



基于 STGAN 的人脸属性编辑改进模型

林志坤^a, 许建龙^b, 包晓安^b

(浙江理工大学, a. 信息科学与工程学院; b. 计算机科学与技术学院, 杭州 310018)

摘要: 人脸属性编辑在美颜 APP 和娱乐领域有重要应用,但现有方法存在生成图像质量不高、属性编辑不够准确等问题,为此提出了一种基于选择传输生成对抗网络(Selective transfer generative adversarial networks, STGAN)的人脸属性编辑改进模型。运用潜码解耦思想,将潜码分解为内容潜码和风格潜码单独操作,提高源域图像和目标域图像的内容编码一致性,从而提高属性编辑准确率;同时运用像素级重构损失和潜码重构损失,在总损失函数中加入像素级限制和潜码重构限制,通过互补作用提高生成图像质量。在 CelebA 人脸数据集和季节数据集上进行实验,该模型相比当前人脸属性编辑主流模型在定性结果和定量指标上均有提高,其中峰值信噪比和结构相似性相比 STGAN 模型分别提高了 6.06% 和 1.58%。这说明该改进模型能够有效提高人脸属性编辑的性能,满足美颜 APP 和娱乐领域的需求。

关键词: 生成对抗网络;人脸编辑;重构图像;潜码解耦

中图分类号: TP391

文献标志码: A

文章编号: 1673-3851(2023)05-0285-08

引文格式: 林志坤, 许建龙, 包晓安. 基于 STGAN 的人脸属性编辑改进模型[J]. 浙江理工大学学报(自然科学), 2023, 49(3): 285-292.

Reference Format: LIN Zhikun, XU Jianlong, BAO Xiao'an. Improved model of face attribute editing based on STGAN [J]. Journal of Zhejiang Sci-Tech University, 2023, 49(3): 285-292.

Improved model of face attribute editing based on STGAN

LIN Zhikun^a, XU Jianlong^b, BAO Xiao'an^b

(a. School of Information Science and Engineering; b. School of Computer Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: Face attribute editing technology has important applications in beauty APPs and entertainment fields. However, the existing methods still have problems such as low-quality and inaccurate editing. To this end, an improved face editing model based on selective transfer generative adversarial networks (STGAN) was proposed. Using the idea of latent code decoupling, the latent code was decomposed into the content latent code and the style latent code, which improved the content-coding consistency of the source domain image and the target domain image, thereby improving the accuracy of attribute editing. In the meanwhile, we used pixel-level reconstruction loss and latent code reconstruction loss, and added pixel-level restrictions and latent code reconstruction restrictions to the total loss function, improving the quality of generated images through complementary effects. Experiments were carried out on the CelebA face dataset and seasonal dataset. Compared with the current mainstream model of face attribute editing, this model has improved both qualitative results and quantitative indicators. Compared

收稿日期: 2022-09-01 网络出版日期: 2023-01-16

基金项目: 浙江省重点研发计划项目(2020C03094)

作者简介: 林志坤(1997—), 男, 浙江温州人, 硕士研究生, 主要从事深度学习、计算机视觉等方面的研究。

通信作者: 许建龙, E-mail: xujianlong126@126.com

with the STGAN model, the peak signal-to-noise ratio and structural similarity index of this model are improved by 6.06% and 1.58%, respectively. This shows that the improved model can effectively improve the performance of face attribute editing and meet the needs of beauty apps and entertainment fields.

Key words: generative adversarial networks; face editing; reconstructed images; latent code decoupling

0 引言

图像属性编辑是指将源域图像映射到目标域,在保留源域图像内容特征的前提下,根据目标域图像的风格特征合成得到新图像。图像属性编辑有多个应用方向,如人脸属性编辑、图像修复^[1]、超分辨率、图像着色^[2]等,其中人脸属性编辑是一个热门方向,它在美颜 APP 和娱乐领域有重要应用。目前图像属性编辑模型主要有基于生成对抗网络(Generative adversarial network, GAN)^[3]和基于自编码器^[4]这两种模型。其中基于 GAN 的图像属性编辑模型按照特性有多种分类,如根据是否需要监督分为有监督和无监督的属性编辑模型,根据处理图像属性的数量又可以分为单一属性和多属性的编辑模型。

Isola 等^[5]提出了一个有监督的图像属性编辑模型;Wang 等^[6]在 Isola 等^[5]的基础上对模型进行了改进,提高了生成图像的分辨率。然而这类模型的训练都需要成对的图像数据集,这在很多任务中是无法实现的,如同张人脸的男女转换,几乎无法提供成规模的同一张人脸的不同性别数据集。Zhu 等^[7]设计的模型实现了无监督的图像属性编辑,通过循环一致性损失对模型进行约束,以保留图像转换过程中的基本特征,摆脱了成对图像数据集的束缚,但这类模型每次训练只能实现特定两个域之间的转换,若要实现多域转换需要训练相应数量的模型,耗时耗力。Anoosheh 等^[8]减少了要实现多域转换所需训练的网络数量,但他们提出的模型仍然要训练多个网络,并且也不能支持多个属性的同时转换。Choi 等^[9]提出了 StarGAN(Star generative adversarial networks)模型,这种模型可以根据输入的标签向量(即属性向量)不同实现不同域的转换,且只要训练一组模型,很好地完成了多域图像属性编辑任务。He 等^[10]提出了 AttGAN(Attribute generative adversarial networks)模型,他们将编码器和解码器结构运用到 StarGAN 模型中,实现了更好的图像转换效果。Liu 等^[11]提出了 STGAN(Selective transfer generative adversarial networks)模型,这种模型用目标属性向量

和源域属性向量的差值作为输入,而不是将原本整个属性向量作为输入,从而提高了图像重构质量;STGAN 模型在 AttGAN 模型的生成器(Generative model, G)中加入选择传输单元(Selective transfer unit, STU),在更多层间加入对称跳跃连接,在提高生成图像质量的同时保证了高属性编辑准确率。

近年来,潜码解耦思想^[12-14]被广泛运用于图像属性编辑,如 Shen 等^[15]提出的 InterFaceGAN(Interpreting face generative adversarial networks)模型。不同于 Liu 等^[16]提出的未分解的潜码一致性,潜码解耦思想将潜码进一步分解为内容潜码与风格潜码。内容潜码用来控制图像的基本内容,如人脸的基本轮廓;而风格潜码用来控制图像的不同风格特性,如人脸的发色、性别、年龄等。虽然上述模型已经可以实现多属性的无监督人脸属性编辑,但是它们生成的人脸图像仍然存在图像质量低、人脸属性编辑不准确等问题。当前人脸属性编辑的应用极为广泛,设计一个能生成高质量、高属性编辑准确率人脸图像的人脸属性编辑模型有着重要价值。

为了满足在美颜 APP 和娱乐领域对高质量人脸属性编辑能力的需求,本文提出了一种基于 STGAN 模型的人脸属性编辑改进模型,并将该模型称为 LEGAN(Loss function enhanced generative adversarial network)。该模型在 STGAN 模型的基础上,针对人脸属性编辑不准确问题,运用潜码解耦思想,促进源域图像和目标域图像的内容编码一致性,从而提高人脸属性编辑准确率;针对生成图像质量不高的问题,在 STGAN 模型基础上同时运用像素级重构损失和潜码重构损失,通过互补作用,进一步提高生成图像质量。

1 人脸属性编辑改进模型构建

本文建立了 LEGAN 模型,在 CelebA 人脸数据集^[17]上进行训练,训练后得到的模型具有较强的人脸属性编辑能力。运行时将所需处理的人脸图像和所需转换的属性向量输入模型,模型在属性向量

的指导下通过生成器生成所需的目标人脸图像。对比 STGAN 模型,本文提出的 LEGAN 模型最大的改进点是使用了潜码解耦合思想并改进了损失函数,在人脸图像属性编辑任务中能够处理得到质量更高、属性编辑更准确的人脸图像。

1.1 模型结构

LEGAN 的模型结构如图 1 所示,该模型由生成器、鉴别器、分类器组成,其中生成器由编码器 G_{enc} 和解码器 G_{dec} 组成。在训练阶段,输入的人脸图像 x^a 经过 G_{enc} 处理后得到潜码 z^a 。转换目标图

时,将目标域人脸图像的属性向量 b 和源域人脸图像的属性向量 a 的差值与潜码 z^a 一同输入 G_{dec} ,经过 G_{dec} 处理后得到目标人脸伪造图像 $x^{\hat{b}}$ 。将 $x^{\hat{b}}$ 输入分类器与鉴别器,分别计算属性分类准确率和图像的真实度。将 $x^{\hat{b}}$ 输入 G_{enc} 得到潜码 $z^{\hat{b}}$, $z^{\hat{b}}$ 和 z^a 之间进行潜码解耦合约束。将零向量和 z^a 输入 G_{dec} 后得到重构图 $x^{\hat{a}}$, $x^{\hat{a}}$ 与 x^a 之间需要计算像素级重构损失, $x^{\hat{a}}$ 输入 G_{enc} 后得到重构图的潜码 $z^{\hat{a}}$, $z^{\hat{a}}$ 与 z^a 之间需要计算潜码重构损失。

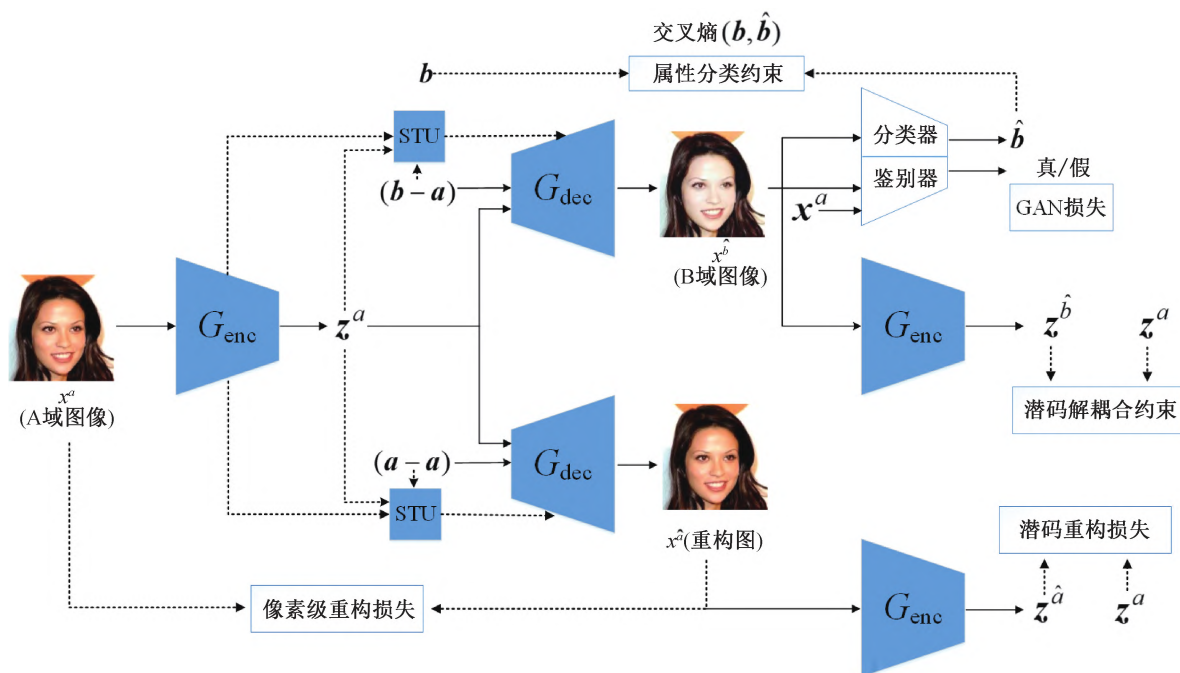


图 1 LEGAN 模型结构示意图

G_{enc} 提取图像潜码的过程可用式(1)~(3)表示:

$$z^a = G_{enc}(x^a) \quad (1)$$

$$z^{\hat{b}} = G_{enc}(x^{\hat{b}}) \quad (2)$$

$$z^{\hat{a}} = G_{enc}(x^{\hat{a}}) \quad (3)$$

G_{dec} 将潜码映射成图像。 G_{enc} 和 G_{dec} 之间采用对称跳跃连接,连接应用在生成器中所有的层。 G_{enc} 之间权重共享, G_{dec} 同理。Liu 等^[16]证明在 G_{enc} 和 G_{dec} 之间添加 STU 单元既能提高生成图像的质量,也能提高生成图像的属性编辑准确率,所以此处的 G_{enc} 和 G_{dec} 不同层之间的连接都经过 STU 单元处理。

1.2 内容编码一致性损失

STGAN 模型使用编码器将图像翻译成潜码,之后用条件向量来改变生成图像的属性。将 STGAN 模型的解码器提取的潜码视为内容编码,

将条件向量视为风格编码,则可在 STGAN 模型中加入内容编码一致性损失来降低不同域之间图像内容编码的差异。对于内容编码一致性损失,本文计算在 $z^{\hat{b}}$ 和 z^a 的 L_1 距离,计算过程可用式(4)表示:

$$L_{ccc} = \|z^{\hat{b}} - z^a\|_1 \quad (4)$$

其中: L_{ccc} 表示内容编码一致性损失。通过最小化 L_{ccc} 可以使不同域的图像经过编码器处理后得到的内容编码趋同,这可以使模型更精确地编辑需要改变的属性,即由条件向量控制的部分,从而提高属性编辑的准确率。属性更精准的控制也能进一步提高重构图像的质量。

1.3 潜码重构损失

为了提高重构能力,常见方法是计算重构图像和输入图像对应像素之间的 L_1 损失或 L_2 损失(像素级重构损失)。例如 STGAN 模型、AttGAN 模型

等,均通过最小化 L_1 损失或 L_2 损失使重构图像和输入图像之间更加相似。本文在像素级重构损失的基础上,计算了输入图像潜码和重构图像潜码之间的 L_1 损失,该损失可用式(5)表示:

$$L_{\text{rec}_2} = \| \mathbf{z}^{\hat{a}} - \mathbf{z}^a \|_1 \quad (5)$$

其中: L_{rec_2} 表示潜码重构损失。通过最小化该损失,可以使输入图像的潜码和重构图像的潜码更加相似。本文实验表明,同时添加像素级重构损失和潜码重构损失可以进一步提高模型的重构能力,详见实验部分。

1.4 总损失函数

本文分别用 $L_{D_{\text{adv}}}$ 和 $L_{G_{\text{adv}}}$ 表示鉴别器和生成器的对抗损失,两个损失可用式(6)–(7)表示:

$$\max_{D_{\text{adv}}} L_{D_{\text{adv}}} = E_{\mathbf{x} \sim D_{\text{adv}}} (D_{\text{adv}}(\mathbf{x}) - E_{\hat{\mathbf{y}} \sim D_{\text{adv}}} (D_{\text{adv}}(\hat{\mathbf{y}})) + \lambda E_{\hat{\mathbf{x}}} [\| \nabla_{\hat{\mathbf{x}}} D_{\text{adv}}(\hat{\mathbf{x}}) \|_2 - 1]^2] \quad (6)$$

$$\max_G L_{G_{\text{adv}}} = E_{\mathbf{x}, \mathbf{a}_{\text{diff}}} (D_{\text{adv}}(G(\mathbf{x}, \mathbf{a}_{\text{diff}}))) \quad (7)$$

其中: \mathbf{x} 表示输入的图像; $\hat{\mathbf{x}}$ 是真实图像和生成图像之间的线性插值; $\hat{\mathbf{y}}$ 表示生成的假图; \mathbf{a}_{diff} 是目标属性向量与原属性向量的差值; D 表示判别器。对抗损失采用 Gulrajani 等^[18] 提出的 WGAN-GP 形式。这里对抗损失以最大化的形式展示,具体实现时加入负号以最小化的形式优化。

本文用 $L_{D_{\text{att}}}$ 和 $L_{G_{\text{att}}}$ 分别表示生成器和鉴别器的分类损失,两个损失可用式(8)–(9)表示:

$$L_{D_{\text{att}}} = - \sum_{i=1}^n [a_s^{(i)} \log D_{\text{att}}^{(i)}(\mathbf{x}) + (1 - a_s^{(i)}) \log(1 - D_{\text{att}}^{(i)}(\mathbf{x}))] \quad (8)$$

$$L_{G_{\text{att}}} = - \sum_{i=1}^n [a_t^{(i)} \log D_{\text{att}}^{(i)}(\hat{\mathbf{y}}) + (1 - a_t^{(i)}) \log(1 - D_{\text{att}}^{(i)}(\hat{\mathbf{y}}))] \quad (9)$$

其中: n 为分类属性的个数; $a_s^{(i)}$ 表示源域第 i 个分类属性向量; $a_t^{(i)}$ 表示目标域第 i 个分类属性向量。

本文用 L_{rec_1} 表示像素级的重构损失,该损失可用式(10)表示:

$$L_{\text{rec}_1} = \| \mathbf{x} - G(\mathbf{x}, \mathbf{0}) \|_1 \quad (10)$$

其中: $\mathbf{0}$ 是零向量。生成重构图像时因为源域的属性向量即为目标域的属性向量,故差值输入为零向量。

最后总的损失函数可用式(11)–(12)表示:

$$\min_G L_G = -L_{G_{\text{adv}}} + \lambda_1 L_{G_{\text{att}}} + \lambda_2 L_{\text{rec}_1} + \lambda_3 L_{\text{rec}_2} + \lambda_4 L_{\text{ccc}} \quad (11)$$

$$\min_D L_D = -L_{D_{\text{adv}}} + \lambda_5 L_{D_{\text{att}}} \quad (12)$$

其中: L_G 表示生成器的损失; L_D 表示鉴别器的损失; $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5$ 是超参数。

2 实验和结果分析

本文在人脸数据集和季节数据集上进行了实验。为验证内容编码一致性损失和潜码重构损失在 STGAN 模型中的有效性,本文设计并进行相关的消融实验。

2.1 数据集和设置

本文选择 CelebA^[17] 作为人脸数据集。CelebA 人脸数据集有数据量大、多样性强、标注详细等优点。在 CelebA 中本文选择 178×218 对齐处理过的数据集。该数据集总共含有 202599 张人脸图像,每张图像都有对应的属性标注。在数据集分配方面本文将前 182000 张图像分为训练集,182001 ~ 182637 分为验证集,其余的分为测试集。属性方面本文选择秃头、刘海、黑发、金发、棕发、浓眉、眼镜、性别、嘴闭合、上唇胡须、络腮胡、苍白和年龄等一共 13 种可转换的属性,这涵盖了当前大部分人脸属性编辑模型所实现的属性。本文的模型使用 Adam 优化器($\beta_1 = 0.5, \beta_2 = 0.999$),Batch_size 大小设置为 32,学习率的衰减与 STGAN 模型一致。总损失函数的超参数 $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5$ 分别为 25、200、15、5、1。

2.2 定量结果

属性编辑的性能可以从生成图像质量和属性编辑准确率两个方面进行评价,其中图像质量可以用峰值信噪比(Peak signal-to-noise ratio, PSNR)和结构相似性(Structural similarity, SSIM)来衡量。因为同一张人脸的不同属性转换的真实数据难以获取(如性别转换、年龄转换),所以本文采用和 STGAN 模型中一样的两种方法来代替评估。本文将 LEGAN 模型与 StarGAN、AttGAN、STGAN、InterFaceGAN 等模型进行比较,其中用来测试人脸属性编辑的 AttGAN 模型和 STGAN 模型是原作者发布的,而 StarGAN 模型因为其作者发布的模型仅支持 5 个属性的操作,本文按照 StarGAN 模型相同的配置并用其作者 github 上的代码训练了支持 13 个属性的人脸编辑模型来进行比较,InterFaceGAN 模型也是根据其作者在 github 上发布的代码训练得到。

在图像质量方面,本文让目标属性向量与源域属性向量保持一致获得图像的重构结果,并且通过评估重构图像的质量来代替评估模型生成的图像质量。本文在测试集(大约 20000 张图像)测试了

StarGAN、AttGAN、STGAN、InterFaceGAN、LEGAN 等模型的重构图像质量,结果如表 1 所示。从表 1 可以看出,LEGAN 模型的 PSNR、SSIM 指标相比其余模型都更高,相对之下 StarGAN 模型和 AttGAN 模型的重构图像质量明显较弱。虽然 AttGAN 模型将 U-NET 的一层对称跳跃连接运用在生成器里,图像质量相对 StarGAN 模型有所提高,但提高的幅度有限。STGAN 模型因为差异属性向量输入和 STU 单元的运用使得图像质量有较为明显的提高,尤其是 SSIM 指标达到了 0.948 的高分,但它仍然存在改进的空间。通过借鉴潜码解耦思想以及两种重构损失的互补运用,LEGAN 模型在 STAGN 模型的基础上再次提高了图像质量,其中 SSIM 达到了 0.963,相比 STGAN 模型提高了 1.58%。LEGAN 模型的 PSNR 指标为 33.59,相比 STGAN 模型提高了 6.06%。

在属性编辑的准确度方面,本文使用和 STGAN 模型一致的人脸属性分类器来代替评估。该分类器是在 CelebA 数据集上对 13 个属性进行训练得到的,

并且在 CelebA 数据集上的准确率达到 94.5%。StarGAN、AttGAN、STGAN、InterFaceGAN、LEGAN 等模型的属性编辑准确率如表 2 所示。

表 1 不同模型在 CelebA 数据集上的定量结果

模型	PSNR/dB	SSIM	属性编辑准确率/%
StarGAN ^[9]	22.80	0.819	67.15
AttGAN ^[10]	24.07	0.841	65.02
STGAN ^[11]	31.67	0.948	86.05
InterFaceGAN ^[15]	32.13	0.941	85.73
LEGAN	33.59*	0.963*	88.79*

注:表中标“*”的数据表示最佳结果。

从表 2 可以看出,LEGAN 模型的平均属性编辑准确率最高,达 88.79%。对 13 个属性进一步观察,可以看见除了苍白和络腮胡这两个属性,LEGAN 模型的准确率比 STGAN 模型稍差(苍白属性差 0.41%,络腮胡属性差 0.63%),其余的 11 个属性 LEGAN 模型的准确率都最高,尤其金发、浓眉、棕发和年龄等 4 个属性分别提高了 9.32%、4.91%、5.40%和 6.63%。

表 2 不同模型的属性编辑准确率

%

属性	模型				
	SatrGAN ^[9]	AttGAN ^[10]	STGAN ^[11]	InterFaceGAN ^[15]	LEGAN
刘海	92.30	93.02	97.69	96.52	99.67*
秃头	24.37	22.52	59.91	60.98	61.58*
棕发	55.68	51.04	89.95	90.23	95.35*
金发	68.90	40.36	78.96	76.25	88.28*
黑发	59.40	46.34	93.01	93.59	94.02*
上唇胡须	18.79	30.29	38.88	40.41	42.66*
嘴闭合	96.09	96.99	98.52	98.39	99.49*
性别	77.13	83.28	96.69	97.23	97.86*
眼镜	95.72	98.59	99.72	98.86	99.77*
浓眉	50.11	49.21	88.26	85.57	93.17*
年龄	83.86	67.94	86.68	89.82	93.04*
苍白	83.99	84.44	97.04*	96.79	96.63
络腮胡	66.63	81.30	93.37*	89.88	92.74
平均准确率	67.15	65.02	86.05	85.73	88.79*

注:表中标“*”的数据表示最佳结果。

2.3 定性结果

经过训练,本文得到了人脸属性编辑效果较优的模型。本文分别针对单属性和多属性编辑进行对比,对比结果如图 2 所示。

从图 2 中可以直观地看出,LEGAN 模型的生成结果相比其余模型生成的图像质量和属性编辑完成度更高。其中 StarGAN 模型和 AttGAN 模型在多属性编辑时有的属性容易崩坏,比如

StarGAN 模型在变老加唇上胡须的转换中人脸肤色出现异常,AttGAN 模型在变老加唇上胡须的转换中唇上胡须的效果没有得到很好体现。InterFaceGAN 模型在最后一列的金发效果上上色转换得不够完全。LEGAN 模型相比 STGAN 模型生成效果相似,但属性细节处有所提高,比如图 2 最后一列多属性编辑中 LEGAN 模型生成的人脸相比 STGAN 模型更显苍老,且生成的刘海也更完整。

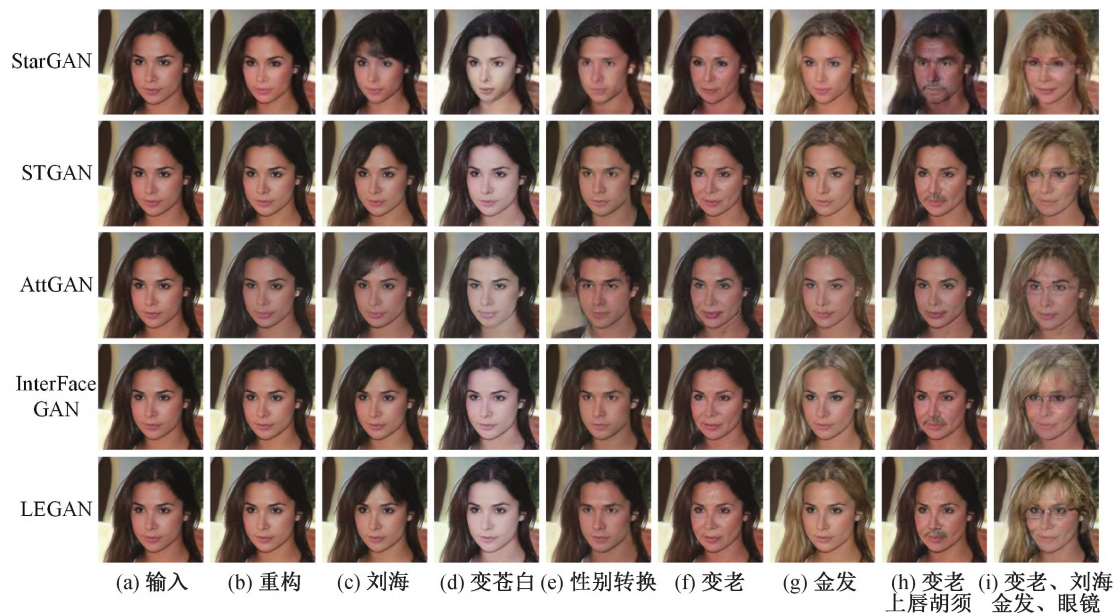


图2 各模型在 CelebA 数据集上的定性结果

2.4 消融实验

在这一部分,本文评估了两个主要组成部分的必要性:内容编码一致性损失和潜码重构损失。在 STGAN 模型的基础上,本文将加入内容编码一致性损失训练得到的模型称为 LEGAN-1,将加入潜码重构损失训练得到的模型称为 LEGAN-2。两个损失都加入训练得到的模型为本文提出的 LEGAN 模型。为了佐证以上两个损失对原始 STGAN 模型的影响,本文还添加了以下实验:STGAN-1,用潜码重构损失代替原始 STGAN 模型的像素级重构损失;STGAN-2,用潜码重构损失代替原始 STGAN 模型的像素级重构损失并加上内容编码一致性损失;STGAN-3,去掉原始 STGAN 模型的像素级重构损失,添加内容编码一致性损失;STGAN-4:去掉原始 STGAN 模型的像素级重构损失。实验结果如表 3 所示。

表3 消融实验结果

模型	PSNR/dB	SSIM	属性编辑准确率/%
STGAN ^[11]	31.67	0.948	86.05
STGAN-1	31.47	0.942	86.50
STGAN-2	32.25	0.944	89.16
STGAN-3	24.79	0.846	90.86*
STGAN-4	24.57	0.831	87.82
LEGAN-1	33.53	0.954	88.42
LEGAN-2	32.75	0.954	88.77
LEGAN	33.59*	0.963*	88.79

注:表中标“*”的数据表示最佳结果。

对比 STGAN、STGAN-1、LEGAN-2、STGAN-4 这些模型的实验结果可以发现,添加潜码重构损

失也能提高模型的重构能力,只是这个能力相比添加像素级重构损失略差。如果将潜码重构损失和像素级重构损失一起使用则可以在 STGAN 模型中起到互补的作用,模型的重构能力相比单独加入像素级重构损失的 STGAN 模型更佳。两两对比 STGAN-2 模型和 STGAN-1 模型、STGAN-3 模型和 STGAN-4 模型、LEGAN-1 模型和 STGAN 模型可以发现,在 STGAN 模型中加入内容编码一致性损失对模型的重构能力、属性编辑能力都有促进作用。对比 LEGAN-1 模型、LEGAN-2 模型和 LEGAN 模型可以发现,在 STGAN 模型中同时加入潜码重构损失和内容编码一致性损失,可以大幅提高模型的重构能力和属性编辑能力。虽然 STGAN-3 模型的属性编辑能力是几个模型中最强的,但它缺乏对模型重构能力的优化,生成的图像质量较低。

2.5 季节转换实验

因为人脸属性编辑和季节转换这类图像转换技术本质上都是图像风格属性的转换,所以本文也在季节数据集上进行了实验,以更好地评价 LEGAN 模型的性能,结果如图 3 所示。

本文所使用的季节数据集^[11]包含四个季节:春季、夏季、秋季和冬季。经过训练的模型需要实现同一风景图像的四个季节转换。本文在季节数据集上训练了 AttGAN 模型、STGAN 模型和 LEGAN 模型,并对它们的定量和定性结果进行比较。在定性结果方面,图 3 显示 LEGAN 模型的转换能力明显优于 AttGAN 模型和 STGAN 模型,尤其是

LEGAN 模型的冬季转换效果相比 STGAN 模型和 AttGAN 模型更自然。在定量结果方面,表 4 显示

LEGAN 模型的 PSNR 和 SSIM 与 STGAN 模型相比分别提高了 5.27% 和 5.15%。

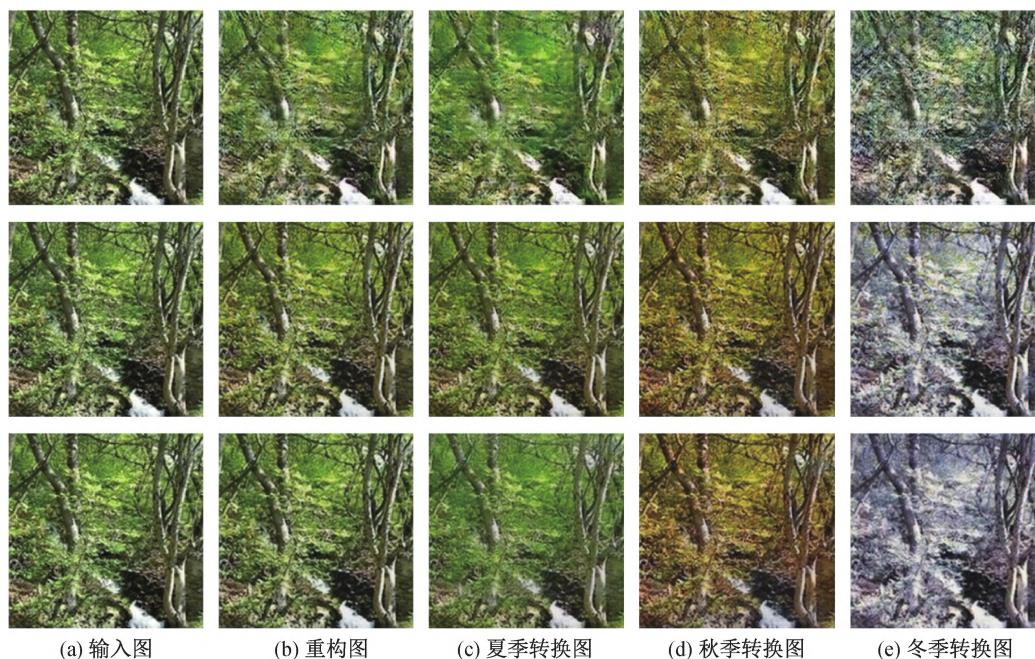


图3 季节转换效果

注:第一行是 AttGAN 模型的季节转换效果;第二行是 STGAN 模型的季节转换效果;第三行是 LEGAN 模型的季节转换效果。

表4 季节转换的定量结果

模型	PSNR/dB	SSIM
AttGAN ^[10]	22.68	0.680
STGAN ^[11]	24.22	0.796
LEGAN	25.50*	0.837*

注:表中标“*”的数据表示最佳结果。

3 结 论

本文提出基于 STGAN 模型的人脸属性编辑改进模型,通过潜码解耦合思想以及两种重构损失的互补运用,得到一个性能更优的人脸属性编辑模型。运用潜码解耦合可以让原 STGAN 模型更精准地编辑需要更改的图像属性,从而提高模型的属性编辑能力。将像素级重构损失和潜码重构损失互补地运用在 STGAN 模型中可以进一步提高模型生成的图像质量。实验结果表明,新模型在 CelebA 人脸数据集上相比 StarGAN、AttGAN、STGAN、InterFaceGAN 等主流模型拥有更优的定量和定性表现。该模型也可以运用在季节变换等图像转换任务里,实验表明新模型在季节数据集中相较 STGAN 等模型也有更好的表现。

由于光照、相机、场景、硬件设备等因素的影响,本文的模型目前还不能准确地处理所有真实数据,而且只能生成分辨率较低的图像。后续研究将进一

步完善数据集,改进模型结构,以训练一个功能更强的模型。

参考文献:

- [1] 曹建芳,张自邦,赵爱迪,等. 增强一致性生成对抗网络在壁画修复上的应用[J]. 计算机辅助设计与图形学学报, 2020, 32(8): 1315-1323.
- [2] 李洪安,郑峭雪,张婧,等. 结合 Pix2Pix 生成对抗网络的灰度图像着色方法[J]. 计算机辅助设计与图形学学报, 2021, 33(6): 929-938.
- [3] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets [C] // Proceedings of Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2014: 2672-2680.
- [4] Kingma D P, Welling M. Auto-encoding variational bayes [EB/OL]. (2014-05-01) [2022-08-31]. <https://arxiv.org/pdf/1312.6114.pdf>.
- [5] Isola P, Zhu J Y, Zhou T H, et al. Image-to-image translation with conditional adversarial networks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 5967-5976.
- [6] 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. Salt Lake City: IEEE, 2018: 8798-8807.
- [7] 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. Venice: IEEE, 2017: 2242-2251.
- [8] Anoosheh A, Agustsson E, Timofte R, et al. ComboGAN: Unrestrained scalability for image domain translation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Salt Lake City: IEEE, 2018: 783-790.
- [9] 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/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 8789-8797.
- [10] 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.
- [11] Liu, M, Ding, Y, Xia, M, et al. STGAN: A unified selective trans-fer network for arbitrary image attribute editing [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 3673-3682.
- [12] Huang X, Liu M Y, Belongie S, et al. Multimodal unsupervised image-to-image translation [C] // Proceedings of the European Conference on Computer Vision. Munich: IEEE, 2018: 179-196.
- [13] 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. Munich: IEEE, 2018: 35-51.
- [14] Lin J, Xia Y, Qin T, et al. Conditional image-to-image translation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 5524-5532.
- [15] 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. Seattle: IEEE, 2020: 9240-9249.
- [16] Liu M Y, Breuel T M, Kautz J. Unsupervised image-to-image translation networks [C] // Proceedings of Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2017: 700-708.
- [17] Liu Z W, Luo P, Wang X G, et al. Deep learning face attributes in the wild [C] // Proceedings of the IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 3730-3738.
- [18] Gulrajani I, Ahmed F, Arjovsky M, et al. Improved training of Wasserstein GANs[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM Press, 2017: 5769-5779.

(责任编辑:康 锋)