

基于属性法向量修改的人脸编辑算法

鞠怡轩¹, 张建海¹, 茅晓阳², 许佳奕^{1*}

(1. 杭州电子科技大学计算机学院, 浙江 杭州 310018; 2. 日本山梨大学工学部, 山梨 甲府 1955-245)

摘要: 根据用户需求修改给定人脸图像, 在美妆造型设计、异质人脸合成等领域具有重要的应用价值。目前已有研究基于生成对抗网络(GAN)模型实现人脸的语义属性迁移或者定性修改, 但少有算法能够对用户指定的属性进行定量修改。针对人脸多个属性的分离(解纠缠)以及定量编辑两大问题, 提出了一个基于属性法向量修改的定量人脸属性编辑算法。首先, 端对端求解出隐藏空间中每一个面部属性的最优控制法向量, 实现人脸多属性的解纠缠; 然后通过调整法向量的移动系数, 控制单个属性的变化程度从而控制该属性的定量修改; 最后通过融合多个单属性控制向量的方式, 达到对人脸多个属性的综合修改效果。实验结果表明, 该算法能够同时适用于生成的虚拟人脸图像和真实人脸照片, 在保持生成图像质量的同时, 实现了人脸属性定量修改的功能。

关键词: 人脸合成; 生成对抗网络; 属性编辑; 定量编辑

中图分类号: TP 391.41

文献标志码: A

文章编号: 0438-0479(2021)04-0701-10

人脸属性编辑的目标是指修改面部图像的单个或多个属性, 进而生成具有目标属性的新人脸。由于其定向修改属性的特性, 人脸属性编辑在人机交互和娱乐社交领域具有广泛应用, 衍生出了人脸妆造编辑和人脸表情编辑等产品。人脸妆造编辑通过更改人脸的妆容, 实现人像美颜, 可以应用于直播与社交领域, 具有巨大的市场潜力。人脸表情编辑不仅用于数字娱乐行业, 通过仿真人脸的不同表情带来趣味性, 还可以用于去除人脸的表情, 辅助提升人脸识别算法的精度。在异质人脸合成领域, 根据用户描述生成罪犯的肖像, 对人脸进行年龄编辑以预测走失儿童长大后的长相研究方向, 在公共安全方面具有极大的应用。因此研究人脸属性编辑, 实现人脸的可控生成具有重要的经济和社会价值。

传统人脸属性编辑算法通常是先找到人脸图像中的关键点, 然后利用人脸关键点通过手工调整或者网格变形等方式编辑人脸。这种方式不仅耗时而且有时会出现人脸扭曲等问题。随着深度学习的发展, 人脸属性编辑通常基于深度生成模型实现。早期的人脸图像生成研究通常使用变分自动编码器(VAE)^[1], 而

现有属性可控的人脸图像生成研究主要是基于生成对抗网络(GAN)^[2]架构, 以解决 VAE 只能生成较模糊人脸的问题。可控人脸生成技术根据其添加控制条件方式的不同, 大致上可以分为以下 3 类: 1) 在生成器的输入层加入条件^[3-4], 生成码由随机噪声和条件两部分组成, 通过改变条件向量实现属性控制。其中条件生成 GAN(cGAN)^[3]和互信息最大化 GAN(InfoGAN)^[4]是两种比较典型的模型。其中, cGAN 模型引入控制向量至生成器和判别器, 生成码与控制向量合并成新的向量作为生成器的输入。由于判别器存在控制条件, 在训练的时候真实数据与条件一起作为过往经验由判别器学习。如果生成器生成的数据不够逼真或者生成的图像与所给条件不符, 则判别器可以根据过往经验进行判别。因此, 为了能够瞒过判别器, 生成器在训练时会根据送入的条件调整其输出。虽然使用 cGAN 模型在一定程度上实现了控制生成, 但是由于在生成器中额外添加了条件, 会出现训练困难和模式崩溃的问题。InfoGAN 将生成码划分为噪声信号和控制向量, 让生成器通过互信息学习的方法控制向量与噪声信号之间的映射, 进而控制生成结果。虽然

收稿日期: 2020-12-16 录用日期: 2021-03-26

基金项目: 浙江警察学院开放课题(2020DSJSYS002)

* 通信作者: xujiayi@hdu.edu.cn

引文格式: 鞠怡轩, 张建海, 茅晓阳, 等. 基于属性法向量修改的人脸编辑算法[J]. 厦门大学学报(自然科学版), 2021, 60(4): 701-710.

Citation: JU Y X, ZHANG J H, MAO X Y, et al. Facial image editing through manipulating attribute normal vector[J]. J Xiamen Univ Nat Sci, 2021, 60(4): 701-710. (in Chinese)



使用 InfoGAN 模型能够对生成人脸进行一定程度的控制,但是在模型训练结束前无法确定控制的内容,控制结果存在随机性. 2) 基于风格迁移^[5-9]的思想,从一幅模板图像中获取条件信息,将获取的条件信息与待修改图像进行融合,生成具有指定信息的新图像. 比较经典的方法有循环 GAN(CycleGAN)^[5]和星形 GAN(StarGAN)^[6]. CycleGAN 模型引入一个生成器将风格转换后的图像再次转换为原风格,并指出在训练好的模型中两者应该相同. 这种基于场景切换的思想,能够用于更改面部的关键部位. 为了应对多风格迁移问题,StarGAN 在生成器中引入风格标签,所有的风格转换都使用同一个生成器,有效地避免了多风格转换导致生成器过多的问题. 本文作者将人脸属性的迁移理解为多模态图像的转换,使用属性分类损失和循环一致性损失较好地平衡了生成图像身份一致性与属性迁移的问题. 基于风格迁移方式的算法,以一幅人脸图像作为模板,然后将该模板的风格信息复制到新生成的样本中. 3) 通过分析生成器中生成码构成的隐空间的具体含义,解构生成器的隐空间,找出具有目标属性且满足控制条件的生成码^[10-18]. InterFaceGAN (interpreting the disentangled face representation learned by GAN)^[10]提出,存在一个线性超平面可以将生成器的隐空间划分为两部分,一边的生成图像具有指定属性,另一边生成图像不具有指定属性,并且生成码距分割平面越远,图像具有属性的程度就越强烈. 基于这种特性,存在一个垂直于分割平面的法向量,使得生成码沿着法向量移动时会越过线性超平面,导致属性发生改变,进而实现修改指定属性的目的. 然而 InterFaceGAN 首先需要借助分类器或者以人工标记的方式计算生成码对应的属性得分,再通过支持向量机(SVM)求解分割平面,从而计算出法向量的值. 该算法无法实现端对端的训练. IcGAN (Invertible conditional GAN)^[13]模型引入两个编码网络,解构出给定图片对应的随机噪声和条件,然后将随机噪声和修改后的条件向量送入 cGAN 重新生成属性编辑后的效果图. 由于 IcGAN 是建立在解构 cGAN 的基础之上,其条件的控制受限于 cGAN,不能很好地进行定量的属性修改.

与常见的方法不同,本文研究的并不是将一张人脸的属性整体性地迁移到另外一张人脸,也不是一次性对人脸的多个属性进行有或没有的定性改变. 为了有效地避免目前条件式 GAN 模型训练的困难,本文针对属性风格迁移的方法需要模板图像问题,通过分析已有生成器的隐空间,解析出控制单个属性变化的

控制向量,再对多个向量进行融合指导生成码的修改,从而实现人脸多属性的解纠缠与细粒度编辑的算法.

1 本文算法

基于 GAN 的人脸图像生成模型,其主要思想是以随机噪声作为生成码,输入生成器中,生成人脸图像. 对人脸图像进行属性编辑,本质上是对生成码的修改. 基于这种思想,为了达到分离单一属性并提供定量修改的目的,本文提出了一种求解法向量来修改生成码,从而实现人脸属性编辑的方法. 人脸属性编辑流程如图 1 所示(以眼镜属性为例). 假设一张佩戴眼镜的人脸经反求解后获得的生成码为 z . 首先由用户指定待修改的属性,使用本文算法训练出法向量以解构出控制眼镜属性变化的部分;然后设置属性对应的修改系数并使用式(1)对生成码进行负向移动以控制属性修改的程度;最后将修改后的生成码输入生成器重新生成,得到一张去除眼镜属性的人脸.

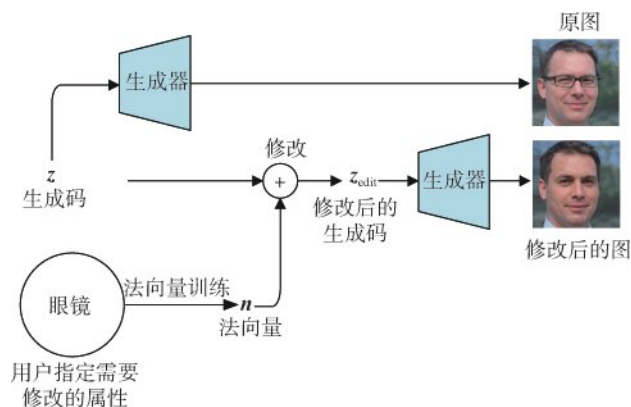


图 1 人脸属性编辑算法的流程

Fig. 1 Architecture of face attribute editing algorithm

1.1 基于属性法向量控制生成码的思想

基于 InterFaceGAN^[10]的思想,用一个线性平面将生成器的隐空间分割成是否包含指定属性的两部分 P_1 和 P_2 ,当生成码 z 沿着垂直于分割平面的法向量移动时能够最快地越过分割平面,并改变指定属性(图 2). 本文使用式(1)对生成码进行单一属性修改.

$$z_{\text{edit}} = z + \alpha n. \quad (1)$$

其中: z 是生成码; z_{edit} 是修改后的生成码; n 是法向量,即控制向量; α 是移动系数. 具体来说,通过选择不同属性所对应的法向量即可实现对不同属性的编辑,通过选择不同的 α 可以控制属性修改的程度,若 α 大于 0 表示添加属性,若 α 小于 0 表示删除属性.

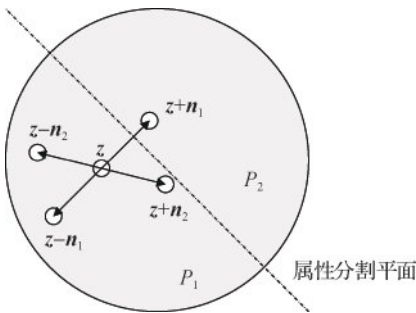


图 2 生成码的编辑
Fig. 2 Editing of latent code

对于单位向量 n_1 和 n_2 , 越垂直于分割平面, 则越接近于法向量, 生成码沿此正方向移动时, 越过分割平面的可能性就越大, 属性就越有可能发生改变. 生成码沿此负方向移动时, 距离分割平面就越远, 图片不具有指定属性的可能性就越大. 所以生成码 z 沿 n_1 正向移动时添加指定属性的可能性比沿 n_2 正向移动时的大, 生成码 z 沿 n_1 负向移动时消除指定属性的可能性比沿 n_2 负向移动时的大.

基于以上思想, 本文设计了损失函数 $V(n)$ 以衡量一个向量与分割平面法向量的距离:

$$V(n) = E[\log(C(G(z-n))) + \log(1 - C(G(z+n)))]. \tag{2}$$

其中: $E[x]$ 为期望函数; G 是生成器; C 是分类器, 用于预测图片具有指定属性的概率. 当 n 为理想中的法向量时, $C(G(z-n))$ 取得最小值, $C(G(z+n))$ 取得最大值, 损失函数整体取得最小值. 即 $V(n)$ 最小时, 可以找到移动生成码前后改变属性程度最大的 n^* , 即为法向量的近似解.

1.2 法向量的求解流程

基于损失函数最小化求解法向量的思想, 本文设

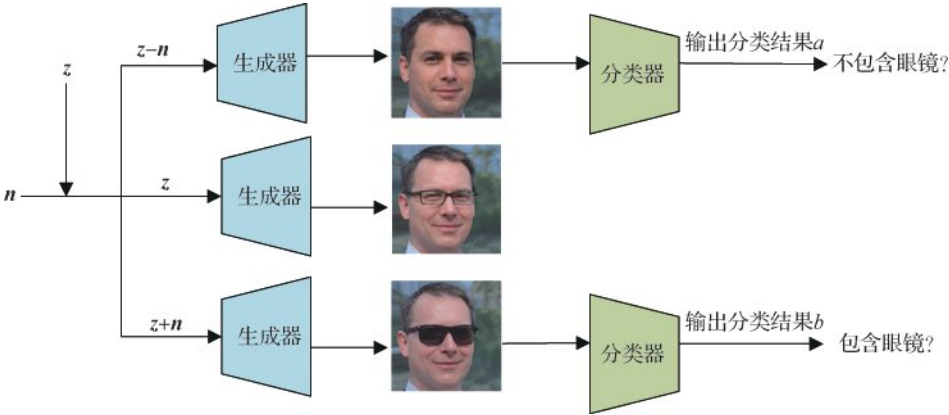


图 3 求解法向量的过程
Fig. 3 Process for solving normal vector

计了一个训练属性的理想法向量的框架(图 3). 主要计算步骤如下所示.

输入: n 作为初始化的零向量.

步骤 1 选取一个随机噪声作为生成码 z .

步骤 2 分别沿着 n 的正负方向移动 z , 生成 $z-n$ 和 $z+n$ 一对生成码.

步骤 3 将 $z-n$ 和 $z+n$ 作为生成器 G 的输入, 产生两张修改后的图片.

步骤 4 将修改后的图片送入分类器 C 进行分类, 得到对应的 a 和 b 两个分类结果.

步骤 5 计算损失值, 优化迭代 n .

步骤 6 回到步骤 1, 直至对任意的 $z, z-n$ 不具有眼镜属性, $z+n$ 具有眼镜属性.

输出: 控制眼镜属性变化的理想法向量的近似解 n^* .

本文通过法向量修改生成码的方式, 可以添加原本不存在的属性, 或者抑制某些不需要的属性. 在步骤 2 中, 使用式(1)对 z 进行修改, 可以分别沿着法向量正负方向移动, 进行添加或抑制两个操作, 以约束优化后的 n 更符合预期. 在步骤 4 中, 预期的分类结果是 a 不包含眼镜属性, b 包含眼镜属性, 在步骤 5 中以此作为两个分类结果的标签值, 采用例如交叉熵等损失函数计算出分类误差. 通过最小化该分类误差, 实现调整 n 值向理想法向量逼近. 当损失值小于一定阈值后, 终止训练, 输出优化后的 n^* .

当 n 训练的足够接近于理想法向量之后, 在实验的多数情况下, 生成码沿着 n 正向移动时能够较好地生成一张具有眼镜属性的人脸, 沿着 n 负向移动时能够较好地生成一张不具有眼镜属性的人脸.

1.2.1 生成器结构

本文使用的生成器为基于风格的 GAN

(StyleGAN)^[19] 中预训练的生成器. 该生成器经由 FFHQ 数据集训练而成, 可输出分辨率为 1 024 的高质量人脸. StyleGAN 的生成器有两种输入方式: 一种是在未经解构的 z 域, 使用 512 维的随机噪声向量作为生成器的输入; 另一种是经过多层全连接解构的域 (即 w^+ 域), 使用 18×512 维的张量作为生成器的输入. 因为经过解构的生成码修改起来更加纯净, 一定程度上避免了使用法向量修改时, 带来其他无关属性的改动, 本文选择 w^+ 域作为生成器的输入.

由于本文算法修改前后的生成码属于生成器的原始隐空间, 所以生成器的质量决定了修改前后图片的质量. 为了保证生成图片的质量, 还可以尝试使用其他^[20-22]能够稳定生成高质量、高分辨率图片的生成器.

由于本文算法修改前后的生成码属于生成器的原始隐空间, 所以生成器的质量决定了修改前后图片的质量. 为了保证生成图片的质量, 还可以尝试使用其他^[20-22]能够稳定生成高质量、高分辨率图片的生成器.

1.2.2 分类器结构

为了对 n 的优化产生正确的指导, 在训练过程中, 通过精确的分类器衡量生成码移动后是否正确改变指定属性, 分类器在训练过程中为 n 的优化提供了指导作用.

本文使用卷积神经网络结构搭建用于判别人脸指定属性是否存在的二分类模型, 其结构如图 4 所示. 网络的输入是一张 1 024 分辨率的彩色图像, 可以直接串联在生成器后面进行判别. 每次卷积时采用 0 填充, 保持卷积后特征图大小不变, 再经过一次最大池化以降低特征图的大小. 在彩色图像经过 5 次的卷积池化操作后, 缩小为 32 分辨率的特征图. 经过一次自适应平均池化后, 将特征图拉伸为向量, 送入后续的几层全连接层. 最终网络经过 softmax 层输出人脸指定属性的二分类结果.

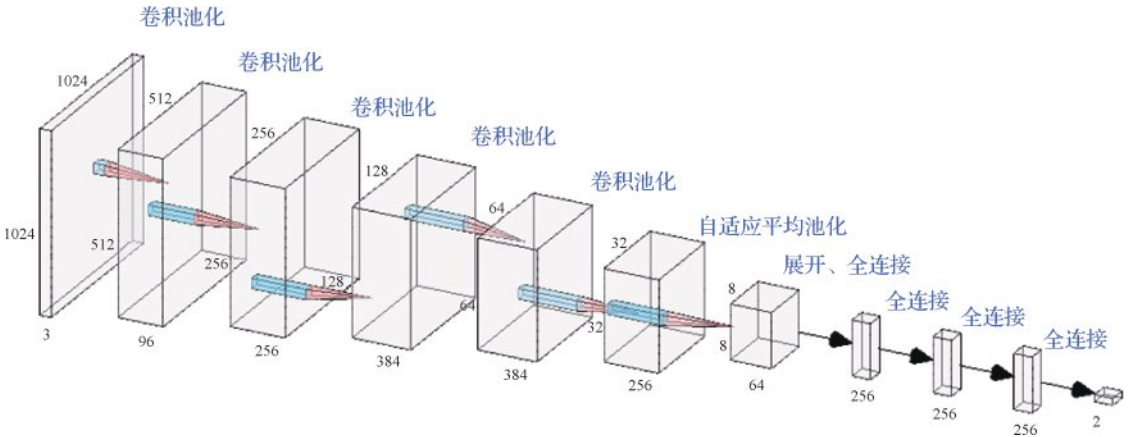


图 4 人脸属性的分类网络

Fig. 4 Face attribute classification network

2 实验结果与分析

本文在分辨率为 1 024 的 CelebA-HQ 数据集 (共计 3 万张图片) 上进行训练以验证本文算法的有效性. 其中训练集和验证集以 8 : 2 的比例划分. CelebA-HQ 的属性分布有 40 个, 由于部分属性不容易辨识, 本文挑选出特征容易辨识的 4 种常见属性: 性别、笑容、眼镜、年龄进行试验.

本文以性别 (男性为正方向)、笑容、眼镜及年龄 (年轻为正方向) 4 种属性为例, 设计如下实验:

- 1) 法向量训练的实验, 评估属性对应法向量的训练效果;
- 2) 单属性与多属性修改的实验, 检验法向量用于人脸单属性和多属性编辑的效果;
- 3) 移动系数 α 与属性修改程度的实验, 分析移动

系数 α 属性对修改程度的影响;

- 4) 人脸属性编辑对图像质量的影响实验, 评估人脸编辑对原始生成器生成质量的影响;
- 5) 真实人脸修改的实验, 将本文算法拓展用于真实人脸的属性编辑.

2.1 法向量训练的实验

为了计算 4 个属性各自对应的法向量, 首先在基于 StyleGAN 框架预训练的生成器的基础上, 训练与各个属性相对应的分类器. 本文算法与 InterFaceGAN 算法在 4 种属性上的二分类训练结果如表 1 所示. 本文算法在各属性上的分类准确率达到 82% 以上, 其中, 眼镜属于硬属性, 易于分类, 准确率比较高; 年龄属于软属性, 分类难度较大, 准确率较低. InterFaceGAN 在各属性上的分类准确率达到 75% 以上. 相比较而言, 本文算法的分类效果更好, 能够更好地辅助法向量的训练.

表 1 4 种人脸属性的分类准确率

Tab. 1 Classification accuracy of 4 kinds of face attributes

算法	准确率/%			
	性别	笑容	眼镜	年龄
本文算法	91.64	91.52	99.50	82.77
InterFaceGAN	84.20	78.50	80.10	75.30

在生成器和分类器预训练完毕后再对法向量的值进行训练,损失函数的变化如图 5 所示. 在训练的开始阶段,生成码沿着 n 正负方向移动后,不能够很

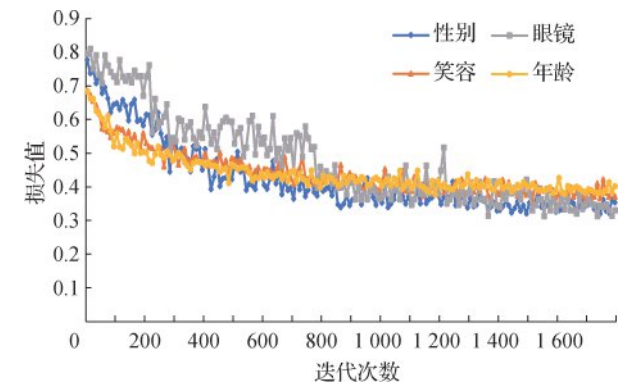


图 5 4 种属性法向量的损失曲线

Fig. 5 Loss curve of 4 kinds of attributenormal vectors

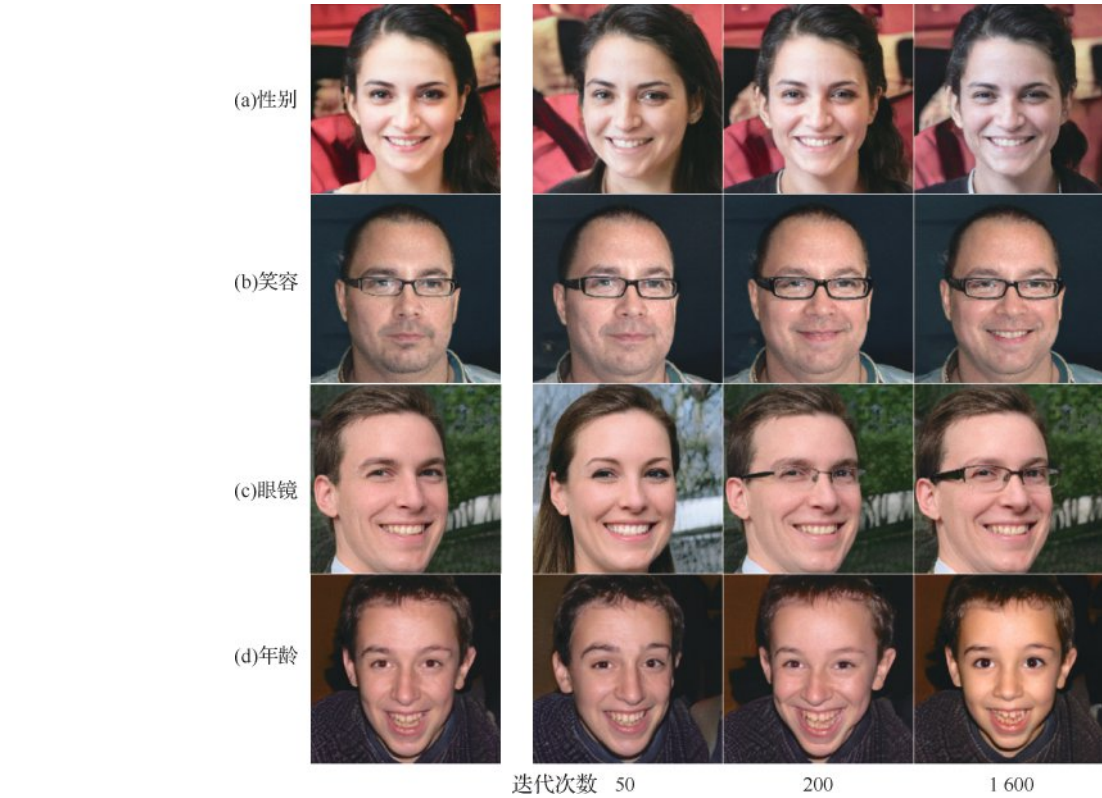


图 6 不同训练迭代次数法向量对应的人脸编辑效果

Fig. 6 Results of face editing with normal vectors under different training iterations

好地产生或消除指定属性,损失值偏高;在经过一定次数的迭代之后, n 逐渐近似于 n^* , 使得大部分生成码在往正方向移动时,产生或者加深了指定属性的特征,在生成码往负方向移动时,消除或者减弱了相关特征;最终分类器的分类损失降低后稳定在一定阈值之内.

为了观察训练时法向量逐步向理想法向量逼近的过程,本文以男性、笑容、眼镜、年龄为例,抽取了迭代次数为 50,200,1 600 的法向量对人脸进行修改,实验结果如图 6 所示. 在迭代次数为 50 时,法向量已经能够初步修改指定属性,但是修改效果不足,而且还会引入比较多的噪声,例如在修改眼镜属性时,此时法向量没有完成修改反而还改变了性别等无关属性. 在迭代次数为 200 时,眼镜属性的特征还需要继续加强,但法向量已经初步具备较稳定的修改效果. 在迭代次数为 1 600 时,性别和年龄属性改变了人物肤质,笑容和眼镜属性强化了特征,使其整体看起来更加满足要求,此时法向量基本接近理想法向量.

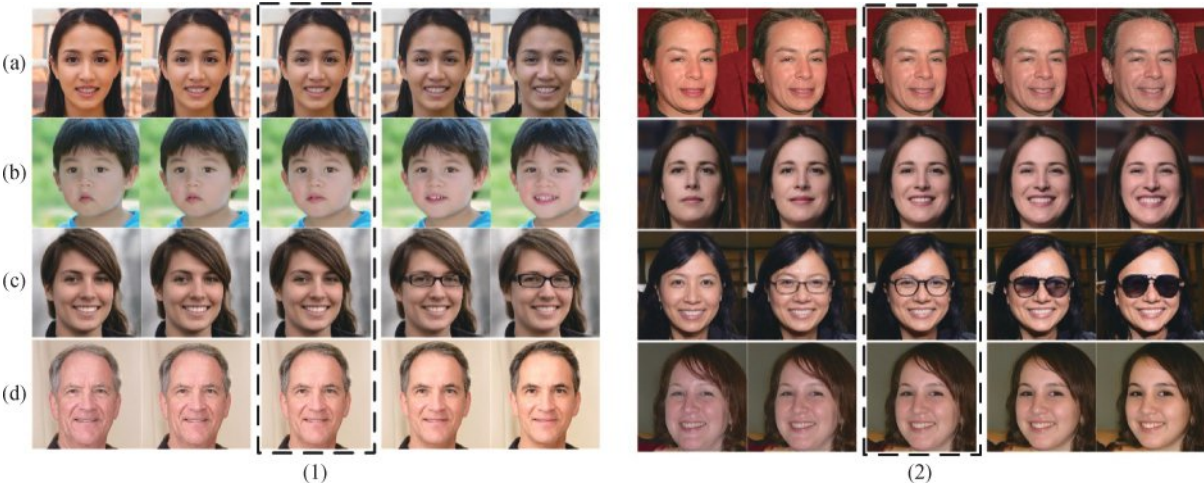
2.2 单属性修改的实验

为了验证本文算法在单属性修改的效果,本文选取了 100 个随机噪声作为生成码,观察了 4 种属性的

法向量使用式(1)修改生成码后的效果. 从整体上来看,法向量能够很好地控制人脸指定属性的从无到有、慢慢加深的过程,生成的结果也比较自然.

其部分结果如图 7 所示. 在往男性方向移动的过程中,头发逐渐变短,面部皮肤逐渐粗糙,整体看起来更加的男性化,往女性方向移动的过程中头发变长,面部逐渐细腻有光泽,整体看起来更加的女性化. 相应地,嘴巴张开及露出牙齿的程度越大,就认为笑的

越厉害. 眼镜的镜框越大,镜片颜色越深,眼镜属性在图像中的显著性越强. 发量越多、脸色越稚嫩,就越显得年轻. 使用本文算法进行单属性修改带来的变化符合人类日常生活经验,比较自然. 这种修改方法能够针对于特定的属性进行修改,在一定程度上避免其他无关联属性的改动. 在属性修改前后,人物面部的整体轮廓及大量细节没有发生太大的变化,整体上看都是针对同一个人的不同属性的表达.



在(1)和(2)两组图中,最中间虚线部分为原图,(a)、(b)、(c)、(d)分别对应男性、笑容、眼镜、年龄法向量在原图上由负方向到正方向逐渐修改的过程.

图 7 单属性修改效果
Fig. 7 Results of single attribute modification

2.3 多属性修改的实验

基于单属性修改的思想,通过使用多个单属性法向量对生成码修改,可以实现多属性修改的效果. 多属性修改公式为

z_{edit} = z + α₁ n₁ + α₂ n₂ + ⋯ + α_n n_n. (3)

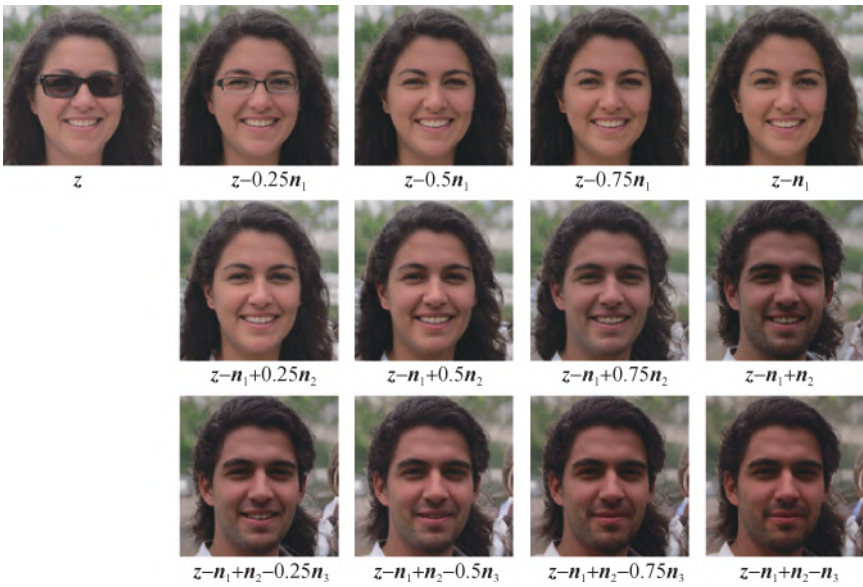
为了验证本文算法多属性修改的有效性,使用式(3)对生成码的性别、笑容及眼镜 3 种属性进行修改,其结果如图 8 所示. 图中每一行相当于一次独立的单属性修改,如第 1,2,3 行分别只在眼镜、性别和笑容上变化. z 对应的原图是一张带眼镜、女性、大笑表情的人脸. 沿眼镜法向量的负方向移动,成功摘除了眼镜,接着使用类似的操作,对原图额外修改了性别和笑容属性,最终得到一张不带眼镜、男性、表情中性的人脸. 从整体上看,这种融合多个单属性法向量为一个混合法向量的方式,能够很好地控制人脸指定属性的编辑.

2.4 移动系数 α 与属性修改程度的实验

为了观察移动系数 α 对属性修改程度的影响,本

文在使用法向量对人脸属性进行编辑时,设定不同的 α 对同一张人脸图像进行修改. 通过调整 α 的值,可以实现对属性的定量修改. 当 α 取值在-1 到 1 之间时,大多数情况下可以实现指定属性的添加或删除. 通过增加移动系数,可以进一步增大修改的幅度. 如图 9 所示,当移动系数 α 取值从 0 达到 1 时,根据修改后的生成码生成的图像开始出现男性的特征. 在移动系数 α 取值达到 2 时,修改后的人脸中男性的特征更加明显,出现了胡子,且随着移动系数的增加大胡子越浓密.

然而,由于不同属性对应的法向量并非正交,相互间存在一定的干扰,当移动系数太大时,在部分情况下可能会放大这些干扰,带来较明显的噪声和色差缺陷. 如图 10 所示,本文在不同移动系数下对指定属性进行了添加. 观察实验发现,各属性在移动系数为 1 时,具有较明显的属性添加效果,且修改后图像的质量良好. 其中眼镜和笑容较好地实现了属性的解纠缠. 而对于年龄与性别属性,当移动系数取值较大(为 3 时),虽然各属性得到了较大强化,但同时引入了较



n_1, n_2, n_3 分别代表眼镜、性别、笑容的法向量.

图 8 多属性修改效果

Fig. 8 Results of multi-attributes modification

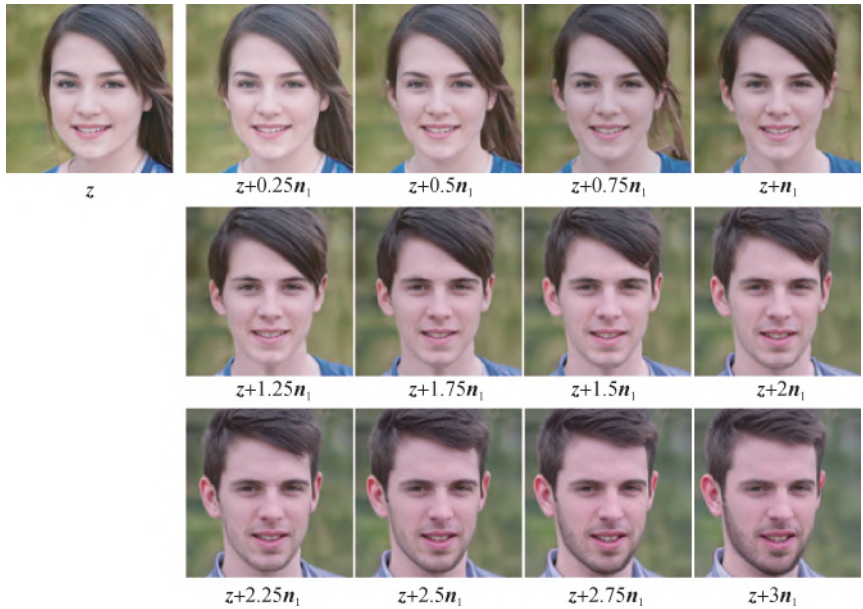


图 9 不同移动系数下性别修改的结果

Fig. 9 Results of gender modification under different displacement factor

大的噪声和色差问题. 本文算法出现色差的原因可能是由于在训练法向量的过程中, 模型将肤色的改变与性别、年龄属性的改变相关联. 例如男性与女性, 小孩与老人在皮肤的颜色与细腻程度等方面存在一定的差距. 后续将尝试通过扩充数据集和改进算法进行改善.

2.5 人脸属性编辑对图像质量的影响实验

为了衡量使用法向量修改人脸前后图像质量的变化, 本文采用衡量真实图像与生成图像之间距离的

FID(Fr chet inception distance)^[23]指标 D 进行评估. 本文选择生成器预训练时使用的 FFHQ 数据集作为真实数据集, 使用法向量对生成器随机生成的 5 000 张样本人脸进行移动系数为 1 的正负属性修改(修改后的图像添加或删除了指定的属性), 构造了由 1 万张修改后的人脸图像构成的生成数据集. 本文计算 FID 的代码由 Github 上的开源项目提供 (<https://github.com/bioinf-jku/TTUR>).

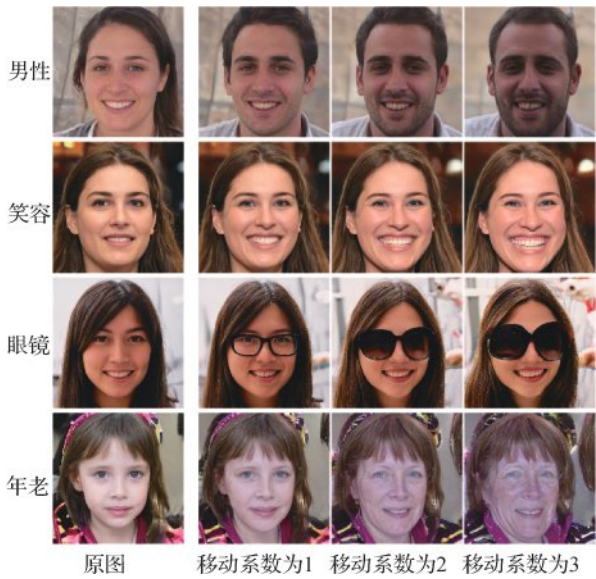


图 10 移动系数对图像质量的影响

Fig. 10 Influence of displacement factor on image quality

本文算法中使用了预训练的生成器,生成器的优劣直接影响人脸图像的生成质量. 本文设计质量因子 Q 用于评估人脸属性编辑对图像质量的影响. Q 的计算如式(4)所示:

$$Q = \frac{D_{\text{before}}}{D_{\text{after}}}. \tag{4}$$

其中, D_{before} 是生成器生成的图像所计算的 FID, D_{after} 是使用法向量修改后的图像所计算的 FID. Q 的值越大

代表生成质量越高,通常情况下 Q 为小于 1 的值,表示 D_{after} 大于 D_{before} ,说明属性修改造成了图像的质量损失.

如图 11 所示,本文算法各属性对应的质量因子在 0.94 以上,其中眼镜属性基本保持原有生成质量(由于计算误差, Q 的值略微大于 1). 实验结果表明本文算法在多数情况下使用法向量进行属性修改仅会造成轻微的质量损失;除了年龄属性外,其他属性的质量因子均优于 InterFaceGAN 算法.

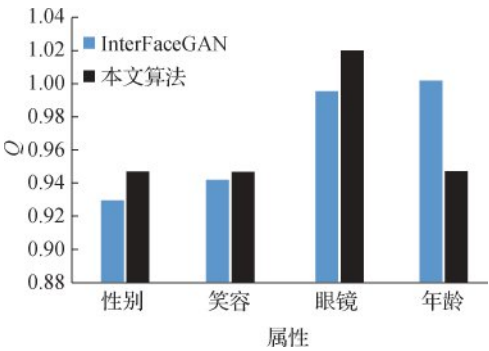


图 11 人脸属性编辑对图像质量的影响

Fig. 11 Influence of face attribute editing on image quality

图 12 展示了本文算法与 InterFaceGAN 算法对同一人脸分别进行笑容、眼镜、性别、年龄等属性修改后的结果. 本文算法与 InterFaceGAN 算法均取得了较好的修改效果,整体修改效果相当. 从笑容属性的修改效果可以看出: InterFaceGAN 算法着重于控制

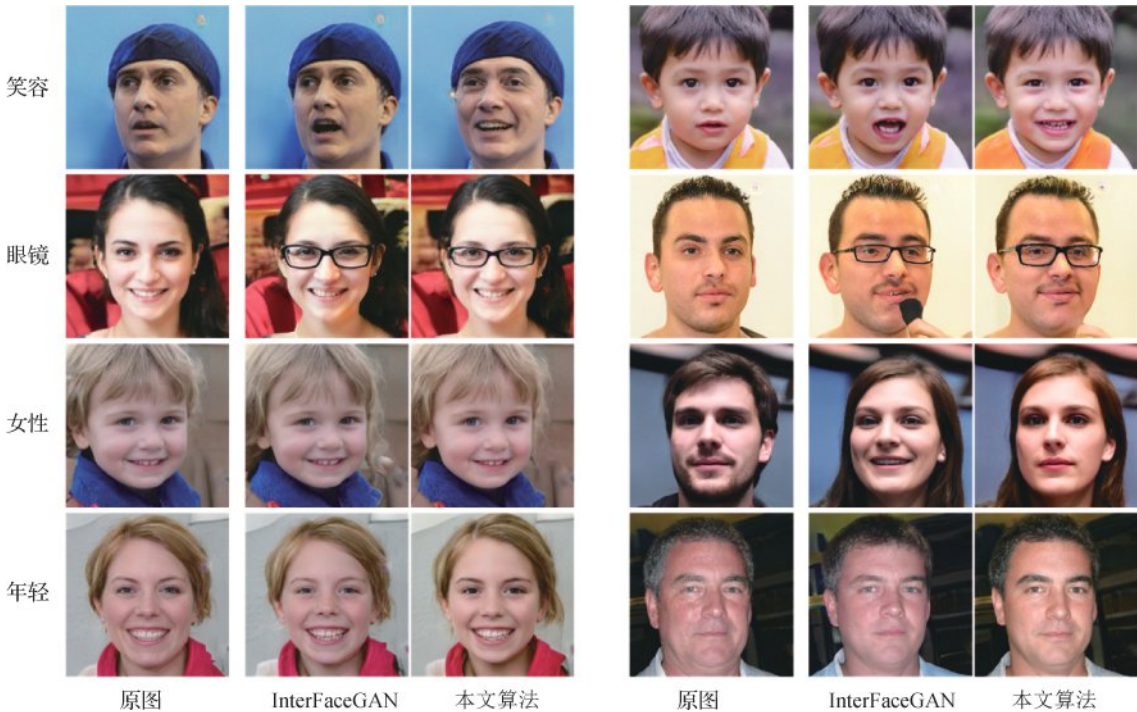


图 12 本文算法与 InterFaceGAN 算法的效果对比

Fig. 12 Comparison results of our algorithm and InterFaceGAN

嘴巴的闭合程度,本文算法能够对嘴巴,面部肌肉,眼睛进行一个整体的调整,相比较而言本文算法修改结果更加自然。

2.6 修改真实人脸属性的实验

目前用于虚拟人脸的属性编辑是以随机噪声作为生成码进行修改,可以通过将真实人脸图片投影至

生成器隐空间,实现本文算法拓展至真实人脸属性编辑。如图13所示,本文通过使用优化生成码的算法^[24],尝试将其论文数据集中奥巴马的肖像,投影至本文所使用生成器的隐空间中。在获得对应的生成码之后,本文尝试使用多属性修改的方法,对映射后的图像进行了修改。



z 为投影到生成器隐空间的生成码; n_1, n_2, n_3 分别代表性别、笑容、眼镜属性对应的法向量。

图13 真实人脸属性编辑效果

Fig. 13 Editing effect of real face image

3 结 论

本文提出了一种端对端的基于属性法向量修改的算法来解决人脸属性分解与定量编辑的问题。本文算法无需借助模板图像,可以通过调整控制法向量的方式达到定量调整人脸属性的目的,具有友好的操作性和良好的实际推广价值。

本文算法有如下创新之处:

- 1) 提出一种端对端的属性控制法向量的计算方法,无需从参考人脸中迁移特征、条件等信息,支持直接通过法向量修改生成码达到修改人脸属性的目的;
- 2) 提供定量的属性编辑功能,不仅可以指定修改单个或者混合属性,还可以定量的控制每个属性改变的程度;
- 3) 无需依赖于 GAN 的生成器架构,本文算法能够适用于常见无条件式生成器,可实现无条件式生成器向有条件式生成器的转换。

为了使人脸属性修改更加纯净,图像质量更加优越,在未来的工作中,将考虑在法向量的训练过程中添加更多限制。为了将所有法向量的求解纳入统一框架,考虑把多个二分类器整合成一个多分类器,实现同时训练多个法向量的目标。

参考文献:

[1] KINGMA D P, WELLING M. Auto-encoding variational

bayes[EB/OL]. [2020-09-01]. <https://arxiv.org/pdf/1312.6114.pdf>.

[2] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]// Advances in Neural Information Processing Systems. Cambridge: MIT, 2014: 2672-2680. pdf.

[3] MIRZA M, OSINDERO S. Conditional generative adversarial nets[EB/OL]. [2020-09-01]. <https://arxiv.org/pdf/1411.1784.pdf>.

[4] CHEN X, DUAN Y, HOUTHOOFT R, et al. InfoGAN: interpretable representation learning by information maximizing generative adversarial nets[C]// Advances in Neural Information Processing Systems. Cambridge: MIT, 2016: 2172-2180.

[5] 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. Venice: IEEE, 2017: 2223-2232.

[6] CHOI Y, CHOI M, KIM M, et al. Stargan: unified generative adversarial networks for multi-domain image-to-image translation[C]// IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 8789-8797.

[7] KIM T, CHA M, KIM H, et al. Learning to discover cross-domain relations with generative adversarial networks[C]// International Conference on Machine Learning. Sydney: ACM, 2017: 1857-1865.

[8] LAMPLE G, ZEGHIDOUR N, USUNIER N, et al. Fader networks: manipulating images by sliding attributes[C]//

- Advances in Neural Information Processing Systems. Venice; IEEE, 2017; 5967-5976.
- [9] SHEN W, LIU R. Learning residual images for face attribute manipulation[C]// IEEE Conference on Computer Vision and Pattern Recognition. Hawaii; IEEE, 2017; 4030-4038.
- [10] SHEN Y, GU J, TANG X, et al. Interpreting the latent space of GANs for semantic face editing[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington; IEEE, 2020; 9243-9252.
- [11] 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. Hawaii; IEEE, 2017; 5810-5818.
- [12] HE Z, ZUO W, KAN M, et al. Attgan: facial attribute editing by only changing what you want [J]. IEEE Transactions on Image Processing, 2019, 28(11): 5464-5478.
- [13] PERARNAU G, VAN DE WEIJER J, RADUCANU B, et al. Invertible conditional GANs for image editing[EB/OL]. [2020-09-01]. <https://arxiv.org/pdf/1611.06355.pdf>.
- [14] ISOLA P, ZHU J Y, ZHOU T, et al. Image-to-image translation with conditional adversarial networks[C]// IEEE Conference on Computer Vision and Pattern Recognition. Hawaii; IEEE, 2017; 1125-1134.
- [15] YIN W, FU Y, SIGAL L, et al. Semi-latent GAN: learning to generate and modify facial images from attributes. [EB/OL]. [2020-09-01]. <https://arxiv.org/pdf/1704.02166.pdf>.
- [16] NATSUME R, YATAGAWA T, MORISHIMA S. RsGAN: face swapping and editing using face and hair representation in latent spaces[EB/OL]. [2020-09-01]. <https://arxiv.org/pdf/1804.03447.pdf>.
- [17] JIANG S, TAO Z, FU Y. Segmentation guided image-to-image translation with adversarial networks[C]// IEEE International Conference on Automatic Face & Gesture Recognition. Piscataway; IEEE, 2019; 1-7.
- [18] JO Y, PARK J. SC-FEGAN: face editing generative adversarial network with user's sketch and color[C]// IEEE International Conference on Computer Vision. California; IEEE, 2019; 1745-1753.
- [19] KARRAS T, LAINE S, AILA T. A style-based generator architecture for generative adversarial networks[C]// IEEE Conference on Computer Vision and Pattern Recognition. California; IEEE, 2019; 4401-4410.
- [20] ALEC R, LUKE M, SOUMITH C. Unsupervised representation learning with deep convolutional generative adversarial networks[EB/OL]. [2020-09-01]. <https://arxiv.org/pdf/1511.06434.pdf>.
- [21] KARRAS T, AILA T, LAINE S, et al. Progressive growing of gans for improved quality, stability, and variation[EB/OL]. [2020-09-01]. <https://arxiv.org/pdf/1710.10196.pdf>.
- [22] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein GAN[EB/OL]. [2020-09-01]. <https://arxiv.org/pdf/1701.07875.pdf>.
- [23] HEUSEL M, RAMSAUER H, UNTERTHINER T, et al. GANs trained by a two time-scale update rule converge to a local nash equilibrium[C]// Advances in Neural Information Processing Systems. Venice; IEEE, 2017; 6626-6637.
- [24] ABDAL R, QIN Y, WONKA P. Image2styleGAN: how to embed images into the styleGAN latent space? [C]// IEEE International Conference on Computer Vision. Seoul; IEEE, 2019; 4432-4441.

Facial image editing through manipulating attribute normal vector

JU Yixuan¹, ZHANG Jianhai¹, MAO Xiaoyang², XU Jiayi^{1*}

(1. School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China;

2. Division of Engineering, University of Yamanashi, Kofu 1955-245, Japan)

Abstract: Modification of given facial images according to users' demands has secured wide applications in areas such as makeup design, criminal portrait. Current generative adversarial network (GAN) based algorithms can transfer or qualitatively change semantic attributes, but can hardly attain the quantitative change of user-specified attributes. Aiming at these problems of attributes separation (entanglement) and quantitative editing, we propose a quantitative facial attribute-editing algorithm based on attribute normal vector modification. First, the optimal control vector corresponding to each attribute is solved end-to-end in the hidden space, and each attribute is entanglement. Then, by adjusting the displacement factor of normal vector, the variation degree of single attribute is controlled. Finally, through fusing multiple control vectors, the comprehensive modification effect of changing hybrid facial attributes is obtained. Experimental results show that the proposed method can be applied to quantitatively editing both the composite and real face images, while maintaining the quality of resulting images.

Keywords: face synthesis; generative adversarial network (GAN); attribute editing; quantitative editing