

基于 StarGAN 和类别编码器的图像风格转换*

许新征^{1,2}, 常建英¹, 丁世飞^{1,2}



¹(中国矿业大学 计算机科学与技术学院, 江苏 徐州 221116)

²(矿山数字化教育部工程研究中心(中国矿业大学), 江苏 徐州 221116)

通信作者: 许新征, E-mail: xuxinzh@163.com

摘要: 图像风格转换技术已经融入到人们的生活中, 并被广泛应用于图像艺术化、卡通化、图像着色、滤镜处理和去遮挡等实际场景中, 因此, 图像风格转换具有重要的研究意义与应用价值. StarGAN 是近年来用于多域图像风格转换的生成对抗网络框架. StarGAN 通过简单地下采样提取特征, 然后通过上采样生成图片, 但是生成图片的背景颜色信息、人物脸部的细节特征会与输入图像有较大差异. 对 StarGAN 的网络结构进行改进, 通过引入 U-Net 和边缘损失函数, 提出了用于图像风格转换的 UE-StarGAN 模型. 同时, 将类别编码器引入到 UE-StarGAN 模型的生成器中, 构建了融合类别编码器的小样本图像风格转换模型, 实现了小样本的图像风格转换. 实验结果表明: 该模型可以提取到更精细的特征, 在小样本的情况下具有一定的优势, 以此进行图像风格转换后的图片无论是定性分析还是定量分析都有一定的提升, 验证了所提模型的有效性.

关键词: 图像风格转换; 生成对抗网络; StarGAN; U-Net; 类别编码器

中图法分类号: TP391

中文引用格式: 许新征, 常建英, 丁世飞. 基于 StarGAN 和类别编码器的图像风格转换. 软件学报, 2022, 33(4): 1516–1526. <http://www.jos.org.cn/1000-9825/6482.htm>

英文引用格式: Xu XZ, Chang JY, Ding SF. Image Style Transferring Based on StarGAN and Class Encoder. Ruan Jian Xue Bao/ Journal of Software, 2022, 33(4): 1516–1526 (in Chinese). <http://www.jos.org.cn/1000-9825/6482.htm>

Image Style Transferring Based on StarGAN and Class Encoder

XU Xin-Zheng^{1,2}, CHANG Jian-Ying¹, DING Shi-Fei^{1,2}

¹(School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China)

²(Engineering Research Center of Mining Digitalization of Ministry of Education (China University of Mining and Technology), Xuzhou 221116, China)

Abstract: The image style transferring technology has been widely integrated into people's life, and it is widely used in image artistry, cartoon, picture coloring, filter processing, and occlusion removal of the practical scenarios, so image style transferring has an important research significance and application value. StarGAN is a generative adversarial network framework for multi-domain image style transferring in recent years. StarGAN extracts features through simple down-sampling, and then generates images through up-sampling. Nevertheless, the background color information and detailed features of people's faces in the generated images are quite different from those in the input images. In this study, by improving the network structure of StarGAN, after analyzing the existing problems of the StarGAN, a UE-StarGAN model for image style transferring is proposed by introducing U-Net and edge-promoting adversarial loss function. At the same time, the class encoder is introduced into the generator of UE-StarGAN, and a small sample image style transferring model is designed to realize the small sample image style transfer. The results of this experiment show that the model can extract more detailed features, have some advantages in the case of small sample size, and to a certain extent, the qualitative and quantitative analysis results of the images can be improved after the image style transferring, which verifies the effectiveness of the proposed model.

* 基金项目: 国家自然科学基金(61976217, 61976216)

本文由“面向开放场景的鲁棒机器学习”专刊特约编辑陈恩红教授、李宇峰副教授、邹权教授推荐.

收稿时间: 2021-06-01; 修改时间: 2021-07-16; 采用时间: 2021-08-27; jos 在线出版时间: 2021-10-26

Key words: image style transferring; generative adversarial network; StarGAN; U-Net; class encoder

近些年, 人工智能(artificial intelligence, AI)^[1,2]作为类脑智能计算领域的一个重要研究方向, 取得了突飞猛进的发展. 由卷积神经网络(convolutional neural network, CNN)^[3]衍生出的各种网络结构被先后提出来, 并引起了国内外专家学者的广泛关注, 卷积神经网络已经广泛应用到计算机视觉^[4]、自然语言处理^[5]、语音识别、信息检索、推荐系统、多媒体等多个领域, 在工业界和学术界掀起了神经网络研究的浪潮, 促进了人工智能的发展.

在深度学习中, 图像风格转换最早是基于卷积神经网络来实现的, 但由于对训练样本要求较高而且训练速度慢, 所以得到的图像转换效果并不理想. 在 2014 年, 由 Goodfellow 等人提出的生成对抗网络(generative adversarial networks, GAN)^[6]以其强大的数据生成能力受到广泛关注, 成为人工智能领域的一项重要研究成果. 尤其是 GAN 在图像分辨率^[7-9]、图像压缩、图像风格转换、文本到图像的生成^[10]、视觉计算、语音和语言处理等方面都有很突出的表现且有着巨大应用前景, GAN 成了计算机视觉和图像处理领域炙手可热的研究热点之一. GAN 的提出, 在图像风格转换领域有了突出的表现, 可以用较小的数据集来完成 GAN 的无监督训练. GAN 采用监督学习的方法来做无监督学习任务, 使用判别器监督学习, 但最终使用生成器学习得到真实数据分布或密度的预估生成新的图像. 针对于图像风格转换, 人们将越来越多的注意力放到了 GAN 上, 它是目前生成模型中最好的网络架构, 并且衍生出了一系列优秀的图像风格转换模型, 如 StyleGAN^[11], Pix2Pix^[12], CycleGAN^[13], DiscoGAN^[14], DualGAN^[15]等. GAN 是由一个生成器和一个鉴别器两部分组成, 用于学习真实样本数据的概率分布, 两者都是在对抗博弈的思想下进行训练. 生成器从输入的噪声中捕捉真实数据样本的潜在分布, 努力生成让判别器认为是真的假图片, 判别器会尽最大努力判别输入图片的真假. 二者通过训练不断优化提高各自的生成能力和判别能力, 直至达到纳什均衡^[16]训练结束, 即判别器无法判别生成器生成的以假乱真的图片.

在实际的生产生活中, 图像风格的转换有着广泛的应用, 比如手机滤镜、图像艺术化、卡通动画制作^[17]、在线隔空试衣购物、试妆卸妆、去遮挡^[18]、样本数据集扩充等. 人脸面部表情^[19]的研究在计算机视觉、认知科学领域有着非常重要的意义, 在娱乐、社交领域、人脸识别领域都有广泛的应用. 但由于人脸表情数据集过少, 制约了深度神经网络的表达能力, 导致在实际训练模型过程中精度不足、图像的局部细节不明显, 可以通过图像风格转换扩展数据集. 在煤矿井下, 由于光照条件差、粉尘较多, 监控视频设备采集到的图像往往存在模糊、亮度低、细节不清晰的问题, 可以通过图像风格转换增强图像的分辨率, 提升图片亮度, 为矿井的安全提供便利. 这些图像风格转换的应用不仅推动了科学技术的发展, 还提高了人们的生活水平, 关键是大大减少了人力物力财力成本. 目前, 图像风格转换方面的工作已经取得了飞跃性的进步, 并且产生了不错的效果, 但是仍然存在一些缺陷与不足, 有着十分大的空间等待我们去探索. 因此, 利用生成对抗网络进行图像风格转换非常具有研究价值.

本文的主要贡献如下:

- 1) 在 StarGAN 模型的基础上, 结合 U-Net 网络中的跳跃连接和边缘损失函数, 提出了 UE-StarGAN 模型用于图像风格转换;
- 2) 引入了类别编码器到 UE-StarGAN 中, 提出了融合类别编码器的小样本图像风格转换模型 CUE-StarGAN, 接下来引入 Mish 激活函数对融合模型进一步优化.

1 相关工作

基于生成对抗网络的图像风格转换, 由于生成对抗网络强大的生成能力并且在图像风格转换上有着不错的效果, 所以这是目前比较常用且流行的一种风格转换方法^[20]. 此方法与传统的卷积神经网络的图像风格转换不同, 该方法可以接受一整类的图像输入, 从大量的输入图像中学习它们的数据分布和共同特征, 并基于这些特征生成尽可能与这些图像相似的图像. 而不再像传统的方法只输入一张图像学习其风格, 大大提高了

图像风格转换的效率.

Li 等人^[21]在生成对抗网络中引入了马尔可夫随机场, 通过对抗训练的方式来训练生成模型, 增强了生成图像的真实感. 随后衍生出一系列基于生成对抗网络的图像风格转换模型, 如 DiscoGAN, DualGAN, Pix2Pix, CycleGAN, StyleGAN 等. 其中, DiscoGAN 和 DualGAN 基于机器翻译的对偶学习思想. 2017 年, Kim 等人^[14]为了寻求跨域之间的转换关系以及解决模型崩溃的问题, 提出了采用对称结构的思想完成图像风格转换的 DiscoGAN 模型; 接着, Yi 等人^[15]提出了基于对偶学习和 L1 范数的无监督图像风格转换 DualGAN. 2017 年, Phillip 等人^[12]提出了 Pix2Pix 有监督图像风格转换模型, 在原始 GAN 的基础上加入了条件来控制生成图像的风格, 实现了马到斑马、苹果到橘子、白天到黑夜、灰度图到彩色图等图像风格转换效果. Pix2Pix 对 DCGAN 的生成器和判别器做了改进, 利用 U-Net 提升细节, 并且利用 PatchGAN 来处理图像的高频部分, 缺点是训练数据必须是成对的, 因此需要大量的成对图片用于网络模型训练. 但是在很多情况下, 成对的数据集很难获取到. 2018 年, Zhu 等人^[13]提出了 CycleGAN 图像风格转换模型, 利用循环一致性损失解决了训练数据必须成对的问题. 该网络不需要成对的数据, 只需要输入和输出的两个数据集即可. CycleGAN 首先将图像从 A 域转换到 B 域, 然后再将 B 域转换到 A 域, 通过这样一个循环将转换前后的图片进行配对. 这就类似于有监督学习, 能够有效地提升转换效果. 2019 年, Karras 等人^[11]提出的 StyleGAN 主要改变了生成器的结构, 实现无监督地生成可控性强的图像. 该模型并没有追求生成更加逼真的图像, 而是致力于提高网络对生成图像的精确控制能力. 在生成对抗网络的基础上不断改进, 衍生出了这些优秀的对抗训练模型, 从输入数据、网络结构、损失函数等方面不断改进, 成功实现了无监督的图像风格转换.

2 基础理论

2.1 StarGAN

2018 年, 由 Choi 等人^[22]提出的 StarGAN 网络模型可以用于解决多域之间图像风格转换问题, 它只需要训练一个生成器便可以完成多域图像风格转换. StarGAN 的网络结构如图 1^[22]所示, 与传统的生成对抗网络结构相似, 由生成器 G 和判别器 D 两个模块组成. 其中, 图 1(b)过程表示将源域图像和目标域标签输入到生成器中生成假图像. 然后通过图 1(c)将假图像和原始域标签作为输入重建生成原始图片.

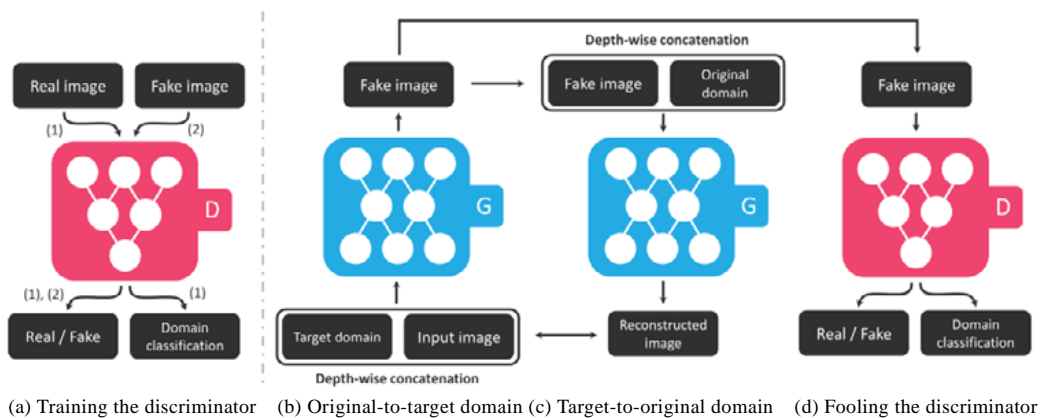


图 1 StarGAN 网络结构^[22]

生成器尽最大努力生成与真实图像无法区分的图像, 判别器 D 用于辨别图像的真假, 并将其分类为目标域. 对抗损失函数如公式(1)所示:

$$L_{adv}(G,D)=E_x[\log D_{src}(x)]+E_{x,c}[\log(1-D_{arc}(G(x,c)))] \quad (1)$$

其中, x 为输入图像, c 为目标域标签. 生成器根据目标域标签 c 生成图片. 图 1(c)过程的重建损失函数如公式(2)所示:

$$L_{rec} = E_{x,c,c'}[\|x - G(G(x,c),c')\|_1] \quad (2)$$

其中, c' 为源域标签, 根据生成的假图片和源域标签重建源域图片. 判别器 D 学习辨别真假图像, 与此同时, 针对输入的多种类别图像, 判别器还负责将输入的图像分类到相对应的域中. 判别真假的损失和域分类损失如公式(3)和公式(4)所示:

$$L_{cls}^f = E_{x,c}[-\log D_{cls}(c | G(x,c))] \quad (3)$$

$$L'_{cls} = E_{x,c'}[-\log D_{cls}(c' | x)] \quad (4)$$

最终, 生成器分别由对抗损失、重建损失和分类损失组成, 判别器由对抗损失和分类损失组成, 分别如公式(5)和公式(6)所示:

$$L_G = L_{adv} + \lambda_{cls} L_{cls}^f + \lambda_{rec} L_{rec} \quad (5)$$

$$L_D = -L_{adv} + \lambda_{cls} L'_{cls} \quad (6)$$

其中, λ_{cls} , λ_{rec} 为超参数.

2.2 U-Net和边缘损失函数

U-Net 最大的特点就是跳跃连接, 将特征图在通道维度进行拼接形成更厚的特征. U-Net 融合浅层特征的时候, 并不是一个简单的求和操作而是采用叠加的方法. 它有大量的特征通道, 这些通道允许网络将浅层范围信息传播到高分辨率层. 上采样可以填补图片信息但仍然会丢失部分信息, 所以需要与左边的高分辨率特征图连接. 上采样过程中, 右边的特征图越来越抽象, 因为随着卷积层数的增多, 提取到的特征图片更加高效和抽象, 所以就与左边的高分辨率特征图进行连接, 这相当于在高分辨率特征图和更抽象特征之间做了一个折中.

通过跳跃连接的操作 U-Net 融合了输入图像的很多细节信息, 有助于还原下采样所带来的信息损失问题. 上采样过程中, 通过跳跃连接融合下采样输出的特征图. 从另一个角度来看, 这是将不同尺寸的特征融合在了一起, 也属于多尺度特征融合. 比如最后一个上采样, 它的特征既有第 1 个卷积层的输出, 也有上采样的大尺度输出, 整个网络中共有 4 次融合过程.

边缘损失函数为了保留清晰的图像边缘而提出. 在整个训练过程中, 判别器 D 的任务是判定输入图像的真假. 然而, 通过实验观察到, 仅仅训练判别器 D 来区分生成的图像和真实的图像是不够的. 因为在图像风格转换过程中, 清晰的边缘也是重要特征, 但这些边缘在整个图像中所占的比例通常非常小. 因此, 对于有正确的底纹输出但没有清晰再现图像边缘的情况, 可能会混淆采用标准损失函数训练的判别器. 所以, 针对判别器 D 提出了边缘损失函数. 首先, 需要对原始的输入图像做一个简单的去边预处理, 将处理过的照片输入判别器进行判别, 可以提高判别器对图像边缘的判别能力. 判别器检测到生成图片边缘的异常, 会将这种情况反馈给生成器, 生成器通过迭代优化可以生成更高质量的照片. 得到的边缘损失函数如公式(7)所示:

$$L_{edge} = E_{x'}[\log(1 - D(x'))] \quad (7)$$

其中, x' 是原始数据集经过去边处理后的输入图像. 最终, 判别器 D 的损失函数如公式(8)所示, 由对抗损失函数、分类损失函数和边缘损失函数这 3 部分组成:

$$L_D = -L_{adv} + \lambda_{cls} L'_{cls} + \lambda_{edge} L_{edge} \quad (8)$$

其中, λ_{cls} , λ_{edge} 是域分类损失和边缘损失的超参数. 分类损失函数 L'_{cls} 如公式(9)所示:

$$L'_{cls} = E_{x,c'}[-\log D_{cls}(c' | x)] \quad (9)$$

其中, x 是输入的真实图像, c' 是源域标签. 判别器 D 在源域标签引导下, 将输入图像恰当地分类到对应的域.

2.3 类别编码器

在图像风格转换过程中, 生成器提取特征图各个通道的均值和方差, 最终影响生成图像的风格.

UE-StarGAN 的生成器采用的实例归一化, 就是对 H , W 做归一化求得均值和方差, 提高图像风格转换效果, 加速模型的收敛. 但是 UE-StarGAN 只局限于源域图像之间的风格转换并且需要大量的训练样本, 所以我

们结合类别编码器来实现小样本的图像风格转换. 类别编码器的作用就是将一组 N 类图像映射到类潜在编码, 类潜在编码解码为均值和方差输入到生成器中来影响图像风格转换效果.

类别编码器将一组目标图片作为输入, 它可以利用输入的各种不同类的少量目标图片和生成器一起来训练模型. 也就是说, 可以通过类别编码器来提取图像的风格, 生成器用于提取输入源图像本身的纹理结构和内容, 最后将源图像转换为不同风格的目标类图像. 类别编码器的网络结构如图 2 所示, 类别编码器由 5 个卷积层和一个平均池化层组成, 每个卷积层后都有一个 ReLU 激活函数. 类别编码器首先将 N 个类别的图像分别映射成 N 个中间潜在变量, 然后对这 N 个中间潜在变量进行平均, 以产生最终类潜在编码 Z_y .

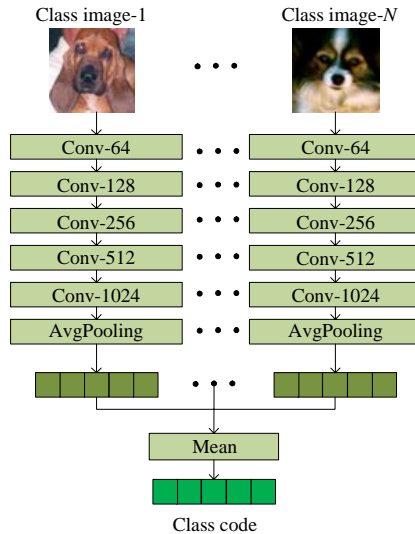


图 2 类别编码器网络结构

为保证生成图像具有目标图像的特点, 还提出了一个特征匹配损失来保证输入和输出的相似性. 特征匹配损失函数如公式(10)所示:

$$L_{FM}(G) = E_{x, y_1 \dots y_k} \left[\left\| F(\bar{x}) - F\left(\sum_k \frac{F(y_k)}{k}\right) \right\| \right] \quad (10)$$

其中, x 是输入的源图像, \bar{x} 是输出图像, $y_1 \dots y_k$ 为输入的目标类图像, F 表示特征提取器. 采用 L1 正则化最小化特征损失来保证输出图像和目标图像的相似性.

3 所提模型和算法

3.1 基于改进StarGAN的图像风格转换模型

针对图像风格转换以后背景颜色失真和图像边缘细节模糊的问题, 在 StarGAN 的基础上构建了一种基于 U-Net 和边缘损失函数的图像风格转换模型(image style transferring model based on U-Net and edged-smoothed, UE-StarGAN). 整个生成对抗网络模型包含生成器和判别器两部分. 借鉴了 U-Net 网络模型中的跳跃连接的思想, 对 StarGAN 中的生成器进行改进, 同时在判别器中引入了边缘损失函数来提高它对生成图像边缘细节的判别能力. 改进后的生成器网络结构如图 3 所示.

如图 3 所示, 生成器由 4 个卷积层、6 个残差模块和 4 个反卷积层组成, 通过跳跃连接将上采样中的特征图和反卷积特征图融合, 有助于弥补还原下采样所带来的信息损失. 生成器的对抗损失函数如公式(11)所示:

$$L_{adv}(G, D) = E_x[\log D_{src}(x)] + E_{x,c}[\log(1 - D_{src}(G(x, c)))] \quad (11)$$

其中, x 为输入的源域图像, c 是目标域标签. 生成器在目标标签的引导下, 可以将输入图像转换为不同风格化的图像.

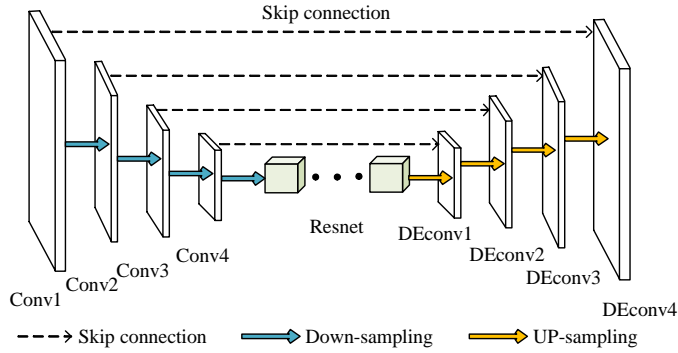


图3 UE-StarGAN 生成器网络结构

3.2 融合类别编码器的小样本图像风格转换模型

无监督图像风格转换模型在跨域风格转换的改进上非常成功,但是从基于新类的少数样本中学习图像风格还是很受限的,因为从自身的先验知识进行推断完全超出了它们的能力范围.它们需要从几万或几十万张新类别图像中学习样本数据的分布,进而生成目标类的风格化图像,所以它们不支持小样本的图像风格转换.针对此问题,本文在 UE-StarGAN 的基础上,构建了一个融合类别编码器的图像风格转换模型(image style transfer model merging class encoders, CUE-StarGAN).整个网络模型包含生成器和判别器两部分.

借鉴类别编码器中的类别编码思想,对 UE-StarGAN 中的生成器进行改进. CUE-StarGAN 的生成器网络结构如图4所示,生成器可以看成由 UE-StarGAN 生成器和类别编码器两部分组成.将源类图像和目标类图像作为输入图像到生成器网络中.源类图像经过上采样提取到了源类图像的主要内容特征,目标类图像经过类别编码器提取到具有一定特征风格的类别潜在编码.最后,通过残差模块将类别潜在编码送入到解码器以控制图像风格.残差模块由自适应实例标准化(adaptive instance normalization, AdaIN)和残差块组成,其中,AdaIN 将每个通道中的样本数据标准化为零均值和单位方差,然后通过仿射变换获取全局外观信息以控制生成图像的风格.

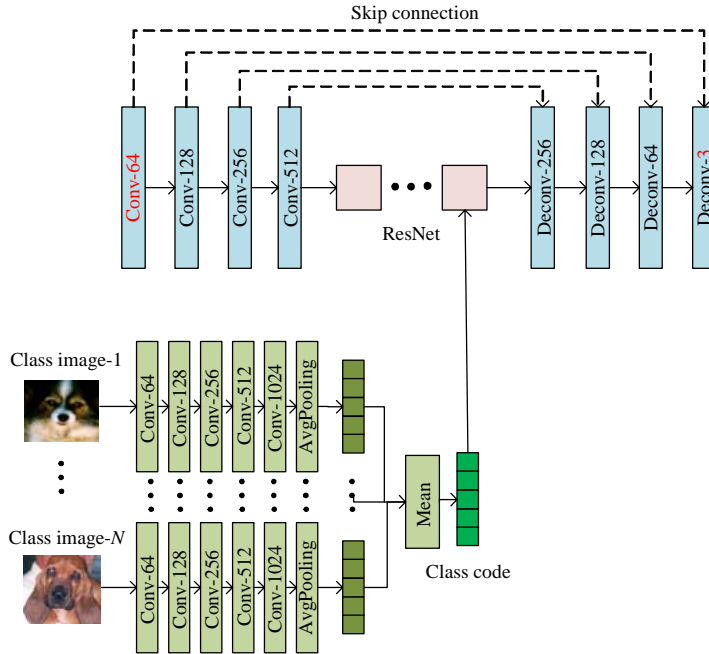


图4 CUE-StarGAN 的生成器网络结构

生成器的损失函数分别由对抗损失函数、重建损失函数和特征匹配损失函数这 3 部分组成, 其中, 对抗损失函数和重建损失函数如公式(12)和公式(13)所示:

$$L_{adv} = E_x[\log D(x)] + E_{x, y_1 \dots y_k}[\log(1 - D(\bar{x}))] \quad (12)$$

$$L_R = E_{x, c}[\|x - G(\bar{x}, c')\|_1] \quad (13)$$

其中, x 是输入的源图像, \bar{x} 是输出图像, $y_1 \dots y_k$ 为输入的目标类图像, c' 是源域标签. 最终生成器的对抗损失函数如公式(14)所示:

$$L_G = L_{adv} + \lambda_R L_R + \lambda_{FM} L_{FM} \quad (14)$$

其中, λ_R , λ_{FM} 分别为重建损失函数和特征匹配损失函数的超参数.

4 实验结果和分析

4.1 数据集

在本文实验和测试过程用到的数据集分别是 CelebA^[23], Fer2013 和 ImageNet 中的 Animal Faces 和 North American Birds 数据集. CelebA (Celeb faces attribute)数据集是人脸识别和人脸表情研究领域具有权威性和完整性的名人人脸属性数据集, 其中包括超过 20 万张不同属性的人脸图片, 共计超过 1 万位名人. 并且数据集中的每张图片都做了特征标注, 包含 40 种人脸属性标记、人脸 bbox 标注框、人脸特征的 5 个点坐标. Fer2013 人脸表情数据集是一个灰度图数据集, 主要用于人脸表情变化的研究. 该数据集共有 7 种表情, 分别对应于数字标签 0~6. 这 7 种表情图片共有 35 886 张, 其中, 用于测试的图片有 28 708 张, 公共验证图片和私有验证图片各 3 589 张, 每张图片大小固定为 48×48. Animal Faces 和 North American Birds 数据集是 ImageNet 数据集中的动物脸和北美鸟数据集. ImageNet 数据集是由斯坦福大学李飞飞教授带领创建, 涵盖了超过 2 万多个种类、1 400 万张图片的数据集.

4.2 评价指标

图像风格转换的常用评价指标^[24]有 PSNR^[25], SSIM^[26], IS, FID. PSNR 和 SSIM 直接对图像本身的质量进行量化, IS 和 FID 通常用于评价图像的质量和多样性. 本文选用以上 4 个作为实验的定量评价指标.

PSNR (peak signal-to-noise ratio)即峰值信噪比, 用于评价和衡量两幅图像的失真情况和噪声大小. PSNR 值越大, 就代表失真越少, 生成的图像就越逼真. SSIM (structural similarity)即结构性相似指数, 主要是从结构、对比度和亮度这 3 个方面来衡量两幅图像的相似度. SSIM 越大, 说明生成图像在结构、亮度、对比度方面都更真实.

IS (inception score)即初始分数, 从图像的清晰度和多样性评价生成图像的质量, 只使用生成数据的相关信息. 初始分数值越大, 表明生成的图片多样性越好, 质量越好. FID (Fréchet inception distance)即弗雷歇距离, 从图像的多样性和质量上来衡量网络的性能, FID 主要是通过计算真实数据与生成数据在特征层上的距离来判定生成图片的质量. FID 数值越小, 说明生成数据与真实数据越接近, 生成的图片质量更高.

4.3 实验结果

为了验证所提模型的有效性, 本文实验在深度学习平台 AIOS 上完成. AIOS 操作系统为 Linux, 支持 RTX 2080 Ti, 4 核 CPU, 16GB 的 GPU, 内存 24GB, 支持 PyTorch 框架进行 GPU 加速计算. 在训练过程中, 批处理大小为 16, 迭代次数为 200 000, 生成器中使用实例归一化(instance normalization, IN), 生成器和判别器的学习率均为 0.0001, 采用 Adam 优化器进行训练.

首先, 选用 CelebA 和 Fer2013 数据集对第 3.1 节所提出的 UE-StarGAN 模型进行训练; 然后, 分别从 CelebA 和 Fer2013 数据集中随机选取 2 000 张图片用于测试; 然后, 在不同数据集上与其他图像风格转换模型算法 Pix2Pix, CycleGAN, StarGAN 作比较. 选用了 SSIM 和 PSNR 作为评价指标, 在两个数据集上得到的测试结果的平均值见表 1 和表 2.

表 1 在 CelebA 数据集上 SSIM 和 PSNR 实验结果对比

CelebA	SSIM	PSNR (dB)
Pix2Pix	0.767	21.463
CycleGAN	0.749	20.686
StarGAN	0.788	22.752
UE-StarGAN	0.881	25.653

表 2 在 Fer2013 数据集上 SSIM 和 PSNR 实验结果对比

Fer2013	SSIM	PSNR (dB)
Pix2Pix	0.834	25.085
CycleGAN	0.859	24.107
StarGAN	0.866	25.882
UE-StarGAN	0.879	26.262

表 1 和表 2 给出了在 CelebA 数据集上戴眼镜、金色头发、刘海、性别、年龄和 Fer2013 数据集上正常、生气、微笑、悲伤、惊讶这 5 种不同风格下的结构相似性指数(SSIM)和峰值信噪比(PSNR). UE-StarGAN 模型的 SSIM 和 PSNR 值均高于其他 4 种模型. 验证了 UE-StarGAN 模型性能整体上优于其他的图像风格转换模型. 在 Fer2013 数据集上的可视化对比结果如图 5 所示.

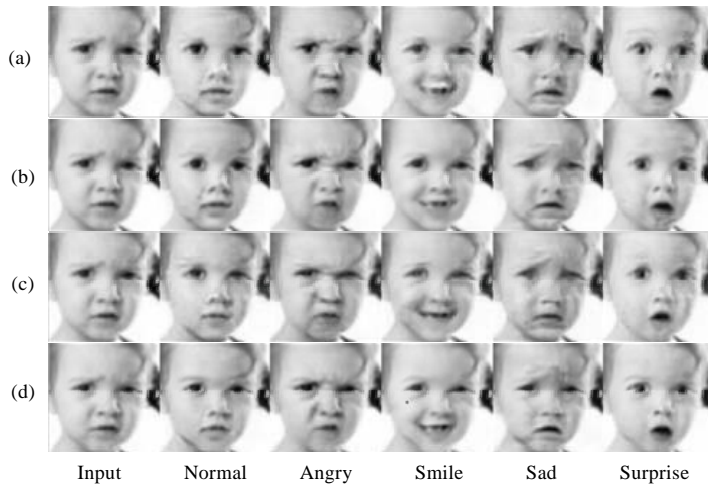


图 5 在 Fer2013 数据集上的对比结果

图 5(a)代表 pix2pix 模型、图 5(b)代表 CycleGAN 模型、图 5(c)代表 StarGAN 模型、图 5(d)代表本文提出的 UE-StarGAN 模型. 从横向来看, 图 5(d)即本文所提出的 UE-StarGAN 模型生成的图像分辨率更高, 在保留面部基本内容信息的基础上, 转换过后的表情更自然、效果更好. 然后是 Pix2Pix 模型较其他 3 种模型转换的效果不是很好, 尤其是嘴巴和眼睛转换过后都变的模糊, 在细节的处理上比较粗糙. 其次, 纵向观察表情为 Angry, Sad 和 Surprised 这些列, 图 5(a)–图 5(c)图像在眼睛、鼻子或者嘴巴等地方也都出现了模糊, 而本文所提方法生成图片则比较清晰, 在细节特征和边缘处上比较细腻.

为了验证第 3.2 节所提 CUE-StarGAN 模型的性能, 将模型在 Amine 和 ImageNet 数据集中的 Animal Faces 和 North American Birds 子数据集上训练并测试. 然后在不同数据集上与其他图像风格转换模型算法 CycleGAN, UNIT, StarGAN, UE-StarGAN 做比较. 选用了 IS 和 FID 作为评价指标: IS 是衡量生成模型的个体特征和整体特征的指标, IS 越大, 说明生成的图像越逼真, 转换效果更好; FID 是一个具有原则和综合性的指标, FID 越低, 说明生成图片的质量越高, 图像多样性更丰富. 在两个数据集上得到的测试结果如表 3 和表 4 所示.

由表 3 和表 4 可知, 在 Animal Faces 和 North American Birds 数据集上, 与经典的 CycleGAN, UNIT 和 StarGAN 模型对比. 从表中结果可以看到: 第 3.1 节提出的 UE-StarGAN 算法的性能优于原始 StarGAN 模型;

其次,第 3.2 节所提 CUE-StarGAN 模型的 IS 值和 FID 值相比于其他模型,IS 指标的值有较高的提升,FID 比其他算法都要低,特别是融合了类别编码器的 CUE-StarGAN 有了较高的提升效果.说明本文算法可以在小样本的条件下生成更逼真的图像以及风格转换过后图像多样性更丰富,验证了 CUE-StarGAN 模型的有效性.

表 3 在 Animal Faces 上的对比结果

Animal Faces	IS	FID
CycleGAN ^[27]	7.43	197.13
UNIT ^[27]	12.14	197.13
StarGAN ^[27]	6.21	198.07
UE-StarGAN	8.96	186.78
CUE-StarGAN	13.75	165.49

表 4 在 North American Birds 上的对比结果

North American Birds	IS	FID
CycleGAN ^[27]	25.28	215.30
UNIT ^[27]	28.28	203.83
StarGAN ^[27]	18.94	260.04
UE-StarGAN	23.76	230.5
CUE-StarGAN	37.43	197.86

接下来展示在 ImageNet 上的 Animal Faces 和 North American Birds 子数据集上的可视化效果,如图 6 和图 7 所示.



图 6 模型在 ImageNet 的 Animal Faces 数据集上的可视化



图 7 模型在 ImageNet 的 North American Birds 数据集上的可视化

图 6 和图 7 中的图(a)和图(b)是输入的分类图像,图(c)为输入的源图像,图(d)是本文所提算法 CUE-StarGAN 转换后的效果图.可以从图 6 中看到: CUE-StarGAN 能够将柴犬转换为白色的宠物狗,且基本保留了柴犬整体内容信息,而在风格上与类别图像相似.图 7 中基本保留了猫头鹰整体内容信息,且在风格上与类别图像相似,将猫头鹰转换为带有北美鸟风格的图片.以上可视化结果进一步验证了所提模型的有效性,可以实现用少量的类别样本完成图像风格的转换.

5 结 论

本文从细节特征提取和小样本两方面入手,以生成对抗网络为基础,对经典的图像风格转换模型 StarGAN 进行改进.首先引入了 U-Net 模块中的跳跃连接,并与 StarGAN 模块相结合进行特征提取生成图像,并在判别器中加入边缘损失函数,使得判别器在处理图像边缘特征时具有更强的判断能力,提出了基于改进

StarGAN 的图像风格转换模型 UE-StarGAN; 其次, 在 UE-StarGAN 的基础上研究了类别编码器, 通过类别编码器学习输入类别图像的风格特征, 利用少量的类别图像完成图像风格的转换, 提出了一种基于类别编码器的小样本图像风格转换模型 CUE-StarGAN. 实验结果表明: 该模型可以提取到更精细的特征, 在小样本的情况下具有一定的优势, 可以较好地实现源图像到类别图像的风格转换. 以此进行图像风格转换后的图片, 无论是定性分析还是定量分析都有一定的提升, 验证了所提模型的有效性.

References:

- [1] Ligeza A. Artificial intelligence: A modern approach. *Applied Mechanics & Materials*, 2009, 263(2): 2829–2833.
- [2] Pollack ME. Artificial intelligence—A modern approach (a review). *AI Magazine*, 1995, 16: 73–74.
- [3] Zhou FY, Jin LP, Dong J. Review of convolutional neural network. *Chinese Journal of Computers*, 2017, 40(6): 1229–1251 (in Chinese with English abstract).
- [4] Lu HT, Zhang QC. Applications of deep convolutional neural network in computer vision. *Journal of Data Acquisition and Processing*, 2016, 31(1): 1–17 (in Chinese with English abstract).
- [5] Li S. Research and development of natural language processing. *Journal of Yanshan University*, 2013, 37(5): 377–384 (in Chinese with English abstract).
- [6] Goodfellow I, Pouget-Abadie J, Mirza M, *et al.* Generative adversarial nets. In: Ghahramani Z, ed. *Proc. of the Advances in Neural Information Processing Systems*. MIT Press, 2014. 2672–2680.
- [7] Zhang D, Shao J, Hu G, *et al.* Sharp and real image super-resolution using generative adversarial network. In: *Proc. of the Int'l Conf. on Neural Information Processing*. Cham: Springer International Publishing, 2017. 217–226.
- [8] Ouyang N, Liang T, Lin LP. Self-attention network based image super-resolution. *Journal of Computer Applications*, 2019, 39(8): 2391–2395 (in Chinese with English abstract).
- [9] Gao Y, Liu Z, Qin PL, *et al.* Medical image super-resolution algorithm based on deep residual generative adversarial network. *Journal of Computer Applications*, 2018, 38(9): 2689–2695 (in Chinese with English abstract).
- [10] Reed S, Akata Z, Yan X, *et al.* Generative adversarial text to image synthesis. *JMLR.org*. 2016.
- [11] Karras T, Laine S, Aila T. A Style-based generator architecture for generative adversarial networks. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. IEEE, 2019. 4401–4410.
- [12] Phillip I, Zhu JY, Zhou TH, *et al.* Image-to-image translation with conditional adversarial networks. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. IEEE, 2017. 5967–5976.
- [13] Zhu JY, Park T, Isola P, *et al.* Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv: 1703.10593*, 2017.
- [14] Kim T, Cha M, Kim H, *et al.* Learning to discover cross-domain relations with generative adversarial networks. In: *Proc. of the 34th Int'l Conf. on Machine Learning (ICML)*. IMLS, 2017. 2941–2949.
- [15] Yi Z, Zhang H, Gong PTM. DualGAN: Unsupervised dual learning for image-to-image translation. *arXiv: 1704.02510*, 2017.
- [16] Ratliff LJ, Burden SA, Sastry SS. Characterization and computation of local Nash equilibria in continuous games. In: *Proc. of the 51st Annu. Allerton Conf. on Communication, Control, and Computing (Allerton)*. 2013. 917–924.
- [17] Chen Y, Lai YK, Liu YJ. CartoonGAN: Generative adversarial networks for photo cartoonization. In: *Proc. of the IEEE/CVF Conf. on Computer Vision & Pattern Recognition*. IEEE, 2018. 9465–9474.
- [18] Qian R, Tan RT, Yang W, *et al.* Attentive generative adversarial network for raindrop removal from a single image. In: *Proc. of the 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2018. 2482–2491.
- [19] Yao NM, Guo QP, Qiao FC, *et al.* Robust facial expression recognition with generative adversarial networks. *Acta Automatica Sinica*, 2018, 44(5): 865–877 (in Chinese with English abstract).
- [20] Chang JY. Image style transferring based on generative adversarial network [MS. Thesis]. Xuzhou: China University of Mining and Technology, 2021 (in Chinese with English abstract).
- [21] Li C, Wand M. Precomputed real-time texture synthesis with Markovian generative adversarial networks. In: *Proc. of the European Conf. on Computer Vision*. Cham: Springer, 2016. 702–716.

- [22] Choi Y, Choi M, Kim M, *et al.* StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation. arXiv: 1711.09020, 2017.
- [23] Yao Z, Zhang BY, Wang ZY. IntersectGAN: Learning domain intersection for generating images with multiple attributes. In: Proc. of the ACM Int'l Conf. ACM, 2019. 1842–1850.
- [24] Wang Z, Bovik AC, Sheikh HR, *et al.* Image quality assessment: From error visibility to structural similarity. IEEE Trans. on Image Processing, 2004, 13(4): 600–612.
- [25] Huynh-Thu Q, Ghanbari M. Scope of validity of PSNR in image/video quality assessment. Electronics Letters, 2008, 44(13): 800–801.
- [26] Zhu XS, Yao SR, Sun B, *et al.* Image quality assessment: Combining the characteristics of HVS and structural similarity index. Journal of Harbin Institute of Technology, 2018, 50(5): 121–128 (in Chinese with English abstract).
- [27] Liu M Y, Huang X, Mallya A, *et al.* Few-shot unsupervised image-to-image translation. In: Proc. of the 2019 IEEE/CVF Int'l Conf. on Computer Vision (ICCV). IEEE, 2019. 10550-10559.

附中文参考文献:

- [3] 卢宏涛, 张秦川. 深度卷积神经网络在计算机视觉中的应用研究综述. 数据采集与处理, 2016, 31(1): 1–17.
- [4] 李生. 自然语言处理的研究与发展. 燕山大学学报, 2013, 37(5): 377–384.
- [5] 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述. 计算机学报, 2017, 40(6): 1229–1251.
- [8] 欧阳宁, 梁婷, 林乐平. 基于自注意力网络的图像超分辨率重建. 计算机应用, 2019, 39(8): 2391–2395.
- [9] 高媛, 刘志, 秦品乐, 等. 基于深度残差生成对抗网络的医学影像超分辨率算法. 计算机应用, 2018, 38(9): 2689–2695.
- [19] 姚乃明, 郭清沛, 乔逢春, 等. 基于生成式对抗网络的鲁棒人脸表情识别. 自动化学报, 2018, 44(5): 865–877.
- [20] 常建英. 基于生成对抗网络的图像风格转换 [硕士学位论文]. 中国矿业大学, 2021.
- [26] 朱新山, 姚思如, 孙彪, 等. 图像质量评价: 融合视觉特性与结构相似性指标. 哈尔滨工业大学学报, 2018, 50(5): 121–128.



许新征(1980—), 男, 博士, 教授, CCF 高级会员, 主要研究领域为机器学习, 数据挖掘, 模式识别.



丁世飞(1963—), 男, 博士, 教授, CCF 杰出会员, 主要研究领域为机器学习, 数据挖掘, 模式识别.



常建英(1996—), 女, 硕士生, 主要研究领域为深度学习, 计算机视觉.