

低分辨率图像的妆容迁移算法^①

林 凯, 陈壹华, 余松森

(华南师范大学 软件学院, 佛山 528225)

通信作者: 陈壹华, E-mail: YihuaChen1314@163.com



摘 要: 现有的妆容迁移算法效果优越, 功能丰富, 但是较少考虑到输入图像为低分辨率的场景. 当高分辨率图像难以获得时, 现有的妆容迁移算法将难以适用, 妆容无法完全迁移. 为此本文提出了一种适用于低分辨率图像的妆容迁移算法, 将包含妆容信息的特征矩阵作为先验信息, 将超分辨率网络与妆容迁移网络结合在一起产生协同效应, 即使输入的图像为低分辨率图像, 也能输出高分辨率的妆容迁移结果, 并且充分保留妆容细节的同时提升姿势和表情的鲁棒性. 由于使用端到端的模型实现妆容迁移与超分辨率, 因此设计了一组联合损失函数, 包括生成对抗损失、感知损失、循环一致性损失、妆容损失和均方误差损失函数. 所提出的模型在妆容迁移与超分辨率的定性实验和定量实验中均达到了先进水平.

关键词: 深度学习; 妆容迁移; 注意力机制; 超分辨率; 生成对抗网络

引用格式: 林凯, 陈壹华, 余松森. 低分辨率图像的妆容迁移算法. 计算机系统应用, 2023, 32(8): 54-66. <http://www.c-s-a.org.cn/1003-3254/9184.html>

Makeup Transfer Algorithm for Low-resolution Images

LIN Kai, CHEN Yi-Hua, YU Song-Sen

(School of Software, South China Normal University, Foshan 528225, China)

Abstract: The existing makeup transfer algorithms are highly effective with rich features, but they seldom take into account the scenarios of the low-resolution input images. When high-resolution images are difficult to obtain, it will be difficult for the existing makeup transfer algorithms to apply and the makeup cannot be fully transferred. In this study, a makeup transfer algorithm applied to low-resolution images is proposed, which uses the feature matrix containing makeup information as prior information and combines the super-resolution network with the makeup transfer network to produce the synergistic effect. The high-resolution makeup transfer results can be delivered even if the input image is a low-resolution one, and the robustness of postures and expressions is improved while the makeup details are fully retained. Since an end-to-end model is adopted to achieve the makeup transfer and super-resolution, a set of joint loss functions are designed, including generative adversarial loss, perceptual loss, cycle consistency loss, makeup loss, and mean square error loss functions. The proposed model attains an advanced level in both qualitative and quantitative experiments on makeup transfer and super-resolution.

Key words: deep learning; makeup transfer; attention mechanism; super-resolution; generative adversarial network (GAN)

妆容迁移算法的目的是将参考图像的妆容迁移到源图像上, 同时保留源图像的身份特征. 它是一项具有

挑战性的任务, 首先需要将妆容从参考图像中提取出来, 再分析源图像的人脸结构, 将其与妆容进行有效地

① 基金项目: 广东省基础与应用基础基金区域联合基金 (2020B1515120089); 广东省高校人工智能重点领域专项基金 (2019KZDZX1033); 广东省佛山市技术创新基金 (2016AG100472)

收稿时间: 2023-01-17; 修改时间: 2023-02-13; 采用时间: 2023-03-03; csa 在线出版时间: 2023-05-22

CNKI 网络首发时间: 2023-05-24

结合,同时应考虑角度、表情、遮挡和阴影等在实际应用中不可控的问题.近年来,随着美妆行业的快速发展,妆容迁移算法变得越来越流行,最主要的应用在于可视化试妆,解决人们挑选妆容的难题,并且丰富妆容的选择,推动相关行业的发展.另外,妆容迁移算法也可在娱乐行业大放异彩,如在短视频平台上制作相关特效,推出软件的新玩法,满足人们的精神文化需求.

传统的妆容迁移算法^[1-6]将图像分解为若干层,如面部结构、颜色和皮肤,然后逐层迁移. Tong 等人^[1]提出的方法包含4个步骤:预处理、妆容映射、外观矫正和眼妆迁移. Du 等人^[3]提出局部妆容迁移,而非基于整个面部的迁移,即先修改参考图像局部区域的亮度和色度通道,并使用图像混合算法将修改后的信息传递到源图像. Lee 等人^[6]使用高斯权重图进行妆容迁移.传统的妆容迁移算法框架和过程复杂,生成的图像真实性不足,对光照和姿势极其敏感,同时参考图像和源图像必须精确对齐,实用性大打折扣.

随着深度学习的广泛应用,出现了不少以生成对抗网络^[7]和 Transformer^[8]为基础的妆容迁移算法^[9-21],它们不仅降低了训练集的要求,还实现了传统的妆容迁移算法所无法实现的功能. Li 等人^[9]提出的 BeautyGAN 是拥有双输入和双输出的生成对抗网络,引入像素级别的直方图损失以实现实例级别的妆容迁移. Gu 等人^[12]提出的 LADN 利用多个重叠的局部对抗判别器解决以往妆容迁移算法无法处理戏剧性妆容的问题. Nguyen 等人^[16]提出的 CPM 使用改进的颜色迁移分支和全新的图案迁移分支来解决无法迁移妆容图案的问题. Deng 等人^[17]提出的 SCGAN 通过将妆容风格编码为组件风格码来解决面部空间错位问题. Wan 等人^[18]提出的 Spatial FAT 引入了 Transformer 模块以实现妆容迁移中高保真的色彩迁移以及精确的形状转换. Sun 等人^[19]提出的 SSAT 使用对称语义感知以实现妆容的迁移与去除. Lyu 等人^[20]提出的 SOGAN 利用 UV 空间中人脸的对称性来减轻妆容迁移时阴影和遮挡的影响.

但是以上方法并未考虑当参考图像和源图像分辨率不足的情况,当高分辨率图像难以获得时,生成的妆容迁移结果仅能迁移部分妆容^[22],这将大大影响算法的实际应用.在前人的相关研究中,一种主要的解决方法是通过将低分辨率的妆容迁移结果进一步做超分辨率,其将妆容迁移和超分辨率视为两个独立的过程^[23],然而此方法并未解决妆容丢失的问题;另外一种解决

方法是考虑如何在低分辨率图像上提取更多的信息以辅助妆容迁移^[22],但是此方法生成图像的质量和分辨率有限.

因此,本文引入了基于先验信息的人脸超分辨率算法.人脸超分辨率是指从低分辨率人脸图像中恢复高分辨率人脸图像的技术.常见的人脸先验信息有人脸关键点,指人脸关键区域的位置;人脸热图,指各点成为人脸关键点的概率热力图;人脸解析图,指人脸图像的语义分割图.这些先验信息提供了人脸五官的位置和人脸结构的属性,利用这些信息可以指导人脸超分辨率网络生成更加合理的结果. Song 等人^[24]和 Jiang 等人^[25]从低分辨率人脸图像中提取人脸关键点,并将人脸裁剪成不同的部分,估计不同部分的高频细节. Bulat 等人^[26]和 Kim 等人^[27]先将低分辨率人脸图像恢复成超分辨率图像后,再建立先验估计网络提取超分辨率图像和高分辨率图像的人脸热图,最后将这两个热图作接近约束. Chen 等人^[28]和 Kalarot 等人^[29]则预先训练一个面部结构先验提取网络以提取人脸解析图,并将其与低分辨率图像进行拼接,作为超分辨率网络的输入.本文的方法也采取类似的步骤.以上现存的研究方法仅考虑常规场景下的人脸超分辨率,并未涉及妆容迁移场景下的人脸超分辨率,因此本文另辟蹊径,将提取出的妆容信息也作为先验信息之一,指导超分辨率过程中妆容的完整生成.

基于以上研究现状,本文将现有的妆容迁移网络 PSGAN^[15]和简单的超分辨率网络进行结合,提出一个基于生成对抗网络的端到端算法,充分利用妆容迁移网络中的妆容信息,将其作为超分辨率网络的先验知识之一,并设计一组联合损失函数,同时实现妆容迁移与超分辨率,生成高分辨率且妆容完整的迁移结果.

1 整体框架

本文假设 X 和 Y 分别表示源图像域和参考图像域,并进一步利用 $\{x^n\}_{n=1,\dots,N}, x^n \in X$ 表示源图像域的示例, $\{y^m\}_{m=1,\dots,M}, y^m \in Y$ 表示参考图像域的示例. 本文的训练数据集来源于公开的 MT 妆容数据集^[9],其不需要成对的数据集,即允许参考图像和源图像有不同的身份特征.因此,模型需学习的函数可以通过如下公式表示:

$$G: \{x_{LR}, y_{LR}\} \rightarrow \tilde{x}_{SR} \quad (1)$$

其中, x_{LR} 表示低分辨率的源图像, y_{LR} 表示低分辨率的

参考图像, \tilde{x}_{SR} 表示超分辨率妆容迁移图像, 即 \tilde{x}_{SR} 拥有 y_{LR} 的妆容风格的同时还保持了 x_{LR} 的身份特征, 并且拥有比 x_{LR} 和 y_{LR} 更高的分辨率. 在模型中, 分辨率放大的倍数为4.

模型包含5个部分, 分别是妆容提取模块、妆容变形模块、妆容应用模块、先验估计模块以及超分辨率

率模块. 前3个模块参考自PSGAN. 模型的网络结构如图1所示, 其中“Conv”表示卷积层(convolutional layer), “IN”表示实例正则化层(instance normalization), “Residual block”表示残差块, “HG block”表示沙漏块(hour-glass block), “DeConv”表示反卷积层(deconvolutional layer).

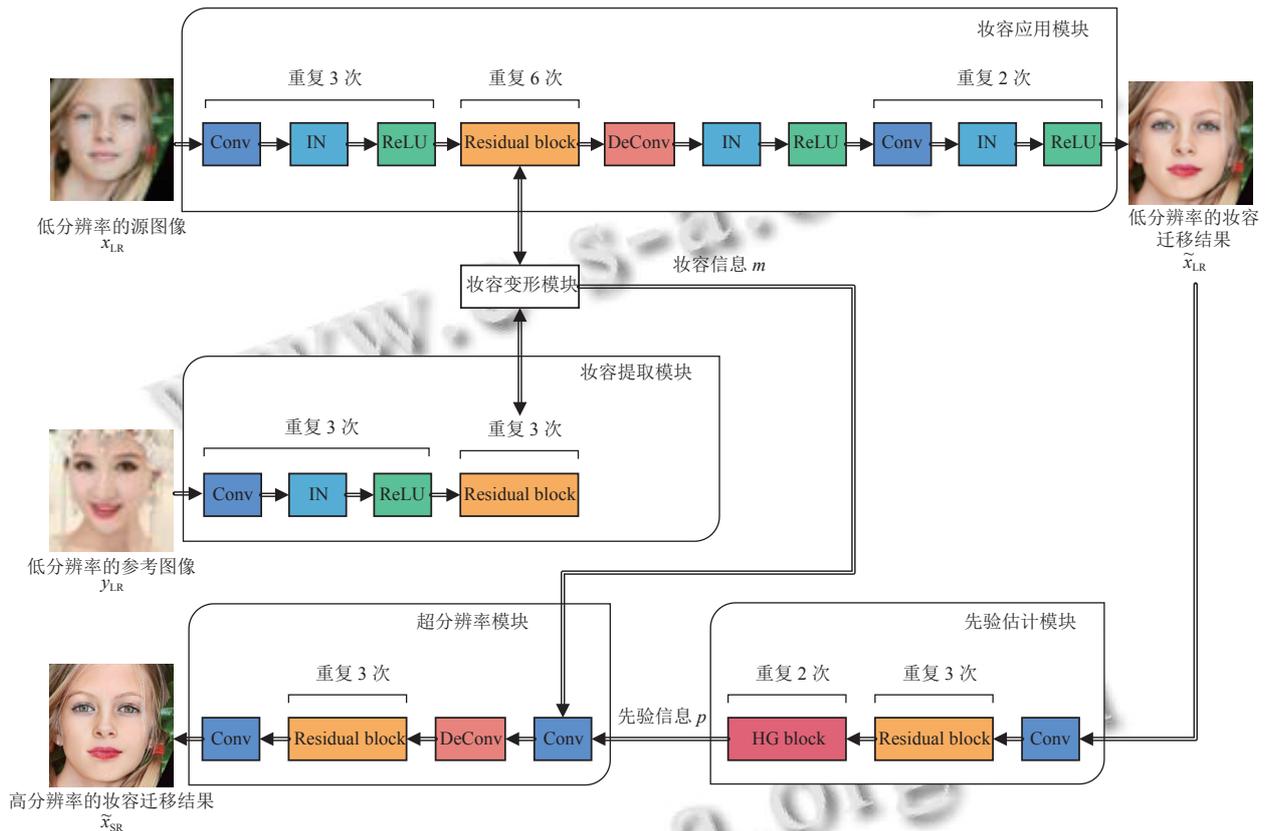


图1 模型的网络结构

模型的基本流程如下: 一方面, 将 y_{LR} 输入到妆容提取模块中, 将其妆容特征与身份特征分离出来, 并将分离出来的妆容特征输入到妆容变形模块中. 另一方面, 将 x_{LR} 输入到妆容应用模块中, 并在妆容变形模块中利用注意力机制对 y_{LR} 的妆容特征进行变形以适应 x_{LR} 的姿势和表情. 接着将变形后的妆容特征输入到妆容应用模块后续的解码器部分生成初步的低分辨率的妆容迁移结果 \tilde{x}_{LR} . 然后通过先验估计模块提取出 \tilde{x}_{LR} 的先验特征 p , 再与提取自妆容变形模块的妆容特征 m 沿特征通道相连接, 强化超分辨率过程中的妆容表达, 最后共同输入到超分辨率模块中生成最终的超分辨率妆容迁移结果 \tilde{x}_{SR} .

1.1 妆容提取模块

妆容提取模块从参考图像 y_{LR} 中提取出妆容风格, 即通过两个 1×1 的卷积层将妆容提取网络的输出特征图 $V_y \in \mathbb{R}^{C \times H \times W}$ 分解为两个具有空间感知性的妆容系数矩阵 $\gamma \in \mathbb{R}^{1 \times H \times W}$ 和偏差矩阵 $\beta \in \mathbb{R}^{1 \times H \times W}$, 用于迁移像素级别的妆容, 实现灵活的局部妆容迁移以及妆容浓度可控的妆容迁移, 其中 C 、 H 和 W 表示特征图的通道数量、高度和宽度. 妆容提取模块使用编码器—瓶颈架构^[30].

1.2 妆容变形模块

由于源图像 x_{LR} 和参考图像 y_{LR} 可能具有不同的姿势和表情, 所以妆容提取模块生成的 γ 和 β 不能直接应

最后, 将 $\gamma' \in \mathbb{R}^{1 \times H \times W}$ 和 $\beta' \in \mathbb{R}^{1 \times H \times W}$ 沿着特征通道维度复制成 $\Gamma' \in \mathbb{R}^{C \times H \times W}$ 和 $B' \in \mathbb{R}^{C \times H \times W}$ 作为妆容应用模块的输入。

1.3 妆容应用模块

妆容应用模块的作用是生成初步的低分辨率妆容迁移结果 \tilde{x}_{LR} , 其网络结构为编码器—瓶颈—解码器架构^[30]. 在瓶颈部分, 将 AMM 生成的 Γ' 和 B' 应用于源图像 x_{LR} 的特征图 $V_X \in \mathbb{R}^{C \times H \times W}$ 生成 V'_X , 公式如下所示:

$$V'_X = \Gamma' V_X + B' \quad (6)$$

最后将 V'_X 输入到后续的解码器部分生成 \tilde{x}_{LR} , 以作为先验估计模块的输入。

1.4 先验估计模块

由于 \tilde{x}_{LR} 的分辨率过低, 妆容细节不足, 先验估计模块将进一步提取其先验信息 p 为后面超分辨率做准备。

先验信息在传统的超分辨率网络中起着重要的作用, 特别是人脸关键点和解析图. 人脸关键点信息包含了人脸轮廓与五官的关键点及位置. 而人脸解析图是人脸图像的语义分割图, 将眼睛、鼻子、嘴唇、皮肤、耳朵和头发等从整张图像中分割出来。

先验估计模块使用沙漏块 (hourglass block, HG block) 估计人脸关键点和解析图. 沙漏块结构曾用于人体姿势估计^[31,32], 其特点在于使用不同尺度的特征去捕捉空间信息, 使网络在训练时更加关注图像的细节. n 阶沙漏块提取了从原始尺度到 $\frac{1}{2^n}$ 尺度的特征. 模型中使用的是四阶沙漏块。

在四阶沙漏块中, 每次下采样之前, 其会分出一个分支路径保留原尺度的信息; 每次上采样之后, 会和之前的分支路径中同样大小的数据相加; 两次下采样之间, 使用 3 个残差块提取特征; 两次上采样之间使用一个残差块提取特征. 在使用沙漏块的前后, 特征图的大小并无发生变化. 通过使用沙漏块, 先验估计模块可以充分提取不同粒度的特征以准确生成先验特征 p , 为后续生成 \tilde{x}_{SR} 做好铺垫。

先验估计模块的结构如图 1 所示. 在 \tilde{x}_{LR} 输入沙漏块之前先使用卷积层和残差块对其进行预处理. 由于人脸关键点和解析图均为人脸二维形状信息, 所以在先验估计模块中这两个任务的权重除了最后一层都是共享的. 在最后一层, 使用两个独立的 1×1 卷积层来分别生成估计的人脸关键点和解析图。

1.5 超分辨率模块

模型同时输入先验估计模块生成的先验特征 p 以及妆容信息 m 至超分辨率模块中, 输出最终的结果 \tilde{x}_{SR} , 其中 p 与 m 通过特征图的通道串联在一起作为联合先验特征。

m 是妆容应用模块第 4 个瓶颈与妆容变形模块的输出进行运算的结果, 因此引入 m 不仅可以解决超分辨率时妆容丢失的问题, 还可以提高人脸超分辨率过程中姿势和表情的鲁棒性, 其适合作为联合先验信息的先验信息之一。

超分辨率模块的结构如图 1 所示. 首先将联合先验特征输入到一个 3×3 的卷积层进行下采样, 再输入到一个 4×4 的反卷积层进行上采样, 最后使用 3 个残差块和 1 个卷积层生成超分辨率妆容迁移结果 \tilde{x}_{SR} 。

1.6 判别器模块

本文所提出的方法是基于生成对抗网络的, 其包含了生成器和判别器. 生成器会尽量去学习真实数据的分布, 而判别器则会尽量分辨出输入数据的真假以反馈给生成器使之生成更加合理的结果. 前文所介绍的模块均为生成器部分. 在 PSGAN 中, 虽然本身已存在一个判别器, 但它适用于判别妆容迁移图像的真假, 而无法对超分辨率图像的真假进行判别, 因此本文模型又引入了一个新的超分辨率图像判别器以指导生成器生成理想的超分辨率结果。

由于人脸妆容图像包含许多细节, 传统的判别器是不合适的. 因此模型引入了 PatchGAN^[33]. PatchGAN 和普通的 GAN 判别器是有区别的, 普通的 GAN 判别器是将输入映射成一个实数, 即输入样本为真样本的概率. 而 PatchGAN 使用的是全卷积网络. 它将输入映射为一个 $N \times N$ 的矩阵, 即卷积层输出的特征图, 矩阵中的每个值代表每个块为真样本的概率. 从这个特征图可以追溯到原始图像中的某一个位置, 这个位置称为块, 可以看出这个块对最终输出结果的影响. 最后将矩阵中每个值的结果求均值, 即为 PatchGAN 最终输出. PatchGAN 的感受域对应于与输入中的一小块区域, 即矩阵中的每个值对应了 PatchGAN 对输入图像的一小块的判别输出, 这样训练使模型更能关注图像细节。

2 联合损失函数

2.1 生成对抗损失函数

在生成对抗网络中, 通过训练令生成器和判别器

不断博弈,使生成器输出更加真实的图像,同时使判别器更加准确鉴别图像的真假.具体而言,模型使用 G_M 生成 \tilde{x}_{LR} ,并为源图像域 X 与参考图像域 Y 训练两个判别器 D_X 和 D_Y 以分别区分两个域的训练集中的真实图像与 G_M 生成的虚假图像.当判别器的输入为训练集中的真实图像时,期望输出的结果为1,当判别器的输入为生成器生成的虚假图像时,则期望输出的结果为0.因此判别器的对抗损失 $L_{D_M}^{adv}$ 如下所示:

$$\begin{aligned} L_{D_M}^{adv} = & -E_{x_{LR}}[\log D_X(x_{LR})] - E_{y_{LR}}[\log D_Y(y_{LR})] \\ & - E_{x_{LR}, y_{LR}}[\log(1 - D_X(G_M(y_{LR}, x_{LR})))] \\ & - E_{x_{LR}, y_{LR}}[\log(1 - D_Y(G_M(x_{LR}, y_{LR})))] \quad (7) \end{aligned}$$

而生成器 G_M 的对抗损失 $L_{G_M}^{adv}$ 的目的是生成以假乱真的图像以欺骗判别器 D_X 和 D_Y ,因此当生成器生成的虚假图像输入进判别器时,其期望的结果为1,对应的公式如下所示:

$$\begin{aligned} L_{G_M}^{adv} = & -E_{x_{LR}, y_{LR}}[\log(D_X(G_M(y_{LR}, x_{LR})))] \\ & - E_{x_{LR}, y_{LR}}[\log(D_Y(G_M(x_{LR}, y_{LR})))] \quad (8) \end{aligned}$$

另外,模型进一步使用 G_S 生成 \tilde{x}_{SR} ,且使用判别器 D_S 以区分真实的 \tilde{x}_{HR} 图像与虚假的 \tilde{x}_{SR} 图像,原理同上,其对应的判别器损失 $L_{D_S}^{adv}$ 如下所示:

$$L_{D_S}^{adv} = -E_{\tilde{x}_{HR}}[\log D_S(\tilde{x}_{HR})] \quad (9)$$

相应的,生成器 G_S 的损失如下所示:

$$L_{G_S}^{adv} = E_{\tilde{x}_{LR}}[\log(1 - D_S(G_S(\tilde{x}_{LR}, m)))] \quad (10)$$

其中, m 是对应的包含妆容信息的特征矩阵.

2.2 感知损失函数

在模型中无论是生成的是 \tilde{x}_{LR} 还是 \tilde{x}_{SR} ,其图像中的身份特征均是不变的,应与 x_{LR} 是一致的,这需要通过损失函数对此进行约束.因此在对抗损失的基础上引入感知损失,这是为了评估视觉感知相关的特征而引入的损失函数.视觉感知相关的特征不使用像素级别的特征,而是使用神经网络的深层特征.具体而言,模型通过两张图像在ImageNet上预训练的VGG-16模型^[34]中的深层特征之间的欧几里得距离进行约束,并使用L2损失衡量二者之间的差异.妆容迁移感知损失函数 $L_{G_M}^{per}$ 和超分辨率感知损失函数 $L_{G_S}^{per}$ 如下所示:

$$\begin{aligned} L_{G_M}^{per} = & E_{x_{LR}, y_{LR}} \|F_l(G_M(x_{LR}, y_{LR})) - F_l(x_{LR})\|_2 \\ & + E_{x_{LR}, y_{LR}} \|F_l(G_M(y_{LR}, x_{LR})) - F_l(y_{LR})\|_2 \quad (11) \end{aligned}$$

$$L_{G_S}^{per} = \|F_l(G_S(\tilde{x}_{LR}, m)) - F_l(\tilde{x}_{HR})\|_2 \quad (12)$$

其中, F_l 表示VGG-16模型的第 l 层输出.

2.3 循环一致性损失函数

由于训练集缺少三元组数据,即源图像、参考图像和对应的妆容迁移图像组,因此本文使用无监督的方式进行训练.模型引入了CycleGAN^[35]中所提出的循环一致性损失函数,即当输出的图像对再输入到生成器中后,应该输出与原始输入图像接近的图像对,该循环过程如下所示:

$$\begin{aligned} (x_{LR}, y_{LR}) \rightarrow & G_M(x_{LR}, y_{LR}) \rightarrow G_M(G_M(x_{LR}, y_{LR})) \\ \approx & (x_{LR}, y_{LR}) \quad (13) \end{aligned}$$

基于此,针对妆容迁移场景可进一步描述为迁移其他人妆容后,再使用原本自身的妆容进行迁移,生成的图像无论是身份还是妆容均应与最初的一致,因此,提出循环一致性损失函数 $L_{G_M}^{cyc}$ 以约束重建图像,具体如下所示:

$$\begin{aligned} L_{G_M}^{cyc} = & E_{x_{LR}, y_{LR}} \|G_M(G_M(x_{LR}, y_{LR}), x_{LR}) - x_{LR}\|_1 \\ & + E_{x_{LR}, y_{LR}} \|G_M(G_M(y_{LR}, x_{LR}), y_{LR}) - y_{LR}\|_1 \quad (14) \end{aligned}$$

2.4 妆容损失函数

为了实现实例级别的妆容迁移,必须对妆容一致性进行约束,因此模型进一步引入了BeautyGAN^[9]中所提出的妆容损失.对源图像进行化妆可以认为是对源图像特定区域的颜色进行重新分布,即无论是口红、眼影还是粉底,在计算机视觉中化妆过程均可以理解为颜色的变化,因此可以引入直方图匹配以提出像素级别的直方图损失函数 $L_{G_M}^{make}$.

具体而言,模型分别对 x_{LR} 和 y_{LR} 的相同人脸部位执行直方图匹配,然后重新组合结果,将其表示为 $HM(x_{LR}, y_{LR})$. $HM(x_{LR}, y_{LR})$ 作为伪数据,其不仅有与 x_{LR} 相同的人脸部位轮廓,又有与 y_{LR} 相同的颜色分布.妆容损失并不使用整个图像的直方图损失进行计算,而是选择3个重要的化妆部位^[36]应用局部的直方图损失,分别是口红、眼影和粉底.这是因为并不是所有的部位都需要化妆,即需要相同的颜色分布,比如头发或者背景等,这些与妆容是无关的,需要将其分开来.另外,妆容不仅体现在全局,它也是由若干个局部妆容组合在一起的.因此,妆容损失使用语义分割网络BiSeNet^[37]生成对应的人脸解析图,将人脸指导掩码表示为 $M = FP(x)$.对于每个输入图像 x ,预训练的BiSeNet将生成表示几个面部位置的人脸指导掩码 M ,包括嘴唇、眼睛、面部皮肤、头发和背景等.最后,寻找对应嘴唇、眼睛和

面部皮肤的掩码,并对这3个标签生成二进制掩码,分别表示成 M_{lip} 、 M_{eye} 和 M_{face} 。除此之外,眼影并没有在 M 上进行标记,因为化妆前的人脸图像并不存在眼影。但是眼影的掩码 M_{shadow} 是必需的,因此根据 M_{eye} 计算出包裹住眼睛的两个矩形区域,然后排除眼睛区域以及部分头发和眉毛区域,将其作为眼影的掩码 M_{shadow} 。妆容损失函数 $L_{G_M}^{make}$ 如下所示:

$$L_{G_M}^{make} = E_{x_{LR}, y_{LR}} \|G_M(x_{LR}, y_{LR}) - HM(x_{LR} \circ M_{item}^1, y_{LR} \circ M_{item}^2)\|_2 \quad (15)$$

$$M^1 = FP(x_{LR}) \quad (16)$$

$$M^2 = FP(y_{LR}) \quad (17)$$

其中, M_{item} 包括 M_{lip} 、 M_{shadow} 和 M_{face} , \circ 运算符表示元素的智能乘积。

2.5 均方误差损失函数

模型也引入了均方误差损失在像素级别上衡量两个图像之间的距离。假设给定的训练集为 $\{\tilde{x}_{LR}^n, m^n, \tilde{x}_{HR}^n, p_{gt}^n\}_{n=1}^N$, 其中 N 是训练集的大小, \tilde{x}_{HR}^n 是低分辨率图像 \tilde{x}_{LR}^n 的高分辨率的真实图像, m^n 是对应于 \tilde{x}_{LR}^n 的包含妆容信息的特征矩阵, p_{gt}^n 是真实的人脸先验信息, 因此均方误差损失函数如下所示:

$$L_{G_S}^{MSE} = \frac{1}{2N} \sum_{n=1}^N \left[\|\tilde{x}_{HR}^n - G_S(\tilde{x}_{LR}^n, m^n)\|^2 + \lambda_p \|p_{gt}^n - p^n\| \right] \quad (18)$$

其中, p^n 是第 i 个图像估计的先验信息, λ_p 是估计的先验信息的权重。

2.6 联合损失

通过整合本节上述损失函数, 判别器损失 L_D 和生成器损失 L_G 可以表示为:

$$L_D = \lambda_M^{adv} L_{D_M}^{adv} + \lambda_S^{adv} L_{D_S}^{adv} \quad (19)$$

$$L_G = \lambda_M^{adv} L_{D_M}^{adv} + \lambda_S^{adv} L_{D_S}^{adv} + \lambda_{G_M}^{per} L_{G_M}^{per} + \lambda_{G_S}^{per} L_{G_S}^{per} + \lambda_{G_M}^{cyc} L_{G_M}^{cyc} + \lambda_{G_M}^{make} L_{G_M}^{make} + \lambda_{G_S}^{MSE} L_{G_S}^{MSE} \quad (20)$$

其中, λ_M^{adv} 、 λ_S^{adv} 、 $\lambda_{G_M}^{per}$ 、 $\lambda_{G_S}^{per}$ 、 $\lambda_{G_M}^{cyc}$ 、 $\lambda_{G_M}^{make}$ 和 $\lambda_{G_S}^{MSE}$ 分别是不同损失项的权重。

3 实验结果与分析

3.1 数据集预处理

模型的实验涉及妆容迁移领域和超分辨率领域, 因此本文对MT数据集^[9]做进一步处理以更适合模型

进行训练。MT数据集为当前最大的妆容数据集之一, 数据集中有3834张分辨率大小为 361×361 的女性图像, 其中包括1115张素颜图像和2719张妆容图像。它包含各种各样的种族、姿势、表情和杂乱的背景。另外, 它的妆容图像中也包含多种类型的妆容, 包括烟熏眼妆、浮华妆容、复古妆容、韩式妆容和日式妆容等, 且淡妆到浓妆均有涉及。MT数据集中也有一些裸妆图像, 为了方便起见, 将它们划分到了素颜图像集里。本文随机选择素颜图像 x_{HR} 和妆容图像 y_{HR} 使用PSGAN^[15]进行妆容迁移, 并将迁移结果通过双三次插值将其分辨率大小调整为 256×256 , 以作为训练集中的真实输出值 \tilde{x}_{HR} 。接着, 将其对应的素颜图像 x_{LR} 和妆容图像 y_{LR} 的分辨率大小通过双三次插值调整为 64×64 作为训练集中的输入值。另外, 将 x_{LR} 和 y_{LR} 输入到PSGAN中输出妆容迁移结果 \tilde{x}_{LR} 后, 再将其输入到BiSeNet^[37]中输出人脸解析图的真实值 $\tilde{x}_{LR_parsing}$, 并使用百度AI开放平台的API生成201个人脸关键点的真实值 $\tilde{x}_{LR_keypoint}$, 最终形成一个五元数据集 $\{x_{LR}, y_{LR}, x_{LR_parsing}, \tilde{x}_{LR_keypoint}, \tilde{x}_{HR}\}$ 。

本文训练集包含5000对五元数据集, 测试集包含100对五元数据集。

3.2 实验设置

模型实现的是 $\times 4$ 的超分辨率, 因此在将分辨率为 64×64 的源图像和妆容图像输入到网络中之前, 先通过双三次插值将它们的分辨率放大到 256×256 后再进行后向传播。本文使用RMSprop算法训练, 初始的学习率为 2.5×10^{-4} , 训练的批大小为1。损失函数中权重的设置为: $\lambda_p = 1$ 、 $\lambda_M^{adv} = 10^{-3}$ 、 $\lambda_S^{adv} = 10^{-3}$ 、 $\lambda_{G_M}^{per} = 10^{-5}$ 、 $\lambda_{G_S}^{per} = 10^{-1}$ 、 $\lambda_{G_M}^{cyc} = 10^{-2}$ 、 $\lambda_{G_M}^{make} = 10^{-3}$ 和 $\lambda_{G_S}^{MSE} = 1$ 。本文使用VGG-16模型中的relu5_3层计算感知损失函数。模型在NVIDIA Tesla P100 GPU上训练了200个批次, 花费了120 h。后文将针对模型的妆容迁移与超分辨率效果分别进行实验。

3.3 模型结果展示

图3展示了模型的超分辨率妆容迁移结果。图中分别展示了姿势和表情差别小和姿势表情差别大的情况下的生成的结果, 每种情况的第1列表示 64×64 的源图像, 第2列表示 64×64 的参考图像, 第3列表示 256×256 的妆容迁移结果。从图3可以看到, 模型无论是在姿势和表情差别小或者大的情况下, 即使输入图像的分辨率较低, 也均能稳定输出生动且无伪影的高

分辨率结果. 本文将从妆容迁移的角度和超分辨率的角度, 通过定性实验、定量实验以及消融实验对结果

进行分析, 以证明框架的合理性, 并对原因进行理论上的剖析. 后文将详细介绍这些实验.



图3 模型结果展示

3.4 妆容迁移实验

妆容迁移实验中对比的基线模型不仅包括一般的图像翻译模型, 即 DIA^[38] 和 CycleGAN^[35], 还包括特定的妆容迁移模型, 即 BeautyGAN^[9]、PairedCycleGAN^[11]、BeautyGlow^[10]、LADN^[12]、PSGAN^[15]、Spatial FAT^[18] 和 SCGAN^[17]. 为了保证实验的公平性, 本文均选择使用官方提供的代码和模型进行对比.

图像翻译模型能将图像中的内容从一个图像域转换到另外一个图像域, 将原始图像中的某个属性转换成另外一个属性, 即图像间的跨域转换. 但是若图像翻译模型应用在妆容迁移任务上, 将无法完成实例级别的妆容迁移. 妆容迁移模型则可以实现实例级别的妆容迁移, 其可以看作是实现局部的图像翻译任务, 即应用于特定的化妆部位, 难度相对于图像翻译任务更大. 为了全面地对比效果, 这两种类别的模型均将与其进行比较.

评价方法包含两种: 其一是定性比较, 凭肉眼对不同模型生成的图像细节进行观察, 从多个视觉角度综合考虑不同模型的优劣; 其二是问卷调查, 通过问卷调查系统让众多的受试者去答题, 并根据受试者提交的结果进行统计分析以评价不同模型的优劣. 由于妆容迁移任务目前暂无统一公开的定量评价方法, 因此并未对此设计相关实验.

(1) 定性实验

图4展示了本文模型与其他模型正脸且无明显表情的妆容迁移的对比结果. 图4展示了两个示例, 每个示例的第1行表示的是除本文模型之外的输入和输出图像, 每个示例的第2行表示的是本文模型的输入和

输出图像, 本文模型输入的图像分辨率仅有 64×64 , 远不如其他模型输入图像的 256×256 的分辨率.

从图4中可以看到, 在第2个示例中 DIA 生成的结果中头发的颜色明显地发生了变化, 在妆容迁移时发色应是不变的, 这是图像翻译模型应用于妆容迁移时存在的缺陷之一, 即无法控制局部的变化. 同样在第2个示例中 CycleGAN 生成的结果中妆容并不与参考图像的妆容一致, 甚至唇妆直接丢失, 这是图像翻译模型另外的一个缺陷, 即无法迁移实例级别的妆容, 仅能在素颜域和妆容域之间转换. 在第1个示例中 BeautyGlow 生成的结果瞳孔颜色也受到了参考图像的影响, 这并不应该被迁移, 在第2个 BeautyGlow 的示例中也存在同样的问题, 这表明了 BeautyGlow 并没有较好地保留源图像的身份特征. LADN 生成结果的背景伪影较多, 已经无法还原源图像如头发等背景了. Spatial FAT 生成的结果虽然视觉细节丰富, 但是妆容的伪影较为明显, 其更像使用美颜后的结果, 而非真实上妆后的结果. SCGAN 生成的结果是令人满意的, 并没有前面其他模型存在的缺陷, 但是生成的结果却比源图像和参考图像更加模糊, 甚至比输入图像分辨率仅只有 64×64 的本文模型的结果还模糊, 这在重视视觉效果的实际应用场景中是不可接受的. 而本文模型生成的结果与 PSGAN 相似, 因为本文模型的训练数据集是来源于 PSGAN 的结果, 虽然输入图像分辨率较低, 但本文模型生成的结果仍旧生动逼真, 相比于 PSGAN 生成的结果在细节上仅有些许模糊, 妆容细节保留得较好, 与其他高分辨率输入的基线模型相比更胜一筹.

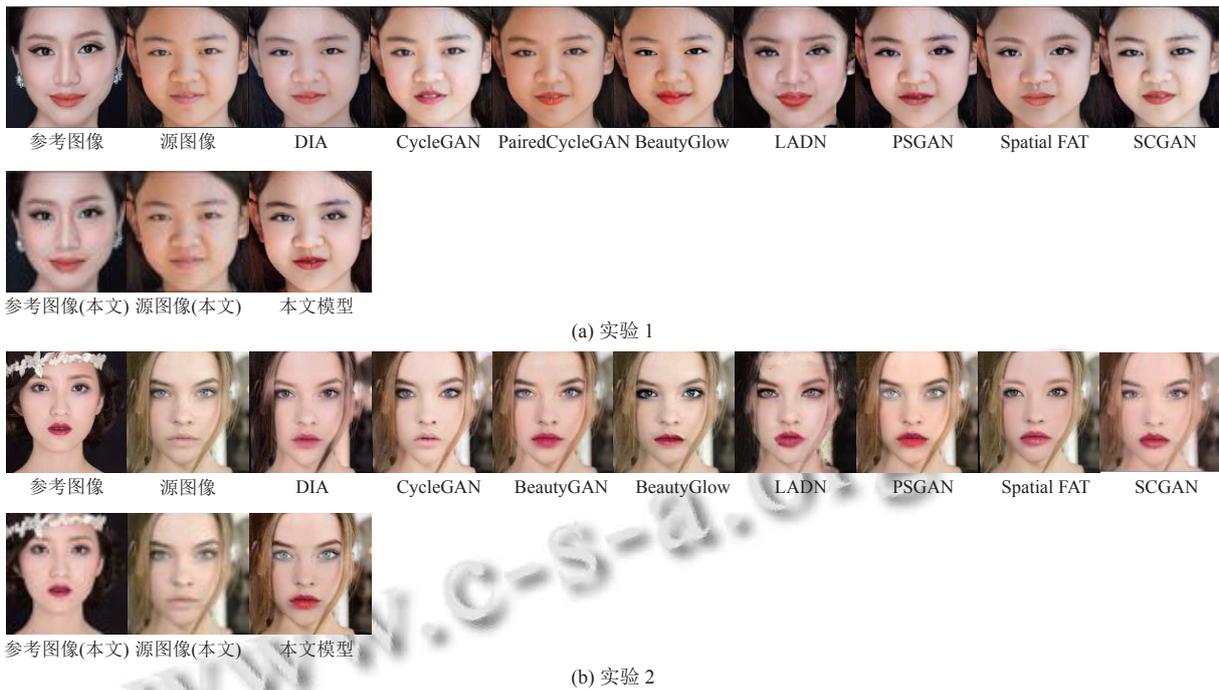


图4 妆容迁移稳定性实验

(2) 定量实验

作为面向化妆行业的应用场景, 市场上大众用户的满意度是衡量一个妆容迁移模型的重要标准, 因此, 本文设计了问卷调查实验, 以对妆容迁移效果进行定量评估。

妆容迁移定量实验要求受试者根据给定的题目中的素颜图像以及参考妆容图像, 以图像真实性和参考妆容相似性为评价标准, 选择选项中妆容迁移效果最好的图像, 选项中包含的图像有 BeautyGAN^[9]、DMT^[14]、LADN^[12]、Spatial FAT^[18] 和本文模型所生成的结果。本实验目的是观察本文模型生成的结果被选择的概率, 从而间接对比出本妆容迁移算法的效果。题目共有 20 道, 受试者共计 50 名。另外, 考虑到不同性别对于妆容的认知与审美不同, 因此将受试者细分为男性和女性, 并分别进行统计。为了方便用户判断, 题目中展示的素颜图像以及参考妆容图像的分辨率大小是 256×256, 并非本文模型的实际输入的分辨率大小 64×64。图 5 展示了妆容迁移定量实验的结果。

从图 5 中可以看出在 5 种模型中有 40% 左右的男性和女性受试者均认为本文模型的结果最优, 证明即使从妆容迁移算法中最主要的且对美感要求更高的女性用户的角度考虑, 也已经超越了其他 4 种模型。虽然本文模型的输入图像的分辨率较低, 但是仍然能在

传统的妆容迁移网络中拥有一定的优势, 为拓宽妆容迁移的应用场景提供了一个良好的参考, 也为化妆行业的发展新增了一定的动力。

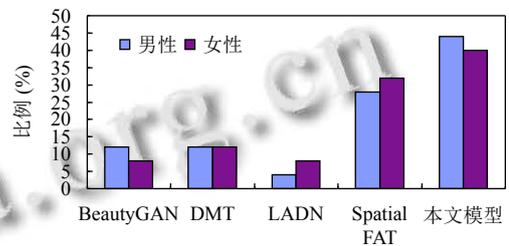


图5 妆容迁移定量实验

3.5 超分辨率实验

本文的超分辨率实验将本文模型生成的结果与多种模型进行对比, 包括人脸复原模型, 比如 FSRNet^[39]、HiFaceGAN^[40] 和 GPEN^[41], 还有 JPEG 格式去伪影模型, 比如 ARCNN^[42], 以及图像复原模型, 比如 DeblurgANv2^[43] 和 SRFBN^[44]。因为其他模型的输入是低分辨率的图像, 而不是妆容迁移模型输入的源图像和参考图像, 所以需制作新的训练数据集。本文将 64×64 的源图像和参考图像输入到 PSGAN 中, 输出的结果作为其他模型的输入, 将 256×256 的源图像和参考图像输入到 PSGAN 中, 输出的结果作为其他模型的真实输

出. 所有的模型均使用此数据集进行训练, 所使用的代码均来源于官方代码, 以保证实验的公平性.

评价方法包含两种: 其一是定性比较, 通过肉眼直接观察不同模型输出的图像, 判断其与真实输出之间的差距, 以此来判断不同模型的效果; 其二是定量实验, 本文使用常见的超分辨率评价指标, 如峰值信噪比 (peak signal to noise ratio, PSNR)、结构相似性 (structural similarity, SSIM)、平均结构相似性 (mean structural similarity, MSSIM)、学习感知图像块相似度 (learned perceptual image patch similarity, LPIPS)、弗雷切特起始距离 (Frechet inception distance, FID) 和自然图像质量评估器 (naturalness image quality evaluator, NIQE).

(1) 定性实验

图6展示了本文模型的超分辨率定性实验的结果. 图中对每种模型均展示了两个示例, 每行的第一张图像表示输入的低分辨率图像, 每行的最后一张图像表示HR图像, 即测试集中的真实GT图像.

从图6中可以看到, ARCNN和DeblurGANv2的结果虽然比输入图像更加清晰, 但是对于更加先进的

方法分辨率提升的还是不足. SRFBN的结果过于平滑, 无法恢复清晰的人脸器官和纹理细节. FSRNet使用本文的训练集训练之后输出的结果出现了明显的白点. FSRNet可能更加适用于 32×32 或者更低的分辨率, 对于更高的分辨率超分辨率的效果有限. 而HiFaceGAN输出的结果虽然分辨率更加高, 但是颗粒感过强, 视觉效果有限. GPEN的效果优异, 但是正由于其出色的性能, 导致结果对妆容做了进一步处理, 比如眼妆和唇妆的颜色更深, 与GT图像不符合, 这在妆容迁移中是难以接受的, 例如当用户只想使用淡妆为自己的人脸上妆, 但是妆容迁移的结果却是浓妆, 将会不符合用户的心理预期. 另外, GPEN输出的结果中瞳孔光线暗淡, 失去神采, 对于重视用户外貌的应用场景来说是个缺陷. 而本文模型输出的图像更加接近真实GT图像, 因为本文模型使用了妆容信息作为先验知识, 强化了妆容的表达, 恢复了绝大部分细节, 比如瞳孔和睫毛. 另外, 高频细节, 比如头发也很好地被恢复. 对于爱美的用户来说, 这些细节的质量是很重要的, 会影响她们对于该妆容的评判.



图6 超分辨率定性实验

(2) 定量实验

超分辨率定量实验的结果如表1所示. 从表1可以得知本文模型在PSNR、SSIM和MSSIM这3个指标上并没有取得最优的结果, 但是这些指标与人类的视觉感知相关程度较低, 仅从像素空间上呈现出相似度. 而本文模型在LPIPS这个与人类视觉感知相关程度较高的指标上取得了最优的结果, 这表明本文模型的输出图像在视觉感知上最接近真实的情况. 同时, 本

文模型的FID和NIQE也获得了最优的结果, 证明本文模型的输出图像分别与真实人脸图像和自然图像的分布最接近. 超分辨率定量实验表明了本文模型在超分辨率领域也能同其在妆容迁移领域一样取得先进的成果, 证实了妆容信息作为先验知识的强大性能.

3.6 消融实验

本文为了进一步验证包含妆容信息的妆容矩阵在超分辨率过程中发挥的作用, 设计了一个消融实验, 分

别从定性和定量的角度表明消融妆容矩阵后对本文模型结果产生的影响. 具体而言, 消融实验消融了妆容矩阵 γ' 和 β' , 即直接使用妆容应用模块第4个瓶颈的结果作为超分辨率模块的先验信息之一.

(1) 定性实验

图7展示了消融定性实验的结果. 图7每行均表

示1个示例, 共展示了3个示例. 第3列表示消融妆容矩阵 γ' 和 β' 后模型输出的结果, 第4列表示未消融妆容矩阵 γ' 和 β' 后模型输出的结果. 从图7可以看到在消融妆容矩阵后, 妆容细节变模糊, 颜色变浅, 妆感发生了变化. 在超分辨率过程期间, 由于缺乏妆容矩阵的指导导致了这些不符合预期的结果.

表1 超分辨率定量实验以及消融实验

模型	PSNR \uparrow	SSIM \uparrow	MSSIM \uparrow	LPIPS \downarrow	FID \downarrow	NIQE \downarrow
ARCNN	30.91	0.8986	0.8918	0.1581	131.25	48.762
DeblurGANv2	29.54	0.8779	0.8720	0.2383	156.16	43.854
SRFBN	32.81	0.9217	0.9179	0.1421	137.42	50.413
FSRNet	25.77	0.8655	0.8712	0.1786	134.49	52.761
HiFaceGAN	24.57	0.7874	0.7932	0.1868	131.28	51.499
本文模型 (消融妆容矩阵)	27.54	0.8771	0.8945	0.0892	124.90	38.044
本文模型	29.13	0.8851	0.9004	0.0828	123.16	37.969



(a) 参考图像 (b) 源图像 (c) 消融妆容矩阵 (d) 未消融妆容矩阵

图7 消融定性实验

(2) 定量实验

表1最后两行展示了消融定量实验的结果. 从表中可以看到, 仅使用两个简单的妆容矩阵 γ' 和 β' 即可使所有的超分辨率指标得到一定的提升, 从另外一个角度证明了妆容矩阵在超分辨率过程中所发挥的重要性.

4 结论与展望

由于传统的妆容迁移算法不适用于低分辨率的输入图像, 因此本文提出的模型以扩大妆容迁移的应用场景. 模型的输入虽然是低分辨率的图像, 但是输出的是拥有相对完整的妆容细节的高分辨率图像, 其核心是使用包含妆容信息的特征矩阵作为先验信息之一以强化超分辨率过程中的妆容表达. 在超分辨率过程中

不仅可引导完整妆容的迁移, 也可提升超分辨率时姿势和表情的鲁棒性. 为了实现端到端的超分辨率妆容迁移, 也设计了一组联合损失函数对过程进行约束. 本文在妆容迁移和超分辨率方面均设计了定性实验和定量实验, 证明了模型在两个领域的先进性. 最后, 通过消融实验证明了使用简单的妆容信息即可实现对超分辨率中妆容细节表达的优化, 证明了妆容信息的重要性以及有效性. 在未来的工作中, 考虑到模型由于受到训练集分辨率的限制导致提升的分辨率有限, 可尝试建立全新的高分辨率的妆容数据集进行训练.

参考文献

- 1 Tong WS, Tang CK, Brown MS, *et al.* Example-based cosmetic transfer. Proceedings of the 15th Pacific Conference on Computer Graphics and Applications (PG 2007). Maui: IEEE, 2007. 211–218.
- 2 Guo D, Sim T. Digital face makeup by example. Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami: IEEE, 2009. 73–79.
- 3 Du H, Shu LQ. Makeup transfer using multi-example. Proceedings of the 2012 International Conference on Information Technology and Software Engineering. Springer, 2013. 467–474.
- 4 Khan A, Ahmad M, Guo YD, *et al.* Digital makeup from Internet images. Optik, 2018, 158: 590–601. [doi: 10.1016/j.ijleo.2017.12.047]
- 5 Xu L, Du YZ, Zhang YM. An automatic framework for example-based virtual makeup. Proceedings of the 2013 IEEE International Conference on Image Processing.

- Melbourne: IEEE, 2013. 3206–3210.
- 6 Lee JY, Kang HB. A new digital face makeup method. Proceedings of the 2016 IEEE International Conference on Consumer Electronics (ICCE). Las Vegas: IEEE, 2016. 129–130.
 - 7 Goodfellow I, Pouget-Abadie J, Mirza M, *et al.* Generative adversarial networks. Communications of the ACM, 2020, 63(11): 139–144. [doi: [10.1145/3422622](https://doi.org/10.1145/3422622)]
 - 8 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.
 - 9 Li TT, Qian RH, Dong C, *et al.* BeautyGAN: Instance-level facial makeup transfer with deep generative adversarial network. Proceedings of the 26th ACM International Conference on Multimedia. Seoul: ACM, 2018. 645–653.
 - 10 Chen HJ, Hui KM, Wang SY, *et al.* BeautyGlow: On-demand makeup transfer framework with reversible generative network. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 10034–10042.
 - 11 Chang HW, Lu JW, Yu F, *et al.* PairedCycleGAN: Asymmetric style transfer for applying and removing makeup. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 40–48.
 - 12 Gu Q, Wang GZ, Chiu MT, *et al.* LADN: Local adversarial disentangling network for facial makeup and de-makeup. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 10480–10489.
 - 13 Kips R, Gori P, Perrot M, *et al.* CA-GAN: Weakly supervised color aware GAN for controllable makeup transfer. European Conference on Computer Vision. Glasgow: Springer, 2020. 280–296.
 - 14 Zhang HL, Chen WQ, He H, *et al.* Disentangled makeup transfer with generative adversarial network. arXiv:1907.01144, 2019.
 - 15 Jiang WT, Liu S, Gao C, *et al.* PSGAN: Pose and expression robust spatial-aware GAN for customizable makeup transfer. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 5193–5201.
 - 16 Nguyen T, Tran AT, Hoai M. Lipstick ain't enough: Beyond color matching for in-the-wild makeup transfer. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 13300–13309.
 - 17 Deng H, Han C, Cai HM, *et al.* Spatially-invariant style-codes controlled makeup transfer. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 6545–6553.
 - 18 Wan ZY, Chen HR, An J, *et al.* Facial attribute transformers for precise and robust makeup transfer. Proceedings of the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2022. 3113–3122.
 - 19 Sun ZY, Chen YX, Xiong SW. SSAT: A symmetric semantic-aware transformer network for makeup transfer and removal. Proceedings of the 36th AAAI Conference on Artificial Intelligence. AAAI, 2022. 2325–2334.
 - 20 Lyu YM, Dong J, Peng B, *et al.* SOGAN: 3D-aware shadow and occlusion robust GAN for makeup transfer. Proceedings of the 29th ACM International Conference on Multimedia. ACM, 2021. 3601–3609.
 - 21 Yang CY, He WR, Xu YQ, *et al.* EleGANt: Exquisite and locally editable GAN for makeup transfer. Proceedings of the 17th European Conference on Computer Vision. Tel Aviv: Springer, 2022. 737–754.
 - 22 Organisciak D, Ho ESL, Shum HPH. Makeup style transfer on low-quality images with weighted multi-scale attention. Proceedings of the 25th International Conference on Pattern Recognition (ICPR). Milan: IEEE, 2021. 6011–6018.
 - 23 Ren YQ, Sun YQ, Wu D, *et al.* A new makeup transfer with super-resolution. Proceedings of the 26th International Conference on Neural Information Processing. Sydney: Australian National University, 2019. 59–68.
 - 24 Song YB, Zhang JW, He SF, *et al.* Learning to hallucinate face images via component generation and enhancement. Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne: IJCAI, 2017. 4537–4543.
 - 25 Jiang JJ, Yu Y, Hu JH, *et al.* Deep CNN denoiser and multi-layer neighbor component embedding for face hallucination. Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm: IJCAI, 2018. 771–778.
 - 26 Bulat A, Tzimiropoulos G. Super-FAN: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with GANs. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 109–117.
 - 27 Kim D, Kim M, Kwon G, *et al.* Progressive face super-resolution via attention to facial landmark. Proceedings of the 30th British Machine Vision Conference. Cardiff: BMVC, 2019. 192.
 - 28 Chen CF, Li XM, Yang LB, *et al.* Progressive semantic-

- aware style transformation for blind face restoration. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 11891–11900.
- 29 Kalarot R, Li T, Porikli F. Component attention guided face super-resolution network: CAGFace. Proceedings of the 2020 IEEE Winter Conference on Applications of Computer Vision. Snowmass: IEEE, 2020. 359–369.
- 30 Choi Y, Choi M, Kim M, *et al.* StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 8789–8797.
- 31 Chen Y, Shen CH, Wei XS, *et al.* Adversarial PoseNet: A structure-aware convolutional network for human pose estimation. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 1221–1230.
- 32 Newell A, Yang KY, Deng J. Stacked hourglass networks for human pose estimation. Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 483–499.
- 33 Isola P, Zhu JY, Zhou TH, *et al.* Image-to-image translation with conditional adversarial networks. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 5967–5976.
- 34 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv: 1409.1556, 2014.
- 35 Zhu JY, Park T, Isola P, *et al.* Unpaired image-to-image translation using cycle-consistent adversarial networks. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 2242–2251.
- 36 Liu S, Ou XY, Qian RH, *et al.* Makeup like a superstar: Deep localized makeup transfer network. Proceedings of the 25th International Joint Conference on Artificial Intelligence. New York: IJCAI, 2016. 2568–2575.
- 37 Yu CQ, Wang JB, Peng C, *et al.* BiSeNet: Bilateral segmentation network for real-time semantic segmentation. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 334–349.
- 38 Liao J, Yao Y, Yuan L, *et al.* Visual attribute transfer through deep image analogy. ACM Transactions on Graphics, 2017, 36(4): 120.
- 39 Chen Y, Tai Y, Liu XM, *et al.* FSRNet: End-to-end learning face super-resolution with facial priors. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 2492–2501.
- 40 Yang LB, Wang SS, Ma SW, *et al.* HifaceGAN: Face renovation via collaborative suppression and replenishment. Proceedings of the 28th ACM International Conference on Multimedia. Seattle: ACM, 2020. 1551–1560.
- 41 Yang T, Ren PR, Xie XS, *et al.* GAN prior embedded network for blind face restoration in the wild. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 672–681.
- 42 Dong C, Deng YB, Loy CC, *et al.* Compression artifacts reduction by a deep convolutional network. Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015. 576–584.
- 43 Kupyn O, Martyniuk T, Wu JR, *et al.* DeblurGAN-v2: Deblurring (orders-of-magnitude) faster and better. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 8877–8886.
- 44 Li Z, Yang JL, Liu Z, *et al.* Feedback network for image super-resolution. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 3862–3871.

(校对责编: 牛欣悦)