

一种多通路的分区域快速妆容迁移深度网络^{*}

黄妍^{1,2}, 何泽文^{1,2}, 张文生^{1,2}

¹(中国科学院 自动化研究所, 北京 100190)

²(中国科学院大学, 北京 100049)

通讯作者: 张文生, E-mail: zhangwenshengia@hotmail.com



摘要: 妆容迁移是指把参考妆容迁移到素颜人脸上,并保持其上妆风格的一种任务.它提供了快速高效的候选妆容可视化的解决方案,得到了学术界和工业界的广泛关注.为了解决真实同人异妆数据的缺失,以及现有妆容迁移方法没有充分考虑人与人的五官差异而导致的迁移脸部结构丢失等问题,提出了一种多通路的分区域快速妆容迁移深度网络模型.具体而言,首先在人脸关键点检测的基础上,完成端到端的人脸校准;再利用通路差异的损失函数,根据眼影、唇膏、粉底的区域妆容特点优化网络;最后通过泊松融合、多通路的输出生成上妆结果.该模型具有存储空间小、生成速度快的优点,在保证人脸结构不变的同时,使得迁移后的眼影更均衡,唇膏色彩更保真,粉底迁移更精细.在国际通用 VMU 和 DLMT 美妆数据库上进行实验研究,结果表明,该方法取得了更协调的视觉效果、更快的上妆速度、更多样的同人异妆和异人同妆的迁移风格,优于对比方法.

关键词: 妆容迁移;深度网络;区域校准;多通路

中图分类号: TP391

中文引用格式: 黄妍,何泽文,张文生.一种多通路的分区域快速妆容迁移深度网络.软件学报,2019,30(11):3549-3566. <http://www.jos.org.cn/1000-9825/5835.htm>

英文引用格式: Huang Y, He ZW, Zhang WS. Fast multi-way regional makeup transfer deep network. Ruan Jian Xue Bao/Journal of Software, 2019,30(11):3549-3566 (in Chinese). <http://www.jos.org.cn/1000-9825/5835.htm>

Fast Multi-way Regional Makeup Transfer Deep Network

HUANG Yan^{1,2}, HE Ze-Wen^{1,2}, ZHANG Wen-Sheng^{1,2}

¹(Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China)

²(University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract. Makeup transfer is a task that can transfer the reference-makeup to the before-makeup face, where makeup style is maintained. It provides a fast and efficient solution for visualizing candidate makeups, receiving extensive academical and industrial attention. However, the difference, such as the human face, eyebrow distance, and lip shape is more or less ignored by recent works, leading to a problem that the facial structures are lost. Moreover, the lack of datasets which are composed of before-makeup and after-makeup images also provides additional difficulties. To this end, this study proposes a regional fast makeup transfer multi-way network. Specifically, the key points of faces are firstly detected and these points are used to align different faces in an end-to-end style. Then, three way-specific losses are used to optimize the makeup transfer network jointly. These losses are designed with the apparent properties of the eyeshadow, the lipstick, and the foundation, which shows better makeup results. Finally, poisson matting is used to fuse multi-way outputs. Compared to the previous works, the proposed model requires smaller storage space and has faster speed. It can keep the structure of the original face and lead to balancer eyeshadow, more vivid lipstick color, and more detailed foundation make-up. The proposed model is estimated on two universal makeup transfer datasets (i.e., VMU and DLMT). The experimental results show that this proposed method achieves a better visual effect. Besides, it also outperforms the alternatives in terms of makeup speed, the effect of transferring different style to the

* 基金项目: 国家自然科学基金(61472423, 61432008, U1636220)

Foundation item: National Natural Science Foundation of China (61472423, 61432008, U1636220)

收稿时间: 2018-08-22; 修改时间: 2019-01-17; 采用时间: 2019-03-11

same person, and transferring the same style to the various people.

Key words. makeup transfer; deep network; region alignment; multi-way

随着经济的快速发展和社会的进步,美妆产业蓬勃发展,如何帮助人们快速而精准地找到个性化的美妆产品逐渐成为研究热点.以 TAAZ(<http://www.taaz.com/>)、MAKEOVER(<https://www.maybelline.com/virtual-makeover-makeup-tools>)、美图美颜(<https://corp.meitu.com/business/>)为代表的虚拟上妆系统能够帮助天猫美妆、唯品会美妆等公司提升用户体验,使用户能够从海量的美妆产品中进行高效的挑选.作为虚拟上妆系统的核心技术——妆容迁移技术,也受到越来越多的关注.该技术旨在将参考妆容迁移到素颜人脸上,并保持其上妆风格.理想的妆容迁移技术应达到:在保持素颜妆后图像脸部结构不变的同时,尽可能地体现参考妆容的风格.

在实际应用中,妆容迁移主要存在两个难点:其一是人与人的脸型、眉眼距离、嘴唇形状等存在差异,导致生成的妆容图像失真,且难以保持参考妆容风格;其二是妆容式样繁杂,而真实的同人异妆数据难以获得,造成高质量的标注数据有限.为了解决这些问题,现有的妆容迁移方法主要分为以下两类:基于传统算子^[1-3]的妆容迁移和基于深度网络^[4,5]的妆容迁移.传统算子方法能够有效地克服人脸结构上的差异.它们首先根据人脸关键点进行校准,再建立映射模型,从而实现妆容迁移,然而该方法建模复杂,计算量大,只能处理素颜和参考妆容肤色相近且背景单一的图像对,限制了其应用场景.基于深度网络的方法在同人异妆数据缺失的情况下有较好的表现.它们通过训练网络,将素颜和参考妆容作为输入,直接获得迁移后的妆容图像.与传统方法相比,它无需在提取人脸关键点和建立映射模型等方面进行精巧的设计,而且受益于深度网络强大的表示能力,在拍摄条件多变和妆容式样繁杂场景下具有较强的鲁棒性^[5],因而逐渐受到学者们的广泛关注.

基于深度网络进行妆容迁移的早期工作来自于 Liu 等人^[5],他们在 Gatys 等人^[6]提出的风格迁移方法的基础上,采用全卷积网络^[7]得到脸部语义分割,并把人脸划分为眼影、嘴唇和脸颊上妆区域,按区域地迭代更新素颜图像的像素值,但每次训练只适用于某一图像对,无法满足实时性要求.与之不同,对于风格迁移任务,Johnson 等人^[8]通过离线地训练卷积神经网络,可以进行已知风格的迁移,而且于在线阶段只需一步就能得到高质量的迁移图像.该方法给我们提供了一种快速妆容迁移的思路,但是其受制于人脸结构的差异和现有妆容对齐数据库的稀缺.如果直接使用 Johnson 等人^[8]的方法,就无法实现多样化的人脸妆容迁移.在此基础上,Liao 等人^[4]在深度网络抽取的特征层上,使用近邻搜索和变形,实现了多样化迁移.Luan 等人^[9]利用图像的语义分割尝试在人脸图像上进行精细的风格迁移.但是这些方法都没有在满足实时性的同时对人脸的眼影、唇膏、粉底等区域进行单独考虑,也缺少对这些区域的整体协调优化.

受上述方法的启发,我们构建了快速妆容迁移(fast makeup transfer,简称 FMaT)模型,实现端到端的妆容迁移.该模型根据眼影、唇膏、粉底的妆容特点,设计了 3 种语义损失函数优化卷积神经网络,从而无需迭代优化像素值,实现了妆容的快速迁移.同时,我们设计了分区域的多通路卷积神经网络.该网络首先将素颜和参考妆容图像对作为输入,端到端地提取人脸关键点完成校准,然后根据校准结果分区域地进行妆容迁移,在保证人脸结构的同时,使得迁移后的眼影更均衡,唇膏色彩更保真,粉底迁移更精细,从而实现多样化的妆容迁移.

本文的主要贡献如下:(1) 构建了端到端的 FMaT 模型,设计了分区域的浅层多通路卷积神经网络(region makeup transfer network,简称区域上妆迁移网络,也称 RMT-net).该结构具有存储空间小、生成速度快的优点,能实现妆容的快速迁移.(2) 针对眼影、唇膏、粉底不同的区域,设计了 3 种语义损失函数,使得迁移后的眼影更均衡,唇膏色彩更保真,粉底迁移更精细.(3) 实现妆容区域的端到端校准,使得不同妆容位置在局部迁移得到优化的同时,也具有相对协调的位置组合,从而更好地保持参考妆容的风格.

本文第 1 节总结妆容迁移相关研究工作进展.第 2 节从端到端人脸校准、多通路卷积网络和语义损失函数这 3 个方面来介绍本文构建的 FMaT 模型.第 3 节给出相关的实验数据和设置,并展示定性实验和定量实验结果.第 4 节进行总结与展望,并对未来研究方向进行初步的探讨.

1 相关工作

我们的模型关键技术涉及上妆网络和妆容区域校准等技术.本节主要阐述基于传统算子的妆容迁移、基于

深度网络的妆容迁移和妆容区域校准等相关工作。

基于传统算子的妆容迁移,首先需要对面部人脸和参考妆容进行精确的语义分割,再结合梯度算子等,实现妆容迁移。针对妆容迁移要保持素颜妆后妆前脸部结构一致等难点,其中脸部结构是判别人脸生物特征的主要依据。学者们提出了一系列解决方法,主要包括有监督和无监督两类。对于有监督方法,如 Tong 等人^[1]通过计算素颜到化妆的映射,把参考妆容迁移到素颜人脸上。但是该方法对数据集的要求较高,需要与一个人妆前、妆后的照片很好地对齐,素颜和参考妆容的肤色相近,且图像的背景单一。Scherbaum 等人^[2]用人脸的三维形变模型^[10],找到素颜到化妆的映射,此法也需要妆前、妆后对齐的高质量数据集。对于无监督方法,如 Guo 等人^[3]先用 Active Shape Model^[11]检测部分人脸关键点,然后用 Thin Plate Spline^[12]对参考妆容进行人脸变形,以校准素颜人脸。再把图像分为脸部结构层、皮肤细节层和颜色层,用梯度算子进行融合。Xu 等人^[13]在 Guo 等人^[3]的基础上,利用人脸关键点检测算法和肤色高斯混合模型,实现人脸区域的自动分割。Li 等人^[14]提出把图像分解为 3 个内在成分层,然后根据参考妆容的成分层和物理反射模型来改变素颜图像的对应层。与上述方法不同,本文采用的方法另辟蹊径,既无需妆前妆后的对齐人脸样本,也无需对图像做分解建模,极大地降低了建模复杂度和数据集要求。

基于深度网络的妆容迁移是基于深度网络的图像风格迁移的一个分支,两者都是在深度网络的基础上设计特定损失函数来优化生成的风格图像,使图像的高层信息明确。深度模型一般需要大量的同一内容不同风格图像,但标记成本太高,所以一些学者基于无监督深度网络的图像风格迁移做了相关研究,主要包含两大类^[15]:基于图像像素迭代的风格迁移和基于神经网络迭代的风格迁移。第 1 类研究是用已经训练好的深度神经网络优化结构损失和风格损失,直接迭代更新图像像素完成风格迁移,代表工作是 Gatys 等人^[6]。Gatys 等人^[6]提出图像的结构和纹理可以由卷积网络的不同层分别表达,其中高层代表图像的结构语义信息,底层则代表风格纹理信息。Liu 等人^[5]在通用框架^[6]基础上,实现人脸妆容迁移。第 2 类研究则是通过训练神经网络来实现风格迁移,代表工作是 Johnson 等人^[8]和 Liao 等人^[4]。Li 等人^[16]利用预训练网络抽取内容特征和风格特征,对编码后图像进行特征转换,再解码得到输出。Chang 等人^[17]基于 cycleGAN^[18]训练得到互为对偶的卸妆网络和上妆网络,实现了妆容迁移。本文采用类似 Johnson 等人^[8]的框架,提出专用妆容迁移算法,能够依据素颜人脸和不同参考妆容,经网络融合生成不同的妆后图像,解决了 Johnson 等人^[8]只能迁移到特定风格且无法实现多样化的人脸妆容迁移等问题。

妆容区域校准涉及人脸语义解析、人脸变形等技术。人脸语义解析技术包括人脸关键点检测、脸部语义分割等子课题。对于人脸关键点检测,山世光等人^[19]提出根据脸部纹理分布和可变模板提取脸部特征点;Kazemi 等人^[20]采用级联的回归树来实现 ms 级的准确检测;face++人工智能开发平台(<https://www.faceplusplus.com.cn/>)还推出了 83-face-landmark 和 103-face-landmark 等检测工具。对于脸部语义分割,毋立芳等人^[21]基于曲线拟合人脸轮廓实现分割;一些基于深度神经网络的方法,从 Long 等人^[7]提出的全卷积网络(FCN),到谷歌提出的目前性能最优的语义分割模型 DeepLab-v3+^[22],其分割准确率逐步增加。人脸变形技术包括形变映射模型、基于深度神经网络的区域变形等子课题。对于形变映射模型,徐岗等人^[23]将其分为 4 类,即基于体、曲面、曲线、点的变形技术;龚勋等人^[24]选择与目标人脸最相关的主成分作为形变模型的基空间,再采用全局和局部双重形变框架来完成人脸形状建模;王进等人^[25]对图像轮廓多边形内的点进行映射,实现人脸变形。对于基于深度神经网络的方法,如 Chen 等人^[26]提出用余弦距离度量相似区域,再进行变形;Liao 等人^[4]在深度网络特征层采用 Barnes 等人^[27]的近邻区域旋转和缩放方法,完成输入人脸到目标人脸的变形。考虑到算法对实时性的要求,本文不考虑采用深度网络的方法获取像素级别的语义分割信息和变形校准妆容区域。

2 快速妆容迁移模型

本文提出的快速妆容迁移模型基于人脸关键点完成素颜与参考妆容的端到端校准,并提出了一种新的区域上妆迁移网络和 3 种语义损失函数,分别对眼影、唇膏、粉底进行妆容迁移。该方法整体框架如图 1 所示,包含训练和测试两个过程。

在训练过程,输入素颜和参考妆容的图像对,以及各自的人脸关键点,将参考妆容进行变形,完成妆容区域的校准。然后分区域地构造脸部、嘴唇、眼影样本数据集。接下来对 3 种样本对构成的数据集分为三通路,训练

独立的区域上妆迁移网络,包括眼影上妆迁移网络,粉底上妆迁移网络和唇膏上妆迁移网络.模型利用预训练的16层人脸分类网络(VGG-face network)^[28]作为损失网络,对不同区域的上妆迁移网络计算对应的结构损失 \mathcal{L}_c 和风格损失 \mathcal{L}_s .通过三通路的分区域上妆网络的学习获得最终的FMaT模型.

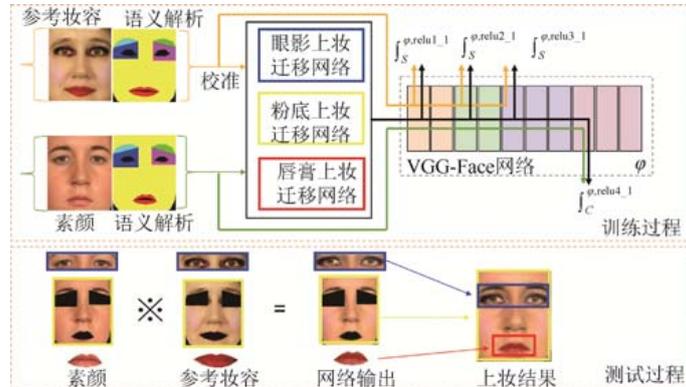


Fig.1 Framework of fast makeup transfer model

图1 快速妆容迁移模型整体框架

在测试过程,输入素颜和参考妆容的图像对,以及各自的人脸关键点,同样对参考妆容进行变形,以对齐妆容区域.然后分别提取二者的眼影、脸部、嘴唇区域,输入到三通路上妆迁移网络以生成对应的妆后眼影、脸部、嘴唇,最后采用泊松融合^[29]得到最终的上妆结果.

快速妆容迁移模型的关键步骤包括:(a) 构建素颜和参考妆容的人脸语义解析,并对参考妆容人脸进行变形以对齐妆容区域;(b) 区域上妆迁移网络的设计;(c) 3种语义损失函数的表达式.下面详细介绍这3个关键步骤.

2.1 妆容区域的校准

这部分主要介绍端到端妆容区域校准采用的人脸语义解析,以及人脸变形的的方法.

将参考脸迁移到素颜脸上生成精细的妆容,需要参考脸和素颜脸的各妆容子区域一一对应,为此,得到人脸的精确语义解析就尤为重要.我们采用的方法是对人脸分区域逐个实现妆容迁移,只需获取脸部区域间的分隔关系即可.首先通过68-face-landmark^[20]的方法获得关键点,据此将人脸划分为7个区域,如图2所示.然后分区域构造3种样本对,依次提取图像对中脸颊、眉毛等区域(图2中黄色、亮蓝色和绿色区域)构成脸部样本对,眼影区域(图2中深蓝色和紫色区域)构成眼影样本对,唇膏区域(图2中红色区域)构成嘴唇样本对.值得注意的是,由于眼珠和嘴巴内部的牙齿没有妆容,不会对该区域进行上妆,所以将这两处归为背景区域,且不参与区域上妆迁移网络的训练.



Fig.2 A face parsing result

图2 人脸语义解析结果

获取脸部区域的语义信息后,再对参考妆容做变形,从而校准参考妆容与素颜的各个子区域.为了满足算法实时性要求,本文方法需要参考脸与素颜脸在进入三通道上妆迁移网络之前就完成图像的相似区域匹配变形,因此我们基于关键点人脸做 Delaunay 三角剖分^[30],再运用仿射变换函数^[31]来完成参考妆容人脸的变形和校准,如图3所示.



Fig.3 Process of face warp and alignment
图3 人脸变形与校准过程

在 Delaunay 三角剖分中,任意三角形的外接圆内不包含任何点.在二维图像中,采用评估行列式来判断点 H 是否在点 E 、 F 、 G 的外接圆内^[30]:

$$\begin{vmatrix} x_E & y_E & x_E^2 + y_E^2 & 1 \\ x_F & y_F & x_F^2 + y_F^2 & 1 \\ x_G & y_G & x_G^2 + y_G^2 & 1 \\ x_H & y_H & x_H^2 + y_H^2 & 1 \end{vmatrix} > 0 \quad (1)$$

其中,点 E 、 F 、 G 按逆时针排列, (x_E, y_E) 、 (x_F, y_F) 、 (x_G, y_G) 、 (x_H, y_H) 分别是点 E 、 F 、 G 、 H 的坐标.当且仅当点 H 在外接圆内时,公式(1)成立.

对素颜和参考妆容图像进行三角剖分后,选取参考妆容的某个三角形,计算将其 3 个顶点映射到素颜图像中对应三角形的 3 个顶点的仿射变换 MAP ,即求解公式(2);再根据公式(3),将参考妆容的该三角形内所有像素进行变换得到变形图像.直到参考妆容中所有三角形完成仿射变换,最终得到校准后的参考妆容.

$$\begin{bmatrix} x_0^B & x_1^B & x_2^B \\ y_0^B & y_1^B & y_2^B \end{bmatrix} = MAP \cdot \begin{bmatrix} x_0^R & x_1^R & x_2^R \\ y_0^R & y_1^R & y_2^R \\ 1 & 1 & 1 \end{bmatrix} \quad (2)$$

其中, (x_0^R, y_0^R) 、 (x_1^R, y_1^R) 、 (x_2^R, y_2^R) 参考妆容某个三角形的顶点坐标, (x_0^B, y_0^B) 、 (x_1^B, y_1^B) 、 (x_2^B, y_2^B) 是素颜图像中对应三角形的顶点坐标,求解公式(2)得到仿射变换矩阵 MAP ,矩阵大小为 2×3 .

$$\begin{bmatrix} x^{RW} \\ y^{RW} \end{bmatrix} = MAP \cdot \begin{bmatrix} x^R \\ y^R \\ 1 \end{bmatrix} \quad (3)$$

其中, (x^R, y^R) 是参考妆容三角形内所有像素点的坐标, (x^{RW}, y^{RW}) 是参考妆容的该三角形内的点进行变换后的坐标.

2.2 区域上妆迁移网络

本文提出一种通用的区域上妆迁移网络,其作用是将参考脸的某区域的妆容迁移到素颜人脸的对应区域上,以生成妆后效果.结合现实化妆性质,眼部化妆使用在眼睛周围,化妆区域相对脸部较小,眼影等化妆品作用是提升个人神情气质,遮盖原有肤色.脸部的化妆主要使用在脸颊、额头和鼻翼两侧,粉底等化妆品作用是遮住斑点痘印等瑕疵.嘴唇的化妆使用在唇部,唇膏等化妆品作用是保持或提亮唇纹.根据眼部、脸部和嘴唇的不同性质和作用位置,算法维护三通路的上妆迁移网络,它们分别作用于眼部、脸部和嘴唇区域.下面以唇膏上妆迁移网络为例,如图 4 所示,该网络的作用是把参考人脸的唇膏妆容迁移到素颜人脸的嘴唇上.

唇膏上妆迁移网络结构包括 3 部分:两个输入端和一个输出端,线性相加的融合层,以及卷积层.首先,输入输出端都是 RGB 图像.为了实现任意的素颜和参考妆容图像对的风格迁移,网络设置两个输入端.具体来说,两个输入分别是素颜的嘴唇 lip_B 和参考妆容人脸中与之对齐的嘴唇 lip_{RW} ;一个输出则是嘴唇上妆后图像 $output(lip)$.

其次,线性相加融合层负责将参考妆容特征和素颜特征进行融合,如图 4 中绿色方块所示.从输入端开始, lip_B 和 lip_{RW} 分别进行卷积得到两个特征图谱,即素颜嘴唇的结构语义信息和参考唇妆的妆容风格信息,然后将两个特征图谱相加来完成融合,最后将融合后的单个特征图谱再卷积得到输出,即素颜的嘴唇上妆后图像.

区域上妆迁移网络一共有 4 个卷积层,均采用 3×3 卷积核,每层的通道数如图 4 所示.融合层之前是一层卷

积层,融合层之后是3层卷积层,这是考虑到网络实现的是区域上妆,不需要过大的网络容量.后续3.2节中的实验也证明融合层后采用3层卷积层足以完成素颜上妆任务.鉴于算法的实时性要求,上妆迁移网络设定为4层卷积层.总的来说,网络的第1层是为了实现人脸结构特征和妆容风格特征的分隔,网络的第2层~第4层是对融合后特征进行深度加工,以生成更逼真的妆容图像.

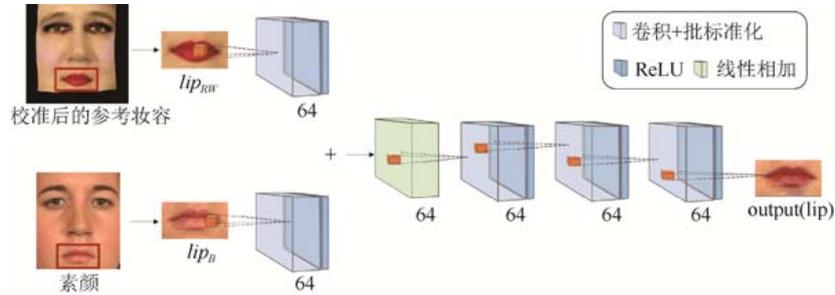


Fig.4 Structure chart of the lipstick RMT-net

图4 唇膏上妆迁移网络的结构示意图

用 $RMTnet_W^{lip}$ 表示唇膏上妆迁移网络,下标 W 表示网络权重.图4可以公式化表示为

$$output(lip) = RMTnet_W^{lip}(lip_B, lip_{RW}) \quad (4)$$

此外,眼影上妆迁移网络和粉底上妆迁移网络的结构与唇膏上妆迁移网络相同,区别只是输入端的样本对不同,粉底上妆迁移网络输入的是素颜和参考妆容的脸部样本对 $face_B$ 和 $face_{RW}$,而眼影上妆迁移网络输入的是眼影区域样本对 $eyeshadow_B$ 和 $eyeshadow_{RW}$.特别地,由于左眼眼影的上妆性质和右眼眼影的上妆性质是类似的,所以本文只训练左眼眼影的上妆迁移网络,再用于右眼眼影上妆.

2.3 妆容迁移损失函数

如前所述,本文方法以素颜人脸图像 B 和参考妆容人脸图像 R 为输入,采用三通路妆容迁移网络把 R 的妆容迁移到 B 的脸上,并且最终输出的上妆结果既要保持 B 的结构,也要拥有 R 的妆容风格,同时尽量逼真.为此分别设计结构损失、妆容风格损失、颜色损失和平滑损失函数来优化生成的图像.对于前面两种损失函数,借助预训练网络来进行衡量.不同于 VGG^[32] 分类网络,这里采用的是在人脸分类数据集 Labeled Faces in the Wild^[33] 和 YouTube Faces^[34] 上训练好的 16 层 VGG-Face 网络^[28].因为 VGG-Face 人脸分类网络的训练集包括了素颜人脸和化妆人脸,因此完全适用于人脸领域的的数据.后续实验也说明选择 VGG-Face 能够得到更好的结果.下面详细阐述利用该预训练网络来计算模型的结构损失和妆容风格损失,以及在 RGB 空间计算颜色损失和平滑损失.

1) 结构损失函数^[6]

对给定输入素颜 B 和参考妆容 R ,为了使输出图像 O 和 B 在视觉上是同一人,即保持语义结构信息,我们借助预训练 VGG-Face 网络^[28] 来衡量二者的结构特征相似度.具体来说,将 O 和 B 分别输入 VGG-Face 网络并提取第 l 层特征图谱,那么对应于该层结构损失函数表达式如下.

$$\mathcal{L}_C^{\phi,l} = \frac{1}{N_l D_l} \sum_{ij} (F_l[O] - F_l[B])_{ij}^2 \quad (5)$$

这个公式计算 O 和 B 在网络 ϕ 第 l 层的特征图谱的平均差值.其中, ϕ 代表 VGG-Face 网络, N_l 是第 l 层特征的通道(channel), D_l 是特征图谱高与宽的乘积,那么第 l 层特征图谱表示为 $F_l \in \mathbb{R}^{N_l \times D_l}$, $(F_l)_{ij}$ 表示第 i 个通道在位置 j 的激活.

2) 妆容风格损失函数^[6,9]

对给定输入素颜 B 和参考妆容 R ,为了使输出 O 的妆容风格与 R 的尽可能地一致,基于二者输入 VGG-Face 网络中提取的特征图谱, O 的妆容风格信息和 R 之间的相似度可通过计算 Gram 矩阵得到.具体来说,第 l 层的妆

容风格损失函数表达式如下.

$$\mathcal{L}_S^{\phi,l} = \frac{1}{N_l^2} \sum_{ij} (G_l[O] - G_l[RW])_{ij}^2 \quad (6)$$

$$G_l[\cdot] = F_{l,a}[\cdot] F_{l,a}[\cdot]^\top, F_{l,a}[O] = F_l[O] \mathcal{M}_{l,a}[B], F_{l,a}[RW] = F_l[RW] \mathcal{M}_{l,a}[B]. \quad (7)$$

其中,第 l 层特征图谱的 Gram 矩阵用 $G_l[\cdot]$ 表示, a 是人脸语义解析 M 的某个区域. 因为本文在训练三通路妆容迁移网络之前, 根据素颜 B 的人脸关键点位置, 对参考妆容 R 进行变形, 得到校准后的参考妆容 RW , 所以公式(7)中第 3 个式子是由 $F_{l,a}[R] = F_l[R] \mathcal{M}_{l,a}[B]$ 改写的.

3) 颜色迁移损失函数

除了迁移风格之外, 输出 O 和参考妆容 R 的脸色要尽可能相似. 因此直接计算输出 O 和参考妆容 R 在 RGB 空间的 L1 损失来衡量颜色差距.

$$\mathcal{L}_{color} = \frac{1}{3 \times D} \sum_{ij} |O - R|_{ij} \quad (8)$$

其中, D 是图像高与宽的乘积. 后续实验证明了除妆容风格损失之外再加上颜色损失的必要性.

4) 平滑损失函数

为了让输出图像 O 保持平滑, 用全变分损失(total variation loss)^[35]来衡量平滑程度.

$$\mathcal{L}_w = \frac{1}{3 \times D} \|\nabla_h O - \nabla_w O\| \quad (9)$$

其中, D 是图像高与宽的乘积, $\nabla_h O, \nabla_w O$ 分别表示对图像的高、宽方向求导. 平滑损失的主要作用是去除可能的椒盐等噪声.

结合公式(5)~(9), 区域上妆迁移网络的语义损失函数表达式为

$$\mathcal{L}_{total} = \alpha \mathcal{L}_C^{\phi,l} + \beta \sum_{l=1}^L \mathcal{L}_S^{\phi,l} + \gamma \mathcal{L}_{color} + \lambda \mathcal{L}_w \quad (10)$$

本文增加了输入图像的语义解析, 根据粉底、唇膏、眼影妆容的性质差异, 对每个语义解析区域设置独立的损失函数及权重. 具体来说, 对粉底上妆迁移网络, 结构损失在网络 ϕ 的 relu5_1 层衡量, 妆容风格损失在 relu1_1、relu2_1、relu3_1 层衡量; 对唇膏上妆迁移网络, 结构损失在网络 ϕ 的 relu4_1 层衡量, 妆容风格损失在 relu1_1、relu2_1、relu3_1 层衡量, 为保持嘴唇纹理将去掉平滑损失; 对眼影上妆迁移网络, 结构损失在网络 ϕ 的 relu4_1 层衡量, 妆容风格损失在 relu1_1 层衡量^[5]. 此外, 三通路网络损失函数采用了不同权重设置, 后续实验(第 3.2 节)证明了这种策略的有效性.

3 实验

3.1 实验数据集和设置

本文采用两个美妆数据集, VMU^[36]和 DLMT^[5].

VMU(virtual makeup)数据集^[36]包含 51 个白人女性, 每位女性包含一张素颜图像和 3 张不同妆容风格的图像, 总计 204 张图像. 如图 5(a)第 1 行所示, 用 TAAZ 的虚拟化妆工具, 为每位素颜女性生成 3 种虚拟化妆: (1) 只使用口红; (2) 只使用眼妆; (3) 涂上口红、粉底和眼妆等整套化妆品. 数据集(<http://antitza.com/makeup-datasets.html>)已经进行过人脸裁剪, 如图 5(a)第 2 行所示, 图像高宽 150×130. 我们将原始图像放置在高宽 224×224 的零矩阵中心处.

DLMT^[5]数据集有 1 000 张素颜图像, 1 000 张化妆图像, 包括亚洲人、欧美人等, 整个数据集约 2 000 人. 部分素颜图像带有裸妆或者淡妆, 如图 5(b)第 1 行所示. 化妆图像包括彩妆、烟熏妆等妆容风格, 如图 5(b)第 2 行所示, 图像高宽 224×224. 值得注意的是, DLMT 不具有 VMU 那样的同人异妆图像对.

实验对数据集做如下切分, 把 VMU 数据集分为 40 人的训练集, 11 人的测试集, 因此 VMU 的训练集包含 40 张素颜、120 张参考妆容, 测试集包含 11 张素颜、33 张参考妆容. DLMT 数据集采用同 Liu 等人^[5]一样的切分方

式:随机抽取 850 张素颜图像和 450 张参考妆容图像作为训练集,随机抽取 100 张素颜图像和 500 张参考妆容图像作为测试集,剩余的 50 张素颜图像和 50 张参考妆容图像作为验证集.最终的网络优化的超参数为:粉底上妆迁移网络的损失函数权重 $[\alpha_f, \beta_f, \gamma_f, \lambda_f]$ 是 $[200, [50, 50, 50], 1, 1]$.眼影上妆迁移网络的损失函数权重 $[\alpha_e, \beta_e, \gamma_e, \lambda_e]$ 是 $[0.3, [4e8], 300, 100]$.唇膏上妆迁移网络的损失函数权重 $[\alpha_{lip}, \beta_{lip}, \gamma_{lip}, \lambda_{lip}]$ 是 $[0.3, [1e5, 1e5, 1e5], 100, 0]$.上妆迁移网络迭代 1 000 次,采用学习率为 1×10^{-4} 的 Adam^[37]进行梯度优化,实验使用单个 12GB GTX TITAN X GPU.



Fig.5 Some examples of the VMU and the DLMT beauty datasets

图 5 美妆数据集 VMU 和 DLMT 的部分示例

第 3.2 节从 3 个方面的实验来进行自身模型对比,分别是损失网络采用 VGG-Face^[28]网络的原因,区域上妆迁移网络的卷积层数,以及损失函数的权重设置对最终结果的影响.第 1 个实验是为了验证图像的结构和风格信息可以分离和组合^[6],类似于 Gatys 等人^[6]和 Johnson 等人^[8]的结构重构和风格重构实验,分别在深度网络 VGG-Face^[28]和 VGG-ImageNet^[38]的特征层上优化损失函数,迭代更新输入图像的像素值.第 2 个和第 3 个实验以唇膏上妆迁移网络为例,为了便于观察实验的输出图像,取嘴唇周围的区域作为网络的输入.第 2 个实验,探究不同深度的区域上妆迁移网络对上妆结果的影响,分别采用 4 层、7 层、10 层的唇膏上妆迁移网络,展示它们的唇妆上妆过程来说明.本文最终的上妆结果采用的是 4 层卷积层的区域上妆迁移网络,如图 4 所示.在图 4 中线性相加融合层之后继续增加卷积层,将得到 7 层、10 层的网络结构.第 3 个实验,通过对损失函数设置不同的权重,验证损失的重要性的探究损失函数的权重值对上妆结果的影响.表 1 是公式(10)的 4 种损失的不同权重设置实验,分别对结构损失、妆容风格损失、颜色损失、平滑损失的权重参数设置不同的数值,控制其余 3 个权重参数值不变,将当前权重设置成不同的值.

Table 1 Weight setting experiments of lipstick RMT-net's four losses

表 1 唇膏上妆迁移网络的 4 种损失的权重设置实验

实验名称	变量值	
	自变量	控制变量
结构损失 \mathcal{L}_c^s	$\alpha_{lip} = [0, 0.3, 0.8]$	$[\beta_{lip}, \gamma_{lip}, \lambda_{lip}] = [[1e5, 1e5, 1e5], 100, 0]$
妆容风格损失 \mathcal{L}_s^o	$\beta_{lip} = \begin{bmatrix} 0 & 0 & 0 \\ 1e3 & 1e3 & 1e3 \\ 1e5 & 1e5 & 1e5 \end{bmatrix}$	$[\alpha_{lip}, \gamma_{lip}, \lambda_{lip}] = [0.3, 100, 0]$
颜色损失 \mathcal{L}_{color}	$\gamma_{lip} = [0, 100]$	$[\alpha_{lip}, \beta_{lip}, \lambda_{lip}] = [0.3, [1e5, 1e5, 1e5], 0]$
平滑损失 \mathcal{L}_v	$\lambda_{lip} = [0, 100]$	$[\alpha_{lip}, \beta_{lip}, \gamma_{lip}] = [0.8, [1e5, 1e5, 1e5], 100], [\alpha_{lip}, \beta_{lip}, \gamma_{lip}] = [0.3, [1e5, 1e5, 1e5], 100]$

第 3.3 节为了展示同一种妆容的浓淡变化风格,设计控制妆容浓淡实验.考虑两种妆容浓淡控制方法:第 1 种如 Liu 等人^[5]的方法在训练阶段调整不同损失的权重比例来影响最终的妆容生成模型;第 2 种如 Li 等人^[16]的方法不改变训练阶段,只对测试阶段的超参数进行调节.本文采用第 2 种方法,对人脸语义解析的上妆区域样本对,采用不同大小的核图像进行形态学腐蚀,再把样本对输入相应的区域上妆迁移网络中,最终融合^[29]得到上妆结果.具体来说,选取的核分别是(3,6,9),且纵横一致.

第 3.4 节为评估本文方法的有效性,与更多的算法进行了定性实验和定量实验对比.一方面,将本领域现有方法^[3,5,6,9,16]的上妆结果与本模型的结果做直观的视觉对比,包括同人异妆和异人同妆,来做定性(qualitative)分析实验;另一方面,参考 PSNR、SSIM、协方差、生成速度等指标^[8,16],将现有方法^[6,9,16]的结果与本文进行量化对比,来做精确的定量(quantitative)分析实验.此外,Liu 等人^[5]、Luan 等人^[9]和 Chang 等人^[17]采用了基于真人评价的打分,但限于成本因素,调查范围较小;不同人存在的年龄、性别、职业、地域等方面的不同将使得评价完全迥异,进而影响最终评估结果的准确性.因此本文的定量实验是参考 Johnson 等人^[8]和 Li 等人^[16]采用的指标来对结果进行定量分析.我们使用有真实标签的样本对(素颜图像 B ,素颜的真实妆后图像 $B\text{-Truth}$), $B\text{-Truth}$ 和 B 是同一个人, $B\text{-Truth}$ 展现 B 妆后协调的状态.由于 VMU 训练集中没有运用训练标签对,我们把一部分训练集也作为测试样本进行测试.把 $B\text{-Truth}$ 作为参考妆容,使用上妆网络生成 B 的上妆结果($B\text{-result}$),最后根据 $B\text{-Truth}$ 和 $B\text{-result}$ 计算评价指标.PSNR(peak signal to noise ratio)是峰值信噪比,取值 0~100,值越大代表图像质量越好,它可以衡量与 $B\text{-Truth}$ 相比, $B\text{-result}$ 的图像质量的优劣程度.SSIM(structural similarity index)^[39]是结构相似性,它从亮度(均值),对比度(标准差),结构(协方差)3 方面,衡量两幅图像的相似度,取值 0~1,值越大, $B\text{-result}$ 与 $B\text{-Truth}$ 的相似度越高.平均协方差的计算方法是将 $B\text{-result}$ 与 $B\text{-Truth}$ 分别输入 VGG-Face^[28]网络中,并统计 relu1_2、relu2_2、relu3_2、relu4_2、relu5_2 层特征图谱的协方差,取对数(底数为 10)后求和除以 5 得到其平均值,平均协方差值越小, $B\text{-result}$ 与 $B\text{-Truth}$ 的相似度越高.PSNR、SSIM、log(Cov)反映在定量层面上,上妆结果和真实协调的妆后状态的相似程度,反映美妆后在视觉上的协调程度.

3.2 妆容迁移模型实验

损失网络实验,如图 6(a)是用 VGG-Face^[28]网络对素颜图像进行结构重构,第 2 列~第 5 列图像分别是在 relu2_1 层、relu3_1 层、relu4_1 层和 relu5_1 层进行结构重构,迭代 200 次.同理用 VGG-ImageNet 网络进行结构重构得到图 6(b).图 7(a)是用 VGG-Face 网络对素颜图像进行妆容风格重构,第 2 列图像是在 relu1_1 和 relu2_1 层上进行重构,第 3 列图像是在 relu1_1、relu2_1 和 relu3_1 层上进行重构,第 4 列图像是在 relu1_1、relu2_1、relu3_1 和 relu4_1 层上进行重构,第 5 列图像是在 relu1_1、relu2_1、relu3_1、relu4_1 和 relu5_1 层上进行重构,迭代 500 次.同理用 VGG-ImageNet 网络进行妆容风格重构得到图 7(b).

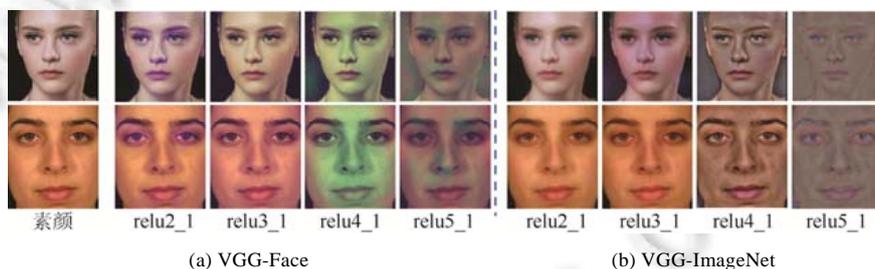


Fig.6 Content reconstructions of the before-makeup images

图 6 素颜图像的结构重构结果

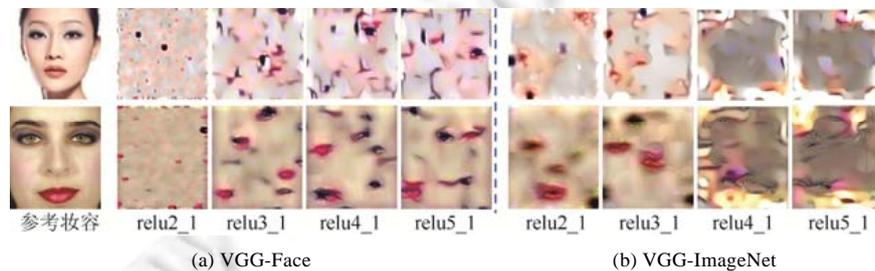


Fig.7 Makeup style reconstructions of the reference-makeup images

图 7 参考妆容图像的妆容风格重构结果

由于 VGG-Face^[28]人脸分类网络与 VGG-ImageNet^[38]分类网络相比,前者对人脸语义信息更敏感,所以损失网络采用 VGG-Face^[28]网络.对比图 6(a)与图 6(b),两者都是随着网络层数加深,丢弃图像的像素信息,保留结构信息.注意到图 6(a)中 relu4_1 用 VGG-Face 重构出现大面积的绿色,这是因为 VGG-Face 网络在 relu 层之前没有批量归一化处理层,导致图像像素值的溢出.随着增大迭代和调节权重,能消除图 6(a)中 relu4_1 和 relu5_1 的绿色,如本节图 8 和图 9 实验结果所示.对比图 7(a)与图 7(b),VGG-Face 网络对人的眼睛、嘴唇等语义信息有增强作用.VGG-Face 网络没有批量归一化处理层,这使得人脸的语义信息放大.



Fig.8 Process of the different layers of lipstick RMT-net
图 8 不同层数的唇膏上妆迁移网络的唇妆上妆过程

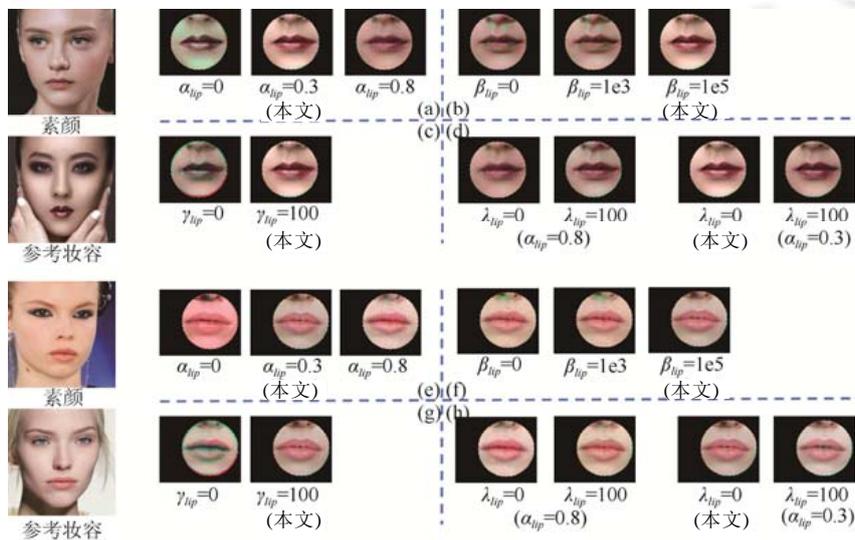


Fig.9 Outputs by setting the weight of lipstick RMT-net's current loss to different values
图 9 将唇膏上妆迁移网络当前损失的权重设置成不同的值,得到的唇妆上妆结果

区域上妆迁移网络的深度探索实验结果如图 8 所示,对 3 种网络结构的上妆迁移网络进行 1 000 次迭代,分别取第 0、200、400、600、800、1 000 次迭代的输出图像.随着网络的加深,素颜嘴唇的上妆图像越来越细化,且颜色更偏向参考妆容的唇妆,而 10 层网络的输出似乎更倾向于把参考妆容的唇妆颜色加深.4 层网络的第 1 000 次迭代得到的图像,与 7 层网络或 10 层网络第 400、600 次迭代得到的图像相近.4 层网络多次迭代后,能消除图像大面积的绿色区域.整体而言,增加上妆迁移网络的卷积层将使得网络迭代 1 000 次后得到的上妆图像差别不大.本文采用 4 层的网络结构,不仅占用内存小,能够加快测试过程的生成速度,也能得到较好的上妆结果.

损失权重设置实验结果如图 9 所示,我们得到 3 个结论.(1) 如图 9(a)和图 9(e)所示,结构损失可以去除图像大面积的单一色区域,如绿色、红色.增大结构损失的权重,网络会学习到素颜图像的结构信息,如嘴唇纹理.但是当权重过大时,网络会放大这种纹理,形成噪声.(2) 颜色损失作为辅助函数,能够帮助网络学习到参考妆容的颜色.对比图 9(b)的 $\beta_{lip}=[0,0,0]$ 和图 9(c)的 $\gamma_{lip}=0$ 的图像,如果缺少妆容风格损失,网络无法完成上妆的细化,如果缺少颜色损失,网络难以学习到参考妆容的颜色.(3) 如图 9(d)和图 9(h)所示,平滑损失把图像中的高频信号保留,去除可能的椒盐等低频噪声信号.考虑到结构损失的权重值以及唇膏上妆的化妆品性质是保留或提亮素颜嘴唇的纹理,所以唇膏上妆迁移网络把平滑损失置 0.粉底、眼影上妆的化妆品性质是遮瑕,所以这两个上妆迁移网络保留平滑损失.

3.3 妆容浓淡实验

在 DLMT 数据集上,本文算法实验得到的上妆浓淡结果和 Liu 等人^[5]进行对比,一共展示了两组样本对的上妆结果,如图 10 所示.输入的素颜和参考妆容如图 10(a)的第 1 列,图 10(a)中的后 3 列和(b)中的 3 列分别是 Liu 等人^[5]和本文方法得到的结果.具体来说,第 1 行和第 3 行是眼影从淡到浓的上妆结果,第 2 行和第 4 行是唇膏从淡到浓的上妆结果.通过图 10(a)和图 10(b)对比可以看出:(1) 在眼影上妆方面,Liu 等人^[5]是根据参考妆容的眼影语义区域,对素颜进行 TPS^[12]变形得到的眼影上妆区域;本文则是根据素颜的人脸关键点,对素颜进行区域划分得到眼影上妆区域.所以本文得到的眼影上妆区域更符合素颜的当前形态,比 Liu 等人^[5]得到的眼影上妆结果更对称均衡.(2) 在唇膏上妆方面,Liu 等人^[5]更偏向保留素颜人脸的嘴唇纹理,本文更偏向获得参考妆容唇妆的光泽感.因为本文增加了颜色损失,学习到的唇妆更保真.

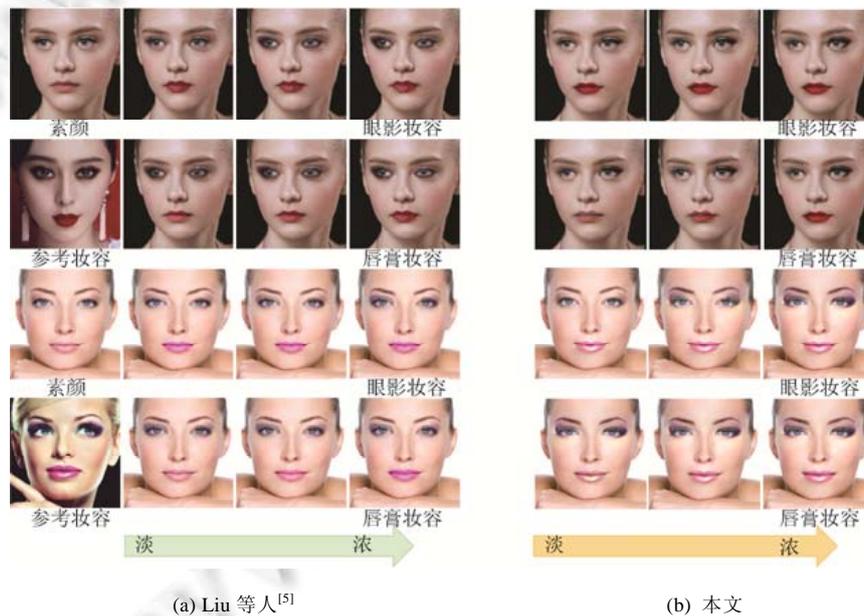


Fig.10 Contrast experiments of the lightness of the makeups on the DLMT dataset

图 10 DLMT 数据集,妆容浓淡对比实验

我们在 VMU 数据集上展示同一种妆容的浓淡变化风格,如图 11 展示了两个样本对在眼影、唇膏区域的妆容浓淡控制结果.其中,图 11 第 1 列是两组样本对的素颜与参考妆容,后面 3 列代表由淡变浓的妆容结果;第 1 行和第 3 行代表眼影的结果,第 2 行和第 4 行代表唇膏的结果.



Fig.11 Experiments of the lightness of the makeups on the VMU dataset

图 11 VMU 数据集,妆容浓淡实验

3.4 定性和定量实验

定性分析实验的结果如图 12 和图 13 所示.图 12 中将本文方法与 Guo 等人^[3]、Gatys 等人^[6]、Liu 等人^[5]在 DLMT 数据集上的结果进行对比,其中,第 3 列~第 5 列展示的 Guo 等人^[3]、Gatys 等人^[6]、Liu 等人^[5]的上妆结果均来自文献[5].图 13 中将本文方法与 Gatys 等人^[6]、Luan 等人^[9]、Li 等人^[16]在 VMU 数据集上的结果进行对比,其中,第 3 列~第 5 列展示的 Gatys 等人^[6]、Luan 等人^[9]、Li 等人^[16]的上妆结果是利用对应的代码(Gatys 等人^[6]:<https://github.com/jcjohnson/neural-style>,Li 等人^[16]:<https://github.com/sunshineatnoon/PytorchWCT>,Luan 等人^[9]:<https://github.com/luanfujun/deep-photo-styletransfer>),重新训练后得到的测试结果.

针对 DLMT 数据集,图 12 的对比表明:(1) Guo 等人^[3]的 4 组样本对都没有把参考妆容的唇妆正确迁移,眼影的迁移实际是在“画”淡妆;(2) Gatys 等人^[6]只实现了轻微的妆容迁移,且生成的图像存在一些噪声;(3) Liu 等人^[5]实现了眼影和唇妆的迁移,但是部分唇妆存在嘴唇变形,左、右眼影有明显的不对称,如第 5 列的第 1 行和第 3 行;(4) 本文的方法实现了眼影和唇妆的迁移,与 Liu 等人^[5]的第 1 行和第 3 行结果对比,本文实现的眼影妆容更对称、自然.与所有对比模型的结果相比,本文实现的唇妆更保真.

针对 VMU 数据集,图 13 的对比表明:(1) Luan 等人^[9]虽然实现了整体妆容风格的迁移,但得到的上妆结果不太真实;(2) Gatys 等人^[6]生成了较真实的人脸,但没有实现眼妆的迁移,且基本没有实现对唇妆颜色的迁移;(3) Li 等人^[16]的上妆结果中明显存在多余的脸部高光,且没有实现眼影迁移;(4) 本文则实现了整体妆容迁移,且所得的上妆结果的眼影更均衡,唇膏色彩更保真,粉底迁移更精细.第 1 行和第 3 行实现了脸部腮红的迁移,第 2 行实现了参考妆容整个肤色的迁移.前者没有实现肤色的迁移是因为训练集的皮肤颜色,浅红色皮肤的样本占多数,当训练网络时,网络会更少关注白色皮肤.

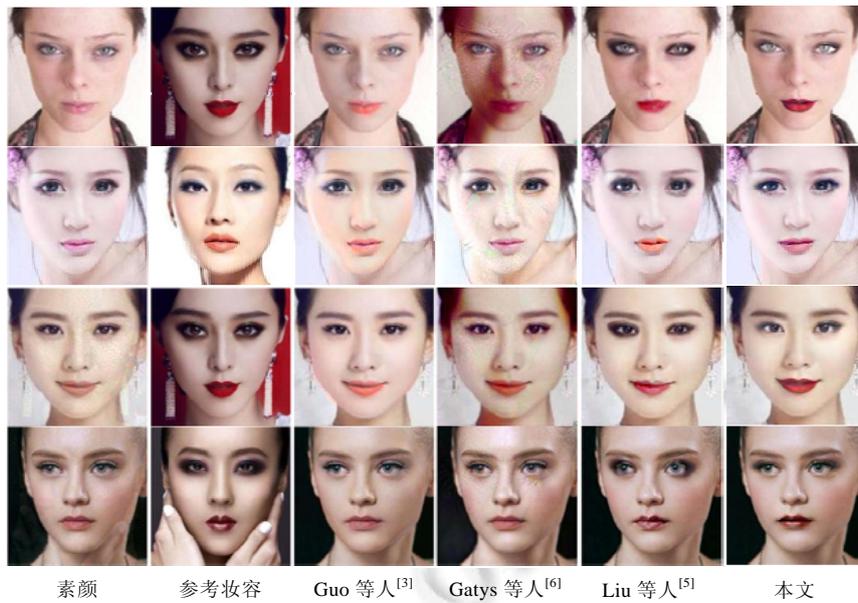


Fig.12 Qualitative comparisons of results on the DLMT dataset
图 12 DLMT 数据集,定性实验对比结果

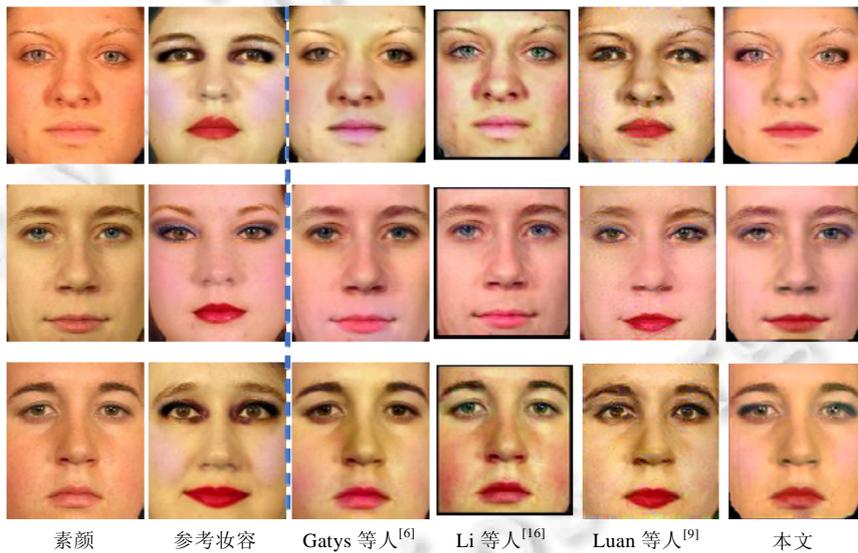


Fig.13 Qualitative comparisons of results on the VMU dataset
图 13 VMU 数据集,定性实验对比结果

同人异妆和异人同妆实验结果如图 14~图 17 所示.其中,图 14 和图 15 分别是本文方法将不同素颜人脸迁移同种妆容在 DLMT 数据集、VMU 数据集上的结果.图 16 和图 17 分别是本文方法将同一素颜人脸迁移不同类型妆容在 DLMT 数据集、VMU 数据集上的结果.由此,相对于 Liu 等人^[5]的方法,本文方法在唇妆、眼影、粉底这 3 个方面均存在优势.



Fig.14 Different people with the same makeup on the DLMT dataset

图 14 DLMT 数据集,异人同妆对比实验



Fig.15 Different people with the same makeup on the VMU dataset

图 15 VMU 数据集,异人同妆实验



Fig.16 The same person with different makeups on the DLMT dataset

图 16 DLMT 数据集,同人异妆对比实验



Fig.17 The same person with different makeups on the VMU dataset

图 17 VMU 数据集,同人异妆实验

在唇妆迁移方面:(1) Liu 等人的结果中,一部分能保持素颜嘴唇的纹理,另一部分则存在变形扭曲.此外对于露齿的素颜嘴唇,Liu 等人^[5]采用人脸分割网络能够较好地实现唇妆迁移,如图 14(a)的第 2 行第 2 列.(2) 本文方法能够较好地保持素颜的嘴唇形状,但是对于露齿素颜的效果还不好,如图 14(a)的第 3 行第 2 列.这主要是由于本文采用的人脸关键点检测技术^[20]难以精确检测嘴唇内部边缘,尤其是在图像分辨率较低的情况下.后续工作将考虑训练一个小型的嘴唇分割网络,进而提高嘴唇内部检测精度.

在眼影迁移方面:(1) Liu 等人^[5]的结果会出现明显的左眼影妆容浓于右眼影,且左、右眼影不对称,如图 14(b)的第 2 行第 1 列~第 3 列,图 16(b)的第 2 行第 5 列.部分结果则出现眼睛的轻微变形,如图 16(b)的第 2 行第 4 列.但是,Liu 等人^[5]能够实现眼睛下方卧蚕部位的妆容迁移,如图 14(b)的第 2 行.(2) 在本文方法的结果中,左、右眼影更对称协调,但是眼睛下方卧蚕部位未完成妆容迁移,如图 14(b)的第 3 行第 4 列.原因是,本文的眼妆上妆区域是依靠人脸关键点划分的,没有参考妆容眼影区域的细节语义信息,而 Liu 等人^[5]的人脸分割网络能得到参考妆容眼影区域的语义信息.后续工作将对眼睛下方的眼影迁移进行改进.

在粉底迁移方面,本文方法在 VMU 上的结果更加精细自然,如图 15 和图 17 所示.但是,有少数参考妆容的肤色无法迁移,如图 17(a)的第 2 行第 4 列.这是因为参考妆容的白色粉底在 VMU 数据集中占少数,而黄色和浅粉色的粉底占多数,所以粉底上妆迁移网络更擅长于实现颜色占多数的粉底妆容的迁移.

整体来说,本文的唇膏、眼影、粉底上妆迁移比其他方法的效果要好,尽管不平衡样本和嘴巴内部边缘检测^[20]带来了少许不足.但从定性对比实验来看,本文提出的上妆迁移网络结构及针对唇膏、眼影、粉底的特点分别设计的损失函数是行之有效的.

定量实验从 VMU 数据集中挑选出有标签的 28 个样本对进行评估,每对样本均为同一人妆前妆后的脸部图像,这样可以避免不同个体样本对存在的脸部结构差异产生的影响.图 18 展示了各个方法在两对样本上的上妆结果以及计算得到的 PSNR 值和 SSIM 值;然后又分别统计了 10 个样本对,以及 28 个样本对的平均 PSNR 值和 SSIM 值.可以看出,本文方法的 SSIM 值较高,上妆结果与参考妆容相似,而且上妆结果不仅完成妆容迁移,还保留了素颜图像的脸部结构信息.

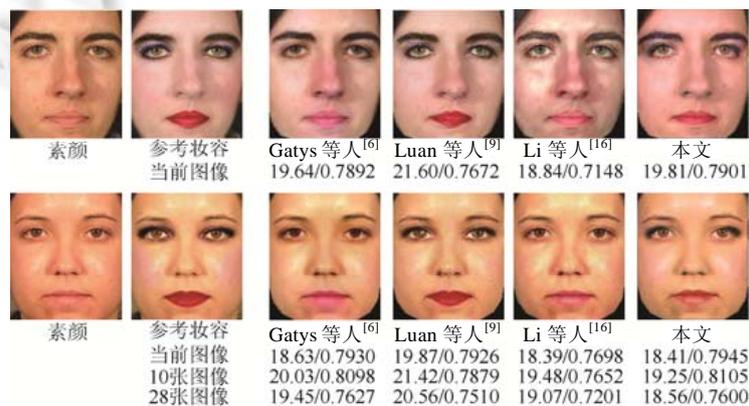


Fig.18 Quantitative comparisons of results by computing PSNR/SSIM on the label-pairs of the VMU dataset

图 18 定量对比实验,VMU 数据集中有标签的样本对的 PSNR/SSIM 值

如表 2 所示,我们统计了 28 个样本对完成上妆结果的耗时.我们的平均生成时间是 0.23s,其中妆容区域校准的时间是 0.038 2s,占总时间的 16.6%.本文方法分别比 Gatys 等人快 801 倍,比 Luan 等人^[9]快 370 倍,比 Li 等人^[16]快 2 倍,这表明我们使用的妆容迁移方法更为迅速.而且与其他模型相比,本文方法的平均协方差更小,上妆结果和参考妆容相似度更高.

Table 2 Results of quantitative, the average generation time and the average covariance of 28 pairs
表 2 定量评价结果,28 个样本对的平均生成时间和平均协方差

	Gatys 等人 ^[6]	Luan 等人 ^[9]	Li 等人 ^[16]	本文
Time (s)	184.30	85.07	0.42	0.23
log(Cov)	6.315 3	6.341 0	6.310 0	6.308 8

4 总结与展望

本文构建了端到端的 FMaT 模型,设计并实现了分区域的浅层多通路卷积神经网络 RMT-net.根据眼影、唇膏、粉底的不同妆容性质,设计了 3 种对应的语义损失函数,最后,在 DLMT 和 VMU 数据集上进行实验并进行定性定量分析,得到的主要结论如下.

1) 在 DLMT 和 VMU 数据集上,本文提出的模型生成上妆结果比较自然,眼影妆容与其他模型相比更对称,更协调.唇膏色彩更保真,更能保持素颜图像的嘴唇形状.粉底妆容在实现更精细迁移的同时,保留素颜图像的脸部结构.本文所提出的方法优于 5 种对比方法^[3,5,6,9,16].

2) 本文所提出的 FMaT 模型能够实现任意妆容风格的迁移,同时取得了明显提升的视觉效果、更快的上妆速度、更好的同人异妆和异人同妆的迁移风格,其中上妆速度比 Gatys 等人^[6]快 801 倍,比 Luan 等人^[9]快 370 倍,比 Li 等人^[16]快 2 倍.通过定性定量分析,本文方法优于对比方法.

3) 与基于深度网络^[5,6,9,16]的 4 种方法相比,本文实现妆容区域的端到端校准,使得不同妆容位置在局部迁移得到优化的同时,也具有相对协调的位置组合,从而更好地保持参考妆容的风格.

本文提出的 FMaT 模型实现了眼影、唇膏、粉底区域的快速上妆,且获得了比其他模型更好的效果.未来将从以下 3 个方面展开工作:(1) 在深度网络的卷积层上,探索用类似最近邻搜索^[4]的方式或求解最小余弦值^[26]等方法实现人脸校准和变形;(2) 面向妆容迁移任务,探索更好的量化评价指标,比如对整体妆容颜色搭配的协调性进行定量衡量;(3) 研究妆容推荐算法,为不同结构的素颜人脸推荐更合适的参考妆容.

References:

- [1] Tong WS, Tang CK, Brown MS, Xu YQ. Example-based cosmetic transfer. In: Proc. Pacific Conf. on Computer Graphics and Applications, 2007. 211–218. [doi: 10.1109/PG.2007.31]
- [2] Scherbaum K, Ritschel T, Hullin M, Thormählen T, Blanz V, Seidel HP. Computer suggested facial makeup. Computer Graphics Forum, 2011,30(2):485–492. [doi: 10.1111/j.1467-8659.2011.01874.x]
- [3] Guo D, Sim T. Digital face makeup by example. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. IEEE, 2009. 73–79. [doi: 10.1109/CVPRW.2009.5206833]
- [4] Liao J, Yao Y, Yuan L, Hua G, Kang SB. Visual attribute transfer through deep image analogy. ACM Trans. on Graphics, 2017, 36(4):Article No.120. [doi: 10.1145/3072959.3073683]
- [5] Liu S, Ou XY, Qian RH, Wang W, Cao XC. Makeup like a superstar: Deep localized makeup transfer network. In: Proc. of the Int'l Joint Conf. on Artificial Intelligence. AAAI Press, 2016. 2568–2575.
- [6] Gatys LA, Ecker AS, Bethge M. Image style transfer using convolutional neural networks. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2016. 2414–2423. [doi: 10.1109/CVPR.2016.265]
- [7] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2016,39(4):640–651. [doi: 10.1109/TPAMI.2016.2572683]
- [8] Johnson J, Alahi A, Li FF. Perceptual losses for real-time style transfer and super-resolution. In: Proc. of the European Conf. on Computer Vision. Springer-Verlag, 2016. 694–711. [doi: 10.1007/978-3-319-46475-6_43]
- [9] Luan F, Paris S, Shechtman E, Bala K. Deep photo style transfer. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. IEEE, 2017. 6997–7005. [doi: 10.1109/CVPR.2017.740]
- [10] Blanz V, Vetter T. A morphable model for the synthesis of 3D faces. In: Proc. of the SIGGRAPH'99. ACM, 1999. 187–194. [doi: 10.1145/311535.311556]

- [11] Milborrow S, Nicolls F. Locating facial features with an extended active shape model. In: Proc. of the European Conf. on Computer Vision. Berlin, Heidelberg: Springer-Verlag, 2008. 504–513. [doi: 10.1007/978-3-540-88693-8_37]
- [12] Bookstein FL. Principal warps: Thin-plate splines and the decomposition of deformations. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1989,11(6):567–585. [doi: 10.1109/34.24792]
- [13] Xu L, Du Y, Zhang Y. An automatic framework for example-based virtual makeup. In: Proc. of the IEEE Int'l Conf. on Image Processing. IEEE, 2014. 3206–3210. [doi: 10.1109/ICIP.2013.6738660]
- [14] Li C, Zhou K, Lin S. Simulating makeup through physics-based manipulation of intrinsic image layers. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. IEEE, 2015. 4621–4629. [doi: 10.1109/CVPR.2015.7299093]
- [15] Jing Y, Yang Y, Feng Z, Ye J, Yu Y, Song M. Neural style transfer: A review. arXiv preprint arXiv:1705.04058, 2017.
- [16] Li Y, Fang C, Yang J, Wang Z, Lu X, Yang MH. Universal style transfer via feature transforms. In: Advances in Neural Information Processing Systems. NIPS, 2017. 386–396.
- [17] Chang H, Lu J, Yu F, Finkelstein A. Pairedcyclegan: Asymmetric style transfer for applying and removing makeup. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. IEEE, 2018. 40–48. [doi: 10.1109/CVPR.2018.00012]
- [18] Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proc. of the Int'l Conf. on Computer Vision. IEEE, 2017. 2242–2251. [doi: 10.1109/ICCV.2017.244]
- [19] Shan SG, Gao W, Chen XL. Facial feature extraction based on facial texture distribution and deformable template. Ruan Jian Xue Bao/Journal of Software, 2001,12(4):570–577 (in Chinese with English abstract). http://www.jos.org.cn/jos/ch/reader/view_abstract.aspx?file_no=20010412&flag=1
- [20] Kazemi V, Sullivan J. One millisecond face alignment with an ensemble of regression trees. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. IEEE, 2014. 1867–1874. [doi: 10.1109/CVPR.2014.241]
- [21] Wu LF, Shen LS, Kong X, Zhu XJ. Face segmentation based on curve fitting. Chinese Journal of Computers, 2003,26(7):893–896 (in Chinese with English abstract).
- [22] Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proc. of the European Conf. on Computer Vision. Springer-Verlag, 2018. 833–851. [doi: 10.1007/978-3-030-01234-2_49]
- [23] Xu G, Wang GZ, Chen XD. Free from deformation and its application. Journal of Computer Research and Development, 2010, 47(2):344–352 (in Chinese with English abstract).
- [24] Gong X, Wang GY. 3D face deformable model based on feature points. Ruan Jian Xue Bao/Journal of Software, 2009,20(3):724–733 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/3317.htm> [doi: 10.3724/SP.J.1001.2009.03317]
- [25] Wang J, Pan ZG, Ye CQ. An image deformation algorithm based on polygon of outline. Journal of Computer Research and Development, 1999,36(11):1358–1363 (in Chinese with English abstract).
- [26] Chen TQ, Schmidt M. Fast patch-based style transfer of arbitrary style. arXiv preprint arXiv:1612.04337v1, 2016.
- [27] Barnes C, Shechtman E, Finkelstein A, Goldman DB. PatchMatch: A randomized correspondence algorithm for structural image editing. ACM Trans. on Graphics, 2009,28(3):Article No.24. [doi: 10.1145/1576246.1531330]
- [28] Parkhi OM, Vedaldi A, Zisserman A. Deep face recognition. In: Proc. of the British Machine Vision Conf. 2015. [doi: 10.5244/C.29.41]
- [29] Pérez P, Gangnet M, Blake A. Poisson image editing. ACM Trans. on Graphics, 2003,22(3):313–318. [doi: 10.1145/1201775.882269]
- [30] Guibas L, Stolfi J. Primitives for the manipulation of general subdivisions and the computation of voronoi diagrams. ACM Trans. on Graphics, 1985,4(2):74–123. [doi: 10.1145/282918.282923]
- [31] Berger M. Geometry I. In: Universitext. Springer-Verlag, 1987.
- [32] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: Proc. of the Int'l Conf. on Learning Representations. ICLR, 2015. Article No.24.
- [33] Huang GB, Mattar M, Berg T, Learned-Miller E. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In: Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition. HAL, 2008.
- [34] Wolf L, Hassner T, Maoz I. Face recognition in unconstrained videos with matched background similarity. In: Proc. of the IEEE

- Conf. on Computer Vision and Pattern Recognition. IEEE, 2011. 529–534. [doi: 10.1109/CVPR.2011.5995566]
- [35] Aly HA, Dubois E. Image up-sampling using total-variation regularization with a new observation model. IEEE Trans. on Image Processing, 2005,14(10):1647–1659. [doi: 10.1109/TIP.2005.851684]
- [36] Dantcheva A, Chen C, Ross A. Can facial cosmetics affect the matching accuracy of face recognition systems? In: Proc. of the 5th IEEE Int'l Conf. on Biometrics: Theory, Applications and Systems. IEEE, 2012. 391–398. [doi: 10.1109/BTAS.2012.6374605]
- [37] Kingma DP, Ba LJ. Adam: A method for stochastic optimization. In: Proc. of the Int'l Conf. on Learning Representations. ICLR, 2015. Article No.11.
- [38] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang ZH, Karpathy A, Khosla A, Bernstein M, Berg AC, Li FF. Imagenet large scale visual recognition challenge. Int'l Journal of Computer Vision, 2015,115(3):211–252. [doi: 10.1007/s11263-015-0816-y]
- [39] Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Trans. on Image Processing, 2004,13(4):600–612. [doi: 10.1109/TIP.2003.819861]

附中文参考文献:

- [19] 山世光,高文,陈熙霖.基于纹理分布和变形模板的面部特征提取.软件学报,2001,12(4):570–577. http://www.jos.org.cn/jos/ch/reader/view_abstract.aspx?file_no=20010412&flag=1
- [21] 毋立芳,沈兰荪,孔潇,朱旭娟.基于曲线拟合的人脸分割算法.计算机学报,2003,26(7):893–896.
- [23] 徐岗,汪国昭,陈小雕.自由变形技术及其应用.计算机研究与发展,2010,47(2):344–352.
- [24] 龚勋,王国胤.基于特征点的三维人脸形变模型.软件学报,2009,20(3):724–733. <http://www.jos.org.cn/1000-9825/3317.htm> [doi: 10.3724/SP.J.1001.2009.03317]
- [25] 王进,潘志庚,叶澄清.一种基于轮廓多边形的图像变形算法.计算机研究与发展,1999,36(11):1358–1363.



黄妍(1993—),女,广东信宜人,硕士,主要研究领域为深度学习,机器学习.



张文生(1965—),男,博士,研究员,博士生导师,CCF专业会员,主要研究领域为人工智能,机器学习,大数据知识挖掘,概率图模型表示与推理,Deep Learning,精密感知与智能控制,三维数字物理仿真,嵌入式视频图像处理.



何泽文(1992—),男,博士生,主要研究领域为深度学习,机器学习,目标检测与跟踪.