DOI:10.13878/j.cnki.jnuist.2019.06.004

黄菲[1]　高飞[1,2]　朱静洁[1]　戴玲娜[1]　俞俊[1]

基于生成对抗网络的异质人脸图像合成:进展与挑战

摘要

异质人脸图像合成旨在生成逼真、可识别的多种视觉形态的人脸肖像,包括画像、漫画等多种模态.异质人脸图像合成在公共安全和数字娱乐领域具有广泛的应用前景和重要的研究价值,已成为当前研究热点之一.近年来,随着生成对抗网络的发展以及其在多种图像风格转换任务中的成功,研究人员利用生成对抗网络构建了多种异质人脸图像合成的新方法.本文简要回顾了异质人脸图像合成的发展历史,并从异质人脸图像合成的应用进展、模型结构、性能评估、数据集和定性分析等方面综述了该领域最新的关键技术的发展情况,展望了异质人脸图像合成面临的挑战以及其关键技术的发展趋势.

关键词

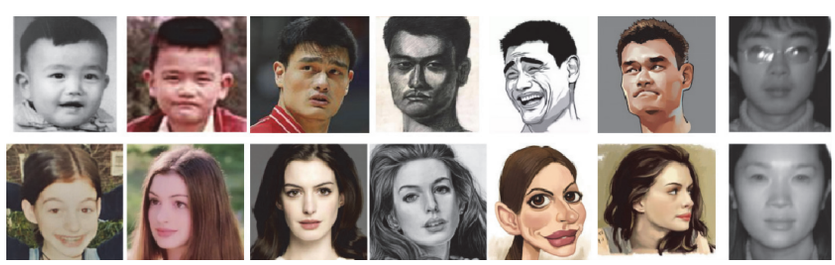
生成对抗网络;异质人脸图像合成;图像风格转换;深度学习;数字艺术

中图分类号 TP183;TP391.41

文献标志码 A

0　引言

人脸图像在人类生活中具有广泛的应用和重要的研究意义.随着科学技术的发展,实际生活中出现了各种各样的图像采集传感器,因而存在着不同的人脸图像形态,如应用在身份认证方面的人脸照片、用于刑侦追捕等领域的人脸画像,应用在数字娱乐中的人脸漫画等.这些图像类型可以成为不同的人脸图像域(如图1所示).异质人脸图像合成旨在设计数学模型使计算机能够基于某一给定域图像,自动生成自然、逼真的其他域人脸图像,包括画像合成、漫画合成、年龄合成、超分辨率重建、人脸美颜等,已成为当今的研究热点之一[1].



a　b　c　d　e　f　g

图1　异质人脸图像举例　a—c.不同年龄段的照片;d.画像(素描);e.漫画;f.插画;g.近红外图像

(所有图片由网络搜索得到,其中(g)取自NIR Face数据库)

Fig. 1　Illustration of heterogeneous face images,(a)-(c)faces of different ages,(d) sketches,(e) caricatures,(f) illustration,and (g) near infrared (NIR) face

(All images are downloaded from the Web,and images of (g) belong to the NIR Face database)

异质人脸图像合成在公共安全和数字娱乐领域具有广泛的应用价值.例如在刑侦追捕中,公安部门备有每位公民照片组成的照片数据库,以用来确定犯罪嫌疑人身份,但实际中一般是用画家和目击者合作得到的犯罪嫌疑人的素描模拟画像来进行后续的人脸检索和识别.由于模拟画像和常见的人脸照片在纹理表达上的差异,直接利用传统的人脸识别方法很难取得满意的识别效果.将模拟画像合成为照片或将照片合成为画像可以有效减小他们纹理上的差距,进而大幅提高准确率和破案效率.在儿童走失案中,公安部门能够获取的都是丢失前的人脸照片.然而随着时间的推移,儿童的面貌会随着年龄的

收稿日期 2019-10-15

资助项目 国家自然科学基金(61601158,61971172,61971339,61836002,61702145);中国博士后自然科学基金(2019M653563);浙江省教育厅一般项目(Y201942162,Y201840785)

作者简介

黄菲,男,硕士生,主要研究领域为计算机视觉与机器学习.997257065@qq.com

高飞(通信作者),男,博士,副教授,主要研究领域为计算机视觉与机器学习,涉及图像生成及风格转换、视觉质量评价及增强、医学影像智能分析等课题.gaofei@hdu.edu.cn

1　杭州电子科技大学　计算机学院/复杂建模与仿真教育部重点实验室,杭州,310018

2　西安电子科技大学　综合业务网理论及关键技术国家重点实验室/电子工程学院,西安,710071

增长发生较大的变化,给人脸识别带来了极大的困难.年龄合成方法可以基于年幼时的照片推演出成长后的外貌,从而提升人脸识别的精度.在数字娱乐领域,由于画像、油画、漫画等相对于照片具有更高的的趣味性和艺术性,人们越来越倾向于利用这些类型的图像作为自己在社交网络上的形象.

至今为止,已有大量的相关工作对相关异质人脸图像合成课题进行了研究[1].其中,Wang 等[2]从数学模型的角度对人脸图像超分辨率重建和人脸画像合成进行了深入分析,并将人脸画像合成中的数学方法划分为贝叶斯推断、子空间学习及稀疏表示几个模块.Nguyen 等[3]对二维和三维人脸图像的超分辨率重建方法进行了总结,划分为了浅层手工特征与深度学习方法两类.之后,文献[4-5]将人脸画像合成方法作为异质人脸图像识别的一个子类进行综述,将相关方法划分为了数据驱动和模型驱动两种类型.整体而言,这一划分可以推广到其他异质人脸图像合成任务中.

在数据驱动方法中,主要包含近邻搜索和目标重构两个模块.具体而言,对于给定源图像区域,在大规模源图像块中寻找与其相近的源图像块及相互之间的近邻关系,之后将这一近邻关系应用到近邻源图像块对应的目标图像块上,用于重构给定区域的目标区域.基于数据驱动的方法由于利用真实目标图像块进行重构,通常可以得到较好的合成效果.但由于近邻搜索规模较大、速度较慢,难以满足实时性需求.

在模型驱动方法,则利用数学模型构建源图像(块)到目标图像(块)的数值映射关系.这类方法通常包含特征提取和机器学习两个模块.首先,提取可以有效表征图像块内容、纹理等复杂信息的特征表达.然后,利用机器学习方法从大规模“源图像-目标图像”中训练得到不同模态之间的映射关系.早期的工作中,研究人员主要从两个模块进行探索,提出了多种富有启发性的工作.模型驱动方法,通常具有很高的计算效率,但受到图像特征和学习方法的限制,合成图像通常视觉质量较差.

随着深度学习的发展,研究人员在纹理生成方面取得了很大进步,并且扩展至图像内容生成领域.近年来,生成对抗网络(generative Adersarial Network,GAN)[6-8],特别是条件生成对抗网络(conditional Generative Adversarial Networks,cGAN)[9]的出现,为异质图像合成开辟了新的路径.条件生成对抗网络可以在给定条件(可以是文本、属性向量或源图像等)下生成对应的目标图像,其已在基于文本的图像生成[10]和图像风格转换[11-14]等领域得到了广泛应用.在对应的任务中,其生成图像包含了非常逼真的视觉效果和内容细节.受此启发,基于生成对抗网络的异质人脸图像合成逐渐引起大家重视,并涌现出大量的新方法,取得了突破性进展.

目前,异质人脸图像合成的应用才刚刚走出实验室,还处于起步阶段,已经有一些较为成功的案例.例如,西安电子科技大学研究团队与警方合作,利用人脸画像合成成功辅助警方破获了重大案件.腾讯优图实验室将年龄合成算法用于跨年龄人脸识别,已成功寻回多名走失多年的儿童.在数字娱乐领域,现在已经有了多种异质人脸图像合成 APP,比如陌陌推出的 AI 换脸视频制作软件 ZAO 可以更改人脸身份及几何结构,俄罗斯无线实验室开发的 FaceApp 可以进行年龄合成以及多种属性的更改,旷世 Face++开发了人脸美颜功能,供用户在线试用.这些应用都在社交媒体上引起了巨大反响,提升了用户进行艺术感知和创作的兴趣与热情.相关应用的截图举例如图 2 所示.

图 2 现有异质人脸图像合成应用 APP 示例,左图为 ZAO 应用截图,右图为 FaceAPP 应用截图

Fig. 2 Illustration of existing heterogeneous face synthesis applications,left:ZAO;right:FaceAPP

本文与之前工作的不同在于:1)据笔者所知,现在尚未有工作对于画像合成、漫画合成、年龄合成、人脸美颜等多种异质人脸图像合成任务进行总结和分析,本文的工作相比之下更为全面,并且对所有任务下的思想和模型进行了抽象和总结;2)之前的综述工作并未对基于生成对抗网络的异质人脸图像合成工作进行对比分析,本文重点对比分析了这类方法在模型结构和思想之间的异同;3)本文对异质人脸图像合成图像的质量评价方法进行了总结和分析,指出了当前评价方法的局限性和发展趋势.

1 背景介绍

1.1 任务描述

在异质人脸图像合成任务中,通常包含源图像域(如人脸照片)X、目标图像域(如人脸画像)Y.异质人脸图像合成的任务旨在学习源图像域到目标图像域的数值映射关系:$F:X\to Y$.给定一幅源图像$x\in X$,异质人脸图像合成模型可以预测其对应的目标图像$y\in Y$.在现有的异质人脸图像合成工作中,通常源图像和对应的目标图像是成对出现的.这时,可以考虑基于有监督学习的方式训练得到异质人脸图像合成模型.在很多任务中,可能难以得到"源图像-目标图像"对,这时可以考虑采用无监督学习或半监督学习的方式构建异质人脸图像合成模型.

1.2 生成对抗网络

生成对抗网络由 Goodfellow 等[6]于 2014 年提出,之后在计算机视觉领域引发了研究热潮,被广泛应用于图像生成、图像风格转换、图像超分辨率重建等多种任务中,并且取得了巨大成功[7-8].生成对抗网络通常包含一个生成器 G 和一个判别器 D.其中,判别器 D 旨在正确判断给定图像是否是真实的,即将实际目标图像 y 判断为"真",将生成图像 $G(x)$ 判断为"假";生成器 G 基于输入信息 x(可以是随机向量、属性向量、输入源图像,或其组合)生成逼真的目标图像 $G(x)$,从而使判别器 D 将其判断为"真".在训练过程中,对生成器和判别器进行迭代优化,两者之间以类似"对抗"的形式进行交替迭代优化,最终达到"纳什均衡"状态,两者达到较优的性能,从而生成器可以生成高质量图像.在异质人脸图像合成领域,通常存在一幅输入源图像,研究人员采用了条件生成对抗网络和循环生成对抗网络.本文接下来对两个模型进行简要介绍.

1.2.1 条件生成对抗网络

在条件生成对抗网络中,最具影响力的工作是 Isola 等提出来的 Pix2Pix 模型[9],其在多种图像风格转换任务中都取得了优异的性能,并给相关课题研究人员带来启发,在多个领域取得了巨大进展.条件生成对抗网络模型结构如图 3 所示.给定输入源图像 x,生成器生成 $G(x)$.之后,判别器将"输入图像-生成图像"对$(x,G(x))$判定为假,而将"输入图像-真实图像"对(x,y)对判断为真,并采用对抗损失进行训练.在现有工作中,一般对生成器采用半饱和对抗损失函数.具体而言,生成器和判别器的对抗损失函数分别为

$$L_{\text{adv}}^{D}=-E_{(x,y)\in(X,Y)}[\log(D(x,y))]-E_{x\in X}[\log(1-D(x,G(x)))],\tag{1}$$

$$L_{\text{adv}}^{G}=-E_{(x,y)\in(X,Y)}[\log(D(x,G(x)))].\tag{2}$$

此外,由于存在真实的目标图像,因此还采用了像素域的图像重建损失,即生成图像和真实图像像素之间的 L_1 距离:

$$L_1=\|y-G(x)\|.\tag{3}$$

最终,将对抗损失与重建损失进行加权,对生成器和判别器进行交替迭代优化,直到收敛.

$$D^{*}=\arg\min_{D}L_{\text{adv}}^{D},$$

$$G^{*}=\arg\min_{G}L_{\text{adv}}^{G}+\lambda L_1.\tag{4}$$

此外,Mao 等[15]提出了最小平方生成对抗网络(Least Squares Generative Adversarial Networks,LSGAN),其对应的生成器和判别器的对抗损失函数分别为

$$L_{\text{adv}}^{D}=E_{(x,y)\in(X,Y)}[(1-D(x,y))^2]+E_{x\in X}[D(x,G(x))^2],\tag{5}$$

$$L_{\text{adv}}^{G}=E_{(x,y)\in(X,Y)}[(1-D(x,G(x)))^2].\tag{6}$$

由于 LSGAN 损失函数形式更为简单,且训练较为稳定,因此有很多工作采用 LSGAN 作为基准模型.

1.2.2 循环生成对抗网络

在很多图像转换任务中,难以或无法获得成对

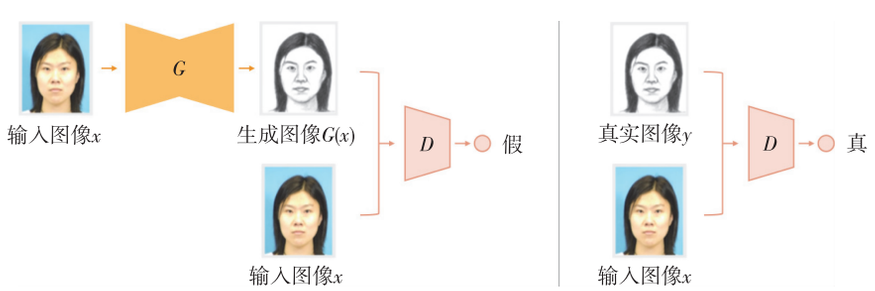


图 3 基于条件生成对抗网络的异质人脸图像合成模型结构示意图[9]

Fig. 3 Architecture of heterogeneous face synthesis via conditional generative adversarial network (cGAN)[9]

的训练样本,因此 Zhu 等提出了循环生成对抗网络(CycleGAN)的思想[16].CycleGAN 的基本框架如图 4 所示.CycleGAN 包含两个生成器 G 和 F,分别模拟图像域 X 和图像域 Y 之间的双向映射关系,即 $G:X\rightarrow Y$ 和 $F:Y\rightarrow X$.此外,还有对应两个图像域的判别器 D_X 和 D_Y. 在训练过程中,除了使用对抗损失之外,CyclegGAN 还采用了重构损失和一致性损失.其中,重构损失是指:

$$L_{\text{rec}}=\|x-F(G(x))\|+\|y-G(F(y))\|, \quad (7)$$

即希望将输入图像 x 依次输入到 G 和 F 后能够重构自身的全部信息;对 y 也相似.一致性损失是:

$$L_{\text{con}}=\|y-G(y)\|+\|x-F(x)\|, \quad (8)$$

即希望将目标图像 y 输入到生成器 G 后能够保持不变;对 x 也相似.最终,将所有损失函数进行加权,对生成器和判别器进行交替迭代优化,直到收敛.在异质人脸图像合成任务中,有时存在真实目标图像,因此可以同时使用式(3)中的 L_1 损失.

2 异质人脸图像合成研究进展

异质人脸图像合成包括画像合成、年龄合成、漫画合成、油画合成、超分辨率重建、人脸美颜等多个任务.其中,图像超分辨率重建旨在基于给定低分辨率图像重构出相对较高分辨率的图像.超分辨率重建对于计算机视觉领域意义重大,研究人员对此开展了大量的研究工作.最近,Wang 等[17]和 Ha 等[18]对于已有的通用图像超分辨率重建方法进行了详尽地总结和回顾,其中包含了人脸超分辨率重建算法.Nguyen 等[3]也对于二维和三维人脸图像的超分辨率重建工作专门进行了总结和分析.因此,本文对于人脸图像超分辨率相关工作不再赘述,而对其余几个课题分别介绍其近期研究进展.

2.1 画像合成

人脸画像合成是指将一幅照片转换为画像的过程,通常利用事先收集好的人脸画像-照片对作为训练集.它在辅助刑侦追捕及数字娱乐方面具有重要作用[19].传统人脸画像合成方法主要可以分为两类:数据驱动的方法和模型驱动的方法[2,4].鉴于深度学习技术的发展,Zhang 等[20]和 Jiao 等[21]提出利用卷积神经网络生成画像,Sheng 等[22]亦提出基于神经网络进行特征学习和画像重构.之后,Zhang 等[23-25]将神经网络与马尔科夫随机场进行结合,通过学习深层特征及构建邻域约束来提升合成质量.

近年来,由于生成对抗网络在多种图像翻译任务中的巨大成功,研究人员开始尝试采用其构建人脸画像合成模型.例如,Wang 等[26]提出首先利用条件生成对抗网络来生成画像,然后利用反向映射方法对生成画像进行后处理,得到最终的生成结果.之后,Wang 等[27]提出使用多对抗网络从低分辨率到高分辨率逐步生成人脸画像.该方法在性能上得到了显著提升,生成画像的纹理效果较好,但仍然存在少量模糊失真.Zhang 等[28-29]针对现有方法对于面部光照敏感的问题,提出基于人脸结构先验及光照映射的画像生成对抗模型,以及多分布约束模型,取得了不错的效果.此外,Chen 等[30]则针对现有人脸照片-画像数据集规模较为有限的情况,利用大规模人脸照片(无对应画像),结合半监督学习思想,以提升深度模型的训练效果.Bae 等[31]则针对多风格人脸画像合成进行了初步尝试,在生成对抗网络中引入了风格分类机制及对应的损失函数,使得模型可以生成多种风格的画像.

此外,研究人员考虑到单个生成模型性能有限,且人脸不同区域的映射关系可能有所不同,因此提出了多支路合成方法[32].例如,Yi 等[33]采用两个生成器,其中一个全局生成器合成人脸整体结构,另外一个局部生成器用来合成眼睛、头发等特定区域的细节,之后利用融合网络对合成结果进行整合,得到最终的合成结果.Zhang 等[34]将数据驱动方法与生成对抗网络方法进行结合,利用数据驱动方法实现

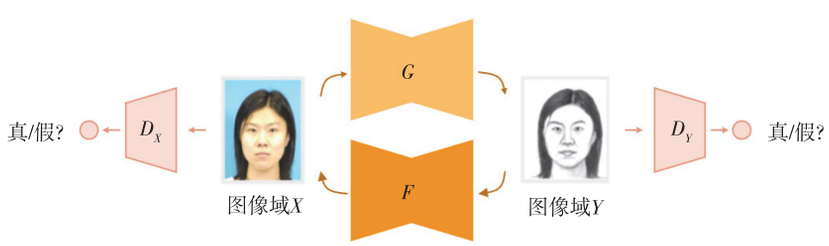


图 4 基于循环生成对抗网络的异质人脸图像合成模型基本框图[16]

Fig. 4 Pipeline of heterogeneous face synthesis via CycleGAN[16]

照片到画像的映射,合成画像初步结果;利用生成对抗网络,实现画像高频信息的合成;然后将两者融合,取得了不错的效果.

Yu 等[35]则将人脸结构先验信息引入到生成对抗网络中,提出了一种基于结构辅助的生成对抗网络(如图 5 所示).在该网络中,首先利用人脸解析网络得到人脸分割结果,用于描述人脸结构;然后,将人脸图像与分割结果分别输入到外观编码器和结构编码器中,并一起输入到解码器中合成对应的输出图像;此外,还引入了结构性重建损失函数,用来提升眼睛等复杂区域的权重,降低面部、头发等区域的权重,从而促使模型生成更加精微的细节.最后,使用了堆叠式生成器和身份损失函数,有效提升了合成图像的质量.该模型在多种合成任务中都取得了优异的合成效果.

最近,Zhu 等[36]认为使用传统生成对抗网络拟合照片与画像之间的映射关系,忽略了两个图像域之间的共有信息,因此提出了一种协作学习框架,将照片/画像首先映射到一个共享的隐藏空间,然后进一步映射到画像/照片,并且提出了协作损失函数,促使两者在隐藏空间具有一致性.Zhang 等[37]则几乎同时提出了相似的工作,与之不同的是其在隐藏空间使用了对抗学习机制来促使照片和画像映射到相同的隐藏空间.这两种方法取得了较好的合成效果,且具有很好的启发性.

2.2 漫画合成

人脸漫画合成是指基于给定人脸照片生成漫画,在游戏制作等数字娱乐领域具有十分广泛的应用[38].研究人员已经在该领域进行了大量的研究[39-42].例如 Akleman[43]提出使用交互式工具来让用户指导漫画的生成.现在的很多手机应用也是基于这一想法,由用户选择人脸组件构成漫画.之后,研究人员开始考虑自动地人脸漫画合成算法.例如,Liang 等[44]提出从训练集中学习漫画原型组件,然后针对目标人脸进行形状调整和纹理风格转换,得到最终结果.Liu 等[45]提出在照片和漫画中学习异质吉布斯模型,用于自动的漫画合成.之后,Chiang 等[46]提出通过分析输入照片的特征,构建人脸漫画.

受条件式生成对抗网络的启发,最近有少量工作将其引入到人脸漫画生成和识别中,取得了一些进展[47].例如,Zheng 等[48]在条件式生成对抗网络的基础上,引入了并行判别器以提升其能力,从而促使生成器合成更好的漫画.Han 等[49]将人脸三维模型和几何形状轮廓图引入到生成对抗网络中,取得了逼真的漫画人脸图像.但由于人脸三维模型的计算复杂度较高且建模难度较大,因此限制了其扩展性.Li 等[50]提出基于人脸照片及关键点图一起用于合成人脸漫画.但人脸关键点图的信息较为有限,使得判别器难以具备较强的判别能力,限制了生成器的性能.最近,Shi 等[51]引入了关键点位移预测机制和人脸识别损失,有效提升了漫画合成的质量.

2.3 年龄合成

现有的年龄合成方法大致可以分为两类:物理模型驱动方法和数据驱动方法[52-53].其中,物理模型驱动是指通过模拟头颅骨和面部肌肉随着年龄的变化机制,预测变化后的人脸结构和外观[54-57].例如,Wu 等[57]基于皮肤的解剖学结构提出了一种三层动态皮肤模型来模拟皱纹.数据驱动方法则不依赖于生物学先验知识,直接从训练数据中挖掘年龄相关的模式[58-60].例如,Shlizerman 等[59]提出一种基于原

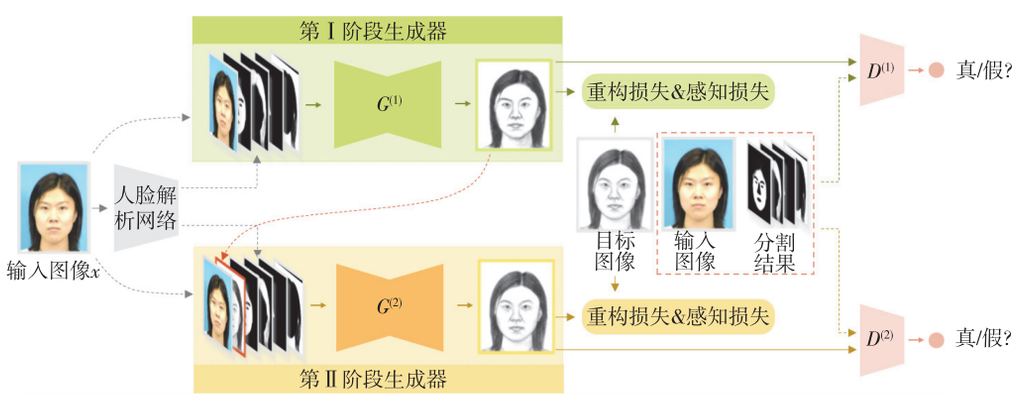


图 5 基于堆叠式结构辅助生成对抗网络的异质人脸图像合成模型基本框图[35]

Fig. 5 Pipeline of heterogeneous face synthesis via stacked composition-aided GANs (SCA-GAN)[35]

型的方法,即利用不同年龄段的人脸图像构建年龄相关的字典,然后利用近邻搜索等技术重构输入人脸对应年龄段的图像.Yang 等[60]提出首先求解多属性分解问题,然后只将年龄相关元素变换到对应目标年龄组实现年龄的合成.这些方法有效提升了合成图像的质量,但会产生鬼脸效应(Ghosting Artifacts).

近年来,随着深度生成网络的发展,研究人员开始尝试将其引入到年龄合成中.现有模型可以大致分为直推式生成和渐进式生成两种方式.

1)直推式生成:是指对于给定输入图像,采用单一模型直接将其映射到目标年龄.例如,Zhang 等[61]利用条件对抗自编码器进行人脸年龄合成,但合成结果不理想,模型只学习到了皱纹等明显特征.Zhou 等[62]将个人职业信息考虑在生成对抗网络中以进行个性化人脸年龄合成.最近,Yang 等[63]基于条件对抗生成网络,提出了一种金字塔结构的判别网络,有效提升了年龄合成的效果.Li 等[64]提出将图像进行小波变换,利用生成对抗网络分别实现全局及局部细节的生成,以提升合成效果.Song 等[65]近来提出了一种 Dual cGAN 模型,同时训练年龄老化和回溯任务,提升了年龄合成的合理性.Kossaifi 等[66]则考虑到人脸不同属性(如种族、性别等)与年龄合成高度相关,因此将人脸属性向量作为合成模型的辅助输入,并使用了小波变换提升细节提升精度.

2)渐进式生成:部分研究人员发现,如果输入人脸与合成人脸之间存在较大的年龄差,使用单一阶段的合成模型难以获得高质量结果[67-69].例如,Wang 等[67]利用循环神经网络(Recurrent Neural Network,RNN)模拟不同年龄段之间人脸的平滑变换.Nhan 等[68]使用残差网络模块构建临近年龄段的合成模型,然后通过将这些模块进行串联实现长年龄间距的合成结果.

2.4 人脸美颜

人脸美学质量,也称为人脸美观度或吸引力,对于人类的社交活动具有广泛的影响力[70-71].人脸美颜[72]亦称为人脸美学质量增强,旨在通过调整人脸肤色[73-74]、人脸结构[75-76]或自动上妆[77]等,提升人脸的吸引力,从而获得更优的社交体验.

现有人脸美颜方法普遍属于数据驱动类型[72],需要美妆后人脸实例作为参考,利用数学模型预测实例信息与输入人脸信息之间的融合参数.例如,Liang 等[73]通过关键点及边缘信息检测皮肤区域,然后通过调整亮度、颜色等进行美化;并在后续工作中将其迁移到云平台中,以满足移动设备的应用需求[74].Leyvand 等[75]通过搜索训练集中与输入人脸相近的图像,预测美化后人脸关键点位置,然后对输入人脸进行美化.

受深度学习在多种计算机视觉任务中的启发,Li 等[76]通过深度神经网络预测美化过程中人脸关键点的位移,用于人脸美颜.Alashkar 等[78]将人脸妆容的区域、类型等考虑在内,并基于经验设定了多种规则,利用深度神经网络预测上妆类型,取得了不错的合成效果.Ou 等[79]则利用图像迁移思想,通过人脸解析获取各个器官的位置信息,将输入人脸转换到参考人脸的妆容风格.

最近,北京航空航天大学和中国科学院大学等机构的研究人员提出了一种姿态稳健型空间可感知式生成对抗网络(Pose-robust Spatial-aware GAN,PS-GAN)[80],可以将输入人脸图像转换到参考人脸的妆容风格.PSGAN 分别采用妆容提取模块和输入人脸编码模块提取参考人脸的妆容以及输入人脸的综合表达,然后输入到妆容转移变形模块,进而利用卸妆-再上妆模块生成化妆后的人脸.整体而言,现在的人脸美颜工作都需要参考妆容图像,且主要局限于自动上妆功能.

2.5 其他异质人脸图像合成

除了上述异质人脸图像合成工作之外,在热红外(近红外)人脸图像合成及通用图像风格转换方面,也有少量基于生成对抗网络的工作.

2.5.1 热红外人脸图像合成

由于热红外图像对于光照不敏感,可以有效提升人脸识别精度,因此已经被广泛应用于现有的人脸识别系统中[81].近期,也有少量热红外人脸图像合成工作.例如,Wang 等[82]利用生成对抗网络将人脸热红外图像转换为人脸照片,并引入人脸关键点检测网络,指导生成器合成更好的细节.Dou 等[83]基于 CycleGAN,结合边缘损失、身份损失等,实现了人脸照片与热红外图像的相互转换.Zhang 等[84]则利用热红外图像生成及对抗学习思想,将热红外图像映射到照片域,以提升人脸照片-近红外图像的跨模态身份识别精度.

2.5.2 通用人脸图像风格转换

通用图像风格转换在给定内容图像和示例风格图像,将内容图像的纹理转换为示例风格,并且保持内容不变[85].相关工作可以追溯到传统的纹理合成

及模拟任务.2016 年,Gatys 等[86]采用卷积神经网络(Convolutional Neural Networks,CNNs)中的图像特征进行图像风格转换,是该领域的开创性工作之一.该方法的优势是只需要一张示例图像,而且可以扩展到不同的风格.研究人员受该工作启发,在图像风格转换任务中取得了大量进展[87-90].关于通用的图像风格转换工作,读者可以参考文献[91-92],其中详细总结了现有的工作进展.本章简要介绍人脸图像相关的工作.

Selim 等[93]将文献[86]扩展到针对人脸油画生成任务进行优化,提升了人脸油画的合成效果.尽管基于神经网络的风格转换技术取得了优异的性能,这些方法通常难以合成纹理上的细节.Fišer 等[94-95]利用非参纹理合成方法,结合人脸分割、姿态、外表等信息综合指导人脸图像的风格转换,取得了非常精细的纹理细节.不过这类方法普遍采用示例图像,基于局部结构搜索实现纹理的迁移,通常效率较低,而且在风格化过程中,会丢失人脸源图像的身份、细微结构等信息.Huang 等[96]针对此问题,提出了一种马尔可夫生成对抗网络(Markovian Generative Adversarial Networks,MGAN)有效提升了计算效率,并将其应用于通用图像和人脸图像的风格转换任务中,在人脸插画、油画分割转换任务中都展现出不错的效果.之后,Li 等[97]利用循环生成对抗网络应用到人脸图像风格转换中.

3 生成对抗网络模型结构

结合上述不同异质人脸图像合成任务的相关工作,本文从输入、生成器、判别器、损失函数四个方面,将其中使用的生成对抗网络模型结构进行抽象总结和对比分析.

3.1 输入

现有方法普遍采用单一输入,即人脸源图像,然后利用生成器将其映射为目标图像.尽管深度学习在理想状态下,应该学习到输入人脸图像对应的全部信息,如人脸的结构、身份、姿态、表情等,并将其应用于映射关系的合成中.但在实际情况下,由于没有对应的辅助信息(包括输入或目标)的指导,而且训练数据的规模有限,无法包含全部可能的人脸属性变种,生成器难以学习到精确的信息表达和映射关系.

因此,采用属性相关的辅助信息来辅助生成器更合理、精确地合成目标图像变得极为可行.例如,Yu 等[35]在输入人脸源图像的同时,将源图像的像素级分割结果作为辅助信息来描述人脸结构信息,并将其输入到生成器中,用于人脸照片-画像的合成.实验结果显示,这一操作有效提升了生成图像的质量.Fišer 等[94-95]则在人脸风格转换中也采用了类似的思路,加入了人脸结构等作为辅助信息.

3.2 生成器结构

在现有工作中,为了提升生成模型的性能,研究人员在条件生成对抗网络和循环生成对抗网络的基础上,对于生成器进行了多种改进.现有的生成器结构大致可以划分为以下几类(模型结构如图 6 所示):

1)单一生成器[9,26-29,84]:即使用单一的生成器,通常为 U-Net 结构,即在编码器和解码器对应层之间增加前向跳接.近期的研究表明,如果将生成器中的卷积层替换为残差模块,通常可以提升合成结果的性能.优于单一生成器的性能受模型深度、复杂度的限制,难以合成高质量图像.

2)堆叠式生成器[35]:也可以成为串联式生成器,即首先使用一个生成器合成初步结果,然后利用后续的生成器进行进一步精细化生成,在图像中添加更多的细节信息,从而合成高质量图像.堆叠式生成器已经成功应用于高分辨率图像合成及文本-图像合成任务[10]中.Yu 等[35]将其应用于人脸照片-画像合成任务中,取得了至今为止最优的合成图像质量.此外,实验结果表明:堆叠的生成器越多,通常可以得到更优的合成效果,但在数据量有限的情况下,训练难度提高,需要防止过拟合现象[35].

3)多支路生成器(全局)[34]:即使用多个(不同结构的)生成器,分别基于输入图像生成对应的合成图像,然后利用融合网络将不同支路的合成结果进行融合,从而得到更高质量的图像.现有工作通常使用两个支路,一个支路用于生成低频信息,另外一个用于合成高频信息.Zhang 等[34]的工作利用该思想将数据驱动方法与生成对抗网络方法结合,取得了不错的效果.

4)多支路生成器(全局+局部)[33]:考虑到人脸不同区域的数值映射关系有所不同,可以针对不同区域采用不同的局部生成器,同时利用全局生成器合成图像的整体结构,最后进行融合得到较好的合成效果.Huang 等[96]利用此思想用于人脸转正.Yi 等[33]将此思想融入到一种新的人脸画像合成任务中.

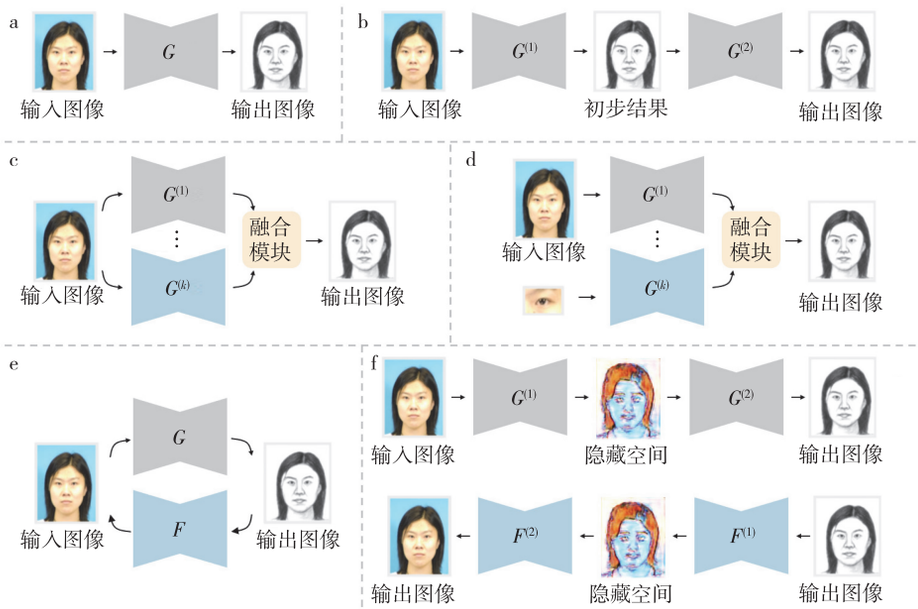


图6 异质人脸图像合成任务中的生成器结构示意图 a.单一生成器;b.堆叠式生成器;c.多支路生成器(全局);d.多支路生成器(全局+局部);e.循环式生成器;f.协作式生成器

Fig. 6 Generators in existing heterogeneous face synthesis,(a) single generator,(b) stacked generators,(c) multi-column generators (global),(d) multi-column generators (global+local),(e) cycle generators,and (f) collaborative generators

5)循环式生成器[16,83,97]:由于在大部分异质人脸图像合成任务中,同时存在照片和其他模态的图像,因此可以基于 CycleGAN 思想,采用循环式生成器同时训练两个合成模型.由于循环式生成器要求合成结果可以逆向恢复输入图像,循环式生成器通常可以提升合成图像与输入图像之间在结构上的一致性.

6)协作式生成器[36-37]:所有上述模型都考虑模拟源图像域到目标图像域的直接映射关系,但源图像目标图像之间应该存在一定的共有信息,因此可以将两者同时映射到一个共享的隐藏空间,消除图像模态信息,然后用于合成目标图像.

3.3 判别器结构

现有异质人脸合成工作中,主要使用了 Isola 等[9]提出的 PatchGAN 结构,即对于给定图像,不对其预测单个标量"真/假"标记,而是对于各个局部区域进行预测,从而得到一个"真/假"标记矩阵.实验表明该方法可以提升局部细节的生成质量.具体而言,按照判别器网络结构,可以将其划分为单一判别器、多尺度判别器和金字塔判别器(如图 7 所示).

1)单一判别器[9]:即针对生成器输出的最终合成图像和真实目标图像,采用单个判别器判断其真假.

2)多尺度判别器[27]:当目标图像分辨率较高时,直接使用 PatchGAN 可能难以对图像整体结构进行约束,因此可以考虑将图像采样到多个分辨率上,然后分别使用判别器判断其真假.此外,也可以考虑让生成器的解码层输出不同尺度的图像,从而促使生成器具有更好的表征能力.这些判别器通常具有相同的结构,只是输入图像的分辨率有所不同.

3)金字塔判别器[63]:在大部分工作中,判别器都是随机初始化,然后通过与生成器交替迭代优化达到较优的判别能力.考虑到特定任务(如年龄合成)中,判别器的特征需要能够有效表征人脸图像的年龄信息,因此 Yang 等[63]首先预训练一个年龄识别网络用于提取多层年龄相关特征,然后在其每一层后面增加一组随机初始化卷积层,从而构建金字塔判别器.在训练过程中,年龄识别网络(灰色条纹模块)保持不变,只训练新增加的卷积层部分.该方法在年龄识别中取得了较优的性能.

此外,基于判别器输入/输出的不同,也可以将现有判别器划分为无监督式判别器、类别辅助判别器、类别辅助多路判别器和条件式判别器(如图 8 所示).

1)无监督式判别器[6]:无监督式判别器是将合成图像单独输入到判别器中,判断其是否逼近真实

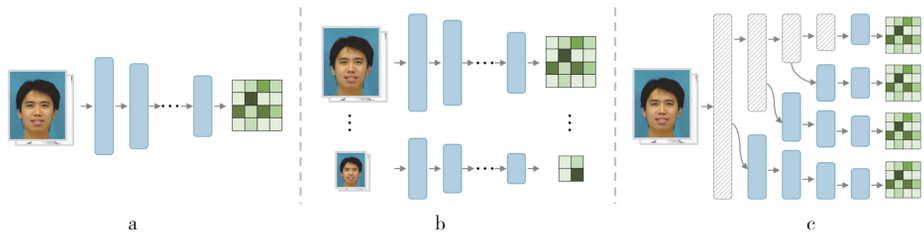


图 7　异质人脸图像合成任务中的判别器结构示意图　a.单一判别器;b.多尺度判别器;c.金字塔判别器

Fig. 7　Discriminator architectures in existing heterogeneous face synthesis methods,(a) single discriminator,(b) multi-scale discriminators,and (c) pyramid discriminators

目标图像,可以用来促使生成更加逼真的图像.其通常可以与条件式判别器共同使用,以提升生成目标图像的视觉质量[14].不过两者之间的平衡,对于图像合成具有较大的影响.

2)类别辅助判别器[31]:对于多风格异质人脸图像合成或者属性辅助的异质人脸合成,可以考虑使用类别辅助判别器,即判别器同时预测输入样本的真假以及所属类别.对于分类损失,可以考虑采用常见的交叉熵损失.这一结构已在多种图像生成任务中取得了可靠的结果[98-99].

3)类别辅助多路判别器[63]:即分别使用判别器来预测真假及类别.其中,类别判别器可以使用有标记数据集提前预训练好,从而使其具有较好的分类性能.

4)条件式判别器[9]:现有工作中,使用较多的为条件式判别器,即将输入图像与合成或真实目标图像串接在一起,然后输入到判别器中.这时判别器可以判断输入图像对是否匹配,能够促使输入图像和合成图像在结构上具有较好的一致性.

3.4　目标函数

在现有基于生成对抗网络的人脸异质图像合成工作中,除了使用 1.2 节所述的对抗损失、重构损失、一致性损失作为目标函数之外,研究人员还提出了以下损失函数,以提升合成图像的性能.

3.4.1　身份损失

由于异质人脸合成希望能够保留输入人脸源图像的身份信息,以保留图像本身的辨识度,因此研究人员引入了身份损失函数.具体而言,研究人员将生成图像和真实目标图像同时输入到预训练好的人脸识别网络中,然后计算两者对应层深度特征之间的欧氏距离.其表示如下:

$$L_{id}=\sum_{l\in S_l}\|\varphi^{(l)}(G(x))-\varphi^{(l)}(y)\|_2^2.$$

当缺少真实目标人脸时,则可以计算生成图像与输入图像之间身份特征的欧式距离,即:

$$L_{id}=\sum_{l\in S_l}\|\varphi^{(l)}(G(x))-\varphi^{(l)}(x)\|_2^2,$$

其中,$\varphi^{(l)}$ 表示身份识别网络提取的第 l 层特征图,S_l 表示选择的特征层集合.

3.4.2　感知损失

此外,也可以将生成图像和真实目标图像同时输入到预训练好的目标识别网络[100]中,然后计算两者对应层深度特征之间的欧氏距离.其表示如下:

$$L_{id}=\sum_{l\in S_l}\|\phi^{(l)}(G(x))-\phi^{(l)}(y)\|_2^2.$$

当缺少真实目标人脸时,则可以计算生成图像与输入图像之间语义特征的欧式距离,即:

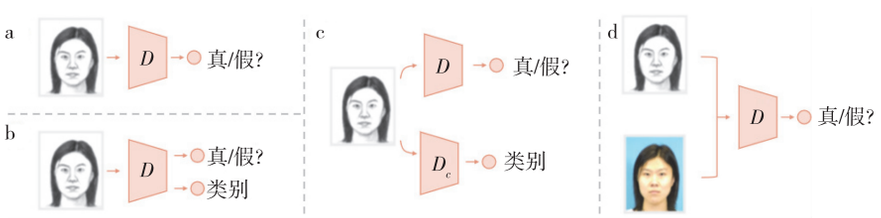


图 8　异质人脸图像合成任务中的判别器功能示意图

a.无监督式判别器;b.类别辅助判别器;c.类别辅助多路判别器[63];d.条件式判别器

Fig. 8　Functions of discriminators in existing heterogeneous face synthesis methods,(a) unsupervised discriminator,(b) discriminators with auxiliary classifier,(c) multi-branch discriminators with auxiliary classifier,and (d) conditional discriminators

$$L_{id}=\sum_{l\in S_l}\|\phi^{(l)}(G(x))-\phi^{(l)}(x)\|_2^2,$$

其中,$\phi^{(l)}$ 表示身份识别网络提取的第 l 层特征图,S_l 表示选择的特征层集合.

3.4.3 结构重建损失

人脸不同区域的合成难度不同,如面部等面积较大的平坦区域相对容易,而眼睛、嘴巴等区域较小,但结构精微,难以合成.然而式(3)中的重建损失,对于不同区域权重相同,可能导致复杂结构难以精细合成.因此,Yu 等[35]提出将人脸划分为眼睛、眉毛、嘴巴、面部皮肤、头发等 8 个区域,提升较小区域的权重,降低较大区域的权重,从而促使生成器生成更加精微的细节.具体而言,对每个区域以其对应面积的逆作为权重,这样等价于各个区域分别计算平均 L_1 重构损失,然后进行求和,即:

$$L_{\text{cmp}}=\sum_{c=1}^{8}\frac{1}{|M^{(c)}|}\|G(x)\otimes M^{(c)}-y\otimes M^{(c)}\|_1,$$

其中,$M^{(c)}$ 是图像分割结果,表示各个像素属于第 c 个区域的概率,$|M^{(c)}|$ 表示第 c 个区域包含的像素总数,$\otimes$ 表示点乘操作.实验结果表明,结构重建损失可以有效提升合成效果.

4 性能评估方法

现有异质人脸图像合成工作中,为了评价合成模型的优劣,通常采用多种方法,具体可以分为:主观评价、保真度评价、可解译度评价和真实度评价.其中,主观评价是指由人类观测者进行标记,保真度指合成图像与对应的真实图像之间的相似度,可解译度指合成图像可以用于识别等任务的能力,真实度指生成图像在内容、纹理等方面与真实图像的相近程度.各类方法又可以按照实际操作的不同进一步划分为几个子类.具体介绍如下.

4.1 主观评价

因为在大部分应用中,人是合成图像的最终使用者,因此人类主观评价结果是评判异质人脸图像合成图像质量优异的最可靠基准.人类观测者在主观评价过程中,通常会综合保真度、可解译度和真实度进行评价,因此也相对更为全面[101].在现有工作中,研究人员通常将合成图像发布在亚马逊劳务众包平台 Amazon Mechanical Turk (AMT)或问卷调查网站上,让大量非专业观测人员进行标记;或在线下招募志愿者对合成结果进行评价;之后将所有评价结果进行综合,作为对应异质图像合成算法的性能测度.按照 2002 年 VQEG 发布的图像主观质量评价建议书[102-103]中的划分标准,现有的主观评价方法主要为激励比较(stimulus comparison)法和单激励(single stimulus)法[104].具体介绍如下.

4.1.1 激励比较法

激励比较法一般分为两种形式:一是成对比较(paired comparison)法[6,8,105-106],即给定观测者两幅图像,观测者需要标记两幅图像的相对质量关系.现有工作中通常采用形容词分类判断法,即观察者需要从特定语义词汇集合中选择一个来描述两个观察图像之间的相对质量关系.这些词汇集合通常用来描述质量差异的存在性和方向性(例如“好”、“相同”、“差”).图像质量激励比较分类量表举例如表 1 所示.二是排序法,即给定一组图像序列,观测者对所有图像之间的相对质量关系进行排序,之后对不同观测者的排序结果进行综合,作为图像之间的相对质量关系描述.

成对比较法具有执行简单、结果直观的优点,但由于可以构建的图像对规模巨大,所以效率较低[107-108],经常用于图像生成模型性能评测.相比之下,排序法可以有效提升主观实验的效率,经常用于异质图像质量评价研究中,用于构建异质图像质量评价数据库和评估质量评价算法的精度[109-110].

表 1 图像质量激励比较分类量表举例

Table 1 Categorization list of stimulus comparison for image quality assessment

数值	形容词
-2	差很多
-1	略差
0	相同
1	略好
2	好很多

在现有工作中,在采用激励比较法时,通常采用强制选择的方式,即观测者被强制要求确定两幅图像中哪一幅优于或劣于另一幅图像.当两幅图像质量相差较小或有其他环境因素干扰时,观测者有可能无法感知到两幅图像的差异.这时,采用强制选择方式就会在标记数据中引入噪声.因此,对每一个图像对,都需要安排多个观测者评判其相对质量,然后进行综合得到最终的相对质量关系描述.

4.1.2 单激励法

单激励法在评测时,测试图像依次显示在屏幕上,观测者通过观察图像评估其质量.单激励法通常

采用形容词或数值分类判断法,即观察者从一组描述图像质量或损伤程度的形容词类别或分数等级中选择最合适的一个作为测试图像的质量描述.其质量量表和损伤量表如表 2 所示.

表 2　图像质量和损伤 5 级数值/类别量表

Table 2　Five grades of image quality or degradation

质量	损伤
5 优	5 不可察觉
4 良	4 可察觉,但不讨厌
3 中	3 稍微讨厌
2 差	2 讨厌
1 劣	1 很讨厌

4.2　客观评价

4.2.1　保真度

图像保真度,是指测试图像与原始图像相比,图像内容或信息的保留程度,或者说是测试图像与原始图像之间的相似程度[101,104].现有异质人脸图像合成工作中采用的保真度指标可以划分为以下两类:一种是通用图像的质量评价(Image Quality Assessment,IQA)方法,另外一种是针对特定人脸合成任务专门设计的专用图像质量评价方法.

4.2.1.1　通用图像质量评价方法

图像质量评价旨在评估图像中由模糊、噪声、压缩等引起的保真度.均方误差(Mean Squared Error,MSE)和峰值信噪比(Peak Signal-to-Noise Ratio,PSNR)由于计算简单、易于理解等原因,已成为图像质量评价的常用指标.然而,研究表明 MSE 与 PSNR 评价结果与人眼主观感受之间一致性较低,因此研究人员提出了很多新型质量评价方法[101].其中在异质人脸图像合成领域应用较广的包括以下 5 种方法:

1) 均方误差(MSE):即原始图像 R 与测试图像 I 在像素域中的平均 L_2 距离,可以表示为

$$E_{\mathrm{MS}}=\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left[I(i,j)-R(i,j)\right]^2,$$

其中 (i,j) 是位置坐标,m 和 n 是图像尺寸.E_{MS} 值越小,表明图像保真度越好.

2) 峰值信噪比(PSNR):描述了图像之间的差异与原始图像自身信息量的比值,其计算为

$$R_{\mathrm{PSN}}=10\cdot\log_{10}\left(\frac{I_{\max}^2}{E_{\mathrm{MS}}}\right),$$

R_{PSN}数值越大,表明图像保真度越好.

3) 结构相似度(Structure Similarity Index Measure,SSIM):2004 年,Wang 等[111]提出了结构相似度方法.该方法基于图像结构变化对于图像质量影响较大这一观察,对图像进行局部归一化作为其结构表示,然后通过比较原始图像与测试图像之间结构信息的相似度,作为图像质量的度量.SSIM 数值在 0 到 1 之间,越大表明测试图像保真度越好.

4) 特征相似度(Feature SIMilarity,FSIM):2011 年,Zhang 等[112]提出的特征结构相似度(Features Similarity Index,FSIM).该方法首先从图像中提取相位一致性特征和梯度幅度特征,然后计算测试图像和原始图像之间的特征相似度,并利用相位一致性进行加权求和,得到图像的最终质量值.FSIM 数值在 0 到 1 之间,越大表明测试图像保真度越好.

5) 视觉信息保真度(Visual Information Fidility,VIF):2006 年,Sheikh 等[113]从信息论的角度出发,将视觉感知过程建模为噪声通道,并将质量感知建模为信息提取过程.基于这一假设,给定测试图像和对应的原始图像,首先估计测试图像中包含的噪声,然后估计两者之间的共有信息,作为测试图像的质量描述,即视觉信息保真度.VIF 越大表明测试图像保真度越好.

4.2.1.2　专用图像质量评价方法

Wang 等[109]研究表明,传统图像质量评价并不适用于合成人脸画像的质量评价,其预测结果与人类主观感受之间一致性较低.Fan 等[110]进而提出了一种结构共现纹理(Structure Co-Occurrence Texture,Scoot)评测指标,同时将局部块空间结构和共现纹理统计特征考虑在内,用于评判合成人脸画像的质量,并且构建了较大规模的合成人脸画像数据集.专用型图像质量评价方法至今仍然处于起步阶段,有必要针对异质人脸图像合成的质量评价问题,进行更加深入、具体的分析,探索与人眼主观感受相一致的专用型图像质量评价方法,以用于合成算法的性能评估和优化.

4.2.2　可解译度

图像的可解译度,也称为可用度,是指人们能够从图像中提取信息的能力,这可以描述图像用于指定任务的能力.现有的异质图像可解译度可以分为身份解译度和属性解译度两类.

1)身份解译度:异质人脸图像合成过程中,需要在生成指定风格模态的同时,尽可能保留输入人脸的身份判别信息,因此现有工作通常利用合成人脸的身份识别或身份匹配精度作为合成图像质量的评

价指标之一.这一指标对应了图像质量中的可解译度或可用度,即其可以用于人脸识别的程度.具体操作中,通常对同一模态的图像进行人脸识别实验.在人脸画像合成实验中,研究人员一般将真实人脸画像作为数据库,将合成人脸画像作为查询样本,然后利用基于 FisherFace[114]、特征脸 EigenFace 或零空间线性判别分析(Null-Space Linear Discriminant Analysis,NLDA)[115]的人脸识别方法,进行身份识别.在具体试验中,通常将测试集的全部样本随机划分为人脸识别模型的训练集和测试集,依此多次重复进行人脸识别的训练-测试实验,统计平均识别精度,作为合成人脸图像可解译度评价指标.此外,研究人员还研发了大量基于深度学习的人脸识别工作,也可以应用于合成人脸的身份识别中.例如,Shi 等[51]将从人脸照片数据与训练好的 SphereFace 网络[116]在人脸漫画图像上进行微调,使其适用于漫画人脸识别.之后,在测试阶段,将其用于计算合成漫画的身份识别精度,作为可解译度的度量.

2)属性解译度:是指可以从图像中识别出目标属性的程度,通常作为任务特定的评价指标.例如,年龄合成的目的是合成目标年龄段的人脸,因此合成人脸需要符合目标年龄段的特点,即可以被人识别为指定年龄段.因此,可以考虑对合成人脸进行年龄估计,计算估计年龄与实际年龄之间的差异或相近程度,作为年龄合成算法的性能指标之一.而在人脸美颜任务中,则可以利用人脸吸引力预测模型评估合成人脸的美观程度,用于对应人脸美颜方法的性能指标.一般情况下,这里采用的属性预测模型可以是在标准数据集上预训练好的,并且具有较高的分类精度.

4.2.3 真实度

为了评价合成人脸图像的真实度,即其是否逼真,现有工作中通常采用 Inception Score (IS)和 Fechit Inception Distance (FID)作为评测指标.IS 和 FID 已经广泛应用于图像生成任务中.特别是 FID 在描述生成图像真实度和多样性方面,与人眼主观感受表现出了较高的一致性.

1)IS:描述了某一特征空间中,生成图像的分布与真实目标图像对应的分布之间的 KL-divergence 距离[8].IS 分数越高,一般可以认为生成图像质量更为真实.但 IS 对使用的深度学习框架高度敏感,而且不能反映过度拟合与模型坍塌现象,因此单独使用 IS 评价图像真实度并不精确.

2)FID:主要描述了某一特征空间中,生成图像的分布与真实目标图像对应的分布之间的推土机距离(Earth Mover Distance,EMD)[8].FID 越小,表明生成图像越逼真.FID 对模型坍塌更加敏感.相比较 IS 来说,FID 对噪声有更好的鲁棒性.因为假如只有一种图片时,FID 这个距离将会相当高.因此,FID 更适合描述生成图像的多样性.

现在,普遍采用 Inception V3 网络[117]的最后一层输出作为图像特征表达,然后用于计算 IS 或 FID.然而,Inception V3 是在 ImageNet 上针对一般图像目标检测任务训练的,用 Inception V3 特征计算 IS 或 FID,然后将其应用于评估合成的异质人脸图像真实度,这是有问题的.有必要针对特定的异质人脸图像类型,训练专门的深度特征提取网络,用于计算 IS 和 FID,以计算合成图像的真实度.

5 数据集及性能分析

现在,除人脸画像合成之外,其他异质人脸图像合成工作相对较少,或普遍缺少统一的基准数据集和评价准则,难以得出可靠的对比分析结论.因此本章节主要介绍相对系统的部分工作,包括人脸画像合成、漫画合成、年龄合成和人脸美颜 4 个方面的数据集,以及对部分合成结果进行性能分析.

5.1 数据集

5.1.1 画像合成数据集

人脸画像合成的任务研究已久,而且已经取得了很大进展.其中应用最为广泛的人脸照片-画像数据集是香港中文大学(Chinese University of Hong Kong,CUHK)的 CUHK Face Sketch (CUFS)数据库[19]和 CUHK Face Sketch FERET (CUFSF)数据库[118].其信息如下:

1)CUFS:总共包含 606 对人脸画像和照片数据,每张照片存在一幅画家手绘的画像.这些照片均为单张正面中性表情证件照,分别来自 3 个子库:CUHK Student 数据库[119](188 张)、AR 数据库[120](123 张)和 XM2VTS 数据库[121](295 张).

2)CUFSF:包含 1 194 对人脸画像-照片,其中照片来自于 FERET 数据库[122],每人有一张照片和一幅画像.CUFSF 数据库中的画像有更多的夸张成分,照片与画像之间存在较为严重的非对齐情况,因此更具挑战性.

5.1.2 人脸漫画数据集

当前规模较大、使用较为广泛的数据集主要有

IIT-CFW[123]和 WebCaricature[124],其信息如下:

1) IIIT-CFW[123]:是一个非可控卡通人脸数据集,包含了 8 928 张不同职业的世界知名人士的注释卡通人脸.它还为交叉模态检索任务提供了 1 000 个公众人物的真实面孔.

2) WebCaricature[124]:网络漫画数据集包含了 252 位名人的 5 974 张照片和 6 042 幅漫画图像,是目前最大的漫画数据集.由南京理工大学研究人员构建,主要用于动漫人脸识别及合成任务.

5.1.3 年龄合成数据集

人脸年龄合成数据集主要为 Morph[125]、CACD[126]和 FGNET[127],相关信息如下:

1) Morph[125]:包含 13 000 多人的 55 000 张独特图像.年龄从 16 岁到 77 岁,平均年龄为 33 岁.每个人的平均图像数量为 4 张,照片之间的平均时间间隔为 164 天,最小为 1 天,最大为 1 681 天.

2) CACD[126]:首先找到超过 200 000 张图像,其中包含所有 2 000 名名人的面部,然后使用基于低层特征的简单重复检测算法来去除重复图像,故该数据集包含 2 000 名名人的 163 446 张图像,这是已知的最大的公开跨年龄数据集.

3) FG-NET[127]:发布于 2004 年,旨在研究由衰老引起的面部外观变化,包含 82 名受试者的 1 002 张图像,年龄从 0 到 69 岁,每张图像含有 68 个手动标注的关键点.

5.1.4 人脸美颜数据集

现有的人脸美颜数据集主要是研究人员自己搜集的,现有工作里提及的数据集主要有以下两个:

1) Multi-Modality Beauty (M^2B)[128]:M^2B 数据库包括 1 240 位女性的面部、着装图像和音频文件,并为每种情态提供相应的吸引力评分.评分范围在 1 到 10 分之间,1 分是最低的美丽水平,10 分是最高的.在我们的模型的训练过程中,只使用了人脸图像方面的信息.

2) Makeup Transfer(MT)[129]:包含 1 115 张源图像和 2 719 张参照上妆图像,有多种姿态和表情变化.

除此之外,还有一些数据都可以考虑作为人脸美颜的辅助数据,用于人脸的美化或人脸美学因素的挖掘和分析.例如,CelebA 数据集[130]也包含近 20 万张人脸图像,每幅图像给出了是否具有吸引力的二值化标签;SCUT-FBP5500 数据集[131]包含了 5 500 张人脸图像,在每张图像都依据其美观程度,给出了大量观测者给出的分数分布、平均分数等标记数据.

5.2 性能分析

5.2.1 画像合成性能分析

我们考虑最新的一些基于生成对抗网络的人脸画像合成工作.表 3 中列出了人脸画像合成数据集及部分工作的性能定量分析数据.在各个数据集上,各个最优性能指标进行了加粗显示.通过对比可以发现,改进型的生成对抗网络,在各个指标上普遍由于原始的 cGAN 和 CycleGAN 模型.整体而言,SCA-GAN 取得了最优的 FID 指标,说明其生成的人脸画像最为逼真;而且其在 CUFSF 数据集上取得了最优的性能.SS-GAN 也取得了较好的性能,说明采用附加数据和半监督学习算法,可以有效提升网络的训

表 3 人脸画像合成数据集及部分工作的性能定量分析

Table 3 Performance analysis of face sketch synthesis datasets and results

数据库	样本总量	训练/测试	方法	SSIM	FSIM	NLDA	FID
CUFS	606	CUHK(88/100) AR(80/43) XM2VTS(100/195)	Pix2Pix[9]		0.711	0.931	43.2
			CycleGAN[16]	0.581	0.723		
			BPGAN[26]		0.691	0.931	86.1
			SS-GAN[30]	0.546	0.726	**0.982**	
			SCA-GAN[35]		0.716	0.957	**34.2**
			Col-Nets[36]	**0.645**			
			MDAL[37]	0.528	**0.728**	0.968	
CUFSF	1194	250/944	Pix2Pix[9]		0.728	0.710	29.2
			CycleGAN[16]		0.706		
			BPGAN[26]		0.682	0.675	42.9
			SS-GAN[30]	**0.409**	0.716	**0.780**	
			SCA-GAN[35]		**0.729**	**0.780**	**18.2**
			MDAL[37]	0.382	0.708	0.671	

练效果.Col-Nets 和 MDAL 方法都基于协作生成思想,也在部分指标上具有优异表现.

图 9 显示了部分人脸画像合成结果,包括 Pix2Pix[9]、CycleGAN[16]、Col-Nets[36]、Col-GAN[36]等方法.可以看出,使用原始生成对抗网络 Pix2Pix 或 CycleGAN 难以合成逼真的细节.对比之下,基于改进的生成对抗网络,普遍可以得到较为逼真的人脸画像.具体而言,SCA-GAN 和 Col-cGAN 在纹理细节上都有不错的表现.Col-Nets 的合成结果局部区域间过渡较为平滑,对比度相对较弱.整体而言,现在的人脸画像合成方法已经取得了较为满意的成果.

5.2.2 漫画合成性能分析

在现有的人脸漫画合成中,普遍使用主观评价方法来评判合成漫画人脸的质量.例如,CariGAN[50]和 WarpGAN[51]都通过让观测者对合成人脸进行打分,评估不同方法合成漫画的视觉质量、身份保持程度或者夸张程度.最后对所有评价进行综合,作为各个方法的性能评估结果.此外文献[48]中还使用了 IS 作为合成漫画人脸的性能指标.而 WarpGAN[51]中则使用了两种人脸识别模型测试合成漫画的可解译度.具体而言,其使用了一种商用的离线人脸匹配模型(Commercial-Off-The-Shelf, COTS)和 SphereFace[116]用来进行照片-照片、手绘漫画-照片、WarpGAN 生成漫画-照片之间的身份匹配任务.表 4 给出了 WarpGAN 在 WebCaricature 数据集上对应的人脸识别精度和主观质量评价结果.可以看出,WarpGAN 在人脸身份保持、几何结构变形和视觉质量方面都取得了不错的效果.

由于人脸漫画与照片之间存在严重的几何变形,而大部分基于生成对抗网络的漫画合成算法缺少有效的几何变形机制,因此难以合成高质量的漫画图像.基于最新的进展,WarpGAN 方法[51]展现出了最优的合成效果.图 10 给出了 CycleGAN 和 WarpGAN 的部分合成结果(所有结果来自文献[51]).可以看出,CycleGAN 难以实现人脸几何结构变形和漫画纹理的生成.相比而言,WarpGAN 有效模拟了人脸的几何形变.这主要是因为在 WarpGAN 中引入了空间变形(Spatial Transformer)模块[132].不过,其在几何形变和纹理生成的细节等方面还存在一定的缺陷,而且存在合成失败的情形,还有待进一步提升.

表 4 部分人脸漫画合成模型在 WebCaricature 数据集上的性能定量分析

Table 4 Performance analysis of existing face caricature synthesis models on the WebCaricature dataset

方法	与照片之间的身份识别精度		主观评价(满分 10 分)	
	COTS	SphereFace	视觉质量	夸张程度
照片	0.948	0.908		
手绘漫画	0.413	0.458	7.70	7.16
WarpGAN 生成漫画	0.790	0.727	5.61	4.87
CycleGAN 生成漫画			2.43	2.27

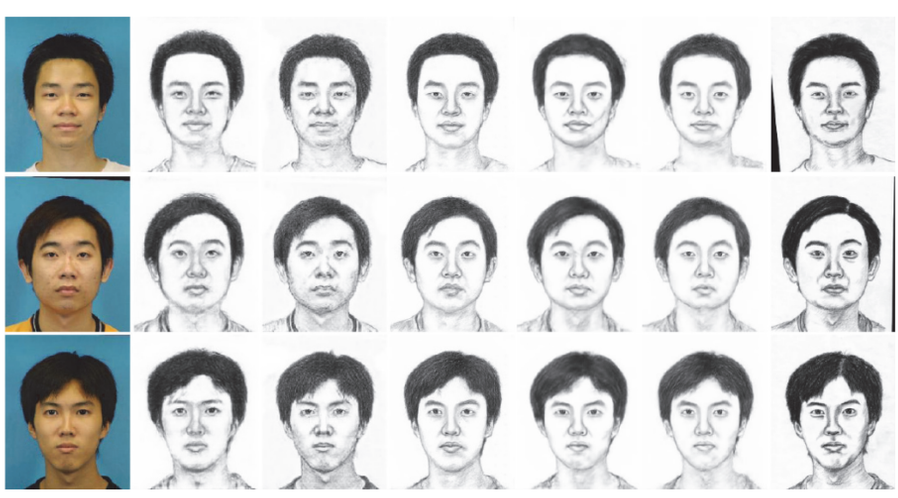


图 9 人脸画像合成结果示例,从左到右依次为输入照片,分别由 Pix2Pix、CycleGAN、SCA-GAN、Col-Nets、Col-GAN 生成的画像,以及真实画像

Fig. 9 Illustration of synthesized face sketches. From left to right: input photo, Pix2Pix, CycleGAN, SCA-GAN, Col-Nets, Col-GAN, and the ground truth sketches

图 10　人脸漫画合成结果示例图,从左到右依次为输入照片、CycleGAN 合成以及 WarpGAN 合成的三个结果(图像取自 WarpGAN[51])

Fig. 10　Illustration of synthesized face caricatures,from left to right:input photo,CycleGAN,three results of WarpGAN[51]

5.2.3　年龄合成性能分析

在年龄合成任务中,通常将年龄划分为几个阶段,然后训练模型使其可以基于输入年龄段图像,合成目标年龄段图像.在具体性能评测中,主要使用了对于合成图像的年龄估计精度作为主要指标.具体而言,一方面可以通过主观实验,让观测者估计合成图像的年龄,然后与目标年龄进行比对;或者,给定不同年龄段的合成图像,判断其对应年龄之间的相对关系,从而评判合成效果.另一方面,可以采用预训练好的模型,预测合成人脸图像的年龄,进而计算预测值与目标值之间的一致性,利用平均精度或混淆矩阵来描述年龄合成算法的性能.此外,合成人脸也需要保持输入人脸的身份信息,因此身份识别精度也是常用指标之一.表 5 和表 6 显示了部分当前算法的年龄解译度和身份解译度评价结果[68].整体而言,现有方法对于成年阶段的年龄合成任务,在年龄解译度和身份解译度方面都取得了不错的性能.

图 11 显示了 Dual cGAN 方法[67]的部分跨年龄

表 5　部分年龄合成模型在 Morph 和 CACD 数据集上的年龄解译度结果[68]

Table 5　Age estimation results of age progression models on the Morph and CACD datasets [68]

年龄组		Morph			CACD		
		31~40	41~50	≥51	31~40	41~50	≥51
估计年龄分布的平均值	真实图像	38.60	47.74	57.25	38.51	46.54	53.39
	合成图像	38.47	47.55	56.57	38.88	47.42	54.05
各年龄组平均年龄之间的差异	CAAE [63]	10.08	15.49	21.42	5.76	11.53	17.93
	GLCA-GAN [66]	0.23	3.61	8.61	1.72	2.07	2.85
	PAG-GAN [65]	0.38	0.52	1.48	0.70	**0.22**	**0.57**
	AFAW-GAN [68]	**0.13**	**0.19**	**0.68**	**0.37**	0.58	0.66

表 6　部分年龄合成模型在 Morph 和 CACD 数据集上的身份解译度结果[68]

Table 6　Face verification results of age progression models on the Morph and CACD datasets[68]　%

	年龄组	Morph			CACD		
		31~40	41~50	≥51	31~40	41~50	≥51
身份认证可信度	~30	95.77	94.64	87.53	93.67	91.54	90.32
	31~40		95.47	98.53		91.74	90.54
	41~50			90.50			91.12
身份认证精度	CAAE [63]	15.07	12.02	8.22	4.66	3.41	2.40
	GLCA-GAN [66]	97.66	96.67	91.85	97.72	94.18	92.29
	PAG-GAN [65]	100.00	98.91	93.09	**99.99**	**99.81**	98.28
	AFAW-GAN [68]	**100.00**	**100.00**	**98.26**	99.76	98.74	**98.44**

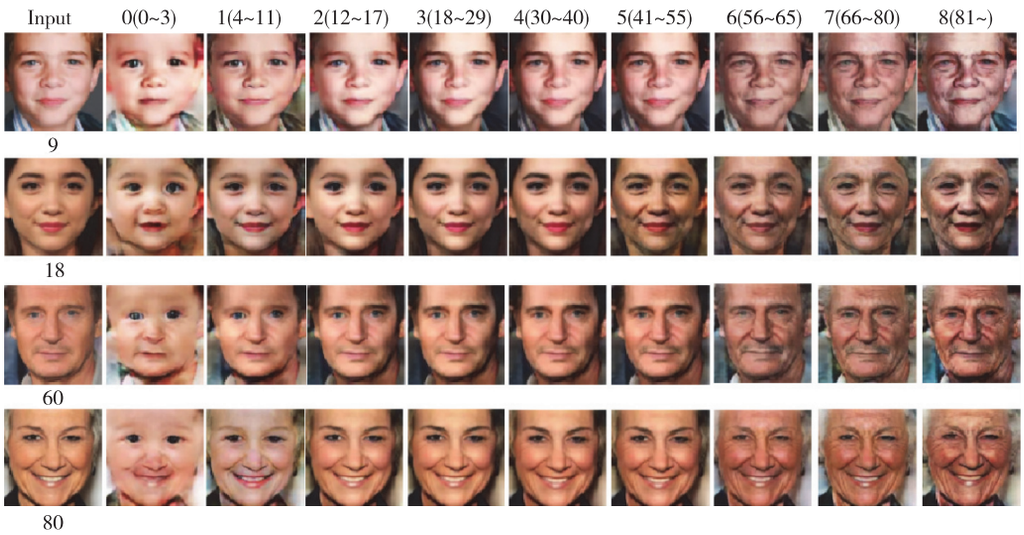


图 11　人脸年龄合成结果示例,最左列为输入人脸图像及其对应的年龄,
其余为 Dual cGAN 方法[67]合成的人脸图像及其对应的年龄段(图像来自文献[67])

Fig. 11　Illustration of face age progression, the first column shows the input image and the corresponding age, and the other columns show synthesized results by Dual cGAN [67] and the corresponding ages (all images are from[67])

段合成结果.可以看出,对于成年阶段的人脸,Dual cGAN 可以合成出较为合理的图像.然而,对于幼年阶段的人脸,合成图像中存在较为严重的结构失真和模糊等.这也是现在的年龄合成模型普遍存在的现象.这是由于大部分人脸在成年之前存在幅度较大的几何变形,但现有的模型普遍是基于卷积操作,难以有效捕捉几何结构上的变形机制.

5.2.4　人脸美颜性能分析

现有的人脸美颜相关工作,主要集中在人脸上妆功能,主要涉及嘴部、面部、眼部等关键部位的妆容.而且,现有工作一般是给定一张人脸源图像,一张上妆参考图像,实现妆容的迁移.图 12 给出了 PSGAN 模型的上妆效果图.可以看出,其可以实现指定部位、指定程度的妆容迁移,而且对于姿态、表情等十分鲁棒.不过,实际应用中,人脸美颜不一定只是上妆,可能还包括光影、色彩、构图的调整.因此,有必要探索更为通用的多功能人脸美颜机制.

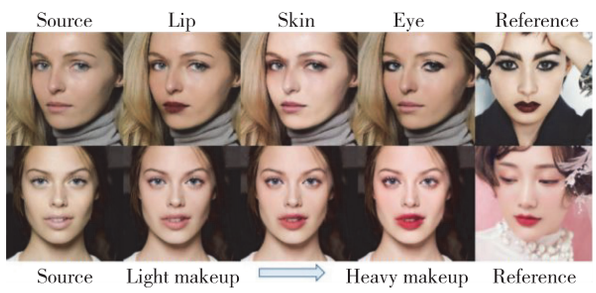


图 12　人脸自动美颜结果示意图,最左列为输入源图像,最右列为参考妆容图像,中间三幅从左到右依次为对嘴唇、皮肤、眼睛上妆,且从淡到浓(所有图像来自 PSGAN[82])

Fig. 12　Illustration of face beautification, the first column show the input faces, the last column show the reference makeup, and the three columns in the middle show the transferred faces with light makeup to heavy makeup (all results are from PSGAN[82])

6　问题与挑战

尽管异质人脸图像合成领域已经取得了巨大进展,现有模型仍然存在一定的局限性.具体而言,异质人脸合成单依然面临着以下挑战:

1)不可控人脸图像:现在的异质人脸合成数据集中的图像普遍为正面的中性表情人脸.而不可控条件下的人脸图像通常在光照、姿态、表情等方面存在巨大差异.当把从标准数据集上训练得到的模型应用到不可控人脸图像时,通常难以合成合理、高质量的异质人脸图像.尽管相关方面已经有了一定探索,但仍然有必要开展更多的工作,进一步提升性能,以推进相关技术的落地和应用.

2)几何形变:异质人脸图像合成面临的变形问题主要有 3 个方面:(a) 图像域迁移过程不希望人脸变形,但标记数据存在变形,限制了迁移模型的学习效率和性能,如画像合成;(b) 迁移后图像与输入图像之间存在小幅度形变,如年龄合成;(c) 迁移后图像与输入图像之间存在大幅度的几何形变,如漫

画合成.现有图像域迁移普遍基于标准卷积操作,难以有效克服人脸几何形变的影响或模拟人脸几何形变的机制,导致迁移后图像存在视觉质量差、辨识度低等问题.如何克服标记数据中几何变形噪声对于合成模型性能的限制,探索人脸几何变形机制,对于提升合成图像的真实度具有重要意义.

3)多风格:现有的异质人脸图像合成通常只考虑一种风格.比如大部分画像合成模型只能合成一种类型,而不同的画家风格会有所不同;现在的人脸美颜也普遍只考虑自动上妆功能,限制了人脸美化的多样性.如何提升多风格的异质人脸图像合成,利用一个模型实现多种风格,甚至多种模态图像的合成,以提升模型的多样性和可靠性,是相关技术理论发展和应用推广的重点和难点.

4)精细生成:现有的异质人脸合成普遍针对相对低分辨率的人脸图像,而且合成图像在纹理细节、精微结构方面存在不足.然而,在物质生活日益丰富的今天,用户对于图像质量的需求逐步提升,如何实现高精度、高质量的人脸图像合成,对于提升用户体验极为重要.

5)视频序列:随着视频媒体及视频社交网络的日益盛行,针对视频序列,探索高效、高质量异质人脸合成变得极为重要.对于人脸视频异质生成,如何结合时空域纹理细节、几何结构等连续性对于合成视频的质量感知至关重要.对此可以参考视频风格迁移相关工作[133-134].

6)性能评估:现在的异质人脸图像合成工作仍然缺少可靠、精准的性能评估准则.在现有工作中,所使用的评估指标本身的可靠性仍然有待进一步验证,而且不同工作采用不同的性能指标,这对相关工作的性能评估以及对比分析造成了巨大困难.针对特定异质人脸图像合成任务,开发专用的图像视觉质量评价方法、异质人脸图像识别模型、跨模态人脸识别方法等,都是未来有待完成的工作.

7 结论

本文对于基于生成对抗网络的异质人脸图像合成工作进行了回顾和总结.首先,针对画像合成、漫画合成、年龄合成和人脸美颜以及其他异质人脸图像合成任务,分别概述了最新的技术进展.然后,从输入、生成器、判别器、目标函数4个方面对现有工作中的生成对抗网络模型进行了归纳.其次,从主观评价、保真度、可解译度和真实度等方面对现有工作中所采用的性能评估方法进行了介绍.之后,对于现有工作中常见的数据集和已有方法的表现进行了总结和分析,指出了现有方法的进展和局限性.最后,对于异质人脸图像合成领域面临的挑战进行了总结.

参考文献

References

[1] 高新波,王楠楠.异质人脸图像合成[M]//张长水,杨强.机器学习及其应用2013.北京:清华大学出版社,2013:77-91
GAO Xinbo, WANG Nannan. Heterogeneous facial image synthesis [M] // ZHANG Changshui, YANG Qiang. Machine learning and its applications 2013. Beijing: Tsinghua University Press, 2013: 77-91

[2] Wang N N, Tao D C, Gao X B, et al. A comprehensive survey to face hallucination[J]. International Journal of Computer Vision, 2014, 106(1): 9-30

[3] Nguyen K, Fookes C, Sridharan S, et al. Super-resolution for biometrics: a comprehensive survey[J]. Pattern Recognition, 2018, 78: 23-42

[4] Wang N N, Zhu M R, Li J, et al. Data-driven vs. model-driven: fast face sketch synthesis[J]. Neurocomputing, 2017, 257: 214-221

[5] 王楠楠,李洁,高新波.人脸画像合成研究的综述与对比分析[J].模式识别与人工智能,2018,31(1):37-48
WANG Nannan, LI Jie, GAO Xinbo. A review and comparison study on face sketch synthesis[J]. Pattern Recognition and Artificial Intelligence, 2018, 31(1): 37-48

[6] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets [C] // Advances in Neural Information Processing Systems, 2014: 2672-2680

[7] Karras T, Aila T, Laine S, et al. Progressive growing of GANs for improved quality, stability, and variation[J]. arXiv Preprint, 2017, arXiv: 1710. 10196

[8] Creswell A, White T, Dumoulin V, et al. Generative adversarial networks: an overview[J]. IEEE Signal Processing Magazine, 2018, 35(1): 53-65

[9] Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1125-1134

[10] Zhang H, Xu T, Li H S, et al. StackGAN++: realistic image synthesis with stacked generative adversarial networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(8): 1947-1962

[11] Johnson J, Alahi A, Li F F. Perceptual losses for real-time style transfer and super-resolution[C] // European Conference on Computer Vision, 2016: 694-711

[12] Huang X, Liu M Y, Belongie S, et al. Multimodal unsupervised image-to-image translation[C] // European Conference on Computer Vision, 2018: 179-196

[13] Li M J, Huang H Z, Ma L, et al. Unsupervised image-to-image translation with stacked cycle-consistent

adversarial networks[C]//European Conference on Computer Vision,2018:186-201
[14] Zhang H,Xu T,Li H S,et al.Stackgan++:realistic image synthesis with stacked generative adversarial networks[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,2018,41(8):1947-1962
[15] Mao X D,Li Q,Xie H R,et al.Least squares generative adversarial networks[C]// IEEE International Conference on Computer Vision (ICCV), 2017:2794-2802
[16] Zhu J Y,Park T,Isola P,et al.Unpaired image-to-image translation using cycle-consistent adversarial networks[C]// IEEE International Conference on Computer Vision (ICCV),2017:2223-2232
[17] Wang Z H,Chen J,Hoi S C H.Deep learning for image super-resolution:a survey[J]. arXiv Preprint, 2019, arXiv:1902.06068
[18] Ha V K,Ren J C,Xu X Y,et al.Deep learning based single image super-resolution:a survey[J]. International Journal of Automation and Computing, 2019, 16(4):413-426
[19] Wang X G,Tang X O.Face photo-sketch synthesis and recognition[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,2009,31(11):1955-1967
[20] Zhang D Y, Lin L, Chen T S, et al. Content-adaptive sketch portrait generation by decompositional representation learning[J].IEEE Transactions on Image Processing,2017,26(1):328-339
[21] Jiao L C, Zhang S B, Li L L, et al. A modified convolutional neural network for face sketch synthesis[J].Pattern Recognition,2018,76:125-136
[22] Sheng B,Li P,Gao C H,et al.Deep neural representation guided face sketch synthesis[J].IEEE Transactions on Visualization and Computer Graphics, 2018, 25(12):3216-3230
[23] Zhang M J,Wang N N,Gao X B,et al.Markov random neural fields for face sketch synthesis[C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence,2018:1142-1148
[24] Zhang M J, Wang N N, Li Y S, et al. Face sketch synthesis from coarse to fine[C]//Thirty-Second AAAI Conference on Artificial Intelligence,2018:7558-7565
[25] Zhang M J,Wang N N,Li Y S,et al.Deep latent low-rank representation for face sketch synthesis[J].IEEE Transactions on Neural Networks and Learning Systems,2019,30(10):3109-3123
[26] Wang N N,Zha W J,Li J,et al.Back projection:an effective postprocessing method for GAN-based face sketch synthesis[J]. Pattern Recognition Letters, 2018, 107:59-65
[27] Wang L D,Sindagi V,Patel V.High-quality facial photo-sketch synthesis using multi-adversarial networks[C]// 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018),2018:83-90
[28] Zhang S C,Ji R R,Hu J,et al.Robust face sketch synthesis via generative adversarial fusion of priors and parametric sigmoid [C] // Proceedings of the 27th International Joint Conference on Artificial Intelligence, 2018:1163-1169
[29] Zhang S C,Ji R R,Hu J,et al.Face sketch synthesis by multidomain adversarial learning[J].IEEE Transactions on Neural Networks and Learning Systems,2019,30(5):1419-1428
[30] Chen C F,Liu W,Tan X,et al.Semi-supervised learning for face sketch synthesis in the wild[C]//Asian Conference on Computer Vision,2019:216-231
[31] Bae S,Ud Din N,Javed K,et al.Efficient generation of multiple sketch styles using a single network[J].IEEE Access,2019,7:100666-100674
[32] Zhang M J,Li J,Wang N N,et al.Compositional model-based sketch generator in facial entertainment[J].IEEE Transactions on Cybernetics,2018,48(3):904-915
[33] Yi R,Liu Y J,Lai Y K,et al.APDrawingGAN:generating artistic portrait drawings from face photos with hierarchical GANs[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019:10743-10752
[34] Zhang M J,Wang R X,Gao X B,et al.Dual-transfer face sketch-photo synthesis[J].IEEE Transactions on Image Processing,2019,28(2):642-657
[35] Yu J,Shi S J,Gao F,et al.Towards realistic face photo-sketch synthesis via composition-aided GANs[J].arXiv Preprint,2017,arXiv:1712.00899
[36] Zhu M R,Li J,Wang N N,et al.A deep collaborative framework for face photo-sketch synthesis[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019,30(10):3096-3108
[37] Zhang S C,Ji R R,Hu J,et al.Face sketch synthesis by multidomain adversarial learning[J].IEEE Transactions on Neural Networks and Learning Systems,2019,30(5):1419-1428
[38] Sadimon S B,Sunar M S,Mohamad D,et al.Computer generated caricature:a survey[C]//IEEE International Conference on Cyberworlds,2010:383-390
[39] Hill M Q, Parde C J, Castillo C D, et al. Deep convolutional neural networks in the face of caricature:identity and image revealed[J].arXiv preprint, 2018, arXiv:1812.10902
[40] Wu Q Y,Zhang J Y,Lai Y K,et al.Alive caricature from 2D to 3D[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition,2018:7736-7345
[41] Zhou J Y,Wu H T,Liu Z C,et al.3D cartoon face rigging from sparse examples[J].The Visual Computer,2018,34(9):1177-1187
[42] Huo J, Li W B, Shi Y H, et al. WebCaricature: a benchmark for caricature recognition[J].arXiv Preprint, 2017,arXiv:1703.03230
[43] Akleman E. Making caricatures with morphing[C]// Visual Proceedings: The Art and Interdisciplinary Programs of SIGGRAPH'97, 1997, DOI:10.1145/259081.259231
[44] Liang L,Chen H,Xu Y Q,et al.Example-based caricature

generation with exaggeration[C]// Proceedings of 10th Pacific Conference on Computer Graphics and Applications,2002:386-393

[45] Liu Z Q,Chen H,Shum H Y.An efficient approach to learning inhomogeneous Gibbs model[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition,2003,DOI:10.1109/cvpr.2003.1211385

[46] Chiang P Y,Liao W H,Li T Y.Automatic caricature generation by analyzing facial features[C]// Proceeding of 2004 Asia Conference on Computer Vision (ACCV2004),2004

[47] Garg J,Peri S V,Tolani H,et al.Deep cross modal learning for caricature verification and identification (CaVINet)[C]// Proceedings of the 26th ACM International Conference on Multimedia,2018:1101-1109

[48] Zheng Z Q,Wang C,Yu Z B,et al.Unpaired photo-to-caricature translation on faces in the wild[J].Neurocomputing,2019,355:71-81

[49] Han X G,Hou K C,Du D,et al.CaricatureShop:personalized and photorealistic caricature sketching[J].arXiv Preprint,2018,arXiv:1807.09064

[50] Li W B,Xiong W,Liao H F,et al.CariGAN:caricature generation through weakly paired adversarial learning[J].arXiv Preprint,2018,arXiv:1811.00445

[51] Shi Y,Deb D,Jain A K.WarpGAN:automatic caricature generation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2019:10762-10771

[52] Shu X B,Xie G S,Li Z C,et al.Age progression:current technologies and applications[J].Neurocomputing,2016,208:249-261

[53] Shu X B,Tang J H,Lai H J,et al.Kinship-guided age progression[J].Pattern Recognition,2016,59:156-167

[54] Ramanathan N,Chellappa R.Modeling shape and textural variations in aging faces[C]// 2008 8th IEEE International Conference on Automatic Face & Gesture Recognition,2008:1-8

[55] Suo J L,Chen X L,Shan S G,et al.A concatenational graph evolution aging model[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,2012,34(11):2083-2096

[56] Todd J T,Mark L S,Shaw R E,et al.The perception of human growth[J].Scientific American,1980,242(2):132-144

[57] Wu Y,Thalmann N M,Thalmann D.A plastic-visco-elastic model for wrinkles in facial animation and skin aging[C]// Proceedings of the second Pacific Conference on Fundamentals of Computer Graphics,1994:201-213

[58] Wang Y H,Zhang Z X,Li W X,et al.Combining tensor space analysis and active appearance models for aging effect simulation on face images[J].IEEE Transactions on Systems,Man,and Cybernetics,Part B (Cybernetics),2012,42(4):1107-1118

[59] Kemelmacher-Shlizerman I,Suwajanakorn S,Seitz S M.Illumination-aware age progression[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2014:3334-3341

[60] Yang H Y,Huang D,Wang Y H,et al.Face aging effect simulation using hidden factor analysis joint sparse representation[J].IEEE Transactions on Image Processing,2016,25(6):2493-2507

[61] Zhang Z F,Song Y,Qi H R.Age progression/regression by conditional adversarial autoencoder[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2017:5810-5818

[62] Zhou S Y,Zhao W Q,Feng J S,et al.Personalized and occupational-aware age progression by generative adversarial networks[J].arXiv Preprint,2017,arXiv:1711.09368

[63] Yang H Y,Huang D,Wang Y H,et al.Learning face age progression:a pyramid architecture of GANs[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition,2018:31-39

[64] Li P P,Hu Y B,Li Q,et al.Global and local consistent age generative adversarial networks[C]// 24th International Conference on Pattern Recognition (ICPR),2018,DOI:10.1109/ICPR.2018.8545119

[65] Song J K,Zhang J Q,Gao L L,et al.Dual conditional GANs for face aging and rejuvenation[C]// Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI),2018:899-905

[66] Kossaifi J,Tran L,Panagakis Y,et al.GAGAN:geometry-aware generative adversarial networks[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition,2018:878-887

[67] Wang W,Cui Z,Yan Y,et al.Recurrent face aging[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2016:2378-2386

[68] Nhan D C,Gia Q K,Luu K,et al.Temporal non-volume preserving approach to facial age-progression and age-invariant face recognition[C]// Proceedings of the IEEE International Conference on Computer Vision,2017:3735-3743

[69] Palsson S,Agustsson E,Timofte R,et al.Generative adversarial style transfer networks for face aging[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW),2018:2084-2092

[70] Sdol E.Trends in cognitive sciences[J].Talk Psychology Journals,2009,19(4):9-11

[71] Shi S J,Gao F,Meng X T,et al.Improving facial attractiveness prediction via co-attention learning[C]// IEEE International Conference on Acoustics,Speech and Signal Processing (ICASSP),2019,DOI:10.1109/ICASSP.2019.8683112

[72] Chen F M,Xiao X H,Zhang D.Data-driven facial beauty analysis:prediction,retrieval and manipulation[J].IEEE Transactions on Affective Computing,2018,9(2):205-216

[73] Liang L Y,Jin L W,Li X L.Facial skin beautification using adaptive region-aware masks[J].IEEE Transactions on Cybernetics,2014,44(12):2600-2612

[74] Liang L Y,Jin L W,Liu D.Edge-aware label propagation

for mobile facial enhancement on the cloud[J]. IEEE Transactions on Circuits and Systems for Video Technology,2017,27(1):125-138

[75] Leyvand T,Cohen-Or D,Dror G,et al. Digital face beautification[C]// ACM SIGGRAPH 2006 Sketches,2006, DOI:10.1145/1179849.1180060

[76] Li J S,Xiong C,Liu L Q,et al. Deep face beautification [C]// Proceedings of the 23rd ACM International Conference on Multimedia,2015:793-794

[77] Guo D,Sim T. Digital face makeup by example[C]// 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009, DOI: 10.1109/CVPR.2009.5206833

[78] Alashkar T,Jiang S Y,Wang S Y,et al. Examples-rules guided deep neural network for makeup recommendation [C]// Thirty-First AAAI Conference on Artificial Intelligence,2017:941-947

[79] Ou X Y,Liu S,Cao X C,et al. Beauty emakeup:a deep makeup transfer system[C]// Proceedings of the 24th ACM International Conference on Multimedia, 2016: 701-702

[80] Jiang W T, Liu S, Gao C, et al. PSGAN: pose-robust spatial-aware GAN for customizable makeup transfer[J]. arXiv Preprint,2019,arXiv:1909.06956

[81] Han H,Jain A K,Wang F,et al. Heterogeneous face attribute estimation:a deep multi-task learning approach [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2018,40(11):2597-2609

[82] Wang Z L, Chen Z Z, Wu F. Thermal to visible facial image translation using generative adversarial networks [J]. IEEE Signal Processing Letters, 2018, 25(8): 1161-1165

[83] Dou H,Chen C,Hu X Y,et al. Asymmetric cyclegan for unpaired NIR-to-RGB face image translation[C]// IEEE International Conference on Acoustics,Speech and Signal Processing (ICASSP), 2019, DOI: 10.1109/ICASSP.2019.8682600

[84] Zhang T,Wiliem A,Yang S Q,et al. TV-GAN:generative adversarial network based thermal to visible face recognition[C]// 2018 International Conference on Biometrics (ICB),2018,DOI:10.1109/ICB2018.2018.00035

[85] Chen C,Ross A. Matching thermal to visible face images using a semantic-guided generative adversarial network [J]. arXiv Preprint,2019,arXiv:1903.00963

[86] Gatys L A, Ecker A S, Bethge M. Image style transfer using convolutional neural networks[C]// IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2016:2414-2423

[87] Huang X,Belongie S. Arbitrary style transfer in real-time with adaptive instance normalization[C]// Proceedings of the IEEE International Conference on Computer Vision, 2017:1501-1510

[88] Gatys L A,Ecker A S,Bethge M,et al. Controlling perceptual factors in neural style transfer[C]// IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2017:3985-3993

[89] Johnson J,Alahi A,Li F F. Perceptual losses for real-time style transfer and super-resolution [M]// Computer Vision-CCV 2016. Cham: Springer International Publishing, 2016: 694-711. DOI: 10.1007/978-3-319-46475-6_43

[90] Li Y J,Fang C,Yang J M,et al. Diversified texture synthesis with feed-forward networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2017:3920-3928

[91] Jing Y C, Yang Y Z, Feng Z L, et al. Neural style transfer:a review[J]. IEEE Transactions on Visualization and Computer Graphics, 2019, DOI: 10.1109/TVCG.2019.2921336

[92] Wu X,Xu K,Hall P. A survey of image synthesis and editing with generative adversarial networks[J]. Tsinghua Science and Technology,2017,22(6):660-674

[93] Selim A,Elgharib M,Doyle L. Painting style transfer for head portraits using convolutional neural networks[J]. ACM Transactions on Graphics,2016,35(4):1-18

[94] Fišer J,Jamriška O,Lukáć M,et al. StyLit:illumination-guided example-based stylization of 3D renderings[J]. ACM Transactions on Graphics (TOG),2016,35(4):92

[95] Fišer J,Jamriška O,Simons D,et al. Example-based synthesis of stylized facial animations[J]. ACM Transactions on Graphics (TOG),2017,36(4):155

[96] Huang R,Zhang S,Li T Y,et al. Beyond face rotation: global and local perception GAN for photorealistic and identity preserving frontal view synthesis[C]// IEEE International Conference on Computer Vision (ICCV), 2017:2439-2448

[97] Li C,Wand M. Precomputed real-time texture synthesis with Markovian generative adversarial networks[C]// European Conference on Computer Vision,2016:702-716

[98] Odena A,Olah C,Shlens J. Conditional image synthesis with auxiliary classifier GANs[C]// Proceedings of the 34th International Conference on Machine Learning-Volume 70,2017:2642-2651

[99] Zhang L M,Ji Y,Lin X,et al. Style transfer for anime sketches with enhanced residual U-net and auxiliary classifier GAN[C]// 2017 4th IAPR Asian Conference on Pattern Recognition (ACPR), 2017, DOI: 10.1109/ACPR.2017.61

[100] Parkhi O M, Vedaldi A, Zisserman A. Deep face recognition[C]// Proceedings of the British Machine vision conference,2015

[101] 高新波.视觉信息质量评价方法[M].西安:西安电子科技大学出版社,2011
GAO Xinbo. Visual information quality assessment[M]. Xi'an:Xidian University Press,2011

[102] International Telecommunication Union Methodology for the subjective assessment of the quality of television pictures[R]. Recommendation ITU-R BT.500-13,2012

[103] Video Quality Expert Group (VQEG). Subjective test plan[R]. Version 3. Geneva, Switzerland: Video Quality Expert Group,2003

[104] 高飞.学习盲图像质量评价方法研究[D].西安:西安

电子科技大学,2015
GAO Fei. Study on learning blind image quality assessment[D].Xi'an:Xidian University,2015
[105] Thurstone L L.A law of comparative judgment[J].Psychological Review,1994,101(2):266-270
[106] Tsukida K,Gupta M R.How to analyze paired comparison data[R].UWEE Technical Report,No.UWEETR-2011-0004,2011
[107] Gao F,Tao D C,Gao X B,et al.Learning to rank for blind image quality assessment[J].IEEE Transactions on Neural Networks & Learning Systems,2017,26(10):2275-2290
[108] Ma K D,Liu W T,Liu T L,et al.DipIQ:blind image quality assessment by learning-to-rank discriminable image pairs[J].IEEE Transactions on Image Processing,2017,26(8):3951-3964
[109] Wang N N,Gao X B,Li J,et al.Evaluation on synthesized face sketches[J].Neurocomputing,2016,214:991-1000
[110] Fan D P,Zhang S C,Wu Y H,et al.Scoot:a perceptual metric for facial sketches[J].arXiv Preprint,2019,arXiv:1908.08433
[111] Wang Z,Bovik A C,Sheikh H R,et al.Image quality assessment:from error visibility to structural similarity[J].IEEE Transactions on Image Processing,2004,13(4):600-612
[112] Zhang L,Zhang L,Mou X Q,et al.FSIM:a feature similarity index for image quality assessment[J].IEEE Transactions on Image Processing,2011,20(8):2378-2386
[113] Sheikh H R,Bovik A C.Image information and visual quality[J].IEEE Transactions on Image Processing,2006,15(2):430-444
[114] Tang X O,Wang X G.Face sketch synthesis and recognition[C]//Proceedings Ninth IEEE International Conference on Computer Vision,2003:687-694
[115] Chen L F,Liao H Y M,Ko M T,et al.A new LDA-based face recognition system which can solve the small sample size problem[J].Pattern Recognition,2000,33(10):1713-1726
[116] Liu W Y,Wen Y D,Yu Z D,et al.SphereFace:deep hypersphere embedding for face recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2017:212-220
[117] Szegedy C,Vanhoucke V,Ioffe S,et al.Rethinking the inception architecture for computer vision[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2016:2818-2826
[118] Zhang W,Wang X G,Tang X O.Coupled information-theoretic encoding for face photo-sketch recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2011,DOI:10.1109/CVPR.2011.5995324
[119] Tang X O,Wang X G.Face photo recognition using sketch[C]//Proceedings International Conference on Image Processing,2002,DOI:10.1109/ICIP.2002.1038008
[120] Martínez A,Benavente R.The AR face database[R].Computer Vision Center Technical Report,No.24,1998
[121] Messer K,Matas J,Kittler J,et al.XM2VTSDB:the extended M2VTS database[C]//Second International Conference on Audio and Video-Based Biometric Person Authentication,1999,964:965-966
[122] Phillips P J,Moon H,Rizvi S A,et al.The FERET evaluation methodology for face-recognition algorithms[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,2000,22(10):1090-1104
[123] Mishra A,Rai S N,Mishra A,et al.IIIT-CFW:a benchmark database of cartoon faces in the wild[C]//European Conference on Computer Vision,2016:35-47
[124] Huo J,Li W B,Shi Y H,et al.WebCaricature:a benchmark for caricature recognition[J].arXiv Preprint,2017,arXiv:1703.03230
[125] Kemelmacher-Shlizerman I,Suwajanakorn S,Seitz S M.Illumination-aware age progression[C]//IEEE Conference on Computer Vision and Pattern Recognition,2014:3334-3341
[126] Chen B C,Chen C S,Hsu W H.Cross-age reference coding for age-invariant face recognition and retrieval[C]//European Conference on Computer Vision,2014:768-783
[127] Lanitis A,Taylor C J,Cootes T F.Toward automatic simulation of aging effects on face images[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,2002,24(4):442-455
[128] Nguyen T V,Liu S,Ni B B,et al.Sense beauty via face,dressing,and/or voice[C]//Proceedings of the 20th ACM International Conference on Multimedia,2012:239-248
[129] Li T T,Qian R H,Dong C,et al.BeautyGAN:instance-level facial makeup transfer with deep generative adversarial network[C]//ACM Multimedia Conference on Multimedia Conference,2018:645-653
[130] Liu Z W,Luo P,Wang X G,et al.Deep learning face attributes in the wild[C]//IEEE International Conference on Computer Vision (ICCV),2015:3730-3738
[131] Liang L Y,Lin L J,Jin L W,et al.SCUT-FBP5500:a diverse benchmark dataset for multi-paradigm facial beauty prediction[C]//24th International Conference on Pattern Recognition (ICPR),2018:1598-1603
[132] Jaderberg M,Simonyan K,Zisserman A.Spatial transformer networks[C]//Advances in Neural Information Processing Systems,2015:2017-2025
[133] Li H Y,Xu X M,Cai B L,et al.Style transfer at 100+ FPS via sub-pixel super-resolution[C]//IEEE International Conference on Multimedia & Expo Workshops (ICMEW),2018,DOI:10.1109/ICMEW.2018.8551538
[134] Huang H Z,Wang H,Luo W H,et al.Real-time neural style transfer for videos[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2017:783-791

南京信息工程大学学报(自然科学版),2019,11(6):660-681
Journal of Nanjing University of Information Science and Technology(Natural Science Edition),2019,11(6):660-681

Heterogeneous face synthesis via generative adversarial networks: progresses and challenges

HUANG Fei[1] GAO Fei[1,2] ZHU Jingjie[1] DAI Lingna[1] YU Jun[1]

1 Key Laboratory of Complex Systems Modelling and Simulation, School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018

2 State Key Laboratory of Integrated Services Networks, School of Electronic Engineering, Xidian University, Xi'an 710071

Abstract Heterogeneous face synthesis aims at generating visually realistic and identity-preserving portraits of different modality, such as sketches, caricatures, etc. Heterogeneous face synthesis is of great significance for both public security and digital entertainment, and has attracted numerous attention. Recently, inspired by the dramatic progress in generative adversarial networks (GANs) and their great successes in image-to-image translation tasks, researchers have proposed a number of new heterogeneous face synthesis methods based on GANs. In this paper, we briefly introduce the development of heterogeneous face synthesis, and detailed recent progresses in terms of developments of applications, architectures of GANs, performance evaluation approaches, datasets, and qualitative analysis. Finally, we summarize the challenges and some prospects of heterogeneous face synthesis.

Key words generative adversarial networks; heterogeneous face synthesis; image style transfer; deep learning; digital art