软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2019,30(4):896−913 [doi: 10.13328/j.cnki.jos.005666]
©中国科学院软件研究所版权所有.
E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

基于区域敏感生成对抗网络的自动上妆算法*

包仁达[1], 庾 涵[1], 朱德发[1], 黄少飞[1], 孙 瑶[1], 刘 偲[2]

[1](中国科学院 信息工程研究所,北京 100093)

[2](北京航空航天大学 计算机学院,北京 100191)

通讯作者: 刘偲, E-mail: liusi@buaa.edu.cn

摘 要: 自动上妆旨在通过计算机算法实现人脸妆容的编辑与合成,隶属于人脸图像分析领域.其在互动娱乐应用、图像视频编辑、辅助人脸识别等多方面起着重要作用.然而作为人脸编辑任务,其仍难以在保证图像的编辑结果自然、真实的同时又很好地满足编辑需求,并且仍有难以精确控制编辑区域、图像编辑前后一致性差、图像质量不够精细等问题.针对以上难点,创新性地提出了一种掩模控制的自动上妆生成对抗网络,该网络利用掩模方法,能够重点编辑上妆区域,约束人脸妆容编辑中无需编辑的区域不变,保持主体信息.同时其又能单独编辑人脸的眼影、嘴唇、脸颊等局部区域,实现特定区域上妆,丰富了上妆功能.此外,该网络能够进行多数据集联合训练,除妆容数据集外,还可以使用其他人脸数据集作为辅助,增强模型的泛化能力,得到更加自然的上妆结果.最后,依据多种评价标准,进行了充分的定性及定量实验,并与目前的主流算法进行了对比,综合评价了所提方法的性能.

关键词: 生成对抗网络;自动上妆;人脸图像编辑;深度学习

中图法分类号: TP391

中文引用格式: 包仁达,庾涵,朱德发,黄少飞,孙瑶,刘偲.基于区域敏感生成对抗网络的自动上妆算法.软件学报,2019,30(4): 896−913. http://www.jos.org.cn/1000-9825/5666.htm

英文引用格式: Bao RD, Yu H, Zhu DF, Huang SF, Sun Y, Liu S. Automatic makeup with region sensitive generative adversarial networks. Ruan Jian Xue Bao/Journal of Software, 2019,30(4):896−913 (in Chinese). http://www.jos.org.cn/1000-9825/5666.htm

Automatic Makeup with Region Sensitive Generative Adversarial Networks

BAO Ren-Da[1], YU Han[1], ZHU De-Fa[1], HUANG Shao-Fei[1], SUN Yao[1], LIU Si[2]

[1](Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China)

[2](School of Computer Science and Engineering, Beihang University, Beijing 100191, China)

Abstract: Automatic makeup refers to the editing and synthesis of face makeup through computer algorithms. It belongs to the field of face image analysis, and plays an important role in interactive entertainment applications, image and video editing, and face recognition. However, as a face editing problem, it is still difficult to ensure that the editing result of the image is natural and satisfies the editing requirements. Makeup still has some difficulties such as precisely controlling the editing area is hard, the image consistency before and after editing is poor, and the image quality is insufficient. In response to these difficulties, this study innovatively proposes a mask-controlled automatic makeup generative adversarial network. Through a masking method, this network can edit the makeup area with emphasis, restrict the area that does not require editing, and maintain the key information. At the same time, it can separately edit the eye shadow, lips, cheeks, and other local areas of the face to achieve makeup on specific areas and enrich the makeup function. In addition, this network can be trained jointly on multiple datasets. In addition to makeup dataset, it can also use other face datasets as an aid to enhance the model's generalization ability and get a more natural makeup result. Finally, based on a variety of evaluation methods, more comprehensive qualitative and quantitative experiments are carried out, the results are compared with the other methods, and the

* 基金项目: 国家自然科学基金(U1536203, 61572493, 61876177)
Foundation item: National Natural Science Foundation of China (U1536203, 61572493, 61876177)
本文由"多媒体数据的知识关联与理解专题"特约编辑蒋树强研究员、刘青山教授、孙立峰教授、李波教授推荐.
收稿时间: 2018-04-16; 修改时间: 2018-06-13; 采用时间: 2018-09-30

performance of the proposed method is comprehensively evaluated.

Key words: generative adversarial nets; automatic makeup; face image editing; deep learning

随着社会经济和科技的飞速发展,社交网络、视频直播平台和视频播放平台得以迅速发展,海量用户不断产出各种图像数据.图像数据以其直观、信息量丰富的特点,在人们的日常生活中发挥着重要的作用.其中,人脸图像因其与人的紧密相关性,受到格外重视.本文研究的问题——自动上妆算法,即自动合成妆容,正是对人脸图像进行妆容编辑.自动上妆在娱乐互动、身份认证等多个领域具有广泛的应用.在娱乐互动领域,自动上妆算法可以融入到众多图像编辑软件、视频直播软件中,进一步丰富软件的功能、趣味性以及与用户的互动性.在身份认证领域,如刷脸支付、身份证校验等,人脸识别技术必须与其他验证方式搭配使用才能满足各类场景的安全性要求,也就是多因素验证.而自动上妆算法,可以辅助人脸识别技术,排除妆容因素对人脸识别产生的干扰.图 1 所示为自动上妆任务示意图,任意输入一张人脸图片以及某种妆容类型,通过自动上妆模型,就可以得到上妆后的效果图.

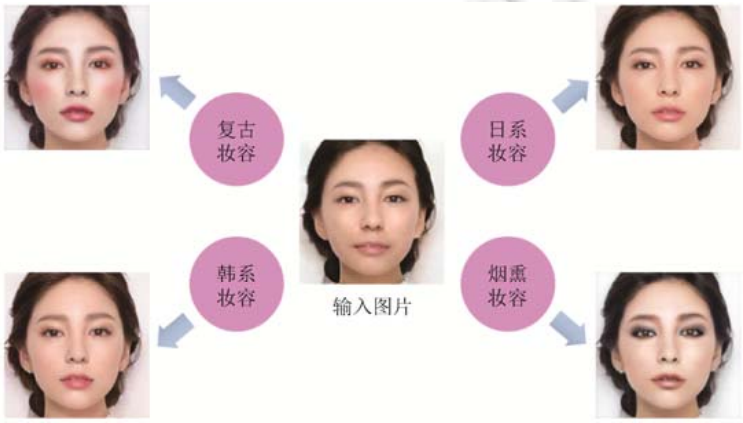

Fig.1 Automatic make-up task

图 1 自动上妆任务示意图

另一方面,随着人工智能深度学习领域的飞速发展,深度学习技术已在计算机视觉、语音识别、自然语言处理等多个研究领域取得了巨大的成功.其中,生成对抗网络更是在图像处理领域取得了令人瞩目的成绩.2014 年,蒙特利尔大学 GoodFellow 等学者发表了论文《Generative Adversarial Nets》,首次提出了 GANs(generative adversarial nets)这一概念,即生成对抗网络.近两年,生成对抗网络在大量不同的计算机视觉任务上取得了显著的成绩,例如:图像生成[1−7]、图像转换[8−12]、图像编辑[13]、超分辨率[14]、人脸图像合成[15−17]、表示学习[18−20]、图像修补[21]、未来预测[22]以及其他如视频生成[23]和 3D 数据场景分类[24]等等.

本文拟将 GANs 用于自动上妆问题的研究.与其他方法比较后可知:

(1) 生成对抗网络最大的优势来自于判别器的引入,让我们不用费尽心思地为一些难以直接用数学公式衡量好坏的任务设计损失函数.

(2) 在人脸图片编辑方面,如在人脸妆容属性编辑等实际应用中,GANs 可以通过模拟妆容特定属性的分布,实现很好的合成效果.

(3) 端到端的图像生成,只需设计一个整体的网络,不需要独立考虑各个部件的设计,模型结构简洁、直观.

(4) 完整地生成一张图片,自适应地为不同的人脸上妆,让整个妆容的搭配更加自然.

然而,作为人脸编辑问题,针对自动上妆任务设计生成对抗网络结构存在以下难点.

- 人脸妆容属性的编辑需要保持人物主体身份信息不变,不会改变人物的脸型、性别等决定人物身份的因素,也不会改变头发、配饰等无需编辑区域,只改变眼影、唇彩、腮红等上妆区域.而生成网络却往往将整张图

© 中国科学院软件研究所 http://www.jos.org.cn

像生成一遍,很难约束无需要改变区域的不变性.

- 一般的人脸编辑问题如改变年龄、改变表情等任务,只需要对人脸整体特征同时进行编辑,然而上妆任务还有局部编辑的需求,如单独针对眼影、嘴唇或脸颊中的某些部位进行特定的局部编辑.这就对自动上妆的方法提出了更高的要求.
- 可供训练的人脸妆容数据集少.训练网络得到的上妆后图片不够真实自然,有时有明显的人工痕迹.

针对以上几个难点,本文提出一种掩模控制的自动上妆生成对抗网络模型,该模型的主要贡献包括:

- 提出掩模控制方法,在编辑人脸妆容属性时,通过掩模约束不需要改变的区域不发生明显的变化.例如,保持头发、背景、配饰等非妆容区域不变,只改变眼影、唇彩、腮红等上妆区域.同时,网络能够保持脸型、性别等决定人物身份的信息不变,保持主体信息.
- 能够独立编辑人脸的眼影、嘴唇、脸颊中的某些局部区域,实现特定区域上妆,丰富了上妆功能.
- 能够进行多数据集联合训练,除了化妆数据集外,还可以使用相关人脸数据集作为辅助,使模型泛化能力更强,生成的上妆结果更加自然.

本文提出的掩模控制的自动上妆生成对抗网络模型包括 3 个子网络:(1) 生成网络;(2) 判别网络;(3) 掩模生成网络,模型的框架如图 2 所示.

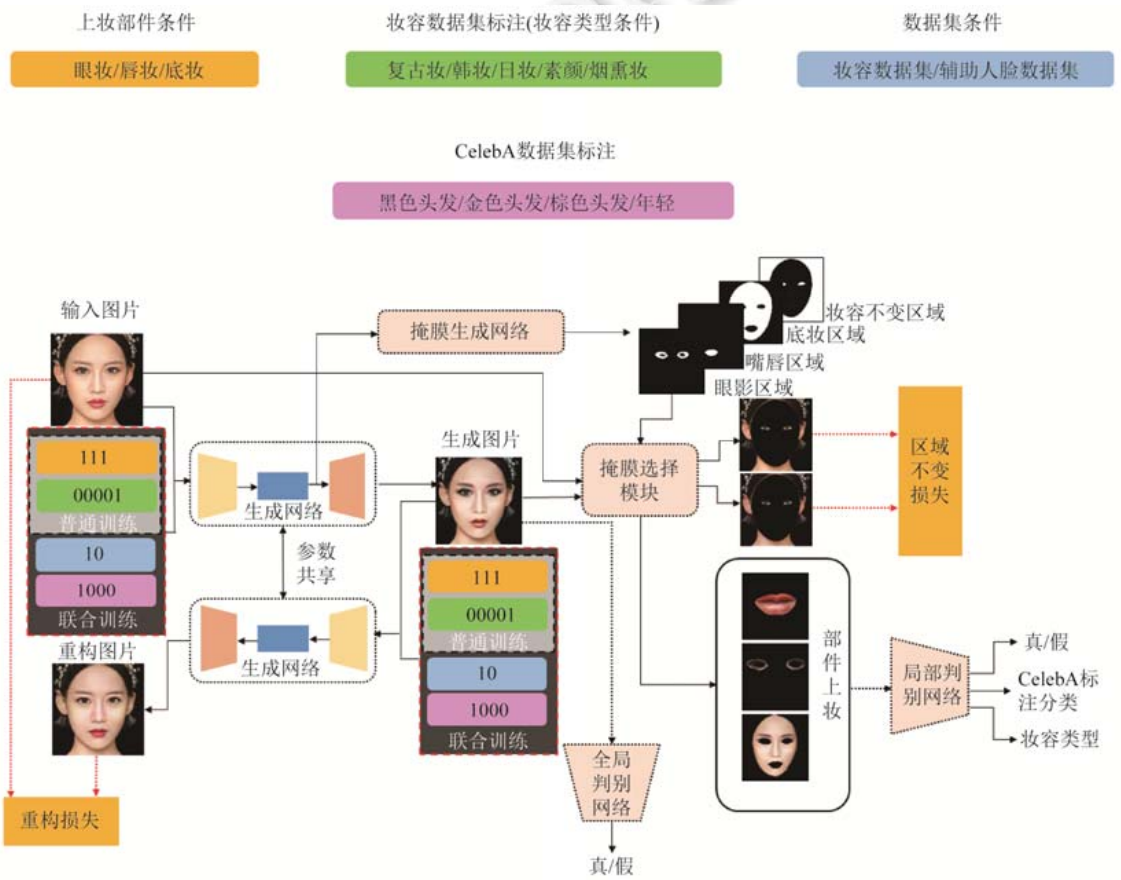

Fig.2 Framework of automatic makeup generative adversarial network using mask

图 2 掩模控制的自动上妆生成对抗网络模型框架图

其中,掩模选择模块具体如图 3 所示.

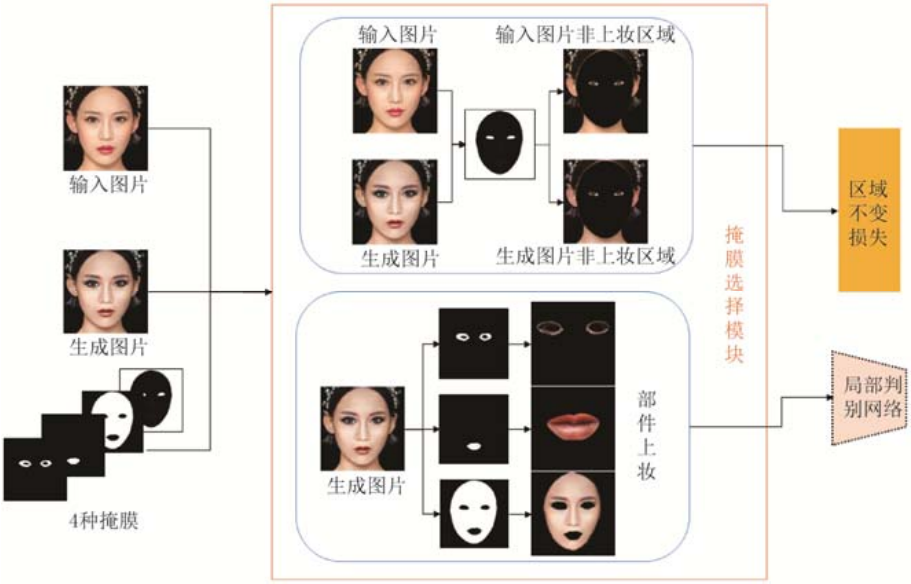

Fig.3 Structrue of mask-choosen module

图 3 掩模选择模块结构

本文第 1 节介绍相关工作.第 2 节介绍实验用到的妆容数据集.第 3 节描述本文提出的掩模控制的自动上妆生成对抗网络模型.第 4 节展示提出的模型与现有主流的生成对抗网络模型在妆容数据集上的实验结果.第 5 节总结工作,并对未来值得关注的研究方向进行初步探讨.

1 相关工作

基于传统方法的自动上妆算法:Guo 等人[4]提出解决美妆迁移任务的方法.他们的方法类似于物理化妆,在保留脸部结构的同时改变颜色和肤色细节.更确切地说,他们首先将素颜人像和参考的带妆人像分解为 3 个层次:面部结构层、皮肤细节层和颜色层.之后,他们将信息从图像的每个图层迁移到另一个图像的相应图层.这一方法的主要优点在于,其仅需要一个参考的带妆人像图像,较为方便和实用.而缺点是它需要将带妆图片的人脸形状扭曲形变成素颜图片的人脸形状,并且需要分成多个区域分别进行处理,处理速度慢,耗费时间长,而且一次只能合成一种类型的妆容.除此之外,Scherbaum 等人[5]提出使用变形人脸模型来进行面部化妆工作.但是,该方法的不足在于需要收集同一个人妆前妆后的成对的图片,要求同一个人妆前妆后图片的基本姿态不变,同时,光照、角度等外在因素也不可有太大的改变,这种数据集的收集有很大难度.Tong 等人[6]提出了一种“美容迁移”程序,将实例对中的妆容风格逼真地转移到另一个人的脸上.但是同样地,该工作对化妆前后图片对也有要求,限制了系统的实用性.Liu 等人[7]提出了一个根据素颜图片自动推荐妆容图片并在素颜图片上进行妆容合成的系统,但这一工作的主要研究点在于妆容推荐.综上,基于传统方法的自动上妆算法对妆前妆后图片的要求比较严格,得到的妆容合成图片的效果也不是很自然.

基于深度学习的图像编辑算法:自动上妆算法属于图像编辑算法的一种,可以与其他图像编辑算法相互借鉴.与自动上妆算法联系最为紧密的两种图像编辑算法是图像风格迁移和图像到图像的转换.图像风格迁移是 2015 年 Gatys 等人[25]提出的一个研究方向.图像风格迁移输入内容图像和风格图像,输出的图像既需要保持内容图像的内容,又需要具有风格图像的风格.之后又出现了许多工作,如 Li 等人[26]即着重于提升算法速度,而 Fiser 等人[27]着重于图形学方法,实现了基于真实图像的面部动画合成.Liao 等人[28]的利用多层特征的约束,在多种迁移任务上都取得了较好的效果.图像到图像的转换则是另一种视觉和图形问题,它的目标是通过成对的图像训练数据集来学习输入图像和输出图像之间的对应关系,即从源域图像转换成输出域图像.图像转换的具

© 中国科学院软件研究所 http://www.jos.org.cn

体内容包括从标签图像得到场景图像[29],从卫星照片得到地图,从黑白图像得到彩色图像,从草图得到具体的图片[30]等.这类工作需要建立起图片域与图片域之间的转换公式,主要针对的是不同域之间的相互转换,而非图像风格迁移中图像个体之间的风格转换.由于生成对抗网络提出了对抗性损失的想法,这种损失迫使生成的图像原则上与真实图像无法区分,所以生成对抗网络在图像生成任务中展现出了极为出色的能力.

基于深度学习的自动上妆算法:近年来,自动上妆技术相关研究引起了更多的关注,然而与其他图像编辑任务相比仍然相当有限,但是随着深度学习领域研究的不断发展,除了传统的自动上妆算法,近几年也出现了多种基于深度学习的自动上妆算法.Li 等人[31]引入对抗网络生成素颜图像用于妆容不变人脸验证.Wang 等人[32]提出了一种基于局部约束字典学习的妆容检测和卸妆框架.Li 等人[33]针对此问题提出了另一种图像分解的方法.妆容迁移也是另一个有吸引力的应用,其与图像风格迁移相似,目标是给定带妆图像作为参考,将带妆图像的妆容迁移到源图像得到生成图片.其中较有代表性的是 Liu 等人[34]提出的一种根据素颜照片推荐合适妆容且将妆容迁移到人脸上的方法.该网络的模型基于 VGG19,输入素颜图像和参考的带妆图像,输出妆容合成后的结果.整体来说,该模型也属于生成模型,但损失函数是人工设计的,与生成对抗网络相比,该模型不能一次性输出多种妆容合成的结果,生成过程也是迭代的而非一次性的,速度比较慢.此外,Chang 等人[35]基于 GANs 和图像分割,提出了循环约束的妆容风格迁移方法,取得了较好的效果.

本文主要解决的问题有别于妆容迁移任务,更接近于多域的图像到图像的转换.在妆容迁移任务中,图像没有定义妆容属性类别,其特点在于图片对图片的转换,可以将多种多样丰富的妆容类别迁移到目标图像,但在迁移时,算法必须要有指定的参考图像.而本文主要的任务为多域的图像转换,即将人脸图像按照指定的类别进行编辑,使原本属于某个域的图像经过编辑后拥有另一个域的妆容特征.其特点在于不需要参考图像,仅需指定一个类别即可.如将某张素颜照片编辑为烟熏妆,此时并不需要某个烟熏妆的参考图片,输入烟熏妆类别对应的属性编码,算法即可自动为素颜照片编辑出相应的烟熏妆效果.

2 妆容数据集介绍

现阶段生成对抗网络对训练数据图片有着较高的要求,如果想要生成质量较好的数据,则对训练数据的数量和质量都有一定要求.对于本文中的自动上妆任务,我们利用了最近新提出的妆容数据集.该数据集基于妆容特点进行了分类标准的划定,使得每种妆容的特点突出,具有排他性.数据集中的图片都进行了妆容类别的标注,使得数据集标签定义精确,类内差距小,类间区别明显.

2.1 数据集类别和数量

图 4 所示为亚洲妆容数据集示意图,其中,图 4(a)所示日系妆容的特点:底色白皙,眉形自然,眼影粉底等妆容很淡,唇色不重,运用较浅的腮红进行修颜.图 4(b)所示韩系妆容的特点:大多为平眉,眉形较粗,眼妆很淡,较少使用浓重的腮红修颜.图 4(c)所示亚洲素颜的特点:基本未化妆,五官平面化.图 4(d)所示亚洲烟熏妆的特点:眼线和眼影在眼窝处漫成一片,眼影面积较大,唇色明显,以黑灰色为主色调.图 4(e)所示复古妆容的特点:一字粗眉,腮红、唇膏、眼影颜色相同或相近,都为粉色或橘色的淡色系.

(a) 日系妆容 (b) 韩系妆容 (c) 亚洲素颜 (d) 亚洲烟熏妆 (e) 复古妆容

Fig.4 Examples of Asia makeup dataset

图 4 亚洲妆容数据集示意图

© 中国科学院软件研究所 http://www.jos.org.cn

本文实验中,亚洲妆容数据集分为训练集和验证集,其中,训练集一共 5 374 张,验证集一共 361 张,具体情况见表 1.

Table 1 Details of Asia makeup dataset

表 1 亚洲数据集详情

	复古妆容	韩系妆容	日系妆容	亚洲素颜	亚洲烟熏妆	总计
训练集	1 260	1 265	1 505	1 064	640	5 374
验证集	69	60	76	95	61	361

图 5 所示为欧洲妆容数据集示意图,其中,图 5(a)所示彩妆妆容的特点:突出深邃而色彩饱和度高的眼影,哑光腮红,唇彩多为红色系等较为浓重的颜色.图 5(b)所示欧洲素颜的特点:基本未化妆,五官立体化.图 5(c)所示欧洲烟熏妆的特点:在亚洲烟熏妆的特点上,眼窝更加深邃,唇彩、眉毛等妆容更加立体和夸张.

(a) 彩妆 (b) 欧洲素颜 (c) 欧洲烟熏妆

Fig.5 Examples of Europe makeup dataset

图 5 欧洲妆容数据集示意图

本文实验中,欧洲妆容数据集分为训练集和验证集,其中,训练集一共 3 815 张,验证集一共 422 张,具体情况见表 2.

Table 2 Details of Europe makeup dataset

表 2 欧洲数据集详情

	彩妆	欧洲素颜	欧洲烟熏妆	总计
训练集	1 341	1 095	1 379	3 815
验证集	148	121	153	422

3 掩模控制的自动上妆生成对抗网络模型

掩模控制的自动上妆生成对抗网络包括 3 个子网络:生成网络 G 网络、判别网络 D 网络和掩模生成网络 S 网络,结构如图 2 所示.网络的输入包含两部分,分别为待上妆人脸图片和条件编码.待上妆人脸图片是 256×256 像素的三通道 RGB 图像.条件编码包括妆容标签、上妆区域标签以及辅助数据集标签和掩码条件标签,其中,后两者用于多数据集联合训练,将在第 3.6 节中加以详细介绍.妆容标签为 One-hot 编码,以亚洲数据集为例,即以五维向量表示复古妆、韩式妆容、日式妆容、素颜、烟熏妆 5 种类别,如(1,0,0,0,0)即代表复古妆.上妆区域标签使用 0/1 向量的形式进行编码,0 代表不上妆,1 代表上妆,通过三维向量表示是否在眼影、嘴唇、脸颊这 3 个区域上妆,如(0,1,1)即代表在嘴唇和脸颊区域上妆,而眼影区域不上妆,通过改变上妆区域标签,即可控制不同区域是否上妆.

3.1 生成对抗网络简介

生成对抗网络:2014 年,蒙特利尔大学 GoodFellow 等学者发表了论文《Generative Adversarial Nets》,并在此文中首次提出了 GANs,即生成对抗网络.典型的生成对抗网络模型由两个模块组成:生成器和判别器.判别器学习区分真假样本,生成器学习如何产生以假乱真的假样本.通常情况下,训练一个深度学习模型都需要损失函数或者有效的评价模型好坏的评价指标.损失函数一般是人为设计的,其在反向传播中更新模型参数.而生成对

© 中国科学院软件研究所 http://www.jos.org.cn

抗网络通过生成器与判别器之间的对抗学习来训练模型,从而合成与所期望图像相似的图像.如图 6 所示,输入一个随机向量 z,经过生成器 G,得到一张图片 $G(z)$,生成器的目的是让 $G(z)$看起来和真实图片一样,所以需要训练一个判别器 D,给出真实图片的样本 x 作为参考,对于真实的图片样本,让 $D(x)$的结果尽量接近 1,对于生成图片 $G(z)$,让 $D(G(z))$的结果尽量接近 0.这里的判别器就充当了评价图片的角色,它能判别图片的真伪,换句话说,就是判别图片来自于真实图片数据集还是生成器生成的假数据.而生成器的任务就是尽可能地生成真假难辨的图片来欺骗判别器,让 $D(G(z))$接近 1.D 和 G 的目标是相反的,这就是生成对抗网络被称为“对抗”的原因.

Fig.6 Framework of GANs

图 6 生成对抗网络模型框架图

条件生成对抗网络:典型的生成对抗网络模型由生成器和判别器组成.判别器学习区分真假样本,生成器学习如何产生以假乱真的假样本.生成对抗网络中生成器的输入是随机向量.在此基础上,为解决 GANs 生成图片不可控的问题,条件生成对抗网络 CGAN 被提了出来.通过向生成网络和判别网络的输入层加入额外的信息约束,CGAN 能够按照输入的条件信息,控制网络产生期望的图像.近几年的相关工作表明:带有类别信息的判别器和生成器能够根据类别信息生成样本[11,36,37].最新的研究还包括根据给定的文字描述来生成特定的图像[38,39].此外,条件生成对抗网络的思想还被成功地应用在域转换[40,41]、超分辨率[14]等领域.

3.2 自动上妆模型生成对抗网络

生成网络 G:生成网络是基于条件生成网络设计的.如图 2 所示,条件生成网络由下采样卷积模块、残差卷积模块以及上采样卷积模块构成.网络输入为人脸图像及目标标签信息,输出为上妆后的人脸图片.将复古妆容的输入图片 X 与期望某个其他属性,如图中的烟熏妆容属性(假设其对应的编码为 Y)送入生成网络 G,G 所生成的图片 $G(X,Y)$既包含 X 的主体信息,又与编码 Y 对应的烟熏妆容属性相符.生成网络 G 主要依靠判别网络 D 及第 3.3 节中的区域不变约束加以监督.此外,其主体信息的保持通过约束 $G(G(X,Y),Y_X)=X$ 实现,其中,Y_X 为生成图片本身的条件编码,图 2 所示为复古妆容.即:将复古妆容的输入图片 X 与目标域属性编码 Y(期望的烟熏妆容属性)级联送入生成网络 G,得到符合烟熏妆容属性的生成图片 $G(X,Y)$,再将生成图片 $G(X,Y)$与源域属性编码 Y_X(输入图片的复古妆容属性)送入生成网络 G,得到重构的输入图片,重构的输入图片应与输入图片保持循环一致性.这一约束也被称为重构约束 L_{rec},即最小化如下目标函数:

$$L_{rec}=E_{(X,Y)\sim P_{x,y},Y\sim p_y}[\| X-G(G(X,Y),Y_X)\|_1] \tag{1}$$

判别网络 D:本文中的判别网络分为两类,一类为全局判别网络 D_{global},另一类为局部判别网络 D_{local}.全局判别网络实际上是一个二分类模型,输入为整张人脸图像,输出为图片真假的分类结果,真实图片对应的标注为“真”,生成器生成的图片对应的标注为“假”.其主要训练目标就是能够正确地判断图片的真假,以向生成网络提供对抗性的监督,从全局特征层面帮助生成网络输出更加自然、真实的图片.局部判别网络则着重于局部特征,

© 中国科学院软件研究所 http://www.jos.org.cn

其输入为脸颊、眼影、嘴唇等上妆区域的局部图像,共有两个训练目标,一是判断局部图像的真假,以生成更加真实的局部上妆结果,二是进行属性分类,即判断输入的局部图片的妆容属性类别,从而能够监督生成网络生成图片的属性是否正确,控制生成网络对局部区域上妆的类别.在判别网络的实际训练过程中,当输入的是真实图片时,判别网络既判断图片的真假,也判断图片的妆容属性类别;当判别网络的输入是生成图片时,判别网络只判断生成图片的真假,而不判断生成图片的属性类别.判别网络的具体优化目标函数将在第 3.4 节中加以详细介绍.

3.3 掩模保持不可变区域

人脸妆容编辑往往只需对图像中局部区域进行修改,同时希望无需编辑的区域在编辑前后保持不变.而生成网络需要生成全局的图像,如果不进行约束,则无需编辑区域难以与原始输入图片保持一致性.例如,在上妆过程中,网络需要保持人物主体身份信息不变,即不会因为上妆改变人物的脸型、性别等决定人物身份的因素,也不会改变诸如头发、背景等非上妆区域,只需编辑眼影、唇彩、腮红等需要上妆的区域.为了解决以上问题,本文基于妆容编辑任务,定义了人脸图像中的可变区域和不可变区域,也可称其为上妆区域和非上妆区域.如眼影、嘴唇、脸颊等区域是可变区域,头发、背景、牙齿、瞳孔等区域是不可变区域.在编辑过程中对不可变区域进行编辑前后一致性的约束,即可使网络集中编辑可变区域特征,保持其他区域信息不变.

可变区域和不可变区域由掩模区分,为得到人脸图像的掩模,我们引入了掩模生成网络和掩模选择模块.

掩模生成网络:掩模生成网络实际上是一个人脸妆容部件语义分割网络,其对人脸图像进行分割得到妆容部件语义分割图,分割图包括:(1) 妆容不变区域,用于指示自动上妆中无需编辑的区域(包括背景,头发,左右耳朵,脖子,左右眼睛,牙齿等);(2) 眼影区域(上妆中眼影的区域,眉毛之下,眼睛周围一圈的皮肤),用于眼妆编辑;(3) 底妆区域(脸部除去眼睛区域、嘴唇区域、眼影区域之外的皮肤区域),用于底妆编辑,包括腮红、粉底、磨皮等; (4) 嘴唇区域(包括上下嘴唇),用于唇妆编辑.实验中,将生成网络 G 上采样之前的特征输入掩模生成网络,得到上述 4 种妆容部件语义分割结果,对应 4 种掩模:(1) 妆容不变区域掩模;(2) 眼影区域掩模;(3) 底妆区域掩模;(4) 嘴唇区域掩模.

掩模选择模块:如图 3 所示,掩模选择模块的功能是选择掩模生成网络的掩模结果.需要注意的是,掩模选择模块在第 3.4 节分部件妆容编辑中还会被用到.掩模选择模块根据输入的上妆部件条件,从眼影区域掩模、底妆区域掩模、嘴唇区域掩模中进行选择,完成不同部件的上妆.值得注意的是,可以同时选择 3 种掩模、3 种掩模中的两种或者一种.而妆容不变区域掩模则与未被选择的上妆部件掩模取并集,作为人脸上妆不变区域的监督,约束生成网络生成的图像中非上妆区域与原图相应区域保持一致.具体来说,妆容不变区域掩模与输入图片和生成图片分别进行操作,得到输入图片中的不变区域和生成图片中的不变区域,进而计算区域不变损失.此外,训练集的样本标注是通过预先训练好的图像语义分割模型,对训练集中所有人脸图像预测人脸部件语义分割图得到的.

用 X 表示原始输入图片, $\widetilde{X_Y}$ 表示生成图片,图像 X 在特定属性 Y 的条件下的可变区域由 V_{XY} 表示,不可变区域由 $\overline{V}_{XY}$ 表示.约束生成图片 $\widetilde{X_Y}$ 和原始输入图片 X 的不可变区域保持一致,即用掩模 M_{XY} 的 0、1 来描述图像 X 特定属性 Y 的可变区域 V_{XY}、不可变区域 $\overline{V}_{XY}$, 如下式:

$$M_{XY}(i,j)=\begin{cases}1, & (i,j)\in\overline{V_{XY}}\\ 0, & (i,j)\in V_{XY}\end{cases} \tag{2}$$

因此,定义图像 X 属性 Y 的损失函数为

$$L_{mask}(\widetilde{X_Y},X)=\left\|\widetilde{X_Y}\times M_{XY}-X\times M_{XY}\right\|_1 \tag{3}$$

其中,×操作代表像素级别的乘法操作,如 $\widetilde{X_Y}\times M_{XY}$ 即代表图片 $\widetilde{X_Y}$ 中每个像素点与掩模 M_{XY} 中对应点相乘.

假设只对一个属性进行编辑,图像掩模记为 M_X,用以下目标函数优化生成网络 G:

$$L_{mask}=E_{X\sim p_X,Y\sim p_Y}\left\|G(X,Y)\times M_X-X\times M_X\right\|_1 \tag{4}$$

© 中国科学院软件研究所 http://www.jos.org.cn

3.4 分部件妆容编辑

生成网络往往只能生成整张图像,无法单独编辑图像的局部区域,而人脸妆容编辑任务又存在着分区域上妆的需求.为此,我们提出了部件条件编码、掩模方法以及局部/全局判别网络,共同实现分部件妆容编辑功能.

在第 3.3 节中提到的掩模生成网络负责生成人脸部件分割图,如图 3 所示,掩模选择模块按照输入网络的部件条件信息,依据掩模结果从生成图像中选择需要上妆的部件对应的分割图,再将分割图送入局部判别网络 D_{local} 进行图像真假的判断以及属性分类,从而在图像真实性和属性正确性两个角度上同时监督生成网络 G 的训练,希望生成网络在部件条件信息和妆容条件信息的共同指导下,对每个需要上妆的部件进行正确的妆容属性编辑.

为了向生成网络 G 提供有效监督,局部判别网络 D_{local} 需要同时进行训练,其优化目标函数如下:

$$L_{D_{local}} = -L_{adv_local} + \lambda_{cls} L^{r}_{cls_local} \tag{5}$$

$L^{r}_{cls_local}$ 表示 D 网络读入真实数据和生成数据时分类的损失函数,λ_{cls} 为其权重,实验时设置为 1.分类损失函数采用常见的交叉熵的形式.L_{adv_local} 表示监督图像真实性的真假分类器的对抗损失函数,依据 WGAN-gp[42] 的方法,近似优化两个数据分布的 Wasserstein 距离,其具体形式为

$$L_{adv_local} = \sum_{i=1}^{N} E_{X\sim P_X, Y\sim P_Y}[D_{local}(X^{i}_{part}) - D_{local}(G(X^{i}_{part}, Y))] + \lambda E_{\hat{X}\sim p_{\hat{X}}}\left[\left(\left\|\nabla_{\hat{X}} D_{local}(\hat{X}^{i}_{part})\right\|_1 - 1\right)^2\right] \tag{6}$$

其中,N 为上妆部件数量, X^{i}_{part} 表示第 i 个需要上妆部件对应的局部图像,梯度惩罚项的系数λ是超参,实验设为 10. $\hat{X}^{i}_{part}$ 是 X^{i}_{part} 和 $G(X^{i}_{part}, Y)$ 连接线段上的随机一点,实验中,在此线段上均匀取值.

除局部判别网络外,生成网络输出的整体人脸图像会被送入全局判别网络 D_{global},该网络只需要判断图像的真假而不必进行分类,目的是保证生成图像的整体真实、自然.所以其优化目标函数为

$$L_{D_{global}} = -L_{adv_global} \tag{7}$$

L_{adv_global} 的具体形式与局部判别网络的对抗损失函数类似,唯一的区别是输入为整张图像而非局部图像,这里不再赘述.综上所述,判别网络的优化目标函数为

$$L_D = L_{D_{local}} + L_{D_{global}} \tag{8}$$

3.5 整体流程

网络使用端到端的生成模式,输入原始待上妆图片和妆容条件、部件条件标签,即可在保持主体身份信息的同时,在指定区域完成目标妆容的编辑,同时尽可能地保持非上妆区域不变.

上妆前后的人脸身份信息保持功能由第 3.2 节中的重构约束 L_{rec} 实现.掩模生成网络负责生成人脸部件分割结果.而掩模选择模块有两个功能,一是在分部件妆容编辑过程中从人脸各个部件的分割结果中挑选需要编辑的部件,送入局部判别网络,局部判别网络再进行局部图片真实性、属性类别正确性的监督;二是在约束非上妆区域不变时,根据上妆部件条件信息,从人脸部件分割结果中选择、合并生成非上妆区域的掩模,约束非上妆区域的一致性.此外,全局判别网络负责监督整体图像生成得自然、真实.

综上所述,网络总体的优化目标函数为

$$\begin{cases} L_D = -L_{adv_global} - L_{adv_local} + \lambda_{cls} L^{r}_{cls_local} \\ L_G = L_{adv_global} + L_{adv_local} + \lambda_{cls} L^{f}_{cls_local} + \lambda_{rec} L_{rec} + \lambda_{mask} L_{mask} \end{cases} \tag{9}$$

其中,λ_{cls}、λ_{mask} 和λ_{rec} 是权重超参,实验时分别设置为 1、10、10,L_{rec} 是重构约束,L_{adv_Global} 和 L_{adv_Local} 为全局对抗损失和局部对抗损失函数. $L^{r}_{cls_Local}$ 表示在训练 D 网络的过程中,D 网络读入真实数据时分类的损失函数,L_{mask} 为不变区域一致性的约束.以上损失函数均在第 3 节中进行了解释.而前文中未提到的 $L^{f}_{cls_Local}$ 实际上是在训练 G 网络时,将 G 网络生成的图片送入 D 网络进行分类的损失函数,该损失函数利用 D 网络的属性分类器,监督 G 网络生成符合属性要求的生成图片.

算法的整体训练流程如下,将原始待上妆图片和妆容条件、部件条件标签共同输入生成网络得到生成图

© 中国科学院软件研究所 http://www.jos.org.cn

片,再将生成图片和原始条件送回生成网络得到重构图片,进行重构监督.同时,通过掩模生成和选择模块挑选出生成图片和原始图片各自的不可变区域,进行区域不变的监督.此外,还有全局判别网络和局部判别网络对图片真实性、属性编辑正确性进行监督.

3.6 多数据集联合训练

搜集、标注较大的特定数据集需要花费大量的人力物力.而一些现有人脸数据集虽没有标注妆容属性,但标注了人脸相关属性.如果能提出一种方法,在标注一个较小的数据集的基础上,利用现有的相关标注的大数据集,如拥有 20 万张图片的 CelebA[43]数据集辅助训练,将极大地提升图片生成的质量.用大数据集上的数据训练,虽然标注不一样,例如 CelebA 数据集包括 40 个与人脸属性相关的标注,包括头发颜色、性别、年龄等属性标注,但是网络能在海量数据中学习到更鲁棒的特征,从而在目标数据集上有很好的表现能力,同时提高网络泛化能力.

因此,我们提出自动上妆任务与相关人脸属性编辑任务联合训练的方法.有妆容数据库 A 和辅助人脸数据库 B(这里是 CelebA 数据集),在训练过程中,两库交替训练.每次迭代分两批数据训练,第 1 批数据来自妆容数据库 A,第 2 批数据来自辅助人脸数据库 A.设 A 库的条件编码为 Y_A,B 库的条件编码为 Y_B,用掩码向量 Y_{mask} 表示数据来源于哪个数据库,Y_{mask} 的长度为数据库的个数,在本实验中,[1,0]和[0,1]分别表示数据来自 A 库和 B 库.综上,设计条件编码 $Y=[Y_A,0,Y_{mask}]$ 或 $Y=[Y_A,0,Y_{mask}]$. 对应的 D 网络的输出增加一个数据库 B 上的属性分类器.当训练 A 库时,D 网络忽略 B 库上的属性分类器,监督图像真假分类器和 A 库上的属性分类器;反之,当训练 B 库时网络忽略 A 库上的属性分类器,监督图像真假分类器和 B 库上的属性分类器.

4 实验结果分析

本节首先介绍最近提出的基于生成对抗网络的图像到图像转换模型 CycleGAN 和 StarGAN,然后介绍自动上妆任务的评价指标,最后将基于 CycleGAN 和 StarGAN 得到的自动上妆结果作为基准实验,得到定性结果和定量结果,再与本文提出的掩模控制的自动上妆生成对抗网络模型得到的结果作这两方面的分析与比较.

4.1 基准实验模型CycleGAN和StarGAN简介

Zhu 等人在 2017 年提出了 CycleGAN 网络[9],CycleGAN 不需要成对的图片就能够实现图片从源域 X 到目标域 Y 的转换.CycleGAN 的目标是学习一个映射 G,使得生成器生成的图片集 $G(X)$的概率分布与目标域的图片集 Y 的概率分布一致.同时,CycleGAN 引入了一致性损失来加强约束,即通过建立一个反向映射 $F:Y\to X$ 以及引进一个循环一致性损失 $F(G(X))\approx X$ 来保证图像的主体信息在编辑过程中不发生较大变化.

不过,CycleGAN 仅能够实现两个域之间的图像到图像的转换,若有多个类别(域)的数据,则需要单独对每一对图像域转换问题建立不同的生成网络,导致了模型训练的低效率和不可扩展性.而 StarGAN[44]可以只用一个模型实现多领域的转换.StarGAN 网络是一个统一的模型框架,网络可以在有多个域的数据集上进行训练,并且只需训练一个单独的模型即可编辑多种属性,将输入图片转换到任意一种所期待的目标图片域.

4.2 实验设置

G 和 D 网络交替优化,G 网络每更新 1 次,D 网络更新 5 次,以获取更好的监督.网络结构遵循 StarGAN 的设计,G 网络有两个步长为 2 的卷积层进行下采样,6 个残差模块以及两个反卷积层进行上采样.D 网络如第 3.4 节所述,使用了 WGAN-gp 优化两个数据分布的 Wasserstein 距离.训练的 batch size 设置为 10,共训练 200 个 epoch,学习率设置为 0.000 1.此外,掩模生成网络为图像分割网络,为了运行时间及空间的需求,实验中删减了模型的部分参数.掩模生成网络的实验结果在下文中加以详细分析.

4.3 掩模生成网络实验结果

掩模生成网络在 Deeplab v2 网络[45]的基础上删减了部分参数,以便于取得更小的内存占用和更快的运行时间.掩模的分割效果将直接影响后续的掩模选择模块以及最终的妆容的编辑质量,为了验证掩模生成网络的

© 中国科学院软件研究所 http://www.jos.org.cn

性能,本文利用 mean IoU,对掩模生成网络进行了定量实验分析,实验结果见表 3,掩模生成网络妆容部件语义分割如图 7 所示.

Tabel 3 Quantitative experiment results of mask generation network

表 3 掩模生成网络定量实验结果

	眼影	嘴唇	面部	其他区域	Mean IoU
分割精度(IoU)	0.746	0.769	0.889	0.853	0.815

Fig.7 Face makeup part segmentation

图 7 掩模生成网络妆容部件语义分割示意图

4.4 定性实验结果

(1) CycleGAN、StarGAN 以及本文所提方法在数据集上的上妆结果比较

实验使用了第 2 节介绍的数据库,分为 5 种或 3 种妆容属性.同时,本文使用了 CycleGAN 和 StarGAN 这两种效果较好的 GANs 图像编辑方法作为基准实验,因为 CycleGAN 只能做两个域之间的转换,所以在所有的属性对之间均训练了模型,以实现多领域之间的图像编辑效果.实验参数设置均是使用文章默认的参数.图 8 和图 9 所示为我们的方法与两种基准实验的定性对比结果.

Fig.8 Automatic makeup result of CycleGAN and StarGAN and our model (Asia makeup dataset). The first column shows the input image, next five columns show the makeup results of different makeup style

图 8 基于 CycleGAN 和 StarGAN 和本文模型的自动上妆结果比较(亚洲妆容数据集).
第 1 列是原始输入图片,后 5 列是 5 种不同妆容的上妆结果

CycleGAN 沿用 LSGAN[46]的思想,使用了最小二乘模型作为损失函数,模型训练比较稳定,生成图片质量较好.不过,其只能做两个域之间的图像转换,所以一次性只能做两种类型妆容之间的转换.StarGAN 是较新的 GAN 模型,能够实现图像多属性之间的相互转换,其利用了 WGAN-gp、PatchGAN、循环保持身份信息、*D* 网

© 中国科学院软件研究所 http://www.jos.org.cn

络分类等特点,生成的图像质量较好.但是,由于上述两种方法均未进行额外约束,其上妆细节和背景生成的效果不佳.我们提出的方法,即掩模控制的自动上妆生成对抗网络,在编辑人脸妆容属性时,只对图像中局部妆容区域作了修改,能够约束无需改变的区域不变换,例如第 3 行中背景及头发颜色特征保持得比较好.另外,本文提出的模型的效果图更加自然,图像中网格、噪声等更少,图像更加清晰.

Fig.9 Automatic makeup result of CycleGAN and StarGAN and our model (Europe makeup dataset). The first column shows the input image, next three columns show the makeup results of different makeup style

图 9 基于 CycleGAN 和 StarGAN 和本文模型的自动上妆结果比较(欧洲妆容数据集).
第 1 列是原始输入图片,后 3 列是 3 种不同妆容的上妆结果

(2) 有无掩模控制的结果比较

人脸妆容编辑中,简单来说,妆容属性对应的可变区域是人脸,除此之外均是不可变区域.如图 10 所示,第 1 列是待编辑图片,第 2 列是在掩模监督下使用原始图片及原始标签重构图片得到的结果,第 3 列是在没有掩模监督情况下直接重构原始图片得到的结果.

观察重构图中不可变区域后发现,如眼睛瞳孔、头发、牙齿、饰物、背景等不可变区域在有掩模监督训练下的质量更加自然.例如第 2 行红圈处掩模法重构图像的头发、背景等颜色保持得更加接近原始图片.第 1 行第 3 列中的右下角红圈处,可以看见无掩模生成图片的噪点、网格更多.又如第 3 行第 3 列额头上的装饰物,无掩模重构法的额饰与原始图片相比,颜色发生了改变,而这是我们并不希望的,与此相反,掩模重构法的额饰却并未改变颜色.此外,第 1 行第 2 列中的唇彩也比第 1 行第 3 列的唇彩更加自然,效果更好.因为模型对图像全局都要进行编辑,难度较大,而引入图像掩模的信息为网络提供了编辑区域的位置信息,有助于编辑好重点变化的区域.综上,掩模法不仅可以更好地保持非上妆区域的信息,对生成图片质量的提高也是很显著的.

(3) 分部件上妆结果

如图 11 所示,第 1 列是原图,第 1 列是底妆区域上妆,可理解为粉底腮红区域的上妆结果,这里用到的粉底腮红属于复古妆的类型,第 2 列是唇部上妆的结果,这里用到的唇部妆容属于复古妆的类型,第 3 列是眼影区域的上妆,这里用到的眼影属于亚洲烟熏妆的类型.

如图 12 所示,第 1 列是原图,第 1 列是底妆区域上妆,可理解为粉底腮红区域的上妆结果,这里用到的粉底腮红属于彩妆的类型,第 2 列是唇部上妆的结果,这里用到的唇部妆容属于欧洲烟熏妆的类型,第 3 列是眼影区域的上妆,这里用到的眼影属于欧洲烟熏妆的类型.分部件上妆极大地方便了用户,使得用户可以根据心意灵活

© 中国科学院软件研究所 http://www.jos.org.cn

地选择不同的上妆部位和对应的上妆类型.

Fig.10 Comparison between reconstructed imgae of network using mask or not. The first column shows the input image, next columns show reconstructed imgaes

图 10 有掩模控制得到的重构图片与无掩模控制得到的重构图片比较. 第 1 列是输入的原始图片,后两列是重构图片

Fig.11 Parts makeup result (I)

图 11 分部件上妆效果示意图(I)

Fig.12 Parts makeup result (II)

图 12 分部件上妆效果示意图(II)

(4) 多库联合训练和单库训练的结果比较

如图 13 所示,第 1 行是只用妆容数据库的结果,第 2 行是妆容数据库和 CelebA 数据库联合训练的模型的结果.第 1 列第 1 行的眉毛眼影处和第 1 列第 2 行的眉毛眼影处比较,第 2 列第 1 行的背景处和第 2 列第 2 行的背景处比较,第 3 列第 1 行的眼影头发处和第 3 列第 2 行的眼影头发处比较,可见多库联合训练生成的图片

© 中国科学院软件研究所 http://www.jos.org.cn

更清晰、真实,没有网格噪声点,我们认为这种扩大数据量的方式对模型表现能力的提高是极有帮助的.

Fig.13 Comparison between results of proposed model training with multiple datasets and results of proposed model training with single datasets

图 13 多库联合训练与单库训练结果比较示意图

4.5 定量实验结果

本节讨论在自动上妆任务上评价指标的选取.由于现在学术界也还未有特别好的方法评价生成图片的质量,目前经常使用的一种方式是在亚马逊劳务众包平台 AMT 上,让人们对生成结果图进行评价并打分.即给定一些生成图片,根据人类视觉感受,选择人们认为质量最佳的图片.但是这种人工打分的方法不是可复现的,所以,本文只将 AMT 评分结果作为辅助.我们从当前主流评价方法中选取一些合理的评价指标,并且结合最近在生成对抗网络模型评价方面取得的进展,最终选取了 FID(Frechet inception distance)和 SSIM 作为评价指标,根据这些评价指标,最终得到了本文中的基准实验和我们提出的方法的定量结果比较.

一个好的模型应该生成这样的样本,即当被分类器评估时,类分布的熵要低.同时,它应该生成各种各样的样本,涵盖所有的类.通过考虑真实数据和生成数据的嵌入差异,可以计算 FID.假设编码层遵循多变量高斯分布,则分布之间的距离被减小到相应高斯之间的 Frechet 距离.FID 是一个合理的指标,因为它在模式坍塌和编码网络选择方面具有鲁棒性,同时能够较好地衡量真实数据和生成数据之间的差异,所以,本文将 FID 作为一个评价指标.FID 利用分类网络输出的图像类别特征分布,计算真实数据和生成数据之间的差异.实验中使用本文的化妆数据训练集上训练的 Inception v3 网络进行特征提取.这是由于本文数据均为人脸图像,主要针对的是人脸化妆任务而非噪声到图像的通用 GANs 任务,当人脸图像作为输入时,通用的分类器得到的类别特征分布过于集中,难以精确地衡量不同人脸妆容之间的差异,所以本文使用了妆容分类器,网络输出类别为数据集对应的妆容类别.而 SSIM 属于图像质量评价全参考方法中比较经典的方法,它分别从亮度、对比度、结构这 3 个方面度量图像相似性,给出了较为客观的评价.所以,最终我们挑选了 FID 和 SSIM 这两种评价指标作为本文中自动上妆任务的评价指标.在实验中,FID 主要衡量真实数据和生成数据的分布相似性,并不需要每张生成图像都有真实图像作为参考.所以在计算 FID 时,验证集作为真实数据,其中的每张图像都会生成全部妆容类别的生成图像,以亚洲数据集的 361 张验证集为例,共得到了 361×5 张生成图像.而 SSIM 直接衡量图像之间的相似性,需要每张生成图像都有对应的真实图像参考,所以每张验证集图像仅生成与本身妆容类别对应的生成图像,即生成过程中保持妆容类别不变,完成“重构生成”,以亚洲数据集为例,共生成了 361×1 张生成图像.定量实验结果见表 4 和表 5.

FID 的数值越低,说明生成图片质量越好;SSIM 数值越高,说明生成图片质量越好.从表中可以看到,与 StarGAN 和 CycleGAN 两种当前主流方法相比,本文方法在自动上妆这个任务上能够取得更好的结果.

© 中国科学院软件研究所 http://www.jos.org.cn

Tabel 4 Comparison of quantitative experiment in Asian dataset

表 4 亚洲数据库定量实验结果对比

	FID(↓)	SSIM(↑)
CycleGAN	28.115 1	0.548 0
StarGAN	35.968 1	0.621 7
Out Method	**24.509 2**	**0.674 4**

Tabel 5 Comparison of quantitative experiment in European dataset

表 5 欧洲数据库定量实验结果对比

	FID(↓)	SSIM(↑)
CycleGAN	24.367 4	0.594 8
StarGAN	28.941 1	0.643 5
Out Method	**16.380 1**	**0.731 2**

4.6 与其他相关工作的对比

本文选取了图像风格转换领域中 Liao 等人[28]的代表性工作,以及妆容迁移领域中 Liu 等人[34]的代表性工作作为对比.这两种方法均需要参考妆容图片,而本文方法则不需要参考妆容图片,所以在实验中,两种相关工作均采用图 14 所示的参考图片,而本文提出的算法则是将妆容编辑目标类别设置为烟熏妆.对比实验结果如图 14 所示.

Fig.14 Comparison with other related works

图 14 与其他相关工作的对比结果

从对比结果中可以看出,Liu 等人[34]的方法较好地将参考图片的眼影、唇彩等妆容迁移到输入图片,但存在妆容编辑较不自然的问题.Liao 等人[28]的方法较好地整体迁移了参考图片的妆容,但背景等与上妆无关的区域也发生了较大的变化.而本文提出的方法则比较自然地将输入图片编辑为了目标妆容烟熏妆的风格.不过,由于并不是依照参考图片进行的妆容迁移,难以编辑出参考图片的特征风格,如特殊的唇彩颜色等.

5 总结与展望

本文基于生成对抗网络,提出了掩模控制的自动上妆方法.该方法在训练生成网络的同时训练一个掩模生成网络,并通过掩模选择模块以及全局和局部的判别网络实现了分部件上妆的功能,且能够保持非上妆区域信息不变.本文使用 SSIM 和 FID 两种主流的评价指标对该方法进行了定性和定量的对比实验,并与两种主流的图像生成模型进行了对比,充分证明了本文方法具有较好的上妆效果和图像生成质量.本文基于生成对抗网络对

© 中国科学院软件研究所 http://www.jos.org.cn

自动上妆进行了深入研究,取得了一定的成果,但在以后的工作中还有改进和优化的方面.

- 尝试对数据进行扩展,如光照、色彩对比度等,增强模型对人脸色彩生成的多样性.
- 尝试着重训练每个类别中有代表性的部分数据,尽可能地突出类别特点,同时兼顾其他数据,增强模型的泛化能力.

References:

[1] Denton EL, Chintala S, Fergus R. Deep generative image models using a Laplacian pyramid of adversarial networks. In: Proc. of the Int'l Conf. on Neural Information Processing Systems. 2015. 1486–1494.

[2] Huang X, Li Y, Poursaeed O, Hopcroft JE, Belongie SJ. Stacked generative adversarial networks. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2017. 2:3.

[3] Zhao J, Mathieu M, LeCun Y. Energy-based generative adversarial network. In: Proc. of the Int'l Conf. on Learning Representations. 2017. 2.

[4] Guo D, Sim T. Digital face makeup by example. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2009. 73–79.

[5] Scherbaum K, Ritschel T, Hullin M, Thormählen T, Blanz V, Seidel HP. Computer-suggested facial makeup. Computer Graphics Forum, 2011,30(2):485–492.

[6] Tong WS, Tang CK, Brown MS, Xu YQ. Example-based cosmetic transfer. In: Proc. of the Computer Graphics and Applications. 2007. 211–218.

[7] Liu L, Xing J, Liu S, Xu H, Zhou X, Yan S. Wow! you are so beautiful today! ACM Trans. on Multimedia Computing, Communications, and Applications, 2014,11(1s):20.

[8] Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2017. 5967–5976.

[9] Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proc. of the IEEE Int'l Conf. on Computer Vision. 2017. 2242–2251.

[10] Kim T, Cha M, Kim H, Lee JK, Kim J. Learning to discover cross-domain relations with generative adversarial networks. In: Proc. of the Int'l Conf. on Machine Learning. 2017. 1857–1865.

[11] Mirza M, Osindero S. Conditional generative adversarial nets. arXiv Preprint arXiv: 1411.1784, 2014.

[12] Resales R, Achan K, Frey B. Unsupervised image translation. In: Proc. of the IEEE Int'l Conf. on Computer Vision. 2003,1: 472–478.

[13] Zhu JY, Krähenbühl P, Shechtman E, Efros AA. Generative visual manipulation on the natural image manifold. In: Proc. of the European Conf. on Computer Vision. 2016. 597–613.

[14] Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, Shi W. Photo-realistic single image super-resolution using a generative adversarial network. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2017. 105–114.

[15] Li M, Zuo W, Zhang D. Deep identity-aware transfer of facial attributes. arXiv Preprint arXiv: 1610.05586, 2016.

[16] Shen W, Liu R. Learning residual images for face attribute manipulation. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2017. 1225–1233.

[17] Zhang Z, Song Y, Qi H. Age progression/regression by conditional adversarial autoencoder. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2017. 4352–4360.

[18] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv Preprint arXiv: 1511.06434, 2015.

[19] Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A, Chen X. Improved techniques for training GANs. In: Proc. of the Conf. and Workshop on Neural Information Processing Systems. 2016. 2226–2234.

[20] Mathieu MF, Zhao JJ, Zhao J, Ramesh A, Sprechmann P, LeCun Y. Disentangling factors of variation in deep representation using adversarial training. In: Proc. of the Conf. and Workshop on Neural Information Processing Systems. 2016. 5040–5048.

[21] Pathak D, Krahenbuhl P, Donahue J, Darrell T, Efros AA. Context encoders: Feature learning by inpainting. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2016. 2536–2544.

© 中国科学院软件研究所 http://www.jos.org.cn

[22] Mathieu M, Couprie C, LeCun Y. Deep multi-scale video prediction beyond mean square error. In: Proc. of the Int'l Conf. on Learning Representations. 2016. 2.

[23] Vondrick C, Pirsiavash H, Torralba A. Generating videos with scene dynamics. In: Proc. of the Conf. and Workshop on Neural Information Processing Systems. 2016. 613–621.

[24] Wu J, Zhang C, Xue T, Freeman B, Tenenbaum J. Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling. In: Proc. of the Conf. and Workshop on Neural Information Processing Systems. 2016. 82–90.

[25] Gatys LA, Ecker AS, Bethge M. A neural algorithm of artistic style. arXiv Preprint arXiv: 1508.06576, 2015.

[26] 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. Cham: Springer-Verlag, 2016. 694–711.

[27] Fišer J, Jamriška O, Simons D, Shechtman E, Lu J, Asente P, Lukac M, Sýkora D. Example-based synthesis of stylized facial animations. ACM Trans. on Graphics (TOG), 2017,36(4):155.

[28] Liao J, Yao Y, Yuan L, Hua G, Kang SB. Visual attribute transfer through deep image analogy. arXiv Preprint arXiv: 1705.01088, 2017. 1, 2, 4, 6, 7, 8.

[29] Sangkloy P, Lu J, Fang C, Yu F, Hays J. Scribbler: Controlling deep image synthesis with sketch and color. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2016. 6836–6845.

[30] Bousmalis K, Silberman N, Dohan D, Erhan D, Krishnan D. Unsupervised pixel-level domain adaptation with generative adversarial networks. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2017,1(2):7.

[31] Li Y, Song L, Wu X, He R, Tan T. Anti-makeup: Learning a bi-level adversarial network for makeup-invariant face verification. arXiv Preprint arXiv: 1709.03654, 2017.

[32] Wang S, Fu Y. Face behind makeup. In: Proc. of the AAAI Conf. on Artificial Intelligence. 2016. 58–64.

[33] 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. 2015. 4621–4629.

[34] Liu S, Ou X, Qian R, Wang W, Cao X. Makeup like a superstar: Deep localized makeup transfer network. In: Proc. of the Int'l Joint Conf. on Artificial Intelligence. 2016. 2568–2575.

[35] 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. 2018. 40–48.

[36] Odena A. Semi-supervised learning with generative adversarial networks. arXiv Preprint arXiv: 1606.01583, 2016.

[37] Odena A, Olah C, Shlens J. Conditional image synthesis with auxiliary classifier GANs. In: Proc. of the Int'l Conf. on Machine Learning. 2017. 2642–2651.

[38] Reed S, Akata Z, Yan X, Logeswaran L, Schiele B, Lee H. Generative adversarial text to image synthesis. In: Proc. of the Int'l Conf. on Machine Learning. 2016. 1060–1069.

[39] Zhang H, Xu T, Li H, Zhang S, Huang X, Wang X, Metaxas D. StackGAN: Text to photo-realistic image synthesis with stacked generative adversarial networks. arXiv Preprint arXiv: 1612.03242, 2016,2(3):5.

[40] Kim T, Cha M, Kim H, Lee JK, Kim J. Learning to discover cross-domain relations with generative adversarial networks. In: Proc. of the Int'l Conf. on Machine Learning. 2017. 1857–1865.

[41] Taigman Y, Polyak A, Wolf L. Unsupervised cross-domain image generation. arXiv Preprint arXiv: 1611.02200, 2016.

[42] Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville AC. Improved training of wasserstein GANs. In: Proc. of the Int'l Conf. on Neural Information Processing Systems. 2017. 5767–5777.

[43] Liu Z, Luo P, Wang X, Tang X. Deep learning face attributes in the wild. In: Proc. of the IEEE Int'l Conf. on Computer Vision. 2015. 2, 4, 6.

[44] Choi Y, Choi M, Kim M, Ha JW, Kim S, Choo J. StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2018. 8789–8797.

[45] Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2018,40(4):834–848.

[46] Mao X, Li Q, Xie H, Lau RY, Wang Z, Smolley SP. Least squares generative adversarial networks. In: Proc. of the IEEE Int'l Conf. on Computer Vision. 2017.

© 中国科学院软件研究所 http://www.jos.org.cn

包仁达(1994－),男,河北石家庄人,硕士,主要研究领域为生成对抗网络,图像语义分割.

庾涵(1992－),女,硕士,主要研究领域为图像语义分割,关键点检测.

朱德发(1994－),男,硕士,主要研究领域为图像及视频合成,迁移学习.

黄少飞(1996－),女,硕士,主要研究领域为基于生成对抗网络的人脸编辑及应用.

孙瑶(1983－),男,博士,副研究员,主要研究领域为密码哈希函数分析,代数方程组求解算法研究,符号计算,计算机视觉.

刘偲(1985－),女,博士,副教授,CCF 专业会员,主要研究领域为深度学习,计算机视觉,多媒体分析.

www.jos.org.cn
www.jos.org.cn