# Face Beautification: Beyond Makeup Transfer

Xudong Liu [†1,2], Ruizhe Wang [‡1], Chih-Fan Chen [‡1], Minglei Yin [†2], Hao Peng [‡1], Shukhan Ng [‡1], and Xin Li [*2]

[1]Oben, Inc
[2]West Virginia University

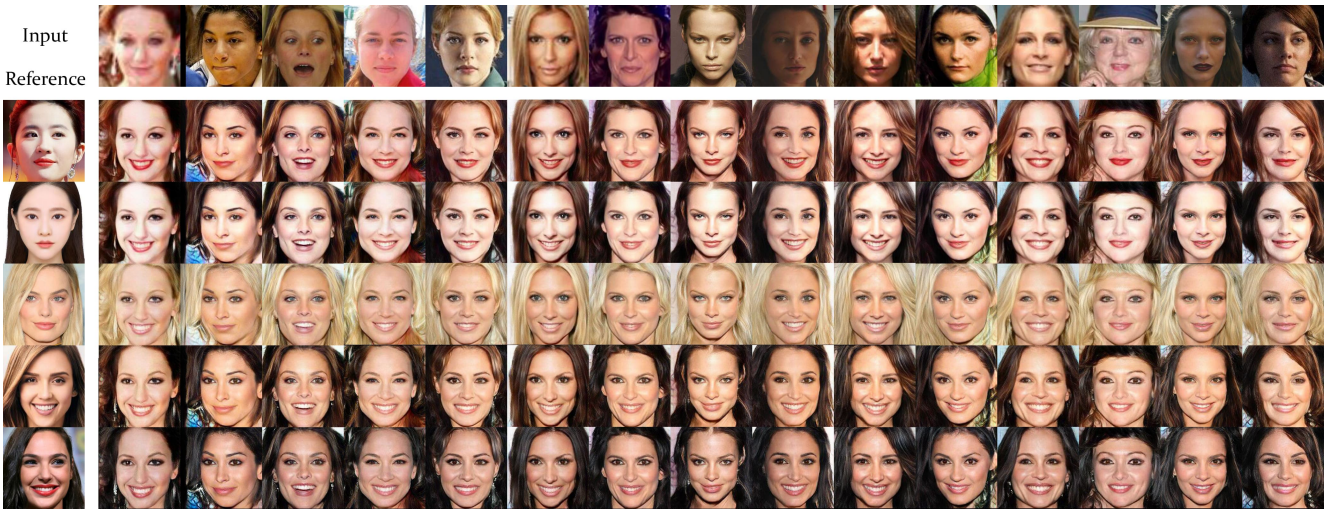


Figure 1: Face beautification as many-to-many image translation: our approach integrates style-based beauty representation with beauty score prediction model and is capable of fine-granularity control.

## Abstract

*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. Similar 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 of 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 referenced beauty style and targeted 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 SCUT-FBP database) into the process of beautification. Unlike makeup transfer, our approach targets at many-to-many (instead of one-to-one) translation where multiple outputs can be defined by either different references or varying beauty scores. Extensive experimental results are reported to demonstrate the effectiveness and flexibility of the proposed face beautification framework.*

## 1. Introduction

Facial appearance plays an important role in our social lives [3]. People with attractive faces have many advantages in their social activities such as dating and voting [23]. It has been found that attractive people enjoy higher chances of getting dating [34], and their partners are more likely to gain satisfaction when compared to dating with less attrac-

† {xdliu,my0033}@mix.wvu.edu
‡ {ruizhe, chihfan, hpeng,shukhan}@oben.com
∗ xin.li@mail.wvu.edu

tive ones [1]. It has also been found that faces could affect hiring decisions and influence voting behavior [23]. Overwhelmed by social fascination with beauty, women with unattractive faces may suffer from social isolation, depression, and even psychological disorders [3, 33, 32, 27, 2]. 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).

The problem of face beautification has been extensively studied by philosophers, psychologists and plastic surgeons. Rapid advances in imaging technology and social media greatly expedited the popularity of digital photos especially selfies in our daily lives. Most recently, virtual face beautification based on the idea of makeup application or transfer has been developed in computer vision communities: PairedCycleGAN [4], BeautyGAN[21], BeautyGlow [5]. Although these existing works have achieved impressive results, we argue that face beautification based on makeup transfer only has fundamental limitations. Without changing important facial attributes (e.g., shape and lentigo), the application of makeup - abstracted by image-to-image translation [44, 16, 20] - can only improve the beauty score to some extent.

A more flexible and promising framework is to formalize the process of face beautification by *one-to-many* translation where the destination can be defined in many different manners. On one hand, we can target at producing a sequence of output images with monotonically increased 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 style). Under this framework, face beautification can be made more flexible - e.g., we can transfer the beauty style from a reference image to reach a specified beauty score, which is beyond the reach of makeup transfer [21, 5].

To achieve this objective, we propose a novel generative adversarial network (GAN)-based architecture in this paper. Inspired by the latest advances in style-based synthesis (e.g., styleGAN[17]) and face beauty understanding from data [24], we propose to integrate both style-based beauty representation (extracted from the reference face) and beauty score prediction (trained on SCUT-FBP database [42]) into the process of face beautification. More specifically, style-based beauty representations will be learned from both inquiry and reference images first via light convolutional neural network (LightCNN) and leveraged to guide the process of style transfer (actual beautification). Then a dedicated GAN-based architecture integrated with reconstruction, beauty and identity loss functions is constructed. In order to have a fine-granularity control of the beautification process, we have invented a simple yet effective reweighting strategy of gradually improving the beauty score in synthesized images until reaching the target (specified by the reference image).

Our key contributions are summarized as follows:

- 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 guiding the process of face beautification.
- A face beauty prediction network based on fine-tuning of LightCNN is trained and integrated into the proposed style-based face beautification network. The prediction module provides valuable feedback to the synthesis module while approaching the desirable beauty score.
- A piggyback trick to extract both identity and beauty features from fine-tuned LightCNN and design of loss functions reflecting the tradeoff between identity preservation and face beautification.
- To the best of our knowledge, this is the first work capable of delivering face beautification results with fine-granularity control (i.e., a sequence of face images approaching the reference one with monotonically increasing beauty scores).
- A comprehensive evaluation shows the superiority of the proposed approach when compared to existing state-of-the-art image-to-image transfer techniques including CycleGAN [44], MUNIT[16], and DRIT[20].

## 2. Related Works

**Makeup and Style Transfer**. Two recent works on face beauty are BeautyGAN [21] and BeautyGlow [5]. In BeautyGlow [5], the makeup features (e.g., eyeshadows and lip gloss) are first extracted from reference makeup images and then transferred to source non-makeup images. The magnification parameter in the latent space can be tuned to adjust the extent of the makeup. In BeautyGAN [21], the issue of extracting/transferring local and delicate makeup information was addressed by incorporating both global domain-level loss and local instance-level loss in an dual input/output GAN.

Face beautification is also related to more general image-to-image translation. Both symmetric (e.g., CycleGAN [44]) and asymmetric (e.g., PairedCycleGAN [4]) have been studied in the literature; the latter was shown effective for makeup application and removal. Extensions of style

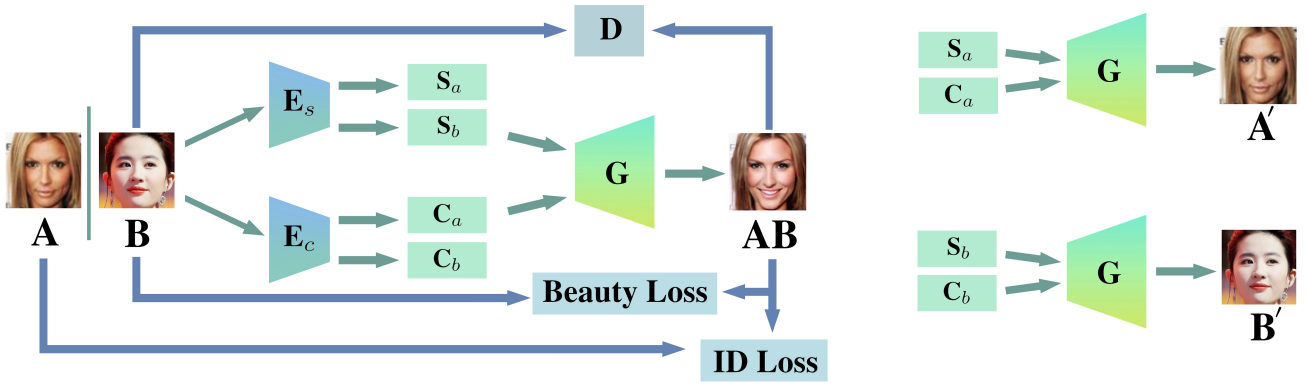


Figure 2: Overview of the proposed network architecture.

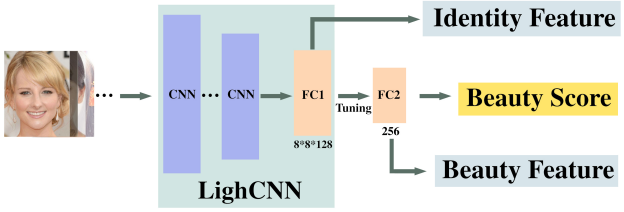


Figure 3: Fine-tuning network for beauty score prediction.

transfer into multimodal domain (i.e., one-to-many translations) have been considered in MUNIT [16] and DRIT [20]. It is also worth mentioning face image synthesis via StyleGAN [17] which has demonstrated super-realistic performance.

**Face Beauty Prediction**. The perception of facial appearance or attractiveness is a classical topic in psychology and cognitive sciences [37, 31, 30]. However, developing a computational algorithm that can automatically predict beauty scores from facial images is only a recent endeavor [9, 11]. Thanks to the public release of face beauty database such as SCUT-FBP [42], there has been a growing interest in machine learning based approaches toward face beauty prediction [10, 43].

## 3. Proposed Method

### 3.1. Facial Attractiveness Theory

Why facial attractiveness matters? From an evolutionary perspective, a plausible working hypothesis is that the psychological mechanisms underlying primates' judgments about attractiveness are consequence of long-period evolution and adaptation. More specifically, facial attractiveness is beneficial to choosing a mate which in turn facilitates the gene propagation [37]. At the primitive level, facial attractiveness is hypothesized to reflect information about an individuals health. Accordingly, conventional wisdom in facial attractiveness research 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 [6]).

In particular, facial attractiveness for young females is a stimulating topic as witnessed by the long-lasting popularity of beauty pageants. In [6], the relation between female facial features and the responses of males was investigated. Based on the male subjects' attractiveness ratings, two classes of facial features (e.g., large eyes, small nose, and small chin; prominent cheekbones and narrow cheeks) are positively correlated with attractiveness ratings. It is also known from the same study [6] that facial features can also predict personality attributions and altruistic inclinations. We opt to focus on face beautification for females only in this work.

### 3.2. Problem Formulation and Motivation

Given a target face (an ordinary 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? Such problem of face beautification can be formulated as two subproblems: *style transfer* and *beauty prediction*. Meantime, an important new insight brought into our problem formulation is that the treatment of face beautification as a sequential process where the beauty score of the target face can be gradually improved by a consecutive 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 stud-

ied in the literature which dated back to content-style separation [36]. The idea of extracting style-based representation (style code) has attracted increasingly more attention in recent years -e.g., [15, 16, 19, 7, 29]. Note that makeup transfer only represents a special case where style is characterized by local features only (e.g., eye-shadow and lipstick). In this work we conceive 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 sharing of learned features between style transfer and beauty prediction allows us to achieve the fine-granularity control over the process of beautification.

### 3.3. Architecture Design

As illustrated in Fig. 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 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 beauty transfer). Assume 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 (identity) encoder $E_c$ and style (beauty) $E_s$ encoder respectively. In order to transfer the beauty style from reference $B$ to target $A$, it is natural to concatenate the content(identity)-based representation $C_a$ with the style(beauty)-based representation $S_b$; and then reconstruct the beautified image $AB$ through a dedicated decoder $G$ defined by

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

The rest of our architecture in Fig. 2 mainly includes two components: a GAN-based module ($G$ pairs with $D$) responsible for style transfer and a module of beauty and identity loss responsible for beauty prediction (please refer to Fig. 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 [38], we propose to integrate an Instance-Normalization (IN) layer after convolutional layers as part of the encoder for content feature extraction. Meantime, a global average pooling and a fully connected layer follow convolutional layers as part of the encoder for beauty feature extraction. Note that we skip IN in beauty encoder because IN would remove the characteristics of original feature representing critical beauty-related information [15] (that's why we keep it within content encoder). To cooperate with beauty encoder and speed up the translation, the decoder is equipped with an Adaptive Instance Normalization (AdaIN) [15]. Additionally, we have adopted the popular multi-scale discriminators [40] with Least-Square GAN (LSGAN) [28] as the discriminator in our GAN module.

Our beauty prediction module is based on fine-tuning an existing LightCNN [41] as shown in Fig 3. Since it's difficult to train a deep neutral network for beauty prediction from the scratch, we opt to work with LightCNN [41] - a pre-trained model for face recognition with millions of face images. Instead, we employ a fine-tuning layer (FC2) to adapt it for beauty score prediction (FC2 plays the role of beauty feature extractor). Meantime, in order to preserve the identity during face beautification, we propose to take the full advantage of our beauty prediction model by piggybacking the identity feature it produced. More specifically, identity feature is generated from the second fully connected layer (FC1) of LightCNN; note that we have only fine-tuned the last fully connected (FC2) for beauty prediction. By using this piggyback trick, we manage to extract both identity and beauty features from one off-shelf model.

### 3.4. Fine-granularity Beauty Adjustment

As we argued before, beautification should be modeled by a continuous process instead of a discrete domain transfer. In order to achieve the fine-granularity control of the beautification process, we propose to formulate a weighted beautification equation by

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

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 fullest-extent beautification when $w_1 = 0, w_2 = 1$. Such 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: replacing $G[E_c(A), E_s(A) + E_s(B)]$ with $G[E_c(A), E_s(B)]$ in the training stage so that we do not need to train multiple weighted models when weights vary. Instead we apply the weighted beautification equation of Eq. (2) for testing directly. In other words, we pretend the beauty feature of the target image $A$ is forgotten during the training but partially exploit it during the testing (since it is less relevant than identity feature). In summary, our fine-granularity beauty adjustment strategy heavily counts on the capability of beauty encoder $E_s$ for reliably extracting beauty representation. The effectiveness of the proposed fine-granularity beauty adjustment can be justified by referring to Fig. 5.

### 3.5. Loss Functions

**Image reconstruction**. Both encoder and decoder need to make sure that target and reference images can be approximately reconstructed from the extracted content/style

representation. Here we have adopted $L_1$-norm for reconstruction loss for the reason of robustness.

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

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

**Adversarial loss**. We apply adversarial losses [12] for matching the distributions of the generated image $AB$ and the target data $B$. In other words, the adversarial loss ensures the beautified face looks as realistic as the reference.

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

where $G(AB)$ is defined by Eq. (1).

**Identity preservation**. To preserve the identity information during the process of beautification, we propose to adopt an identity loss function from the off-shelf face recognition model LightCNN [41] trained on millions of faces. Identity features are extracted from the FC1 layer, which is a $2^{13}$-dimensional vector.

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

where $\mathcal{L}_{\text{ID}}^{A}$ and $\mathcal{L}_{\text{ID}}^{B}$ are responsible for identity preservation, and $\mathcal{L}_{\text{ID}}^{AB}$ aims at preserving the identity after beautification. Note that our objective is to preserve the identity but improve the beauty in the generated image $AB$ as jointly constrained by Eqs. (4) and (5).

**Beauty loss**. In order to leverage the beauty feature from the reference, a beauty prediction model is first used to extract beauty features and then we propose to minimize the $L_1$ distance between the beautified face $AB$ and $B$ as following:

$$\mathcal{L}_{\text{BT}}^{A} = ||f_{bt}(G[E_c(A), E_s(A)]) - f_{bt}(A)||_1,$$
$$\mathcal{L}_{\text{BT}}^{B} = ||f_{bt}(G[E_c(B), E_s(B)]) - f_{bt}(B)||_1,$$
$$\mathcal{L}_{\text{BT}}^{AB} = ||f_{bt}(G[E_c(A), E_s(B)]) - f_{bt}(A)||_1, \quad (6)$$

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

**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 [26], semantic inconsistency is a major issue for such unsupervised many-to-many translation To address this issue, we propose to apply a perceptual loss to minimize the perceptual distance between the beautified face $AB$ and the reference face $B$. This is a modified version from [26], where Instance Normalization [38]is performed on VGG [35] features before computing the perceptual distance.

$$\mathcal{L}_{\text{P}}^{AB} = ||f_{vgg}(G(AB)) - f_{vgg}(B)||_2, \quad (7)$$

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

**Total loss**. Putting things together, we jointly train the architecture by optimizing the following objective function:

$$\min_{E_c,E_s,G}\max_{D} \mathcal{L}(E_c, E_s, G, D) = \lambda_1(\mathcal{L}_{\text{REC}}^{A} + \mathcal{L}_{\text{REC}}^{B}) +$$
$$\lambda_2(\mathcal{L}_{\text{ID}}^{A} + \mathcal{L}_{\text{ID}}^{B} + \mathcal{L}_{\text{ID}}^{AB}) + \lambda_3(\mathcal{L}_{\text{BT}}^{A} + \mathcal{L}_{\text{BT}}^{B} + \mathcal{L}_{\text{BT}}^{AB}) +$$
$$\lambda_4\mathcal{L}_{\text{GAN}}^{AB} + \lambda_5\mathcal{L}_{\text{P}}^{AB} \quad (8)$$

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

# 4. Experimental Setup

## 4.1. Training Datasets

Two datasets are used in our experiments. First, we have used CelebA [25] to conduct the beautification experiment (only female celebrities are considered in this paper). Authors in [24] have found that some facial attributes have a positive impact on beauty perception. So we have followed their findings to prepare our training datasets - i.e., the images containing those 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 (to be beautified) dataset $A$. We have merged the training and validation originate from CelebA as our new training set in order to enlarge the training size, but keep the testing dataset the same as the original protocol [25]. Our finalized training set includes 7195 for $A$ and 18273 for $B$, and testing set has 724 class-$A$ images and 2112 class-$B$ images. Another dataset called SCUT-FBP5500 [22] is used to train our face beauty prediction network. Following their protocol we have used 60% samples (3300 images) as training and the rest 40% (2200) as testing in our experiment.

## 4.2. Implementation details

**Generative model**. Similar to [16], our $E_c$ consists of several strided convolutional layers and residual blocks [14], all convolutional layers are followed by Instance Normalization (IN) [38]. As for $E_s$, a global average pooling layer and a fully connected (FC) layer are followed by the strided convolutional layers. IN layer is removed to preserve the beauty features. Inspired by recent GAN works

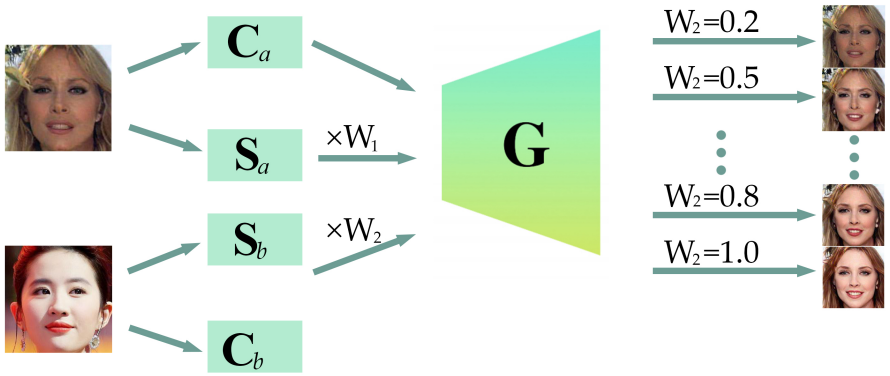


Figure 4: Testing stage for fine-granularity beautification adjustment.

Figure 5: Beauty degree adjustment by controlled beauty representation (the leftmost is the original input, from left to right: light to heavy beautification).

[15, 8, 17] that use affine transformation parameters in normalization layers to better represent style, our decoder $G$ is equipped with the residual blocks as well as Adaptive Instance Normalization (AdaIN). The parameters of AdaIN are dynamically generated by a Multiple Perceptron (MLP) from the beauty codes similar as [16], seeing as following:

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

where $z$ is the activation of the previous convolutional layer, $\mu$ and $\sigma$ are channel-wise mean and standard deviation, $\gamma$ and $\beta$ are parameters generated by the MLP.

**Discriminative model**. We have implemented multi-scale discriminators [39] to guide generative model to generate both realistic and consistent image in a global view. In addition, LSGAN [28] is used in our discriminative model to leverage the image quality.

**Beauty and identity model**. As shown in Fig. 3, we have used an off-shelf face recognition model– LightCNN [41], 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 training process. When tested on the popular CUT-FBP5500 dataset [22], our method achieves the MAE of 0.2372 on testing set, which significantly outperforms theirs (0.2518) [22] in our experiment.

In our experimental setting, the off-shelf LightCNN is considered as the identity feature extractor and the fine-tuning beauty prediction model is used as the face beauty extractor. In order to extract both ID and beauty features using one model, we have taken advantage of the beauty

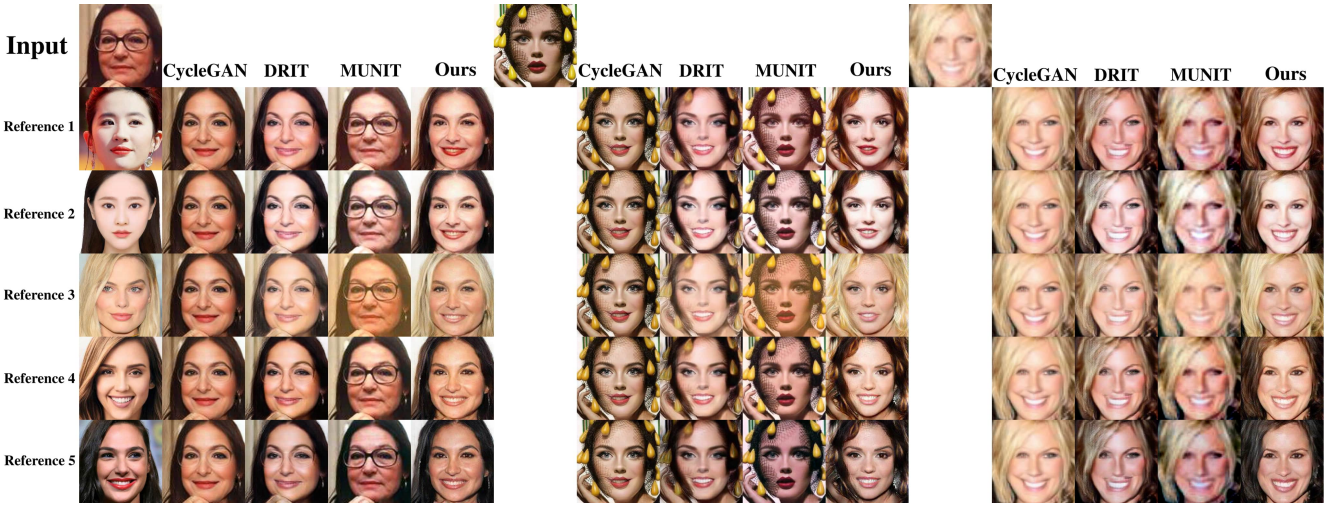


Figure 6: Different reference beautification comparison with baseline models. Top images are original input and the left are five references, noted CycleGAN outputs are the same without reference influence.


Figure 7: Same reference (Reference 1 in Fig 6) beautification comparison with baseline models.

prediction model and extract the beauty feature from the last FC layer (FC2 in Fig 3), and the second to last FC layer (FC1 in Fig 3) as the identity feature outputs. When optimization involves two interacting networks, we have found such piggyback idea is more efficient than jointly training both beautification and beauty prediction modules.

## 5. Experimental Results and Evaluations

### 5.1. Baseline Methods

**CycleGAN** [44] A cycle consistency loss was introduced to facilitate the image-to-image translation, which provides a simple but efficient solution to style transfer from unpaired data.

**DRIT** [19] An architecture projects images onto two spaces: a domain-invariant content space capturing shared information across domains and a domain-specific attribute space. Similar 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.

**MUNIT** [16] A framework for multimodal unsupervised image-to-image translation, where images are decomposed into a content code that is domain-invariant 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 their weakness when applied to reference-based beautification. CycleGAN cannot take advantage of specific references for translation, the outputs lack diversity once training done. DRIT and MUNIT are capable of many-to-many

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

Table 1: User study preference for beautified images.

| 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%** |

Table 2: Average beauty score after beautification.

translation but fail to generate a sequence of correlated images (e.g., faces with increasing beauty scores). By contrast, our model is capable of not only beautifying faces based on a given reference but also controlling the degree of the beautification to fine-granularity, as shown in Fig 5.

### 5.2. Qualitative and Quantitative Evaluations

**User study**. To evaluate the image quality from human's perception, we develop a user study and ask users to vote the most attractive one among ours and the baseline. 100 face images from testing set are submitted to Amazon Mechanical Turk (AMT), and each survey requires 20 users. We collect 2000 data points in total to evaluate human preference. The final results demonstrate the superiority of out model, showing in Table 1.

**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 before and the scale of beauty score is 5 in that dataset. After calculating and averaging the testing images (724), our model outperforms all other methods and gains a 37.11% increase when compared to average beauty score of the original input as shown in Table 2.

### 5.3. Discussions and Limitations

When compared against recently developed makeup transfer such as BeautyGAN [21] and BeautyGlow [5], we note that our approach differs in the following aspect. Similar to BeautyGAN [21], ours assumes the availability of a reference image; but unlike BeautyGAN [21] focusing on local touchup only, ours is capable of transferring both global and local beauty features from the reference to the target. Similar to BeautyGlow [5], ours can adjust the magnification in the latent space; but unlike BeautyGlow [5], ours can improve the beauty score (rather than only increasing the extent of makeup).

Figure 8: Our model is robust to low quality images and small pose variations.

Figure 9: Failed case with artifacts: large occlusions and pose variations.

Both user study and 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 Fig. 8. However, we also notice there are a few typical failed cases in which our model tends to produces noticeable artifacts when the inputs have large occlusions and pose variations (please refer to Fig. 9). 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 an GAN-based architecture. Unlike makeup transfer, our approach targets at many-to-many (instead of one-to-one) translation where multiple outputs can be defined by 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 manage to demonstrate the

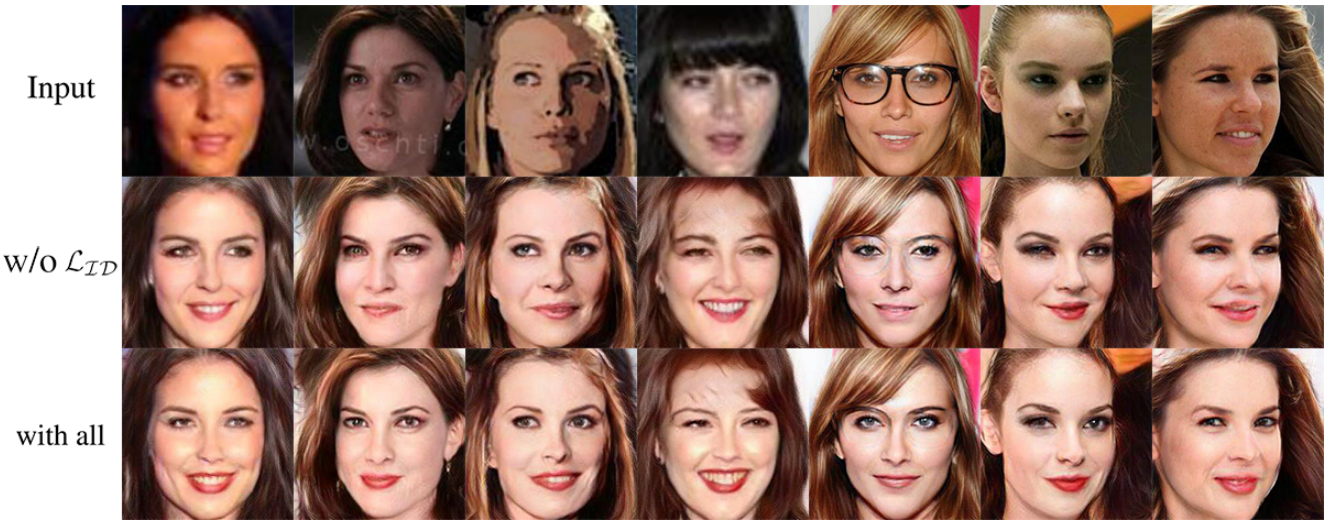


Figure 10: Comparisons with and w/o ID Loss $\mathcal{L}_{\mathcal{ID}}$

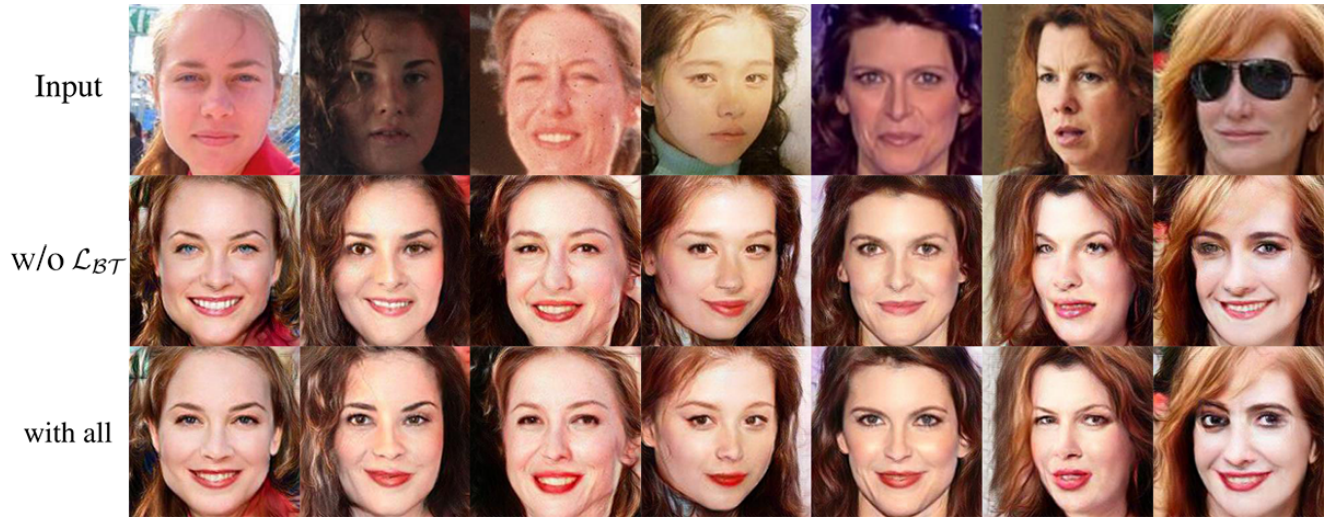


Figure 11: Comparisons with and w/o Beauty Loss $\mathcal{L}_{\mathcal{BT}}$

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

Personalized beautification is expected to attract increasingly more attention in the incoming years. This work we have only focused on the beautification of female Caucasian faces. A similar question can be studied for other populations even though the relationship between gender, race, cultural background and the perception of facial attractiveness has remained under-researched in the literature. How can AI help reshape the practice of personal makeup and plastic surgery is an emerging field for future research.

# 7. Appendix

## 7.1. Ablation Study

To investigate the importance of each loss, we experiment three variants of our model by removing $\mathcal{L}_{\mathcal{ID}}$, $\mathcal{L}_{\mathcal{BT}}$ and $\mathcal{L}_{\mathcal{P}}$, one at a time. See Fig 10, 11 and 12 for visual comparisons. These losses compliment each other and work in harmony to reach the optimum beautification effect. This further demonstrates that our loss functions and architecture are well-designed for the facial beautification task.


Figure 12: Comparisons with and w/o Perceptual Loss $\mathcal{L}_{\mathcal{P}}$

### 7.2. Network Architectures and Hyperparameters

**Generator Architecture.** We adopt our architecture from MUNIT [16]. Following the convention used in Johnson et al.'s Github repositoty *, let $c7s1-k$ denote a $7 \times 7$ convolutional block with $k$ filters and stride 1. $dk$ denotes a $4 \times 4$ convolutional block with $k$ filters and stride 2. $rk$ denotes a residual block that contrains two $3 \times 3$ convolutional blocks. $uk$ denotes a $2\times$ nearest-neighbor upsampling layer followed by a $5 \times 5$ convolutional block with $k$ filters and stride 1. $gap$ denotes a global average pooling layer and $fc$ denotes a fully connected layer. Instance Normalization (IN) [38] is in use to the content (ID) encoder and Adaptive Instance Normalization (AdaIN) [15] to style (beauty) encoder. And we use ReLU activations for generator. The generator architecture is as following:

- Content encoder $E_c$: $c7s1-64$, $d128$, $d256$, $r256$, $r256$, $r256$
- Style encoder $E_s$: $c7s1-64$, $d128$, $d256$, $d256$, $d256$, $gap$, $fc$
- Decoder $G$: $r256$, $r256$, $r256$, $r256$, $u128$, $u64$, $c7s1-3$

**Discriminator Architecture.** For discriminator, we follow CycleGAN's implementation, and use Leaky ReLU with a slop of 0.2 and multi-scale discriminators with 3 scales.

- Discriminator $D$: $d64$, $d128$, $d256$, $d512$

*https://github.com/jcjohnson/fast-neural-style

**Hyperparameters.** The batch size is set as 4 with a single 2080Ti GPU. Our total iteration is 360,000 for a total of around 200 epochs. We use Adam Optimization [18] with $\beta_1 = 0.5$, $\beta_2 = 0.999$ and Kaiming initialization [13]. The learning rate is set as 0.0001 with a 0.5 decay rate in every 100,000 iterations. The style codes from $fc$ has 64 dimension and the loss weights are set as: $\lambda_1 = 10$, $\lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = 1$.

## References


[1] Ellen Berscheid, Karen Dion, Elaine Walster, and G William Walster. Physical attractiveness and dating choice: A test of the matching hypothesis. *Journal of experimental social psychology*, 7(2):173–189, 1971. 2

[2] Eileen Bradbury. The psychology of aesthetic plastic surgery. *Aesthetic plastic surgery*, 18(3):301–305, 1994. 2

[3] Ray Bull and Nichola Rumsey. *The social psychology of facial appearance*. Springer Science & Business Media, 2012. 1, 2

[4] Huiwen Chang, Jingwan Lu, Fisher Yu, and Adam Finkelstein. Pairedcyclegan: Asymmetric style transfer for applying and removing makeup. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 40–48, 2018. 2

[5] Hung-Jen Chen, Ka-Ming Hui, Szu-Yu Wang, Li-Wu Tsao, Hong-Han Shuai, and Wen-Huang Cheng. Beautyglow: On-demand makeup transfer framework with reversible generative network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 10042–10050, 2019. 2, 8

[6] Michael R Cunningham. Measuring the physical in physical attractiveness: quasi-experiments on the sociobiology of female facial beauty. *Journal of personality and social psychology*, 50(5):925, 1986. 3

[7] Chris Donahue, Zachary C Lipton, Akshay Balsubramani, and Julian McAuley. Semantically decomposing the latent spaces of generative adversarial networks. *arXiv preprint arXiv:1705.07904*, 2017. 4

[8] Vincent Dumoulin, Jonathon Shlens, and Manjunath Kudlur. A learned representation for artistic style. *arXiv preprint arXiv:1610.07629*, 2016. 5

[9] Yael Eisenthal, Gideon Dror, and Eytan Ruppin. Facial attractiveness: Beauty and the machine. *Neural Computation*, 18(1):119–142, 2006. 3

[10] Yang-Yu Fan, Shu Liu, Bo Li, Zhe Guo, Ashok Samal, Jun Wan, and Stan Z Li. Label distribution-based facial attractiveness computation by deep residual learning. *IEEE Transactions on Multimedia*, 20(8):2196–2208, 2017. 3

[11] Junying Gan, Lichen Li, Yikui Zhai, and Yinhua Liu. Deep self-taught learning for facial beauty prediction. *Neurocomputing*, 144:295–303, 2014. 3

[12] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014. 5

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015. 10

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 5

[15] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1501–1510, 2017. 4, 5, 10

[16] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal unsupervised image-to-image translation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 172–189, 2018. 2, 4, 5, 6, 7, 10

[17] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4401–4410, 2019. 2, 3, 5

[18] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 10

[19] Hsin-Ying Lee, Hung-Yu Tseng, Jia-Bin Huang, Maneesh Singh, and Ming-Hsuan Yang. Diverse image-to-image translation via disentangled representations. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 35–51, 2018. 4, 7

[20] Hsin-Ying Lee, Hung-Yu Tseng, Jia-Bin Huang, Maneesh Kumar Singh, and Ming-Hsuan Yang. Diverse image-to-image translation via disentangled representations. In *European Conference on Computer Vision*, 2018. 2

[21] 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. In *2018 ACM Multimedia Conference on Multimedia Conference*, pages 645–653. ACM, 2018. 2, 8

[22] Lingyu Liang, Luojun Lin, Lianwen Jin, Duorui Xie, and Mengru Li. Scut-fbp5500: A diverse benchmark dataset for multi-paradigm facial beauty prediction. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 1598–1603. IEEE, 2018. 5, 6

[23] Anthony C Little, Benedict C Jones, and Lisa M DeBruine. Facial attractiveness: evolutionary based research. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1571):1638–1659, 2011. 1, 2

[24] Xudong Liu, Tao Li, Hao Peng, Iris Chuoying Ouyang, Taehwan Kim, and Ruizhe Wang. Understanding beauty via deep facial features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. 2, 5

[25] 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)*, 12 2015. 5

[26] Liqian Ma, Xu Jia, Stamatios Georgoulis, Tinne Tuytelaars, and Luc Van Gool. Exemplar guided unsupervised image-to-image translation with semantic consistency. *arXiv preprint arXiv:1805.11145*, 2018. 5

[27] Frances Cooke Macgregor. Social, psychological and cultural dimensions of cosmetic and reconstructive plastic surgery. *Aesthetic plastic surgery*, 13(1):1–8, 1989. 2

[28] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2794–2802, 2017. 4, 6

[29] Michael F Mathieu, Junbo Jake Zhao, Junbo Zhao, Aditya Ramesh, Pablo Sprechmann, and Yann LeCun. Disentangling factors of variation in deep representation using adversarial training. In *Advances in Neural Information Processing Systems*, pages 5040–5048, 2016. 4

[30] David I Perrett, D Michael Burt, Ian S Penton-Voak, Kieran J Lee, Duncan A Rowland, and Rachel Edwards. Symmetry and human facial attractiveness. *Evolution and human behavior*, 20(5):295–307, 1999. 3

[31] David I Perrett, Kieran J Lee, Ian Penton-Voak, D Rowland, Sakiko Yoshikawa, D Michael Burt, SP Henzi, Duncan L Castles, and Shigeru Akamatsu. Effects of sexual dimorphism on facial attractiveness. *Nature*, 394(6696):884, 1998. 3

[32] Katharine A Phillips, Susan L McElroy, Paul E Keck Jr, Harrison G Pope Jr, and James I Hudson. Body dysmorphic disorder: 30 cases of imagined ugliness. *The American Journal of Psychiatry*, 150(2):302, 1993. 2

[33] Marlene Rankin, Gregory L Borah, Arthur W Perry, and Philip D Wey. Quality-of-life outcomes after cosmetic surgery. *Plastic and reconstructive surgery*, 102(6):2139–45, 1998. 2

[34] Ronald E Riggio and Stanley B Woll. The role of nonverbal cues and physical attractiveness in the selection of dat-

ing partners. *Journal of Social and Personal Relationships*, 1(3):347–357, 1984. 1

[35] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 5

[36] Joshua B Tenenbaum and William T Freeman. Separating style and content with bilinear models. *Neural computation*, 12(6):1247–1283, 2000. 3

[37] Randy Thornhill and Steven W Gangestad. Facial attractiveness. *Trends in cognitive sciences*, 3(12):452–460, 1999. 3

[38] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. 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*, pages 6924–6932, 2017. 4, 5, 10

[39] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8798–8807, 2018. 6

[40] Zongwei Wang, Xu Tang, Weixin Luo, and Shenghua Gao. Face aging with identity-preserved conditional generative adversarial networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 4

[41] Xiang Wu, Ran He, Zhenan Sun, and Tieniu Tan. A light cnn for deep face representation with noisy labels. *IEEE Transactions on Information Forensics and Security*, 13(11):2884–2896, 2018. 4, 5, 6

[42] Duorui Xie, Lingyu Liang, Lianwen Jin, Jie Xu, and Mengru Li. Scut-fbp: A benchmark dataset for facial beauty perception. In *2015 IEEE International Conference on Systems, Man, and Cybernetics*, pages 1821–1826. IEEE, 2015. 2, 3

[43] Jie Xu, Lianwen Jin, Lingyu Liang, Ziyong Feng, Duorui Xie, and Huiyun Mao. Facial attractiveness prediction using psychologically inspired convolutional neural network (pi-cnn). In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1657–1661. IEEE, 2017. 3

[44] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2223–2232, 2017. 2, 7