

基于深度生成模型的人脸编辑研究进展

唐雨潇 王斌君

中国人民公安大学信息网络安全学院 北京 100038

(851637579@qq.com)

摘要 人脸编辑广泛应用于公安追逃、人脸美化等领域,传统的统计学方法、基于原型的方法是解决人脸编辑的主要手段,然而这些传统技术面临着操作难度大、计算成本高等问题。近年来,深度学习快速发展,特别是生成网络的出现,为人脸编辑提供了一种全新的思路,采用深度生成模型的人脸编辑技术具有速度快、模型泛化能力强的优势。为总结近年利用深度生成模型解决人脸编辑问题的相关理论与研究,首先介绍了基于深度生成模型的人脸编辑技术采用的网络框架与原理;然后对该项技术所运用的方法进行详述,将其归纳为图像翻译、在网络内部引入条件信息、操纵潜在空间 3 个方面;最后总结了该项技术所面临的身份一致性、属性解耦、属性编辑精确性的挑战,并指出未来该方向亟待解决的若干问题。

关键词:人脸编辑;生成对抗网络;变分自编码器;深度学习;潜在空间

中图分类号 TP309,TP18

Research Progress of Face Editing Based on Deep Generative Model

TANG Yu-xiao and WANG Bin-jun

College of Information Network Security, People's Public Security University of China, Beijing 100038, China

Abstract Face editing is widely used in public security pursuits, face beautification and other fields. Traditional statistical methods and prototype-based methods are the main means to solve face editing. However, these traditional technologies face problems such as difficult operation and high computational cost. In recent years, with the development of deep learning, especially the emergence of generative networks, a brand new idea has been provided for face editing. Face editing technology using deep generative models has the advantages of fast speed and strong model generalization ability. In order to summarize and review the related theories and research on the use of deep generative models to solve the problem of face editing in recent years, firstly, we introduce the network framework and principles adopted by the face editing technology based on deep generative models. Then, the methods used in this technology are described in detail, and we summarize it into three aspects: image translation, introduction of conditional information within the network, and manipulation of potential space. Finally, we summarize the challenges faced by this technology, which consists of identity consistency, attribute decoupling, and attribute editing accuracy, and point out the issues of the technology that need to be resolved urgently in future.

Keywords Face editing, GAN, VAE, Deep learning, Latent space

1 引言

人脸编辑指改变特定面孔的某些特征(如改变性别、年龄、头发的颜色、人脸朝向等)。这项技术可以使人们将留存脑海中的人脸图像真实化,并进行任意改造,因而在人脸美化、公安刑事侦查追捕犯罪嫌疑人,以及寻找失踪人口等领域都具有广泛的应用场景。不容忽视的是,这项技术在为人类生产生活带来便利的同时,也对个人隐私安全与社会安全提

出了新的挑战,如不法分子将该项技术用于煽动舆情、制造虚假新闻、绕开人脸认证机制等。因此,对人脸编辑技术的研究不仅有着广泛的发展前景,也有助于为人脸伪造带来的安全威胁提供解决思路。

在早期,人脸编辑往往采用基于统计学、基于渐变技术、基于原型或基于物理模型的方法。然而这些方法都具有一定的局限性,例如基于原型的方法利用不同属性间的平均差异进行模式转移,难以满足不同人之间的差异,使得生成的面孔

到稿日期:2021-04-12 返修日期:2021-07-06

基金项目:国家社会科学基金重点项目(20AZD114);CCF-绿盟科技“鲲鹏”科研基金(CCF-NSFOCUS 2020011);中国人民公安大学公共安全行为科学实验室开放课题基金(2020sys08)

This work was supported by the Key Program of National Social Science Foundation(20AZD114), CCF-NSFOCUS “Kunpeng” Scientific Research Fund(CCF-NSFOCUS 2020011) and Open Research Fund of the Public Security Behavioral Science Laboratory, People's Public Security University of China (2020sys08).

通信作者:王斌君(wangbinjun@ppsuc.edu.cn)

不具个性化特点;基于物理模型的方法则是对形状和纹理等进行参数化建模,这通常需要大量的训练样本,计算成本较高。在数据科学持续发展的过程中,人脸编辑被视为一个回归问题,例如,Zhu等^[1]提出一种人脸正面化方法,以人脸作为输入,通过最小化像素级图像重建损失来回归所需的正面图像;Zhang等^[2]提出一种去除人脸眼镜的方法,从佩戴眼镜的人脸图像和对应不佩戴眼镜的人脸图像训练样本中训练出一个多变量线性回归模型,这些方法的性能在很大程度上依赖于配对训练的数据样本质量。

随着深度学习的繁荣,GAN(Generative Adversarial Network)^[3],VAE(Variational Auto-encoder)^[3]等生成模型被提出,通过生成模型,低维数据得以转换到高维图像数据域,人脸生成也变得不再困难。目前最先进的人脸生成模型,如BigGAN^[4],StyleGAN^[5]已经可以创造出现实世界中并不存在的、高分辨率的逼真人脸图像,但由于这些生成模型大多以随机噪声作为输入,生成的图像具有随机性,直接导致生成模型的使用受限。人脸编辑破解了生成模型在应用中的困境,生成模型也为人脸编辑提供了新的思路。与传统人脸编辑方案相比,在生成模型的基础之上采取的人脸编辑不仅可以对生成的优质图像加以利用,而且对特征库的学习完全自主,编辑速度更快。

在生成模型的基础之上,人们开始研究在网络中加入额外的条件信息或进一步理解并运用网络内部规律来实现人脸编辑,试图将随机的“人脸生成”推动到可定向调整面孔形象的“人脸编辑”。据此,本文从3个方面总结了现有的基于深度学习的人脸编辑技术。本文第2节对人脸编辑的基础网络框架与原理进行了阐述;第3节归纳了近年来基于深度生成网络的人脸编辑所采用的具体方法;第4节对人脸编辑需要解决的关键问题进行了探讨;第5节展望了基于深度生成网络的人脸编辑所面临的问题与未来的发展方向;最后总结全文。

2 基础网络框架与原理

2.1 生成网络

在人脸编辑领域,最常用的网络框架仍然采用GAN和VAE的思想,如图1所示。

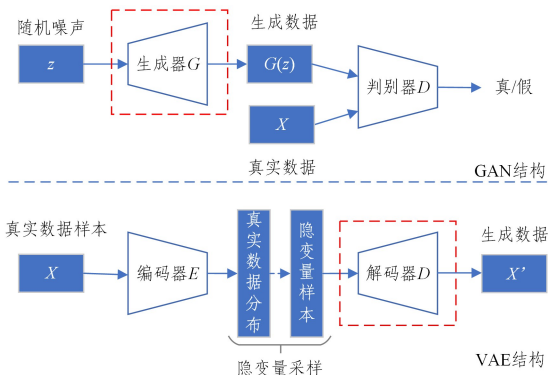


图1 生成对抗网络与变分自编码器

Fig. 1 Generative adversarial network and variational auto-encoder

对。这种模型的典型架构包括两个子网络:生成器 G (generator) 和判别器 D (discriminator),生成器将输入随机变量转换为高维数据,以“欺骗”判别器;判别器评估所生成的高维数据是否源自原始分布,以识别图像真伪。对于生成器 G 来说,如果它生成的图像经过判别器 D 的判定后结果为真,就说明生成模型已经较好地掌握了数据的分布模式,可以产生符合要求的样本;反之则没有达到要求,还需要继续训练。

变分自编码器 (Variational Auto-Encoder, VAE) 是基于变分下界约束得到的编码器-解码器模型对。这种模型的典型架构包括两个子网络:编码器 (encoder) 和解码器 (decoder),编码器利用神经网络表示概率图模型,将输入数据编码成若干正态分布向量以描述高维数据的低维潜在语义空间,随后,在编码分布中随机采样隐变量,使隐变量被约束在原始数据的概率分布中;解码器则将随机隐变量还原成高维数据,实现图像的生成过程。

GAN 的生成器与 VAE 的解码器都能实现低维数据到高维数据的转换,因此,当网络的基本框架采用这种结构时,编辑操作可以在潜在空间中进行;当编辑操作并不着眼于潜在空间时,考虑到网络的输入和输出都是高维图像数据,合成网络也可能采用沙漏型的图像-图像的结构,如简单的卷积-反卷积结构、U-net 结构、上卷积残差结构等,此时,人们只对该整体结构进行训练,并不关心潜在空间。

2.2 人脸编辑的原理

实现人脸编辑时有两种不同的思想,一种是采用额外的条件信息约束生成器生成图像;另一种是学习输入的随机向量与生成图像之间的关系,运用学习到的规律实现人脸编辑。

2.2.1 采用条件进行引导

采用条件引导的思想是实现人脸编辑最常用的方法之一,将“人脸描述”作为条件输入生成模型中,可以实现基于描述生成定制的人脸。条件生成式对抗网络 (CGAN)^[6] 作为原始 GAN 的一种扩展,是引入额外条件约束生成数据的典型模型,其在训练时就对生成器和判别器输入额外的条件,使判别器根据条件信息判断生成器生成的图像是否满足条件,从而使生成器受条件的约束生成相应数据。在人脸编辑中,条件信息往往可以采用不同的表现形式,如特定图像、图像轮廓、图像域或者图像的隐变量等;同时,条件数据的输入区域也有所不同,如可以在输入图像时一同输入条件,或将图像压缩为低维的隐变量后再输入条件;在引入过程中,不仅可以采用直接输入的方式将条件引入网络,还可以采用仿射变换、跳跃连接等方式。

2.2.2 运用隐变量

采用条件引导的方法并不着眼于理解生成数据与输入隐变量之间的关系,故其无法直接通过生成器的输入来控制生成器的输出,例如修改输入隐变量的某个维度以操作人脸的某个属性。但是,特征学习可以解决这一问题。

特征学习指从数据中学习相关的特征并加以使用。在人脸编辑中,人们往往用隐变量作为特征学习的结果,其也可称为特征向量、隐码、潜在编码 (latent code),隐变量的样本空间称为潜在空间、隐空间 (latent space)。隐变量蕴含着控制

人脸生成中不可观测的特征或人脸语义信息,这些特征同时赋予隐变量更加深刻的含义,决定了人脸图像数据的实际分布。理解并运用隐变量中的语义信息,可以从源头上实现人脸属性编辑,例如,对隐变量进行插值运算、沿隐变量中具有语义信息的方向移动等。

3 实现人脸编辑的方案

基于深度学习的人脸编辑方案可归纳为3种类别:1)图像翻译,即通过训练端到端的网络模型直接对图像进行处理,实现图像属性转变;2)在网络内部引入条件信息,着眼于解决在网络中间层与潜在空间的条件输入问题;3)操纵潜在空间,找到图像对应的隐变量,通过探索隐变量语义信息的规律,来操纵潜在空间的方式进行人脸编辑。

3.1 图像翻译

图像翻译旨在通过设计端到端的模型将源域图像转换到目标域图像,通常源域提供图像的内容,目标域提供图像的“风格”(图像风格或属性),在源域内容下实现目标域的“风格”化。在实际应用中,图像翻译可以有效解决图像着色、语义分割、草图到图像合成等各种图像-图像转换问题。采用图像翻译实现人脸编辑通常有两种思路:第一种是属性迁移,即将人脸属性操作视为一种转换过程,由原始图像作为输入,然后输出转换属性后的图像,例如年轻人脸转换成老年人脸;第二种是草图-图像转换,即将草图、热图等轮廓信息作为输入,采用画笔方式控制人脸图像的轮廓、颜色等细节属性,例如模拟画像转换成真人脸。

最典型的图像翻译网络分为需要成对数据集的有监督模型以及不需要成对数据集的无监督学习模型。监督学习模型以 pix2pix^[7] 为代表,其借鉴了 CGAN 的思想,并将源域图像作为输入的约束;无监督学习模型以 cycleGAN^[8] 和 UNIT^[9] 为代表,二者分别采用双映射网络结构以及共享潜在空间的思想。之后的图像翻译网络大多在这两种模型的基础之上进行改进。接下来将从这3个代表模型分别采用的图像翻译方法出发,进一步对图像翻译中人脸编辑的发展进行阐述。

3.1.1 有监督的图像翻译

图像翻译的开山之作——pix2pix^[7] 采用 U-net^[10] 网络结构代替 CGAN 中的生成器(见图2),条件信息能够以草图的形式输入,约束生成具有相应轮廓的图像。pix2pixHD^[11] 在此基础上进行改进,采用多尺度判别器与多层次生成的 coarse2fine 生成器,有效提高了生成图像的分辨率与质量。此后,在人脸编辑领域,控制人脸轮廓的图像翻译网络层出不穷:Faceshop^[12], SC-FEGAN^[13] 等网络借用草图-图像转换模型的原理,将笔画和颜色相结合,以在人脸图像上进行笔画操作的方式巧妙地实现交互式人脸编辑,使用户编辑变得更为便捷。MaskGAN^[14] 则使用语义图代替草图控制人脸轮廓,相比之下,语义图作为高保真度的人脸中间表示具有更灵活的面部操作能力。Deepfacedrawing^[15] 提出手绘草图容易失真问题,并运用了局部线性嵌入方法,能够实现草图到人脸图像的高保真度生成,使手绘草图到人脸图像的生成更加实用化。

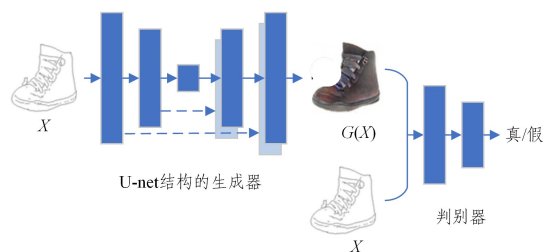


图2 pix2pix 网络结构

Fig. 2 pix2pix structure

然而,由于有监督的图像翻译在网络训练时往往需要成对数据,例如在训练人脸老化网络过程中要求拥有大量同一人物在不同年龄时的图像,故将人脸属性作为风格域的人脸编辑方法难以推进。为克服这一困难,一些网络采用额外模块来确保人脸身份一致,从而将判别器从判断图像配对与否的任务中解放出来。以 IPCGAN^[16] 和 DIAT^[17] 为例,二者通过额外的属性分类损失约束生成器学习属性转换任务,同时利用预训练的特征提取网络对人脸身份强制统一,其中前者用于人脸老化,采用的判别器具有判断生成图像是否为真实人脸及是否处于指定年龄组的能力;后者采用属性转换网络与掩膜网络进行协作,使人脸编辑不会影响到无关区域,进一步保证身份一致。Viazovetsky 等提出一种合成数据的方法^[18],在得到配对数据集后再训练改变人脸属性的图像翻译网络,该网络最终以图像-图像的方式实现了人脸语义控制,如性别互换、衰老/年轻化等,但其存在明显的缺陷,即每一个属性的转换都需要分别训练一个相应的网络来获取数据。

为寻求人脸轮廓与属性域的共同控制,SPADE^[19] 提供了一种新思路,如图3所示。

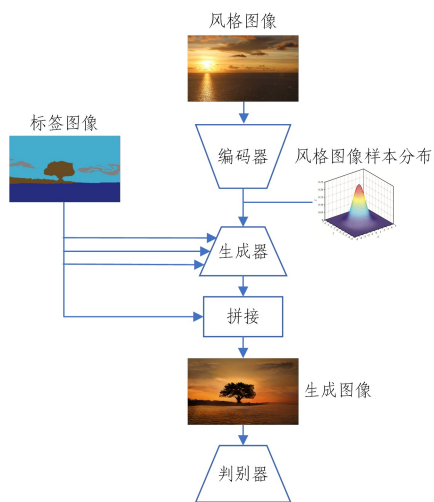


图3 SPADE 的网络结构

Fig. 3 Structure of SPADE

一方面,其以风格图像控制生成图像的风格,通过变分的思想,将生成器的输入约束在风格图像的概率分布中,使输入的隐变量得以表达;另一方面,其以标签图像控制生成图像的轮廓,将标签图像以空间自适应层的方式输入网络,使语义信息贯穿于整个生成模型,用户则可以在标签图像上进行笔画操作,对生成图像轮廓进行调整。在人脸编辑领域,Michi-

GAN^[20] 借鉴该思路实现了多输入条件的发型属性编辑,其采用头发的形状及结构、发型外观、人像背景 3 个模块控制肖像发型。其中,形状和结构信息以条件归一化的形式融入网络,调节各层的激活;外观以初始编码的形式输入网络,取代了传统 GAN 中随机采样的隐变量;背景特征在网络中逐步注入。Hu 等^[21] 在 ATTGAN^[22] 的基础上对网络结构进行改良,在输入草图的同时输入属性相关的条件信息,最终由草图控制人脸轮廓,由输入条件控制属性信息。

3.1.2 基于双映射思想的图像翻译

典型的采用双映射实现的无监督图像翻译网络有 CycleGAN^[8], DiscoGAN^[23] 和 DualGAN^[24], 其网络结构如图 4 所示。其原理都是通过 x 逼近 $F(G(X))$ 的方式约束两个生成器实现域之间的转换。

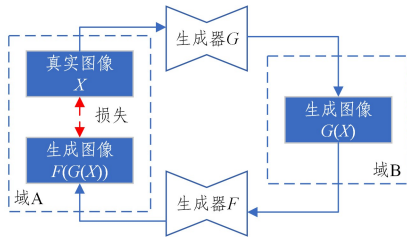


图 4 基于双映射思想的图像翻译原理

Fig. 4 Principles of image translation based on the idea of double mapping

CycleGAN^[8] 具有两个生成器 F 和 G , 在域 A 中的图像通过 G 可以转移到域 B , 而在域 B 中的图像通过 F 可以转移到域 A , 再采用对抗损失与循环一致损失训练网络, 实现不同域之间的图像转换。与之相比, DiscoGAN^[23] 通过重构损失重点解决域之间的关联性与模式坍塌问题, 避免一个域的多种模式映射到另一个域的单一模式, 从而保证图像翻译网络实现一对一的映射。DualGAN^[24] 在生成器中加上域信息 z 作为输入, 用独立的域信息表达映射的方向。在此基础之上, 改进的网络模型有两种不同的发展趋势, 一种是提高模型的“扩展性”, 另一种是提高网络的“针对性”。

由于基础的网络模型只能解决单一任务问题, 针对每一种不同的属性转换都需要重新训练一个网络, 当属性增多时, 所需的模型数量也呈指数型增长, 同时, 在训练一个特定的两两映射时, 难以利用来自其他领域的的数据。因此, 相关学者针对模型的“通用性”对网络进行改进, 从而实现多域图像翻译。StarGAN^[25] 采用了一种跨多个域、多个数据集的训练方法, 以星型的网络结构, 将数据集的选择、人脸的各项面部特征(头发颜色、性别、表情等)以标签的形式与原始图像一同输入, 能够生成具有相应标签属性的人脸图像, 判别器用于检测输出图像的真假及标签对应情况, 因此仅通过一个网络即可实现对多种人脸属性的操控。该网络通过原始图像标签与生成图像对原始图像进行重构, 重构损失不再局限于两个生成器的循环转换上。在此基础上, StarGANv2^[26] 采用映射网络自动生成风格编码以替换原模型的标签, 为目标域提供多类型的风格表示。另一些研究则通过模块组合的方式实现多域图像转换, 由于模块之间相互独立, 故可轻易实现模块的增减

和升级, 提高了模型的灵活性和可扩展性。例如, ModularGAN^[27] 采用了一个由编码器模块、变压器模块、重构模块和判别器模块组成的网络体系结构, 4 个模块可以同时训练, 并在测试时根据特定的图像翻译任务进行重用与组合, 构建相应的 GAN 网络; ComboGAN^[28] 则组合了不同的编码器-解码器对, 可以将额外的域添加到现有的组合模型中。

与此同时, 研究者从不同的应用领域出发, 根据属性的相关特性, 研究出专门针对某特定属性的控制方法, 对特定属性编辑子任务, 训练特定的模型, 使针对该属性的语义控制更为精准。在年龄属性领域, F-GAN^[29] 融合了 GroupGAN 与 FAGAN。其中, GroupGAN 将人脸数据集划分为几个年龄组, 然后在每一对年龄组中训练一个周期模型, 在目标年龄差异大时表现更好; FAGAN 引入了基于学习的年龄估计技术的标准损失, 使网络对年龄的波动更为敏感, 在目标年龄差异小时表现更好。在表情控制领域, G2GAN^[30] 将表情关键点热图作为参考, 将正脸图与相应关键点热图的面部表情融合, 采用双映射的方式实现表情迁移; GANimation^[31] 模型则打破了传统采用离散的表情种类控制人脸语义的方法, 描述了连续流形中的标签, 可以调节人脸表情的变化程度, 但它严重依赖于动作单元的注解; EF-GAN^[31] 在 GANimation^[32] 的基础上, 对人脸的眼睛、鼻子、嘴巴的精细化表情转化与人脸整体表情转化共同进行调整, 这种级联面部表情转换的模型可以产生高真实度的人脸图像。在妆容迁移领域, BeautyGAN^[33] 与 PairedCycleGAN^[34] 采用两张图像作为输入, 其中一张图像作为身份图像, 另一张图像作为妆容参考图像, 以合成特定妆容的人脸。PairedCycleGAN^[34] 训练时将眼、鼻、嘴、眉毛等切成图像块, 最后再用泊松融合, 提高了图像的真实度。

传统双域图像翻译网络模型多采用二元属性对人脸语义进行表达, 无法细粒度控制每个属性的强度, 标签的固定和离散性也使得网络不能产生多样化的输出, 而采用输入人脸图像与期望操纵图像的差值作为相应的残差属性是一种解决方案。例如, RelGAN^[35] 为了实现编辑前后的平滑和真实的插值, 采用相对属性作为条件信息; RAG^[36] 采用残差属性作为条件信息, 又结合了视觉注意力网络, 帮助定位相关区域并改善生成的结果; 除此之外, ResidualGAN 等^[37] 也采用了双映射的方法, 两个转换器分别参照原图像执行反向属性操作, 生成残差图像, 最终的输出图像是残差图像和原图像的像素相加。与对整幅图像进行操作相比, 学习残差图像可以将残差图像的大部分区域抑制为零, 避免了改变与属性无关的区域。

3.1.3 基于潜在空间共享的图像翻译

如图 5 所示, 以 UNIT^[9] 网络为代表的图像翻译方式与前两种不同, 其认为两个域的图像共享潜在空间, 而通过不同的生成器, 又可以将潜在空间中的隐变量转换到两个图像域中。基于这种思想, 网络结构多采用单独的编码和解码器进行组合。之后的研究将图像分解为风格潜在空间(style latent space)和内容潜在空间(content latent space), 并认为不同的图像仅共享内容潜在空间。以 MUNIT^[38], DRIT-GAN^[39] 为例, 二者通过替换不相交的风格潜在空间实现风格迁移, 再由相应域的生成器重构图像, 这样的网络可以在保留

内容信息的同时产生多样化的输出,从而实现多模态翻译。然而这些方法很难扩展到大量的域,因为它们需要对每个域训练一个相应的生成器。

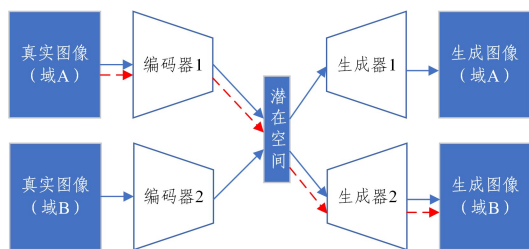


图5 基于潜在空间共享的图像翻译原理

Fig. 5 Principles of image translation based on latent space sharing

3.2 在网络内部引入条件信息

在网络内部或图像的潜在空间中引入额外条件的前提是生成器可以识别条件信息,条件信息也需要包含数据的属性及特征信息。为实现这一目的,有两种方案:第一种是将图像的内容信息与属性信息分离开来,从而通过修改属性信息达到控制人脸属性的目的;第二种是调整生成网络结构,在输出图像后端连接一个属性分类网络,对图像属性进行判别,从而约束生成器学习条件信息。

3.2.1 分离属性信息

将属性内容与其他内容进行分离,再修改隐变量的属性内容可以成功修改图像属性,而此时其他内容得以保持。Fadernetwork^[40]引入一个对抗过程将图像的属性信息从隐变量中分离出来,隐变量将不再包含属性信息,提取出的隐变量与控制属性的变量一同输入编码器,可以控制重构图像的属性。然而,人脸图像的特征与隐变量之间的关系是高度复杂和密切的,简单地对隐变量施加属性独立的约束可能会导致信息丢失,不利于属性的编辑,特别是对于性别、种族等可识别的属性。GeneGAN^[41]通过将输入图像编码成与属性相关的部分和与属性无关部分来学习属性子空间,再从两组具有相反属性的未配对图像中学习对象变形,在学习的子空间中切换两个编码部分实现属性转换。GeneGAN^[41]的不足之处在于其只能通过范例转移单个属性,无法实现模型的多属性传递。为解决这一问题,DNA-GAN^[42]和ELEGANT^[43]采用迭代训练策略,即预定义 n 个属性,并要求由隐变量的不同维度分别对每一个属性进行控制,每一次迭代只对一个属性进行训练,依次进行,所有的属性将会在潜在空间中分离开。然而,迭代训练策略不能有效地学习解纠缠的属性信息,导致同时传递多个样本属性的能力有限。

为将图像的属性特征与身份特征分离开来,有的模型采用两个编码器将图像的属性与内容分别编码。IcGAN^[44]分别训练一个CGAN和两个编码器,图像进行编码之后获取到控制生成图像整体内容的隐变量与控制图片属性的特征向量,修改特征向量后,与隐变量一同输入CGAN网络中,以此控制生成的图片具有相关联的属性;Nitzan等^[45]将预训练的人脸识别网络作为身份编码器,在预训练的生成器下,该模型利用输入图像与生成图像分别通过编码器得到的编码之间的

损失来训练编码器;AGUIT^[46]首先采用两个编码器将人脸编码为风格编码和内容编码,再通过预标注的属性标签与风格编码之间的风格分类损失促使风格编码器学习标签信息,并与身份编码解缠。

3.2.2 接入分类器

对输出图像的属性进行分类,则可以通过分类情况得到相应损失,以此指导网络学习条件信息。ATTGAN^[22]把属性标签与隐变量输入生成器,要求生成器据此生成同一身份但具有与标签对应的不同属性的图像,之后由判别器监督图像的身份一致性,由属性分类器约束生成器生成的图像属性与输入一致。Liu等^[47]将属性向量同时输入潜在空间与判别器,促使判别器监督属性一致性。此外,其采用基于小波包的变换模块,成功处理局部图像纹理,适用于在年龄变换中控制皱纹、笑纹、眼袋等衰老迹象程度等任务。Lifespan^[48]将隐变量与年龄簇编码一同输入生成器,最终能够在期望的年龄簇中生成同一个人的输出图像。该网络模型在采用传统的对抗损失基础之上,还采用自我重建损失、循环重建损失、身份特征损失、年龄向量损失来加强网络对年龄语义的理解。

为使属性语义信息更好地融入模型,也有研究者对引入条件信息的方式进行了改进。由于采用编码-解码结构的网络时,图像将被压缩到“瓶颈模块”(bottleneck)中,这往往会导致图像的某些信息丢失,故经复原的生成图像质量较低。针对该问题,STGAN^[49]使用选择性传输单元处理跨层连接,将属性嵌入模型中,然而这种方式以降低属性控制能力作为代价。还有研究将条件信息与隐变量进行拼接后再输入网络。例如,CAAEE^[50]将原始图像映射到低维空间后,用原始图像的隐变量与以独热编码表示的年龄标签进行拼接作为生成器的输入,该网络模型使用了隐码判别器与图像判别器,由图像判别器来判别其是否属于该年龄段内的真实人脸;EBGAN^[51]提出基于人脸区域修复的编辑方案,将属性编码与原图像的隐变量拼接后再输入生成器合成新的人脸图像,从参考图像中学习相应的信息来完成缺失区域,从而实现基于实例的人脸编辑;PsGAN^[52]是对图像的特征图进行仿射变换,其提出一种姿态和表情鲁棒的妆容迁移网络,采用化妆提取网络将参考图像的妆容分解成两个参数矩阵,再通过化妆应用网络把调整后的矩阵应用到人脸上。StyleRig^[53]以控制参数的形式表达额外条件信息,其设计的DFR网络通过基于三维变形人脸模型(3DMM)^[54]构建的人脸语义模型获取相应属性的控制参数,再使用RigNet将控制参数与原始图像的编码进行融合,从而得到具有该人脸属性的人脸图像隐变量,该模型采用预训练的StyleGAN^[5]将隐变量重构为人脸图像并进行自监督学习;CAFÉ-GAN^[55]则将空间注意力应用于判别器的中层特征,对属性进行分类,并通过利用判别器定位区域的能力,来控制面部编辑只发生在与指定目标属性相关的面部图像部分。

3.2.3 属性分离与接入分类器融合

对于融合属性分离和接入分类器的方法(见图6),CLs-GAN^[56]先通过两个编码器将图像分别编码成属性编码与内容编码,再在STGAN^[49]的基础上,针对编译码结构中编辑

图像容易因跳跃连接而受到原始属性的影响,采用一种上卷积残差网络(Tr-resnet),有选择地从源图像和目标标签中提取信息;同时其认为传统分类器只将源图像作为输入,忽略了生成图像与源图像的属性差异,故引入属性对抗分类器将源图像与生成图像一同输入,以指导生成器。AME-GAN^[57]对属性编码和内容编码进行改进,属性隐变量和背景隐变量分别强制服从高斯分布和均匀分布,使得这些隐变量可以轻松修改和调整图像中的每个细节;为使属性传递的图像更加准确和真实,其开发条件多尺度判别器来区分生成的图像真伪与属性类别。

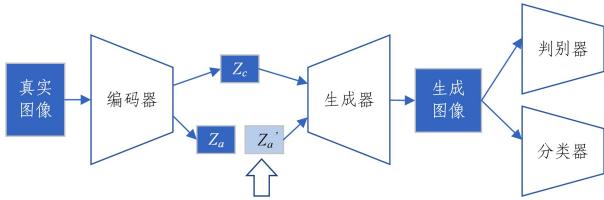


图6 使用属性分离与接入分类器的方法实现人脸编辑

Fig. 6 Use the method of attribute separation and access classifier to realize face editing

3.3 操纵潜在空间

潜在空间中包含着人脸语义信息,因此操纵潜在空间后重构人脸图像,可以在不借助额外条件的情况下实现人脸编辑。具体分为两个步骤:第一步是将图像嵌入低维空间,即寻找生成图像所对应的隐变量,让图像中的语义信息尽可能被隐变量保留;第二步是挖掘隐变量的语义信息,根据隐变量的语义信息规律控制其对应的人脸图像属性。

3.3.1 人脸图像嵌入

GAN的实际应用程序通常依赖于对隐变量的操作,然而,GAN的一个关键限制是缺少对实际样本进行推断的编码机制,也就是说,我们不能通过GAN架构直接推导出给定样本 x 对应的隐变量 z 。为将图像嵌入潜在空间,从训练方式出发,可以将特征学习网络与重构网络共同训练,也可以利用已经成熟的生成模型。从推断机制的方法出发,一种方法是GAN反演,另一种方法是对抗推理。

(1)GAN反演,即根据生成网络生成的图像,逆推出目标图像对应的隐变量。该方法包括直接为单个图像优化隐变量或者训练逆推编码器。

如图7所示,直接为单个图像优化隐变量指在现有预训练生成器的条件下,将目标图像与生成图像的损失最小化,以优化的方式推导出目标图像对应的隐变量。Creswell等^[58]选择一个随机的初始隐变量,计算生成样本与相关真实样本之间的均方误差(MSE),并使用梯度下降法对其进行优化,使像素级重建损失最小化;mGANprior^[59]采用多个隐变量来表达图像语义,以表达出更好的语义,但由于该方法采用大量编码,要求较多的参数,因IC增加了逆向映射的复杂度。在众多生成模型中,StyleGAN^[5]以其高质量的输出与解耦功能被研究者关注,该模型将隐变量 z 通过映射网络转换为隐变量 w ,并分成18个层次输入合成网络,从而实现解耦。这18个层次的潜在空间称为 $W+$ (见图8,其中 w 代表隐变量, W

与 $W+$ 代表潜在空间)。由于该网络可以对生成图像进行更细粒度的语义控制,一些研究便设计了专门针对StyleGAN的嵌入方法:Image2StyleGAN^[60]采用梯度下降算法,将生成的图像嵌入潜在空间 $W+$ 中,在浅层覆盖隐变量则只会改变性别、姿态,在中间层覆盖则会改变头发颜色、眼睛闭合,在深层覆盖则可以控制颜色;Image2StyleGAN++^[61]证明了先优化 $W+$ 后优化 n 的优化顺序可以提高嵌入质量,因为 $W+$ 考虑了轮廓等结构信息, n 存储着细节特征,即更精细化的人脸属性。直接优化的方式具有以下几个问题:1)嵌入过程慢;2)对隐变量的初始化具有较强的敏感性;3)随着迭代进行,隐变量会不断丢失信息。

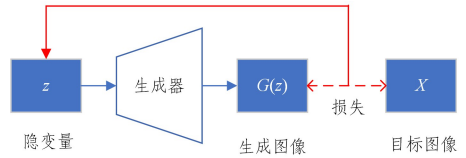


图7 采用直接优化的方式实现GAN反演

Fig. 7 Realize GAN inversion by directly optimization

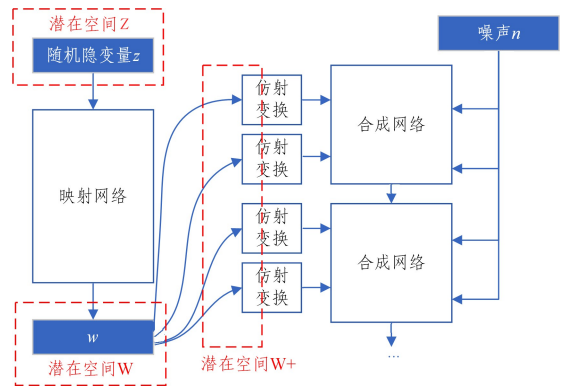


图8 styleGAN结构图

Fig. 8 Structure of styleGAN

为解决上述问题,有研究引入编码器的思想将给定图像映射到潜在空间,即训练逆推编码器(见图9)。

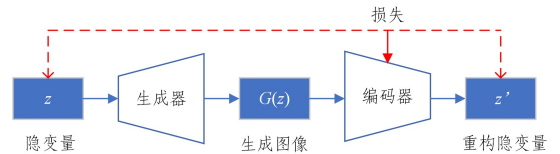


图9 引入编码器的思想实现GAN反演

Fig. 9 Introduce the idea of encoder to realize GAN inversion

VAE/GAN^[62]共享VAE的解码器与GAN的生成器,然而这种结构的重构精度通常比单独使用VAE差,重构后的样品质量往往不如从预先训练好的GAN中采样生成的样品。AEGAN^[63]在预先训练好的GAN模型下,通过VAE算法的原理来微调编码器,该模型不直接使原始隐变量和重构隐变量之间的差别最小,而是尽量使生成的隐变量样本之间的差别最小。Guan等^[64]采用的框架由一个嵌入网络和一个基于优化的迭代器组成,二者形成一个训练循环,在传统的图像/特征级约束之外,引入隐码损失监督网络,最终提高了网络的嵌入质量并加快了迭代器内部的优化步骤。Richardson

等^[65]提出了一个专门针对 StyleGAN 生成网络的编码器,用特征金字塔扩展编码器主干,生成 3 个层次的特征映射,并使用一个简单的中间网络 map2style 提取风格,将图像嵌入 $W+$ 潜在空间,再采用系列损失函数,使编码出的隐变量能通过 StyleGAN 恢复成高质量的图像。

为寻找更优的反演方式,有些模型在训练编码器获取编码后,接着引入直接优化的方式对隐变量进行微调,既解决了直接优化方式中初始隐码的敏感性问题,又提高了编码器面对特定图像的针对性。Zhu 等^[66]提出一种域内 GAN 反演方法,其首先学习一种新的编码器来将给定的图像投影到生成模型的固有潜在空间中,然后提出域正则化的优化方法,通过引入编码器作为正则化器来微调编码器产生的代码,以此更好地恢复目标。

(2) 对抗推理,即利用对抗过程共同学习一个生成网络和一个推理网络。

BiGAN^[67]与 ALI^[68]模型是典型的对抗推理模型(见图 10),生成网络(生成器)将随机的隐变量样本映射到数据空间,推理网络(编码器)则将数据空间中的训练样本映射到隐变量空间,在这两个网络之间进行对抗博弈,并训练一个判别网络来区分生成网络和推理网络之间的潜在/数据空间联合样本(即区分 $(z, G(z))$ 与 $(E(x), X)$)。在此基础上,Big-BiGAN^[69]采用 BigGAN 网络结构作为生成器,使模型在生成高质量图像的同时扩充编码器能够建模的语义。该模型采用了一个稳定版本的联合判别器,进一步提高了模型表示学习的能力。然而,如图 11 所示,对抗性推理通常强调对象的高级表示,而不是以忠实重构样本为目的,在人脸编辑中运用对抗推理获取的隐变量将可能导致图像属性缺失精确性。

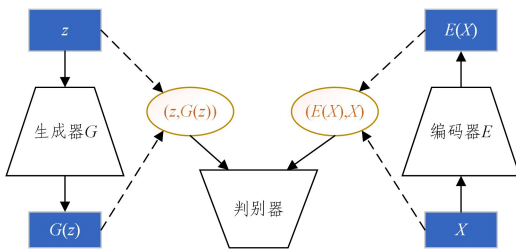


图 10 对抗推理模型结构

Fig. 10 Structure of adversarial inference mode



图 11 BigBiGAN 中的原始图像以及获取隐码后的重构结果

Fig. 11 Original image in BigBiGAN and reconstruction result after obtaining latent code

3.3.2 运用隐变量语义信息

将图像数据嵌入潜在空间后得到其对应的低维隐变量,通过学习隐变量中包含的语义信息,能够重用 GANs 学习到的知识,合理地控制图像生成过程,从而实现广泛的人脸编辑应用。

GAN 的潜在空间通常存在具有某种特定意义的线性方向。找到一般的线性方向后,沿此方向能够增强或减弱线性性质,若线性性质为某人脸属性语义,则将原图像的隐变量沿该方向移动,可以得到该属性的定向改进。最早挖掘潜在空间具有语义意义方向的方法是采用隐变量之差;Radford 等^[70]认为,如果能确定变换后图像的隐变量,就可以计算它与原图像的隐码之差,从而找到与此特定变换对应的潜在空间中的方向;Upchurch 等^[71]从特定属性的图像集所对应的平均特征值中提取平均特征值的差,将其作为属性语义的线性方向。一些模型在生成图像之后接入额外的检测评估函数,根据评估结果对潜在空间信息进行探索:如图 12 所示,InterfaceGAN^[72]在得到分类属性与隐变量的对应关系后,通过 SVM 算法求出分离某个二元语义属性的超平面,平面的法向量将指导特征的变化方向,以沿着该界限的法向量移动隐变量的方式控制图像生成;Jahani 等^[73]借用对象检测器,在潜在空间中寻找与简单图像增强(如缩放或平移)相对应的向量,但是它们只对应于通过自动数据扩充可以实现的简单转换;TL-GAN^[74]在生成图像后接入一个分类网络,以此得到生成图像的隐变量与分类属性之间的对应关系,根据这种对应关系,其采用广义线性模型中的回归算法,得到属性的特征轴,并根据该特征轴来调整隐变量;Ganalyze^[75]作为一个研究图像认知特性的模型框架,利用 MemNet^[76]作为“评估者”网络,在 BigGAN^[4]的潜在空间中搜索可解释方向,以优化潜在空间的方式,使预先训练的模型得分最大化,优化的结果是一个与记忆性增加相对应的方向。除寻找方向的方式之外,近期也有一些工作值得注意:Styleflow^[77]通过评估网络将样式隐变量与对应图像的属性连接起来,再以条件归一化流的正、逆推过程来实现自适应潜在空间的向量操作;Van 等^[78]认为输入数据与其相应的潜在变量都蕴含着特征的域的表达,因此提出了“标签空间”,通过探索标签空间来学习潜在空间与其自身属性(即标签)之间的线性映射。然而,上述模型依赖于对生成图像的进一步学习,这种有监督方式需要人类标签或预先训练的模型,而获得这些模型的成本可能很高。

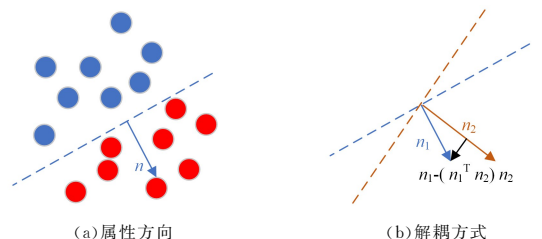


图 12 InterfaceGAN 寻找潜在空间中的属性方向以及两种属性方向的解耦方式

Fig. 12 Finding attribute direction in latent space of InterfaceGAN and way of decoupling two attribute directions

在无监督学习算法的探索中, Voynov 等^[79]学习候选矩阵和分类器,使分类器能够正确识别矩阵中的语义方向; GANspace^[80]基于潜在空间和特征空间的主成分分析,识别出重要的潜在方向,证明了大量的可解释控件可以通过沿主方向的分层扰动来定义,并在 StyleGAN 以及 BigGAN 中进行验证,证明二者都可以通过分层输入来控制属性语义属性; SeFa^[81]方法深入研究了 GAN 的生成机制,并进一步提出了一种通过直接分解预先训练的权值来发现潜在语义的封闭分解算法,进而实现无监督的语义分解; Plumerault 等^[82]用连续变化因素(factors of variation)控制生成器模型,其假设控制图像生成的潜在变量可以分解为离散的变化因素和连续的变化因素,离散变化因素控制图像类别等,而连续变化因素可以控制图像中某物体的大小、位置等。然而,相比之下,无监督与自监督算法更适用于简单的特征,在完成更高级的人脸属性特征调控中具有一定困难。

4 人脸编辑需要解决的关键问题

4.1 身份一致性

人脸身份是一个复杂的高维因素。在人脸编辑过程中,一些网络过度关注特定属性的转换,却往往忽略了保留原始人脸的身份特征,导致重构图像与原始图像身份不一致,甚至在转换中面部变得扭曲、失真。为保证身份一致性,往往采用以下思路:

(1)构建重构损失或采用循环一致损失,避免网络在运算过程中丢失人脸信息。

(2)在输出图像后采用身份一致网络对身份进行判别。例如利用预训练的人脸识别网络作为人脸身份的判定器,对模型的身份一致性进行监督。

(3)避免不必要的区域被编辑。采用掩膜可以区分编辑区域与不变区域;注意力机制可以促进模型对细节特征(如牙齿、人脸边缘)的学习,进而进一步提高结果的真实性;使用目标和原图像之间的差异图作为输入的属性条件,可以保证编辑操作只运用于修改部分。

4.2 属性解耦

解耦能力指在不影响其他因素的情况下控制单个因素或特征的能力。一个适当的解耦表示有利于语义数据混合、下游任务的迁移学习,甚至可解释性。在人脸编辑领域,由于数据集中人脸的属性之间不可避免地存在相关性,导致学习到的变量在潜在空间中高度纠缠,个体维度并不仅仅对应特定的语义特征。例如人脸数据集中,年龄越大的人佩戴眼镜的概率更大,学习该数据集后获得的模型在实现人脸老化时会同时产生为生成人脸佩戴眼镜的效果。对学习到的表示进行解缠处理,使输入的表达可以特定的语义信息进行翻译,能够有效提高翻译的可解释性。为实现属性解耦,往往采用以下思路:

(1)设计损失函数鼓励潜在表示的分解。针对 VAE 可采用增加目标函数、添加隐式^[83]或显式^[84]惩罚的方法实现潜在表示分离;针对 GAN 可采用规则化生成器的方法,如 InfoGAN^[85]以增加输入隐码和输出隐码之间的互信息为目标。在多属性之间的解耦任务中,也可以采用分类器进行迭代训练,使各种单一的属性语义依次解缠。

(2)在潜在空间中,假设存在某个属性对应的方向,采用沿该方向投影的方式可以实现特征解耦。以 InterfaceGAN(见图 12)和 TL-GAN 为例,虽然二者可以通过移动隐变量方向来调整生成人脸的属性,但同时也会导致其他特征发生改变,特别是当移动距离更大时,改变后的隐变量生成的图像基本上完全丢失原图像的属性特征,因此二者采用相同的解耦方式:将某个属性的特征轴映射到一个垂直于另一个属性的特征轴的新方向,以替代原特征轴,消除二者之间的相关性。

(3)GAN 与 VAE 相结合,采用蒸馏学习的方案,将生成模型的优势相结合。

4.3 属性编辑精确性

空间在属性方面需要有足够的表现力,以便提供用户想象的属性更改。每个属性(如“眼镜”)具有许多变体(如太阳镜、圆眼镜和薄眼镜),因此属性编辑精确与否直接关系到操作的灵活度与难易程度。为实现属性编辑精确性,往往采用以下思路:

(1)采用草图、语义图控制轮廓,以绘画方式修改人脸图像的细节方面。

(2)将图像示例作为条件信息,将特定人脸的目标属性应用到原人脸图像上,可以实现定向、精确的属性编辑。

5 总结和展望

图 13 给出了本文总结的基于深度生成网络的人脸编辑技术概览。

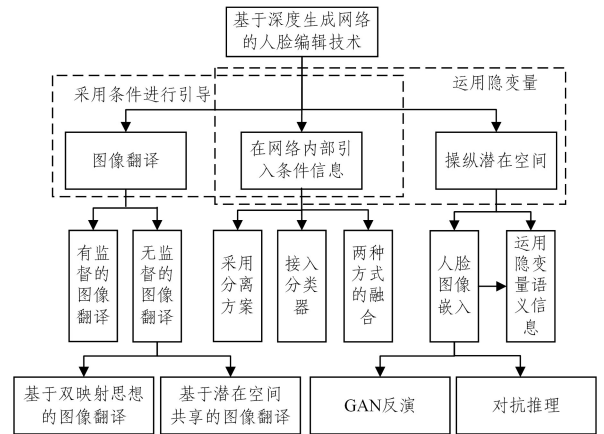


Fig. 13 Overview of face editing technology based on deep generative network

人脸编辑可广泛应用于表情、年龄、眼镜佩戴、秃顶、刘海、肤色、面孔亮度、人脸妆容、人脸朝向等属性的改动。与传统方式相比,基于深度生成模型的人脸编辑具有以下优势:1)扩展性更强,模型通用于各种人脸,无需对单个人脸进行独立操作;2)灵活高效,利用训练好的模型可以对人脸属性实现一键转换;3)高保真度的输出,结合现有生成模型的优势,能够实现真实、自然的人脸重构。

但不容忽视的是,基于深度生成网络的人脸编辑还面临着种种挑战。首先,采用条件引导构造人脸编辑模型的思路存在以下问题:1)条件信息往往采用标签或图像实例进行表达,而标签的获取具有一定难度,手动标注或采用标签网络获取的标签不仅耗费人力,获取的标签也通常具有离散性,无法

对人脸进行连贯的编辑,采用图像实例作为条件信息虽然可以使编辑效果更精确,却不具备灵活性;2)通过手绘草图的方式对人脸细节进行调整往往依赖于作图者的绘画功底与美学修养,一旦绘图失真,将同时导致生成的图像失真。其次,运用隐变量构造人脸编辑模型的思路有一些限制:1)针对图像嵌入的过程,直接编码方法在重建质量方面取得的成果有限,而优化和混合方法代价高昂,需要大量时间来收敛,与此同时,运用隐变量的前提是目标图像所在的数据域被原生成模型的训练数据包含,这导致预训练的生成模型使用受限;2)在潜在空间寻找方向的属性编辑方法只适用于界限分明的属性,如性别、年龄、佩戴眼镜与否等,但对于改变脸型 and 发型等难以分类的属性效果不佳。最后,从整体而言,该项技术面临如下挑战:1)无论采用何种方案,由于每种属性所独有的性质,模型的通用性与针对性难以达到平衡;2)采用编码-解码结构时属性解耦效果更好,但因瓶颈模块(bottleneck)的存在,生成图像的质量往往不高,同时,模型在潜在空间中分离属性和背景的能力往往有限,属性传递过程会相互干扰,这常导致生成图像模糊和出现伪影,而使用预训练的生成模型可以提高输出图像的质量,但往往忽视了潜在空间的耦合问题,因此,模型的解耦程度与保真程度难以兼顾;3)深度学习依赖于大量的数据集,训练过程要求较高的算力,导致获取人脸编辑模型的过程较为复杂。

未来的研究方向可以着眼于以下方面:1)针对采用条件约束的思想,可以利用无监督或数据增强技术^[86],减少对条件标签的依赖;通过元学习、迁移学习等方式让模型在学习任务中积累经验,降低对数据量的需求;在采用草图与语义图进行人脸编辑的任务中,应加强模型的智能化,避免过度依赖于手绘者的绘画功底。2)针对运用隐变量的思想,可以进一步利用已有生成模型的高质量图像生成结果,在确保属性编辑精确性的同时,加强网络的通用性。3)现有人脸编辑网络的属性解耦能力还具有较大的发展潜力,因此可以进一步开发属性解耦程度与保真程度共同发展的网络结构。

结束语 本文介绍了基于深度生成网络的人脸编辑技术进展,从原理和结构出发,分别介绍了采用图像翻译、在网络内部引入条件信息以及操纵潜在空间的人脸编辑方案,并对基于深度生成网络的人脸编辑主要面临的身份一致性、属性解耦、属性编辑精确性3个方面的挑战进行阐述,最后总结了该方法的优缺点与未来的发展方向。总体来说,基于深度生成网络的人脸编辑提高了编辑的便捷性、灵活性与易用性,但也导致个人权益和社会稳定面临威胁。掌握基于深度生成网络的人脸编辑核心技术,是实现高质量人脸编辑及解决人脸编辑安全问题的关键。

参考文献

- [1] ZHU Z, LUO P, WANG X, et al. Recover canonical-view faces in the wild with deep neural networks[J]. arXiv:1404.3543, 2014.
- [2] ZHANG Z, PENG Y. Eyeglasses removal from facial image based on mvlr[M]// The Era of Interactive Media. New York: Springer, 2013:101-109.
- [3] GOODFELLOW I, POUGET-ABADIE J, MIRZAM, et al. Generative adversarial nets[C]// Advances in Neural Information Processing Systems. 2014:2672-2680
- [4] BROCK A, DONAHUE J, SIMONYAN K. Large scale GAN training for high fidelity natural image synthesis[J]. arXiv:1809.11096, 2018.
- [5] KARRAS T, LAINE S, AILA T. A style-based generator architecture for generative adversarial networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019:4401-4410.
- [6] MIRZA M, OSINDERO S. Conditional generative adversarial nets[J]. arXiv:1411.1784, 2014.
- [7] 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.
- [8] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]// Proceedings of the IEEE International Conference on Computer Vision. 2017:2223-2232.
- [9] LIU M Y, BREUEL T, KAUTZ J. Unsupervised image-to-image translation networks[J]. arXiv:1703.00848, 2018.
- [10] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]// International Conference on Medical Image Computing and Computer-assisted Intervention. Cham: Springer, 2015:234-241.
- [11] WANG T C, LIU M Y, ZHU J Y, et al. High-resolution image synthesis and semantic manipulation with conditional gans[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018:8798-8807.
- [12] PORTENIER T, HU Q, SZABO A, et al. Faceshop: Deep sketch-based face image editing[J]. arXiv:1804.08972, 2018.
- [13] JO Y, PARK J, SC-FEGAN. Face Editing Generative Adversarial Network with User's Sketch and Color[C]// Proceedings of the IEEE International Conference on Computer Vision. 2019:1745-1753.
- [14] LEE C H, LIU Z, WU L, et al. Maskgan: Towards diverse and interactive facial image manipulation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:5549-5558.
- [15] CHEN S Y, SU W, GAO L, et al. DeepFaceDrawing: deep generation of face images from sketches[J]. ACM Transactions on Graphics (TOG), 2020, 39(4):1-16.
- [16] WANG Z, TANG X, LUO W, et al. Face aging with identity-preserved conditional generative adversarial networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018:7939-7947.
- [17] LI M, ZUO W, ZHANG D. Deep identity-aware transfer of facial attributes[J]. arXiv:1610.05586, 2016.
- [18] VIAZOVETSKYI Y, IVASHKIN V, KASHIN E. Stylegan2 distillation for feed-forward image manipulation[J]. arXiv:2003.03581, 2020.
- [19] PARK T, LIU M Y, WANG T C, et al. Semantic image synthesis with spatially-adaptive normalization[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019:2337-2346.
- [20] TAN Z, CHAI M, CHEN D, et al. MichiGAN: multi-input-conditioned hair image generation for portrait editing[J]. arXiv:2010.16417, 2020.

- [21] HU M, GUO J. Facial attribute-controlled sketch-to-image translation with generative adversarial networks[J]. *EURASIP Journal on Image and Video Processing*, 2020, 2020(1):1-13.
- [22] HE Z, ZUO W, KAN M, et al. Arbitrary facial attribute editing: Only change what you want[J]. *arXiv*; 1711.10678, 2017.
- [23] KIM T, CHA M, KIM H, et al. Learning to discover cross-domain relations with generative adversarial networks[J]. *arXiv*; 1703.05192, 2017.
- [24] YI Z, ZHANG H, TAN P, et al. Dualgan: Unsupervised dual learning for image-to-image translation[C]// *Proceedings of the IEEE International Conference on Computer Vision*. 2017; 2849-2857.
- [25] CHOI Y, CHOI M, KIM M, et al. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018; 8789-8797.
- [26] CHOI Y, UH Y, YOO J, et al. Stargan v2: Diverse image synthesis for multiple domains[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020; 8188-8197.
- [27] ZHAO B, CHANG B, JIE Z, et al. Modular generative adversarial networks[C]// *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018; 150-165.
- [28] ANOOSHEH A, AGUSTSSON E, TIMOFTE R, et al. Combo-gan: Unrestrained scalability for image domain translation[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2018; 783-790.
- [29] PALSSON S, AGUSTSSON E, TIMOFTE R, et al. Generative adversarial style transfer networks for face aging[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2018; 2084-2092.
- [30] SONG L, LU Z, HE R, et al. Geometry guided adversarial facial expression synthesis[C]// *Proceedings of the 26th ACM International Conference on Multimedia*. 2018; 627-635.
- [31] PUMAROLA A, AGUDO A, MARTINEZA M, et al. Ganimation: Anatomically-aware facial animation from a single image [C] // *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018; 818-833.
- [32] WU R, ZHANG G, LU S, et al. Cascade ef-gan: Progressive facial expression editing with local focuses [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020; 5021-5030.
- [33] LI T, QIAN R, DONG C, et al. Beautygan: Instance-level facial makeup transfer with deep generative adversarial network[C]// *Proceedings of the 26th ACM International Conference on Multimedia*. 2018; 645-653.
- [34] CHANG H, LU J, YU F, et al. Pairedcyclegan: Asymmetric style transfer for applying and removing makeup[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018; 40-48.
- [35] WU P W, LIN Y J, CHANG C H, et al. Relgan: Multi-domain image-to-image translation via relative attributes[C]// *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019; 5914-5922.
- [36] ZHANG H, CHEN W, TIAN J, et al. RAG: Facial Attribute Editing by Learning Residual Attributes [J]. *IEEE Access*, 2019, 7: 83266-83276.
- [37] SHEN W, LIU R. Learning residual images for face attribute manipulation[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017; 4030-4038.
- [38] HUANG X, LIU M Y, BELONGIE S, et al. Multimodal unsupervised image-to-image translation [C] // *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018; 172-189.
- [39] LEE H Y, TSENG H Y, HUANG J B, et al. Diverse image-to-image translation via disentangled representations [C] // *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018; 35-51.
- [40] LAMPLE G, ZEGHIDOUR N, USUNIER N, et al. Fader networks: Manipulating images by sliding attributes [J]. *arXiv*; 1706.00409, 2017.
- [41] ZHOU S, XIAO T, YANG Y, et al. Genegan: Learning object transfiguration and attribute subspace from unpaired data[J]. *arXiv*; 1705.04932, 2017.
- [42] XIAO T, HONG J, MA J. DNA-GAN: Learning disentangled representations from multi-attribute images [J]. *arXiv*; 1711.05415, 2017.
- [43] XIAO T, HONG J, MA J. Elegant: Exchanging latent encodings with gan for transferring multiple face attributes[C]// *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018; 168-184.
- [44] PERARNAU G, VAN DE WEIJER J, RADUCANU B, et al. Invertible conditional gans for image editing [J]. *arXiv*; 1611.06355, 2016.
- [45] NITZAN Y, BERMANO A, LI Y, et al. Disentangling in latent space by harnessing a pretrained generator [J]. *arXiv*; 2005.07728, 2020.
- [46] LI X, HU J, ZHANG S, et al. Attribute guided unpaired image-to-image translation with semi-supervised learning [J]. *arXiv*; 1904.12428, 2019.
- [47] LIU Y, LI Q, SUN Z. Attribute-aware face aging with wavelet-based generative adversarial networks[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019; 11877-11886.
- [48] OR-EL R, SENGUPTA S, FRIED O, et al. Lifespan Age Transformation Synthesis[J]. *arXiv*; 2003.09764, 2020.
- [49] LIU M, DING Y, XIA M, et al. Stgan: A unified selective transfer network for arbitrary image attribute editing[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019; 3673-3682.
- [50] ZHANG Z, SONG Y, QI H. Age progression/regression by conditional adversarial autoencoder[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017; 5810-5818.
- [51] GUO J, LIU Y, QIAN Z, et al. Exemplar-based Generative Facial Editing[J]. *arXiv*; 2006.00472, 2020.
- [52] JIANG W, LIU S, GAO C, et al. Psgan: Pose and expression robust spatial-aware gan for customizable makeup transfer[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020; 5194-5202.
- [53] TEWARI A, ELGHARIB M, BHARAJ G, et al. Stylerig: Rigging stylegan for 3d control over portrait images[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020; 6142-6151.

- [54] EGGER B, SMITH W A P, TEWARI A, et al. 3d morphable face models—past, present, and future[J]. *ACM Transactions on Graphics (TOG)*, 2020, 39(5):1-38.
- [55] KWAK J, HAN D K, KO H. CAFE-GAN: Arbitrary Face Attribute Editing with Complementary Attention Feature [J]. arXiv:2011.11900, 2020.
- [56] LIU Y, FAN H, NI F, et al. ClsGAN: Selective Attribute Editing Model based on Classification Adversarial Network [J]. *Neural Networks*, 2017, 133:220-228.
- [57] XIE D, YANG M, DENG C, et al. Fully-Featured Attribute Transfer [J]. arXiv:1902.06258, 2019.
- [58] CRESWELL A, BHARATH A A. Inverting the generator of a generative adversarial network [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 30(7):1967-1974.
- [59] GU J, SHEN Y, ZHOU B. Image processing using multi-code gan prior [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020:3012-3021.
- [60] ABDAL R, QIN Y, WONKA P. Image2stylegan: How to embed images into the stylegan latent space? [C] // *Proceedings of the IEEE International Conference on Computer Vision*. 2019:4432-4441.
- [61] ABDAL R, QIN Y, WONKA P. Image2StyleGAN++: How to Edit the Embedded Images? [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020:8296-8305.
- [62] LARSEN A B L, SØNDERBY S K, LAROCHELLE H, et al. Autoencoding beyond pixels using a learned similarity metric [C] // *International Conference on Machine Learning*. PMLR, 2016:1558-1566.
- [63] LUO J, XU Y, TANG C, et al. Learning inverse mapping by autoencoder based generative adversarial nets [C] // *International Conference on Neural Information Processing*. Cham: Springer, 2017:207-216.
- [64] GUAN S, TAI Y, NI B, et al. Collaborative learning for faster stylegan embedding [J]. arXiv:2007.01758, 2020.
- [65] RICHARDSON E, ALALUF Y, PATASHNIK O, et al. Encoding in style: a stylegan encoder for image-to-image translation [J]. arXiv:2008.00951, 2020.
- [66] ZHU J, SHEN Y, ZHAO D, et al. In-domain gan inversion for real image editing [J]. arXiv:2004.00049, 2020.
- [67] DONAHUE J, KRÄHENBÜHL P, DARRELL T. Adversarial feature learning [J]. arXiv:1605.09782, 2016.
- [68] DUMOULIN V, BELGHAZI I, POOLE B, et al. Adversarially learned inference [J]. arXiv:1606.00704, 2016.
- [69] DONAHUE J, SIMONYAN K. Large scale adversarial representation learning [J]. arXiv:1907.02544, 2019.
- [70] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks [J]. arXiv:1511.06434, 2015.
- [71] UPCHURCH P, GARDNER J, PLEISS G, et al. Deep feature interpolation for image content changes [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017:7064-7073.
- [72] SHEN Y, GU J, TANG X, et al. Interpreting the latent space of gans for semantic face editing [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020:9243-9252.
- [73] JAHANIAN A, CHAI L, ISOLA P. On the “steerability” of generative adversarial networks [J]. arXiv:1907.07171, 2019.
- [74] SummitKwan. TL-GAN: transparent latent-space GAN [EB/OL]. https://github.com/SummitKwan/transparent_latent_gan.
- [75] GOETSCHALCKX L, ANDONIAN A, OLIVA A, et al. Ganalyze: Toward visual definitions of cognitive image properties [C] // *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019:5744-5753.
- [76] KHOSLA A, RAJU A S, TORRALBA A, et al. Understanding and predicting image memorability at a large scale [C] // *Proceedings of the IEEE International Conference on Computer Vision*. 2015:2390-2398.
- [77] ABDAL R, ZHU P, MITRA N, et al. Styleflow: Attribute-conditioned exploration of stylegan-generated images using conditional continuous normalizing flows [J]. arXiv:2008.02401.
- [78] VAN T P, NGUYEN T M, TRAN N N, et al. Interpreting the Latent Space of Generative Adversarial Networks using Supervised Learning [C] // *2020 International Conference on Advanced Computing and Applications (ACOMP)*. IEEE, 2020:49-54.
- [79] VOYNOV A, BABENKO A. Unsupervised Discovery of Interpretable Directions in the GAN Latent Space [J]. arXiv:2002.03754, 2020.
- [80] HÄRKÖNEN E, HERTZMANN A, LEHTINEN J, et al. GAN-Space: Discovering Interpretable GAN Controls [J]. arXiv:2004.02546, 2020.
- [81] SHEN Y, ZHOU B. Closed-form factorization of latent semantics in gans [J]. arXiv:2007.06600, 2020.
- [82] PLUMERAULT A, BORGNE H L, HUDELOT C. Controlling generative models with continuous factors of variations [J]. arXiv:2001.10238, 2020.
- [83] BURGESS C P, HIGGINS I, PAL A, et al. Understanding disentangling in beta VAE [J]. arXiv:1804.03599, 2018.
- [84] CHEN R T Q, LI X, GROSSE R, et al. Isolating sources of disentanglement in variational autoencoders [J]. arXiv:1802.04942, 2018.
- [85] CHEN X, DUAN Y, HOUTHOOFT R, et al. Infogan: Interpretable representation learning by information maximizing generative adversarial nets [J]. arXiv:1606.03657, 2016.
- [86] MA C G, GUO Y Y, WU P, et al. Summary of Research on Generated Anti-Network Image Enhancement [J]. *Information Network Security*, 2019 (5):10-21.



TANG Yu-xiao, born in 1997, master. Her main research interests include artificial intelligence and face editing.



WANG Bin-jun, born in 1962, Ph. D. professor, Ph.D supervisor, is a member of China Computer Federation. His main research interests include cyber security and artificial intelligence.