the latest advances in style-based synthesis [e.g., styleGAN (Karras et al., 2019)] and face beauty understanding from data (Liu et al., 2019), we propose to integrate both style-based beauty representation (extracted from the reference face) and beauty score prediction (trained on the SCUT-FBP database Xie et al., 2015) into the face beautification process. More specifically, style-based beauty representations will first be learned from both inquiry and reference images *via* a light convolutional neural network (LightCNN) and used to guide the process of style transfer (actual beautification). Then, a dedicated GAN-based architecture is constructed and integrated with the reconstruction, beauty, and identity loss functions. To have fine-granularity control of the beautification process, we have invented a simple, yet effective, reweighting strategy that gradually improves the beauty score in synthesized images until reaching the target (specified by the reference image).

Our key contributions are summarized below.

- A forward-looking view toward virtual face beautification and a holistic style-based approach beyond makeup transfer (e.g., BeautyGAN and BeautyGlow). We argue that

facial beauty scores offer a quantitative solution to guide the facial beautification process.

- A face beauty prediction network is trained and integrated into the proposed style-based face beautification network. The prediction module provides valuable feedback to the synthesis module as we approach the desirable beauty score.
- A piggyback trick to extract both identity and beauty features from fine-tuned LightCNN and the design of loss functions, reflecting the trade-off between identity preservation and face beautification.
- To the best of our knowledge, this is the first work capable of delivering facial beautification results with fine granularity control (that is, a sequence of face images that approach the reference with increasing beauty scores monotonically).
- A comprehensive evaluation shows the superiority of the proposed approach compared to existing state-of-the-art image-to-image transfer techniques, including CycleGAN (Zhu et al., 2017), MUNIT (Huang et al., 2018), and DRIT (Lee et al., 2018).

2. Related works

2.1. Makeup and style transfer

Two recent works on face beauty are BeautyGAN (Li et al., 2018) and BeautyGlow (Chen et al., 2019). In BeautyGlow (Chen et al., 2019), the makeup features (e.g., eyeshadows and lip gloss) are first extracted from the reference makeup images and then transferred to the non-makeup images source. The magnification parameter in the latent space can be adjusted to adjust the extent of the makeup. In BeautyGAN (Li et al., 2018), the issue of extracting/transferring local and delicate makeup information was addressed by incorporating global domain-level loss and local instance-level loss in a dual input / output GAN.

Face beautification is also related to more general image-to-image translation. Both symmetric (e.g., CycleGAN Zhu et al., 2017) and asymmetric (e.g., PairedCycleGAN Chang et al., 2018) have been studied in the literature; the latter was shown to be effective for makeup application and removal. Extensions of style transfer to the multimodal domain (that is, one to many translations) have been considered in MUNIT (Huang et al., 2018) and DRIT (Lee et al., 2018). It is also worth mentioning the synthesis of face images *via* StyleGAN (Karras et al., 2019), which has shown superrealistic performance.

2.2. Face beauty prediction

Perception of facial appearance or attractiveness is a classical topic in psychology and cognitive sciences (Perrett et al., 1998,

1999; Thornhill and Gangestad, 1999). However, developing a computational algorithm that can automatically predict beauty scores from facial images is only a recent endeavor (Eisenthal et al., 2006; Gan et al., 2014). Thanks to the public release of the SCUT-FBP face beauty database (Xie et al., 2015), there has been a growing interest in machine learning-based approaches to face beauty prediction (Fan et al., 2017; Xu et al., 2017).

3. Proposed method

3.1. Facial attractiveness theory

Why does facial attractiveness matter? From an evolutionary perspective, a plausible working hypothesis is that the psychological mechanisms underlying primates' judgments about attractiveness are consequences of long-period evolution and adaptation. More specifically, facial attractiveness is beneficial in choosing a partner, which in turn facilitates gene propagation (Thornhill and Gangestad, 1999). At the primitive level, facial attractiveness is hypothesized to reflect information about an individual's health. Consequently, conventional wisdom in the research on facial attractiveness has focused on *ad hoc* attributes such as facial symmetry and averageness as potential biomarkers. In the history of modern civilization, the social norm of facial attractiveness has constantly evolved and varies from region to region (e.g., the sharp contrast between eastern and western culture Cunningham, 1986).

In particular, facial attractiveness for young women is a stimulating topic, as evidenced by the long-lasting popularity of beauty pageants. In Cunningham (1986), the relationship between the facial characteristics of women and the responses of men was investigated. Based on the attractiveness ratings of male subjects, two classes of facial features (e.g., large eyes, small nose, and small chin; prominent cheekbones and narrow cheeks) are positively correlated with the attractiveness ratings. It is also known from the same study (Cunningham, 1986) that facial features can also predict personality attributions and altruistic tendencies. In this work, we opt to focus on the beauty of the face for women only.

3.2. Problem formulation and motivation

Given a target face (an ordinary one that is less attractive) and a reference face (usually a celebrity one with a high beauty score), how can we beautify the target face by transferring relevant information from the reference image? This problem of facial beautification can be formulated as two subproblems: *style transfer* and *beauty prediction*. Meanwhile, an important new insight brought into our problem formulation is that facial beautification treatment as a sequential process in which the beauty score of the target face can be gradually improved

by successive style transfer steps. As the fine-granularity style transfer proceeds, the beauty score of the beautified target face will monotonically approach that of the reference face.

The problem of style transfer has been extensively studied in the literature, dating back to content-style separation (Tenenbaum and Freeman, 2000). The idea of extracting style-based representation (style code) has attracted increasingly more attention in recent years, e.g. Mathieu et al. (2016), Donahue et al. (2017), Huang and Belongie (2017), Huang et al. (2018), and Lee et al. (2018). Note that makeup transfer only represents a special case where style is characterized by local features only (e.g., eyeshadow and lipstick). In this work, we conceive of a more generalized solution to transfer both global and local style codes from the reference image. The extraction of style codes will be based on the solution to the other problem of beauty prediction. Such a sharing of learned features between style transfer and beauty prediction allows us to achieve fine-granularity control over the process of beautification.

3.3. Architecture design

As illustrated in Figure 2, we use A and B to denote the target face (unattractive) and the reference face (attractive), respectively. The objective of beautification is to translate the image A into a new image AB whose beauty score is Q -percent close to that of B (Q is an integer between 0 and 100 specifying the granularity of the beauty transfer). Assume that both images A and B can be decomposed into a two-part representation consisting of style and content. That is, both images will be encoded by a pair of encoders: content encoder (identity) E_c and style encoder (beauty) E_s , respectively. In order to transfer the beauty style from reference B to target A , it is natural to concatenate the representation based on content (identity) C_a with the representation based on style (beauty) S_b ; and then reconstruct the beautified image \tilde{A} through a dedicated decoder G defined by

$$G(\tilde{A}) = G[E_c(A), E_s(A) + E_s(B)]. \quad (1)$$

The rest of our architecture in Figure 2 mainly includes two components: a GAN-based module (G pairs with D) responsible for style transfer and a beauty and identity loss module responsible for beauty prediction (please refer to Figure 3).

Our GAN module consisting of two encoders, one decoder, and one discriminator aims at distilling the beauty/style representation from the reference image and embedding it into the target image for the purpose of beautification. Inspired by recent work (Ulyanov et al., 2017), we propose to integrate an instance normalization layer (IN) after convolutional layers as part of the encoder for content feature extraction. Meanwhile, a global average pooling and a fully connected layer follow convolutional layers as part of the encoder for

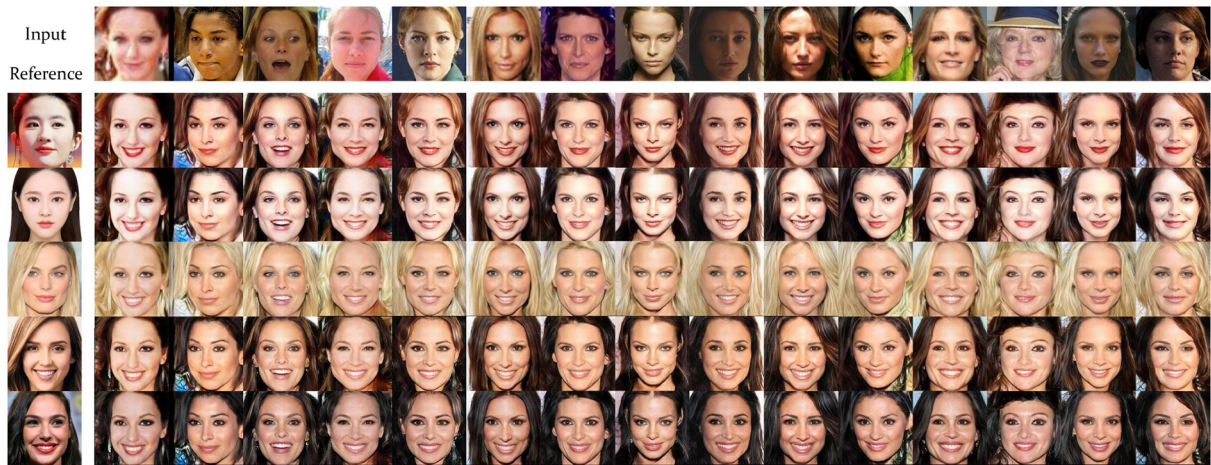


FIGURE 1
Face beautification as many-to-many image translation: Our approach integrates style-based beauty representation with a beauty score prediction model and is capable of fine-granularity control.

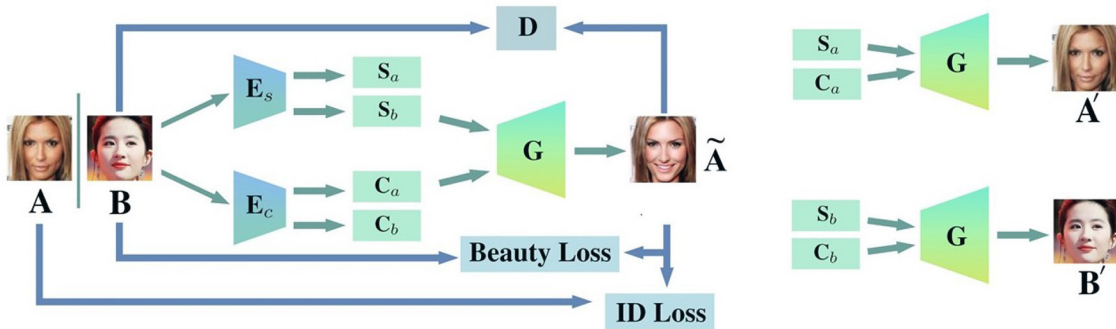


FIGURE 2
Overview of the proposed network architecture (training phase). Left: given a source image A and a target image B, our objective is to transfer the style of target image B to A in a continuous (fine-granularity control) manner. Right: $(S_{a/b}, C_{a/b})$ denote the style-content encoder for images A/B respectively (A', B' denote the reconstructed images of A, B).

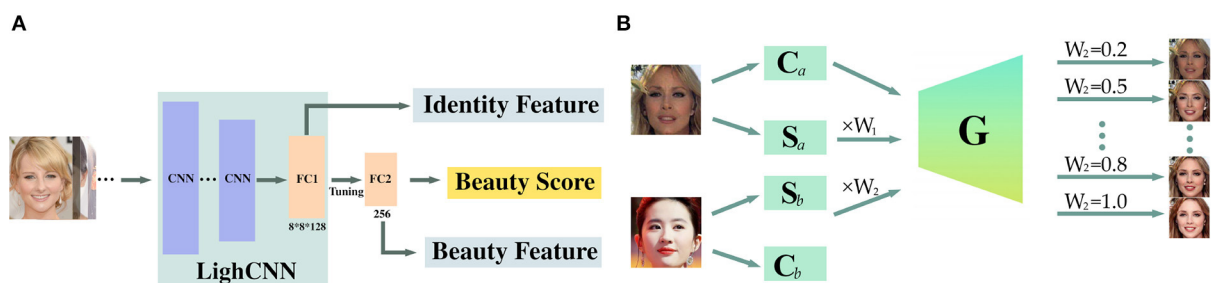


FIGURE 3
(A) Fine-tuning network for beauty score prediction. (B) Testing stage for fine-granularity beautification adjustment.

beauty feature extraction. Note that we skip the layer in the beauty encoder because IN would remove the characteristics of the original feature that represent critical beauty-related information (Huang and Belongie, 2017) (which is why we keep it within the content encoder). To cooperate with the beauty encoder and speed up translation, the decoder is equipped with an Adaptive Instance Normalization (AdaIN) (Huang and Belongie, 2017). Furthermore, we have adopted the popular multiscale discriminators (Wang et al., 2018b) with least squares GAN (LSGAN) (Mao et al., 2017) as the discriminator in our GAN module.

Our beauty prediction module is based on fine-tuning an existing LightCNN (Wu et al., 2018) as shown in Figure 3. It is difficult to train a deep neural network for beauty prediction from scratch with limited labeled beauty score data, we opt to work with LightCNN (Wu et al., 2018), a pre-trained model for face recognition with millions of face images. Instead, we employ a fine-tuning layer (FC2) to adapt it to the prediction of the beauty score (FC2 plays the role of a beauty feature extractor). Meanwhile, to preserve identity during face beautification, we propose taking full advantage of our beauty prediction model by piggybacking on the identity feature it produced. More specifically, the identity feature is generated from the second fully connected layer (FC1) of LightCNN; note that we have only fine-tuned the last fully connected layer (FC2) for beauty prediction. Using this piggyback trick, we managed to extract both identity and beauty features from one standard model.

3.4. Fine-granularity beauty adjustment

As we argued before, beautification should be modeled by a continuous process rather than a discrete domain transfer. To achieve fine-granularity control of the beautification process, we propose a weighted beautification equation by

$$G(\tilde{A}) = G[E_c(A), w_1 E_s(A) + w_2 E_s(B)], \quad (2)$$

where $w_1 + w_2 = 1$ and $0 \leq w_1, w_2 \leq 1$. It is easy to observe the two extreme cases: 1) Eq. (2) degenerates into reconstruction when $w_1 = 1, w_2 = 0$; 2) Eq. (2) corresponds to the most complete beautification when $w_1 = 0, w_2 = 1$. This linear weighting strategy represents a simple solution to adjust the amount of beautification.

To make our model more robust, we have adopted the following training strategy: Replace $G[E_c(A), E_s(A) + E_s(B)]$ with $G[E_c(A), E_s(B)]$ in the training stage so that we do not need to train multiple weighted models when the weights vary. Instead, we apply the weighted beautification equation of Eq. (2) for testing directly. In other words, we pretend that the beauty feature of the target image A is forgotten during training; and partially exploit it during testing (since it is less relevant than the identity feature). In summary, our strategy for fine-granularity

beauty adjustment is highly dependent on the ability of the beauty encoder E_s to reliably extract beauty representation. The effectiveness of the proposed fine-granularity beauty adjustment can be justified by referring to Figure 4.

3.5. Loss functions

3.5.1. Image reconstruction

Both the encoder and decoder need to ensure that target and reference images can be approximately reconstructed from the extracted content/style representation. Here, we have adopted the L_1 -norm for reconstruction loss because it is more robust than the L_2 -norm.

$$\begin{aligned} \mathcal{L}_{\text{REC}}^A &= \mathbb{E}_{a \sim p(a)} [\|G[E_s(A), E_c(A)] - A\|_1], \\ \mathcal{L}_{\text{REC}}^B &= \mathbb{E}_{b \sim p(b)} [\|G[E_s(B), E_c(B)] - B\|_1] \end{aligned} \quad (3)$$

where $\|\cdot\|_1$ denotes the L_1 norm.

3.5.2. Adversarial loss

We apply adversarial losses (Goodfellow et al., 2014) to match the distributions of the generated image AB and the target data B . In other words, adversarial loss ensures that the beautified face looks as realistic as the reference.

$$\tilde{\mathcal{L}}_{\text{GAN}}^A = \mathbb{E}_{\tilde{A}} [\log(1 - D(G(\tilde{A})))] + \mathbb{E}_B [\log D(B)], \quad (4)$$

where $G(\tilde{A})$ is defined by Equation (1).

3.5.3. Identity preservation

To preserve identity information during the beautification process, we propose adopting an identity loss function from the standard light-cognition model LightCNN (Wu et al., 2018) trained on millions of faces. The identity characteristics are extracted from the FC1 layer, which is a vector of dimensions 2^{13} , denoted as f_{id} .

$$\begin{aligned} \mathcal{L}_{\text{ID}}^A &= \|f_{id}(G[E_c(A), E_s(A)]) - f_{id}(A)\|_1, \\ \mathcal{L}_{\text{ID}}^B &= \|f_{id}(G[E_c(B), E_s(B)]) - f_{id}(B)\|_1, \\ \tilde{\mathcal{L}}_{\text{ID}}^A &= \|f_{id}(G[E_c(A), E_s(B)]) - f_{id}(A)\|_1, \end{aligned} \quad (5)$$

where $\mathcal{L}_{\text{ID}}^A$ and $\mathcal{L}_{\text{ID}}^B$ are responsible for the preservation of identity and $\tilde{\mathcal{L}}_{\text{ID}}^A$ aims to preserve identity after beautification. Note that our objective is not only to preserve the identity but also to improve the beauty of the generated image AB as jointly constrained by Equations (4) and (5).



FIGURE 4
Beauty degree adjustment by controlled beauty representation (the leftmost is the original input, from left to right: light to heavy beautification).

3.5.4. Beauty loss

To exploit the beauty feature of the reference, a beauty prediction model is first used to extract the beauty features, and then we propose to minimize the distance L_1 between the beautified face AB and B as follows:

$$\begin{aligned} \mathcal{L}_{BT}^A &= \|f_{bt}(G[E_c(A), E_s(A)]) - f_{bt}(A)\|_1, \\ \mathcal{L}_{BT}^B &= \|f_{bt}(G[E_c(B), E_s(B)]) - f_{bt}(B)\|_1, \\ \mathcal{L}_{BT}^{\tilde{A}} &= \|f_{bt}(G[E_c(A), E_s(B)]) - f_{bt}(B)\|_1, \end{aligned} \quad (6)$$

where f_{bt} denotes the operator extracting the 256-dimensional beauty feature (FC2 as shown in Figure 3).

3.5.5. Perceptual loss

Unlike makeup transfer, our face beautification seeks many-to-many mapping in an unsupervised way, which is more challenging, especially in view of both inner-domain and cross-domain variations. As mentioned in Ma et al. (2018), semantic inconsistency is a major issue for such unsupervised many-to-many translations. To address this problem, we propose the application of a perceptual loss to minimize the perceptual distance between the beautified face AB and the reference face B . This is a modified version of Ma et al. (2018), where instance normalization (Ulyanov et al., 2017) is performed on the features of VGG (Simonyan and Zisserman, 2014) (f_{vgg}) before computing the perceptual distance.

$$\mathcal{L}_p^{\tilde{A}} = \|f_{vgg}(G(\tilde{A})) - f_{vgg}(B)\|_2, \quad (7)$$

where $\|\cdot\|_2$ denotes the L_2 norm.

3.5.6. Total loss

When everything is put together, we jointly train the architecture by optimizing the following objective function.

$$\begin{aligned} \min_{E_c, E_s, G} \max_D \mathcal{L}(E_c, E_s, G, D) &= \lambda_1(\mathcal{L}_{REC}^A + \mathcal{L}_{REC}^B) + \\ &\lambda_2(\mathcal{L}_{ID}^A + \mathcal{L}_{ID}^B + \mathcal{L}_{ID}^{\tilde{A}}) + \lambda_3(\mathcal{L}_{BT}^A + \mathcal{L}_{BT}^B + \mathcal{L}_{BT}^{\tilde{A}}) + \\ &\lambda_4 \mathcal{L}_{GAN}^{AB} + \lambda_5 \mathcal{L}_p^{\tilde{A}} \end{aligned} \quad (8)$$

where $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5$ are the regularization parameters.

4. Experimental setup

4.1. Training datasets

Two datasets are used in our experiments. First, we used CelebA (Liu et al., 2015) to carry out the beautification experiment (only female celebrities are considered in this article). The authors of Liu et al. (2019) have found that some facial characteristics have a positive impact on the perception of beauty. Therefore, we have followed their findings to prepare our training datasets, that is, images containing these positive attributes (e.g., arched eyebrow, heavy makeup, high cheekbone, wearing lipsticks) as our reference dataset B ; and images that do not contain those attributes as our target dataset (to be beautified) A .

We have combined CelebA training and validation sets as our new training set to increase training size, but we keep the test data set the same as the original protocol (Liu et al., 2015). Our finalized training set includes 7,195 for A and 18,273 for B , and the testing set has 724 classA images and 2112 classB images. Another data set called SCUT-FBP5500 (Liang et al., 2018) is used to train our facial beauty prediction network. Following

their protocol, in our experiment, we have used 60% samples (3300 images) as training and the rest 40% (2200) as tests.

4.2. Implementation details

4.2.1. Generative model

Similarly to Huang et al. (2018), our E_c consists of several strided convolutional layers and residual blocks (He et al., 2016), and all convolutional layers are followed by Instance Normalization (IN) (Ulyanov et al., 2017). For E_s , a global average pool layer and a fully connected layer (FC) are followed by the strided convolutional layers. IN layer is removed to preserve the beauty features. Inspired by recent GAN work (Dumoulin et al., 2016; Huang and Belongie, 2017; Karras et al., 2019) using affine transformation parameters in normalization layers to better represent style, our decoder G is equipped with residual blocks and adaptive instance normalization (AdaIN). AdaIN parameters are dynamically generated by a multiple perceptron (MLP) from beauty codes as follows:

$$\text{AdaIN}(z, \gamma, \beta) = \gamma \left(\frac{z - \mu(z)}{\sigma(z)} \right) + \beta \quad (9)$$

where z is the activation of the previous convolutional layer, μ and σ are the mean and standard deviation of the channel, γ and β are the parameters generated by the MLP.

4.2.2. Discriminative model

We have implemented multiscale discriminators (Wang et al., 2018a) to guide the generative model in generating a realistic and consistent image in a global view. Furthermore, LSGAN (Mao et al., 2017) is used in our discriminative model to maximize image quality.

4.2.3. Beauty and identity model

As shown in Figure 3, we have used an existing face recognition model, LightCNN (Wu et al., 2018), which was trained on millions of faces and achieved state-of-the-art performance in several benchmark studies. In order to extract face beauty feature, we do a fine-tuning based on the pre-trained model from LightCNN, the last fully connected (FC2) layer is the learnable layer for beauty score prediction, and all previous layers are kept fixed during the training process. When tested in the popular SCUT-FBP5500 dataset (Liang et al., 2018), our method achieves the MAE of 0.2372 on the test set, which significantly outperforms theirs (0.2518) (Liang et al., 2018) in our experiment.

In our experimental setting, the standard LightCNN is considered as the identity feature extractor and the fine-tuning beauty prediction model is used as the face beauty extractor. To extract both ID and beauty features using one model, we have

TABLE 1 User study preference for beautified images.

Model	Count	Percent
CycleGAN	401	20.05
DRIT	282	14.1
MUNIT	390	19.5
Ours	927	46.35

The bold values indicate the best results.

taken advantage of the beauty prediction model and extracted the beauty feature from the last FC layer (FC2 in Figure 3), and the second to last FC layer (FC1 in Figure 3) as the output of the identity characteristic. When optimization involves two interacting networks, we have found that such a piggyback idea is more efficient than jointly training both beautification and beauty prediction modules. The following hyperparameters are empirically chosen in our experiment: the batch size is set as 4 with a single 2080Ti GPU. Our training has 360,000 iterations for a total of around 200 epochs. We use Adam Optimization (Kingma and Ba, 2014) with $\beta_1 = 0.5$, $\beta_2 = 0.999$ and Kaiming initialization (He et al., 2015). The learning rate is set as 0.0001 with a decay rate of 0.5 in every 100,000 iterations. The style codes from fc have 64 dimensions, and the loss weights are set as: $\lambda_1 = 10, \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = 1$.

5. Experimental results and evaluations

5.1. Baseline methods

5.1.1. CycleGAN

Zhu et al. (2017) a loss of consistency of the cycle was introduced to facilitate image-to-image translation, providing a simple but efficient solution to the transfer of style from unpaired data.

5.1.2. DRIT

Lee et al. (2018) an architecture projects images onto two spaces: a domain-invariant content space capturing shared information across domains and a domain-specific attribute space. Similarly to CycleGAN, a cross-cycle consistency loss based on disentangled representations is introduced to deal with unpaired data. Unlike CycleGAN, DRIT is capable of generating diverse images on a wide range of tasks.

5.1.3. MUNIT

Huang et al. (2018) a framework for unsupervised multimodal image-to-image translation, where images are decomposed into a domain-invariant content code and a style code that captures domain-specific properties. By combining

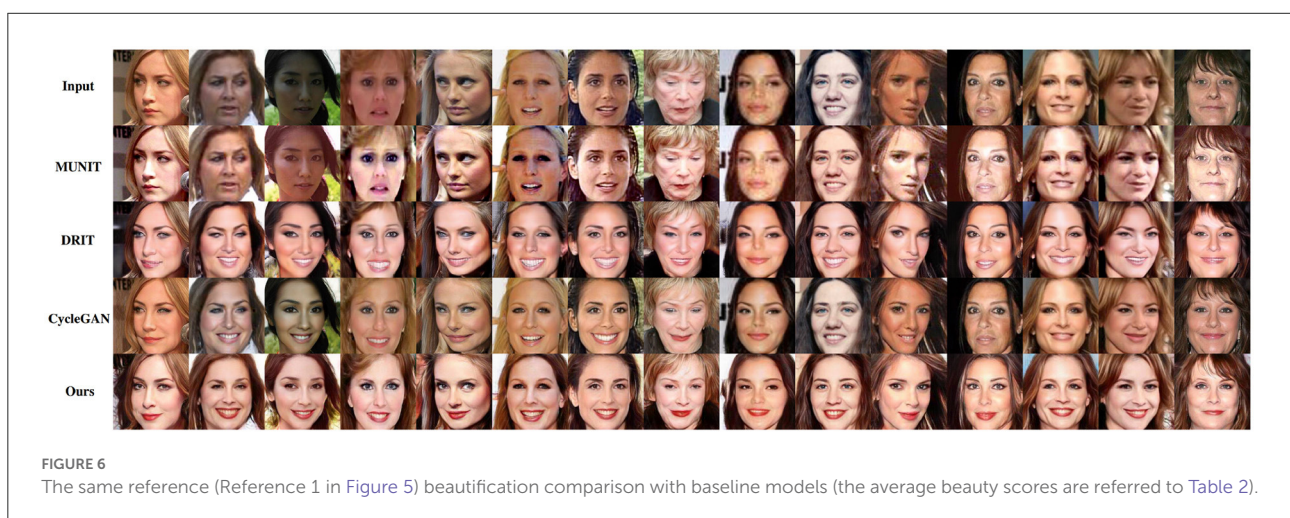
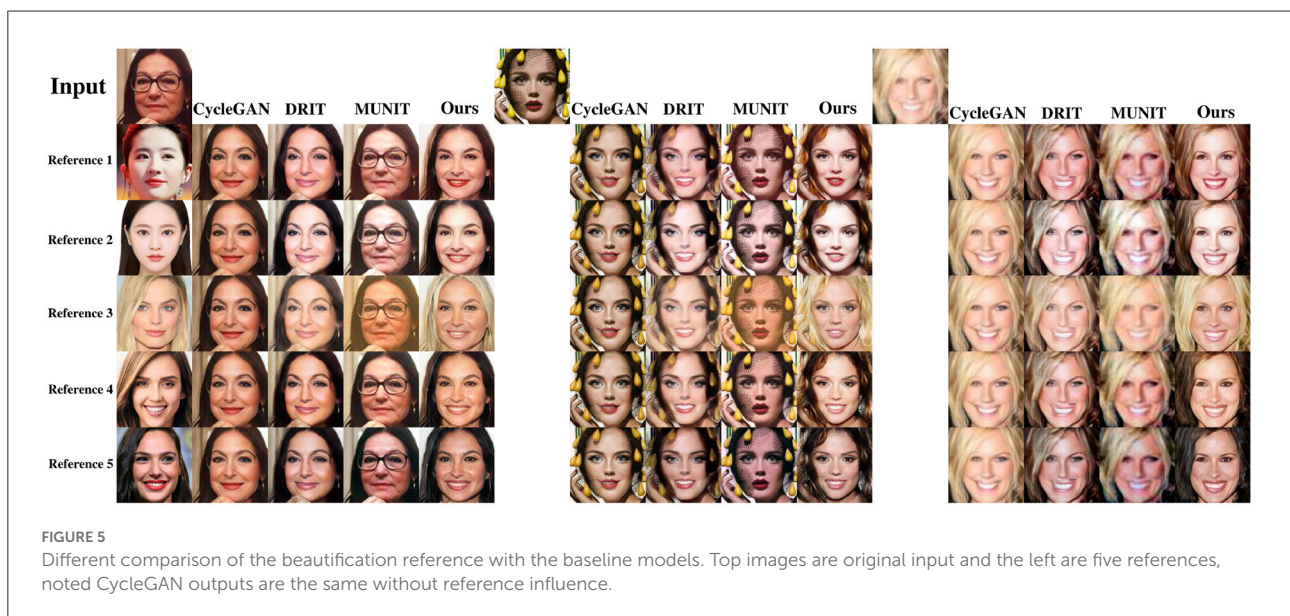
content code with a random style code, MUNIT can also generate diverse outputs from the target domain.

As mentioned in Section 2, all baseline methods have weaknesses when applied to reference-based beautification. CycleGAN cannot take advantage of specific references for translation, and the output lacks diversity once the training has been completed. DRIT and MUNIT are capable of many-to-many translations, but fail to generate a sequence of correlated images (e.g., faces with increasing beauty scores). On the contrary, our model is capable of not only beautifying faces based on a given reference, but also controlling the degree of beautification to fine granularity, as shown in Figure 4.

5.2. Qualitative and quantitative evaluations

5.2.1. User study

To evaluate the image quality of human perception, we develop a user study and ask users to vote for the most attractive among ours and the baseline. One hundred face images from the test set are submitted to Amazon Mechanical Turk (AMT), and each survey requires 20 users. We collected 2,000 data points in total to evaluate human preference. The final results demonstrate the superiority of our model, as shown in Table 1, Figure 5, and Figure 6.



5.2.2. Beauty score improvement

To further evaluate the effectiveness of the proposed beautification approach, we have fed the beautified images into our face beauty prediction model to output the beauty scores. The beauty prediction model is trained on SCUT-FBP as mentioned above, and the beauty score scale is 5 in that dataset. After calculating and averaging the test images (724), our model outperforms all other methods and gains an increase 37.11% compared to the average beauty score of the original input, as shown in Table 2.

5.3. Ablation study

To investigate the importance of each loss, we tested three variants of our model by removing \mathcal{L}_{ID} , \mathcal{L}_{BT} , and $\mathcal{L}_{\mathcal{P}}$, one at a time. See Figures 7–9 for visual comparisons. These losses complement each other and work in harmony to achieve the optimum beautification effect. This further demonstrates that our loss functions and architecture are well-designed for the facial beautification task.

TABLE 2 Average beauty score after beautification.

Model	Beauty score	Gain
Original	0.97	-
CycleGAN	1.15	18.56%
DRIT	1.25	28.87%
MUNIT	1.01	4.12%
Ours	1.33	37.11%

The bold values indicate the best results.

5.4. Discussions and limitations

Compared to recently developed makeup transfer, such as BeautyGAN (Li et al., 2018) and BeautyGlow (Chen et al., 2019), we note that our approach differs in the following aspects. Similar to BeautyGAN (Li et al., 2018), ours assumes the availability of a reference image; but unlike BeautyGAN (Li et al., 2018) which focuses only on local touch-up, ours is capable of transferring both global and local beauty features from the reference to the target. Similar to BeautyGlow (Chen et al., 2019), ours can adjust the magnification in the latent space; but unlike BeautyGlow (Chen et al., 2019), ours can improve the beauty score (rather than only increasing the extent of makeup).

Both the user study and the beauty score evaluation have demonstrated the superiority of our model. The proposed model is robust to low-quality images such as blur and challenging lighting conditions, as shown in Figure 10. However, we also notice that there are a few typical failed cases in which our model tends to produce noticeable artifacts when the inputs have large occlusions and pose variations (please refer to Figure 11). This is most likely caused by poor alignment, i.e., our references are mostly frontal images; while large occlusion and pose variations lead to misalignment.

6. Conclusions and future works

In this paper, we have studied the problem of face beautification and presented a novel framework that is more flexible than makeup transfer. Our approach integrates style-based synthesis with beauty score prediction by piggybacking a LightCNN with a GAN-based architecture. Unlike makeup transfer, our approach targets many-to-many (instead of one-to-one) translation, where multiple outputs can be defined by

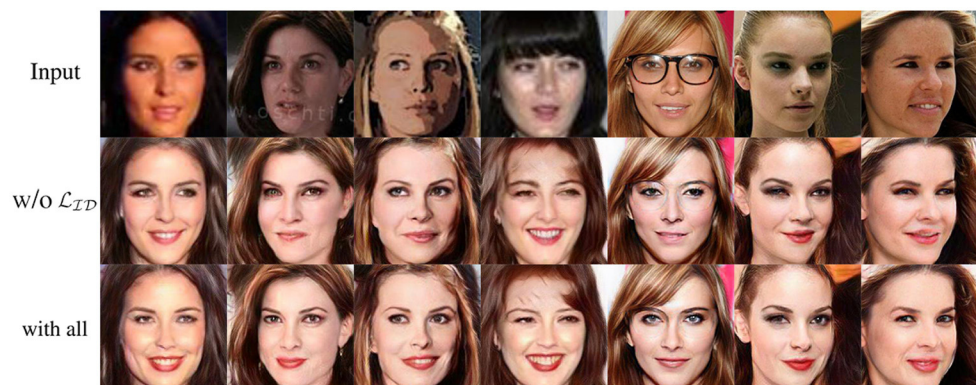
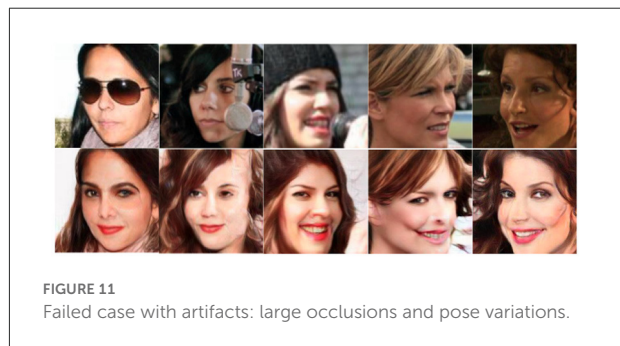
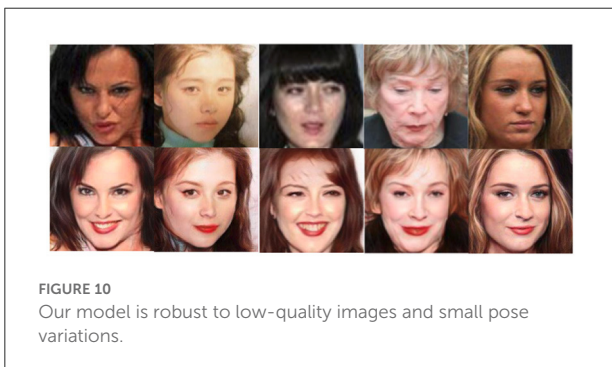
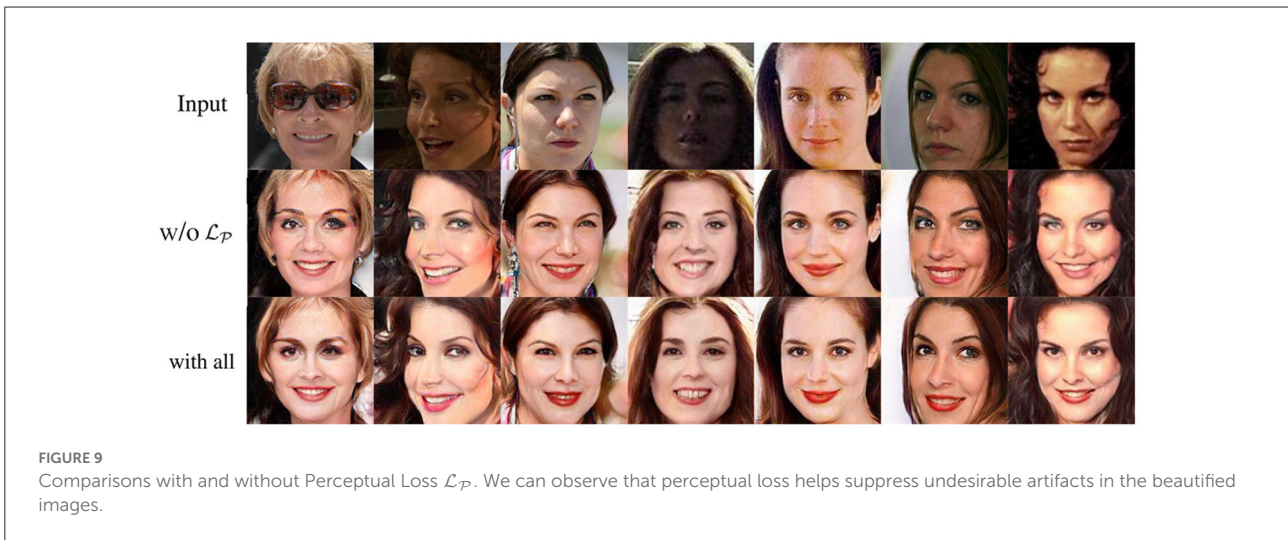
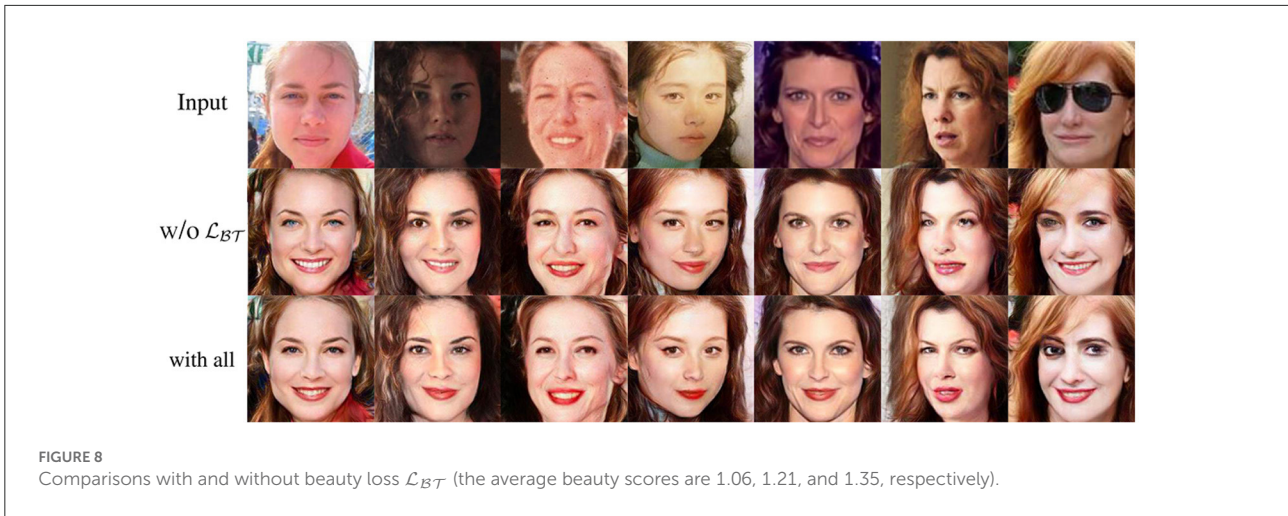


FIGURE 7 Comparisons with and without ID Loss \mathcal{L}_{ID} . Note that the inclusion of ID loss can better preserve the identity information of a given face.



either different references or varying beauty scores. In particular, we have constructed two interacting networks for beautification and beauty prediction. Through a simple weighting strategy, we managed to demonstrate the fine-granularity control of the

beautification process. Our experimental results have shown the effectiveness of the proposed approach both subjectively and objectively.

Personalized beautification is expected to attract more attention in the coming years. In this work, we have focused

only on the beautification of female Caucasian faces. A similar question can be studied for other populations, although the relationship between gender, race, cultural background, and the perception of facial attractiveness has remained underresearched in the literature. How AI can help shape the practice of personal makeup and plastic surgery is an emerging field for future research.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

Ethics statement

Written informed consent was not obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

XuL, RW, and XiL contributed to the conception and design of the study. XuL, RW, HP, MY, and C-FC

performed experiments and statistical analysis. XiL wrote the first draft of the manuscript. All authors contributed to the review of the manuscript and approved the submitted version.

Conflict of interest

Authors XuL, RW, HP, and C-FC were employed by ObEN, Inc.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationship that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Berscheid, E., Dion, K., Walster, E., and Walster, G. W. (1971). Physical attractiveness and dating choice: a test of the matching hypothesis. *J. Exp. Soc. Psychol.* 7, 173–189. doi: 10.1016/0022-1031(71)90065-5
- Bradbury, E. (1994). The psychology of aesthetic plastic surgery. *Aesthetic Plast Surg.* 18, 301–305. doi: 10.1007/BF00449799
- Bull, R., and Rumsey, N. (2012). *The Social Psychology of Facial Appearance*. Springer Science & Business Media.
- Chang, H., Lu, J., Yu, F., and Finkelstein, A. (2018). "Pairedcyclegan: asymmetric style transfer for applying and removing makeup," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT: IEEE), 40–48.
- Chen, H.-J., Hui, K.-M., Wang, S.-Y., Tsao, L.-W., Shuai, H.-H., and Cheng, W.-H. (2019). "Beautyglow: on-demand makeup transfer framework with reversible generative network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Long Beach, CA: IEEE), 10042–10050.
- Cunningham, M. R. (1986). Measuring the physical in physical attractiveness: quasi-experiments on the sociobiology of female facial beauty. *J. Pers. Soc. Psychol.* 50, 925. doi: 10.1037/0022-3514.50.5.925
- Donahue, C., Lipton, Z. C., Balsubramani, A., and McAuley, J. (2017). Semantically decomposing the latent spaces of generative adversarial networks. *arXiv preprint arXiv:1705.07904*. doi: 10.48550/arXiv.1705.07904
- Dumoulin, V., Shlens, J., and Kudlur, M. (2016). A learned representation for artistic style. *arXiv preprint arXiv:1610.07629*. doi: 10.48550/arXiv.1610.07629
- Eisenthal, Y., Dror, G., and Ruppel, E. (2006). Facial attractiveness: beauty and the machine. *Neural Comput.* 18, 119–142. doi: 10.1162/089976606774841602
- Fan, Y.-Y., Liu, S., Li, B., Guo, Z., Samal, A., Wan, J., et al. (2017). Label distribution-based facial attractiveness computation by deep residual learning. *IEEE Trans. Multimedia* 20, 2196–2208. doi: 10.1109/TMM.2017.2780762
- Gan, J., Li, L., Zhai, Y., and Liu, Y. (2014). Deep self-taught learning for facial beauty prediction. *Neurocomputing* 144, 295–303. doi: 10.1016/j.neucom.2014.05.028
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). "Generative adversarial nets," in *Advances in Neural Information Processing Systems* (Montreal, QC), 2672–2680.
- He, K., Zhang, X., Ren, S., and Sun, J. (2015). "Delving deep into rectifiers: surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE International Conference on Computer Vision* (Santiago: IEEE), 1026–1034.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas, NV: IEEE), 770–778.
- Huang, X., and Belongie, S. (2017). "Arbitrary style transfer in real-time with adaptive instance normalization," in *Proceedings of the IEEE International Conference on Computer Vision* (Venice: IEEE), 1501–1510.
- Huang, X., Liu, M.-Y., Belongie, S., and Kautz, J. (2018). "Multimodal unsupervised image-to-image translation," in *Proceedings of the European Conference on Computer Vision (ECCV)* (Munich: ECCV), 172–189.
- Karras, T., Laine, S., and Aila, T. (2019). "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Long Beach, CA: IEEE), 4401–4410.
- Kingma, D. P., and Ba, J. (2014). Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. doi: 10.48550/arXiv.1412.6980
- Lee, H.-Y., Tseng, H.-Y., Huang, J.-B., Singh, M., and Yang, M.-H. (2018). "Diverse image-to-image translation via disentangled representations," in *Proceedings of the European Conference on Computer Vision (ECCV)* (Munich: ECCV), 35–51.

- Li, T., Qian, R., Dong, C., Liu, S., Yan, Q., Zhu, W., et al. (2018). "Beautygan: instance-level facial makeup transfer with deep generative adversarial network," in *2018 ACM Multimedia Conference on Multimedia Conference* (Seoul: ACM), 645–653.
- Liang, L., Lin, L., Jin, L., Xie, D., and Li, M. (2018). "Scut-fbp5500: a diverse benchmark dataset for multi-paradigm facial beauty prediction," in *2018 24th International Conference on Pattern Recognition (ICPR)* (Beijing: IEEE), 1598–1603.
- Little, A. C., Jones, B. C., and DeBruine, L. M. (2011). Facial attractiveness: evolutionary based research. *Philos. Trans. R. Soc. B Biol. Sci.* 366, 1638–1659. doi: 10.1098/rstb.2010.0404
- Liu, X., Li, T., Peng, H., Chuoying Ouyang, I., Kim, T., and Wang, R. (2019). "Understanding beauty via deep facial features," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (Long Beach, CA: IEEE).
- Liu, Z., Luo, P., Wang, X., and Tang, X. (2015). "Deep learning face attributes in the wild," in *Proceedings of International Conference on Computer Vision* (Santiago: ICCV).
- Ma, L., Jia, X., Georgoulis, S., Tuytelaars, T., and Van Gool, L. (2018). Exemplar guided unsupervised image-to-image translation with semantic consistency. *arXiv preprint arXiv:1805.11145*. doi: 10.48550/arXiv.1805.11145
- Macgregor, F. C. (1989). Social, psychological and cultural dimensions of cosmetic and reconstructive plastic surgery. *Aesthetic Plast Surg.* 13, 1–8. doi: 10.1007/BF01570318
- Mao, X., Li, Q., Xie, H., Lau, R. Y., Wang, Z., and Paul Smolley, S. (2017). "Least squares generative adversarial networks," in *Proceedings of the IEEE International Conference on Computer Vision* (Venice: IEEE), 2794–2802.
- Mathieu, M. F., Zhao, J. J., Zhao, J., Ramesh, A., Sprechmann, P., and LeCun, Y. (2016). "Disentangling factors of variation in deep representation using adversarial training," in *Advances in Neural Information Processing Systems* (Barcelona), 5040–5048.
- Perrett, D. I., Burt, D. M., Penton-Voak, I. S., Lee, K. J., Rowland, D. A., and Edwards, R. (1999). Symmetry and human facial attractiveness. *Evolut. Hum. Behav.* 20, 295–307. doi: 10.1016/S1090-5138(99)00014-8
- Perrett, D. I., Lee, K. J., Penton-Voak, I., Rowland, D., Yoshikawa, S., Burt, D. M., et al. (1998). Effects of sexual dimorphism on facial attractiveness. *Nature* 394, 884. doi: 10.1038/29772
- Phillips, K. A., McElroy, S. L., Keck Jr, P. E., Pope Jr, H. G., and Hudson, J. I. (1993). Body dysmorphic disorder: 30 cases of imagined ugliness. *Am. J. Psychiatry* 150, 302. doi: 10.1176/ajp.150.2.302
- Rankin, M., Borah, G. L., Perry, A. W., and Wey, P. D. (1998). Quality-of-life outcomes after cosmetic surgery. *Plast Reconstr. Surg.* 102, 2139–2145. doi: 10.1097/00006534-199811000-00053
- Riggio, R. E., and Woll, S. B. (1984). The role of nonverbal cues and physical attractiveness in the selection of dating partners. *J. Soc. Pers. Relat.* 1, 347–357. doi: 10.1177/0265407584013007
- Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. doi: 10.48550/arXiv.1409.1556
- Tenenbaum, J. B., and Freeman, W. T. (2000). Separating style and content with bilinear models. *Neural Comput.* 12, 1247–1283. doi: 10.1162/089976600300015349
- Thornhill, R., and Gangestad, S. W. (1999). Facial attractiveness. *Trends Cogn. Sci.* 3, 452–460. doi: 10.1016/S1364-6613(99)01403-5
- Ulyanov, D., Vedaldi, A., and Lempitsky, V. (2017). "Improved texture networks: maximizing quality and diversity in feed-forward stylization and texture synthesis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Hawaii, HI: IEEE), 6924–6932.
- Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., and Catanzaro, B. (2018a). "High-resolution image synthesis and semantic manipulation with conditional gans," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT: IEEE), 8798–8807.
- Wang, Z., Tang, X., Luo, W., and Gao, S. (2018b). "Face aging with identity-preserved conditional generative adversarial networks," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Salt Lake City, UT: IEEE).
- Wu, X., He, R., Sun, Z., and Tan, T. (2018). A light cnn for deep face representation with noisy labels. *IEEE Trans. Inf. For. Security* 13, 2884–2896. doi: 10.1109/TIFS.2018.2833032
- Xie, D., Liang, L., Jin, L., Xu, J., and Li, M. (2015). "Scut-fbp: a benchmark dataset for facial beauty perception," in *2015 IEEE International Conference on Systems, Man, and Cybernetics* (Hong Kong: IEEE), 1821–1826.
- Xu, J., Jin, L., Liang, L., Feng, Z., Xie, D., and Mao, H. (2017). "Facial attractiveness prediction using psychologically inspired convolutional neural network (pi-cnn)," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (New Orleans, LA: IEEE), 1657–1661.
- Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE International Conference on Computer Vision* (Venice: IEEE), 2223–2232.