刘鹤,周勇,潘翼,等. 结合多尺度特征和多维注意力的人脸风格转换[J]. 江西师范大学学报(自然科学版),2023,47(1):69-76.
LIU He,ZHOU Yong,PAN Yi,et al. The face style conversion combining multiscale feature fusion and multi-dimensional attention[J]. Journal of Jiangxi Normal University(Natural Science),2023,47(1):69-76.

文章编号:1000-5862(2023)01-0069-08

结合多尺度特征和多维注意力的人脸风格转换

刘 鹤,周 勇*,潘 翼,张金桃
(江西师范大学计算机信息工程学院,江西 南昌 330022)

摘要: 针对 StarGANv2 模型生成的人脸图像存在风格重建效果不佳、人脸纹理不够自然等现象,该文提出结合多尺度特征和多维注意力的人脸风格转换模型. 1) 将多尺度特征融合模块 PSConv 嵌入 StarGANv2 生成器内,提高了模型对图像特征的提取能力;2) 提出了多维注意力模块 MDConv,并将该模块嵌入 StarGANv2 判别器内,从而提高了模型对真假人脸图像的判别能力. 与 StarGANv2 方法在 CelebA-HQ 数据集上进行对比实验的结果表明:该方法生成的人脸图像风格更美观,纹理细节更自然,学习感知图像相似度(LPIPS)的值也得到了提升.

关键词: 人脸风格转换;人脸属性合成;多尺度特征融合;多维注意力
中图分类号: TP 391.4 **文献标志码:** A **DOI:** 10.16357/j.cnki.issn1000-5862.2023.01.09

0 引言

人脸作为身份识别的关键信息,在计算机视觉领域中引起了广泛的关注. 近年来,人脸合成技术[1]发展迅速,在影视娱乐、公安侦查、虚拟现实等应用领域中发挥了较大的作用. 在人脸合成技术中,人脸风格转换源于人脸属性合成,是对人脸的若干属性进行编辑,转换其表现形式,其中属性指具有明确语义信息的特征,如发型、肤色、年龄等. 最初对人脸属性合成往往局限在年龄修改(如人脸老化等[2]),随着生成对抗网络(generative adversarial networks,GAN)[3]的发展,可实现对指定的属性转换.

Li Mu 等[4]构建了属性转换网络和图像增强网络,在保留输入人脸身份前提下,实现了面部属性的平滑合成. Liu Rujie 等[5]借助残差网络 ResNet[6]的思想,通过学习输入图像和生成图像的差异来实现对属性的控制,并且采用 2 个不同的生成器实现特定属性的不同表现形式的互相转换(如佩戴眼镜和未佩戴眼镜等). 但上述方法都只对单个属性进行改动,并且由于属性间具有较强的关联,无关属性容易随之改动,属性合成较难控制.

2018 年,Y. Choi 等[7]提出了 StarGAN,仅用 1 个生成器实现了多个属性之间的转换,并且加入属性分类误差,以实现无关属性的保留,但生成图像精度较差. 随后,Liu Ming 等[8]基于目标属性标签和源属性标签之间的差异,利用选择性传输单元的跳跃连接,显著提高了生成图像的精度和视觉效果. 为了挖掘隐式空间和人脸属性之间的内在联系,Shen Yujun 等[9]提出了 InterFaceGAN,通过子空间投影对人脸属性特征进行解耦,顺利完成了语义化的人脸编辑. 2021 年,Yang Guoxing 等[10]在样式转换器中引入了正交性约束,将属性相关的样式代码与不相关的样式代码分离,取得了较好的属性转换效果. Wang Huipo 等[11]通过迭代遍历非线性潜在空间实现了更平滑的属性转换. 2022 年,S. Khodadadeh 等[12]采用神经网络来改变潜在空间内属性编码,极大程度地保留了输入人脸的身份特征和其他无关

收稿日期:2022-11-23
基金项目:江西省教育厅科学技术研究基金(KJLD14021)和江西省教育厅重点教改课题(JXJG1821)资助项目.
通信作者:周 勇(1971—),男,江西南昌人,副研究员,主要从事数据库、数据挖掘和人工智能方面的研究. E-mail:zhou_yong@126.com

属性.

上述人脸属性合成方法实现了对人脸属性不同精细程度的控制. 为使属性合成更加多样化,借助风格迁移[13]思想,T. Karras 等[14]提出了 StyleGAN,实现了人脸风格融合,并陆续提出了改进版本[15-16],此模型虽获得了较高的生成人脸质量,但在风格混合时混合的是潜在向量生成的人脸风格,无法对指定输入人脸图像进行风格转换. 而 StarGANv2[17]将人脸风格定义为人脸的独特风格外观,可视作多个属性特征的集合,性别用"域"表示,可极大程度保留人脸身份的前提,实现了对输入人脸的域内或跨域多样化风格转换. 然而在实验中发现,StarGANv2 模型生成的图像存在风格重建效果不佳、人脸纹理细节不自然的现象.

1 相关理论

1.1 生成对抗网络

生成对抗网络作为目前最热门的生成模型,不用关注隐藏变量服从任何基础分布,仅通过生成器和判别器彼此对抗博弈进行训练,使得生成器学习到样本的分布,其基本模型结构如图 1 所示. 首先将高斯噪声 z 输入生成器 G 中,生成虚假数据 $G(z)$,然后分别将真实数据 x 和虚假数据 $G(z)$ 输入判别器 D 中,输出判别的真假情况. 其中 G 不断学习数据的分布,争取伪造出难辨真伪的数据; 而 D 的目的是要不断提高自身分辨真假的能力,当鉴别力足够强无法判断数据是真实数据还是生成数据时,就获得了一个学习到真实数据分布的生成器. 在图像转换领域中,GAN 与变分自编码器 VAE[19]、流模型 GLOW[20]等其他生成模型相比,其特有的对抗博弈形式能得到更清晰、更逼真的图像.

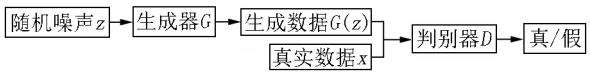

图 1 生成对抗网络基本结构

1.2 多尺度特征融合

在利用卷积神经网络进行特征提取时,为了融合不同尺度的特征信息,常常用到多尺度特征融合. 其中常见网络结构分为 2 种,一种是串行的跳层连接网络结构,如 FCN[21]、U-Net[22]等,其主要思想在于通过融合不同层级的特征图,以便综合低层位置信息和高层语义信息; 另一种是并行多分支网络结构,如 Inception[23]、MixConv[24]等,它是通过多个卷积分支分别对输入数据进行特征提取再融合,这些卷积分支可设置不同的卷积核大小、扩张率等.

在这 2 类结构中,并行结构的优势在于能在同一个卷积层中获取不同大小的感受野,在融合后传入下一层,从而在提高模型的能力和计算量中取得平衡. 如典型结构 MixConv 通过对卷积核施加不同的核大小进行分组卷积,以获取不同尺度的特征信息,但每一组的输出仍为单尺度特征. 而 PSConv(poly-scale convolution)[25]利用特征图分组交换,结合不同扩张率的扩张卷积,使得每组输出都包含多尺度的特征. 因此,本文将多尺度特征融合模块 PSConv 嵌入 StarGANv2 模型中,以实现多尺度特征融合.

1.3 多维注意力

在计算机视觉中,注意力机制的基本思想在于让系统能像人的视觉一样,学会获取需要重点关注的区域,从而提高模型的性能. 按照注意力机制应用的维度,主要分为空间注意力、通道注意力和混合注意力.

近年来,动态卷积使用越来越广泛,注意力机制在 CNN 中得到了较大发展,如 CondConv[26]和 DyConv[27],它们在 n 个卷积核上应用注意力机制,对核空间的整个卷积核赋予动态特性,其精度较高,但忽略了对其他 3 维(输入通道、卷积核空间和输出通道)的注意力关注; 全维动态卷积(ODConv)[28]改善了这个问题,利用并行卷积策略,在沿核空间的所有维度建立注意力机制,但增加了一定的计算量. 为了减少运算负担,同时获取对关键维度的注意力关注,本文提出多维动态注意力模块 MDConv(Multi-dimensional dynamic convolution),并将其嵌入 StarGANv2 模型中,以期获取图像特征的多维注意力.

2 MFMA-StarGANv2 整体结构

本文模型整体结构是基于 StarGANv2 多域风格转换模型,整体架构如图 2 所示,其中 G 是生成器,D 是判别器,E 是风格编码器,F 是映射网络.

在此模型结构中,域代表性别,分为男性和女性,可进行域内或跨域的人脸风格转换. 在 G 转换图像前,需要接收输入图像和风格码. 生成人脸图像方式分为参考图像引导转换和潜码引导转换,这 2 种转换方式的区别在于风格码的来源不同. 前者利用的风格码是通过将参考图像和对应性别(0 代表

女性,1 代表男性) 输入 E 中生成而来的,使得生成图像具备了与参考图像类似的风格; 后者利用的风格码是通过将潜码和指定性别输入 F 生成而来的,使得生成图像具有对应性别的随机风格. E 和 F 都有 2 条输出分支,在训练或测试中,E 和 F 通过输入性别来确定输出分支,从而生成目标性别的风格码.

如图 2 所示,生成图像采用了参考图像引导转换方式,实现了跨域人脸风格转换,使得生成图像具备了输入人脸图像重要的结构特征(如脸型、五官等) ,既保留了“身份 ID”,同时也具备与参考人脸图像类似的风格特征(如发型、肤色等) .

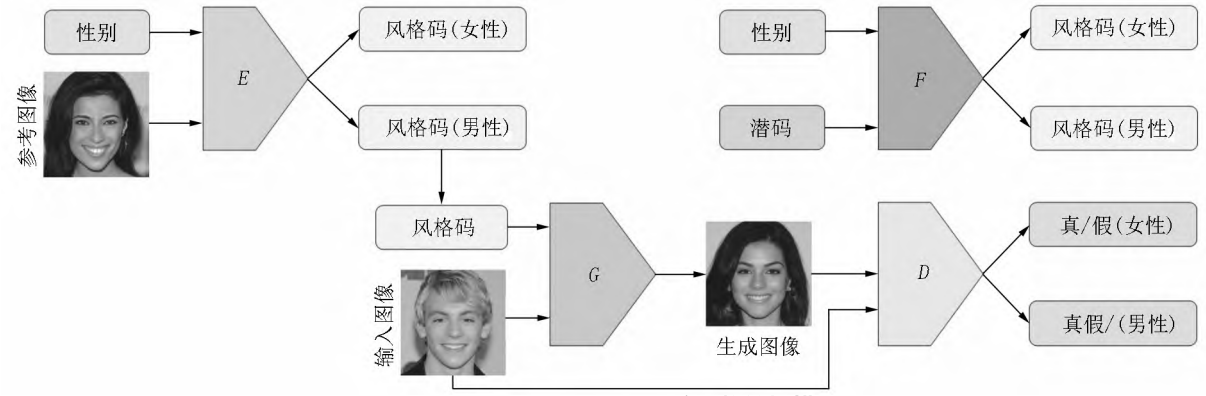

图 2　MFMA-StarGANv2 人脸风格转换模型

3　MFMA-StarGANv2 生成器

3.1　MFMA-StarGANv2 生成器结构

在进行图像转换前,生成器需要对输入图像进行下采样操作,目的是解耦人脸风格特征和人脸结构特征. 此时,融合不同尺度的人脸特征具有重要意义,因此将 PSConv 模块嵌入 StarGANv2 模型生成器中间层内,如图 3 所示. 为了对输入图像风格进行转换,下半部分中间层和上采样层借助了风格迁移方法 AdaIN [13] ,通过在每层输入内容图特征和风格图特征(在图 3 中用风格码 style 表示) ,以实现多层次的风格表征.

3.2　PSConv 模块

PSConv 模块包含 3 个卷积分支,分别用 gdConv、shift_gdConv 和 mask_Conv 表示(见图 4) .

输入特征图 $X \in \mathbf{R}^{W\times H\times C}$ (其中 $W=H$) 按通道分成了标号为 1 ~ 4 的 4 组. 虚线框内包含对应卷积分支的卷积核,3 个分支分别进行卷积,最后将卷积结果相加,为了保持 3 个卷积前后尺寸一致,设 stride 为 1,kernel_size 为 3,padding 的值保持和 dilation 的值一致. 其中 gdConv 和 shift_gdConv 的共同点在于都使用了分组卷积和扩张卷积 [29] ,并包含了 $C/4$ 个 $3\times3\times C$ 的卷积核. 但不同之处是在 shift_gdConv 进行卷积操作之前,将 X 的前 2 组和后 2 组进行交换,得到特征图 X'. 本文设置不同的扩张率 2 和 3. 这 2 部分卷积输出结果分别为 g_i 和 s_i,$i\in\{1,2,3,4\}$,分别代表对输入的第 i 个特征分组的卷积输出结果.

mask_Conv 将卷积核按个数分成了标号为①~④的 4 组,每组包含 $C/4$ 个 $3\times3\times C$ 的卷积核. 为了减少计算负担,屏蔽了一半卷积核参数,使其权值为 0,图中浅色方块代表被屏蔽的区域. 因此,标号为①、③的卷积核分别对应输出标号为 m_{13}、m'_{13} 的 2 个特征分组,代表着对输入的 1、3 特征分组的联合卷积结果; 标号为②、④的卷积核分别对应输出标号为 m_{24}、m'_{24} 的 2 个特征分组,代表着对输入的 2、4 特征分组的联合卷积结果.

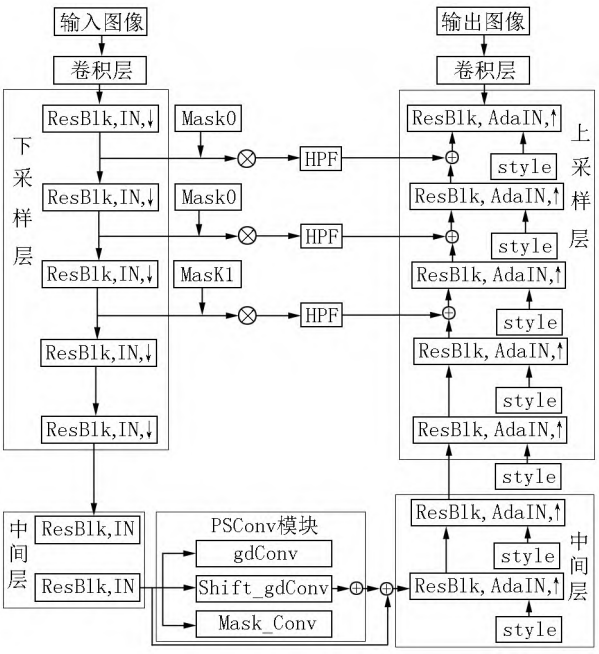

图 3　MFMA-StarGANv2 生成器结构

因此,3 个卷积结果按特征值位置相加输出尺寸仍为 $W\times H\times C$ 特征图,如下式所示:

$$Y=\mathrm{gdConv}(X)+\mathrm{shift\_gdConv}(X')+\mathrm{mask\_Conv}(X). \quad (1)$$

综上可知,通过对 shift_gdConv 设置分组交换以及与 gdConv 不同的扩张率,融合了 1、3 特征分组不同尺度的特征以及 2、4 特征分组不同尺度的特征;通过对 mask_Conv 设置部分参数屏蔽,减少了训练参数,并进一步增强了 1、3 特征分组间联系和 2、4 特征分组间联系. 这些操作对位置相距较远的特征分组实现了多尺度的特征融合,便于生成器后续对人脸图像进行完整性重建.

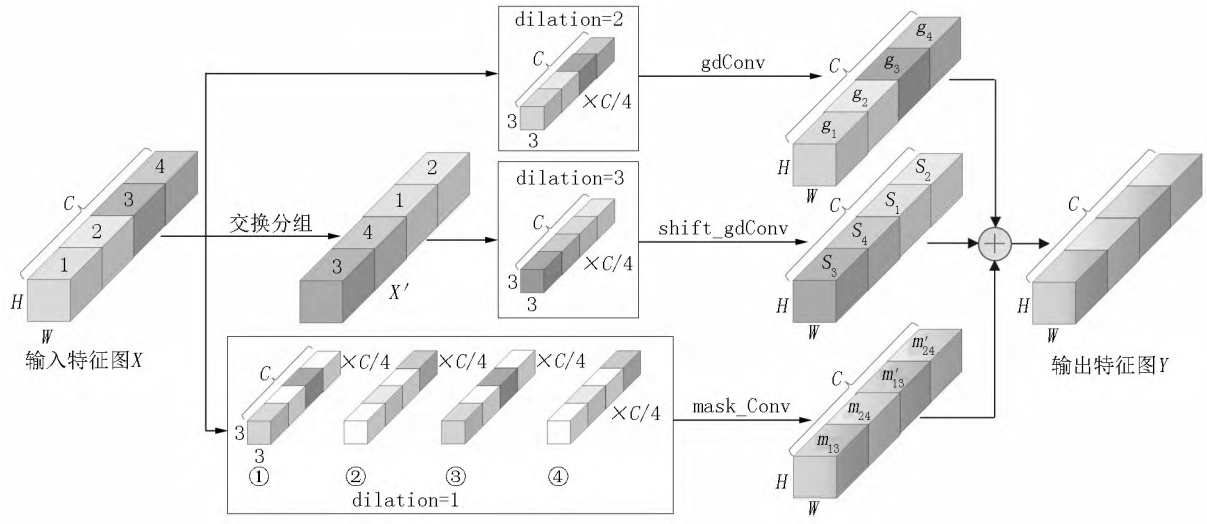

图 4 PSConv 模块内部结构

4 MFMA-StarGANv2 判别器

4.1 MFMA-StarGANv2 判别器结构

由于人脸结构的复杂性,判别器在进行人脸真假预测时,联系图像上下文特征进行综合判断具有关键作用. 随着下采样的进行,深层下采样生成的特征通道代表人脸的高层语义特征,此时对输入通道和输出通道进行注意力关注更有意义;但若下采样层过深则会使特征图空间太小,不利于捕获空间域注意力. 因此本文选择在第 5 个下采样层之后加入 MDConv 模块,判别器内部结构如图 5 所示.

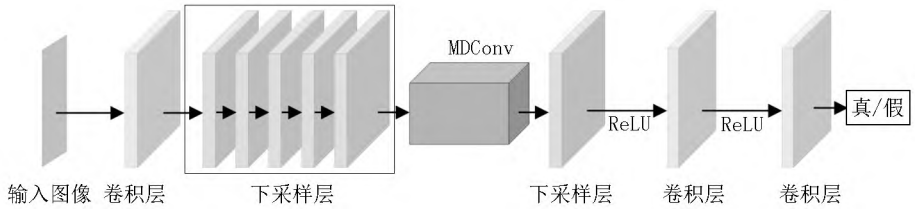

图 5 MFMA-StarGANv2 判别器结构

4.2 MDConv 模块

图 6 为 MDConv 模块的内部结构图,其中 $X\in\mathbf{R}^{h\times w\times c_{\mathrm{in}}}$ 代表输入特征图,首先进行过程①的全局平均池化(GAP),再通过 FC 层、ReLU 层将 X 压缩到具有 C_{Att} 长度的特征向量(为了降低计算复杂度,此处设定压缩比为 1/16). ②~④等 3 个 FC 层分别生成输入通道的注意力值 $\boldsymbol{\alpha}_i\in\mathbf{R}^{C_{\mathrm{in}}}$、$k\times k$ 空间位置处注意力值 $\boldsymbol{\alpha}_s\in\mathbf{R}^{k\times k}$、输出通道的注意力值 $\boldsymbol{\alpha}_0\in\mathbf{R}^{C_{\mathrm{out}}}$.

首先对输入通道维度进行注意力关注(如过程⑤所示). 将输入特征图 X 与输入通道的注意力值 $\boldsymbol{\alpha}_i$ 按对应通道相乘,输出结果为

$$Y_1=X\odot\boldsymbol{\alpha}_i. \quad (2)$$

然后对上述输出结果 Y_1 进行空间维度的注意力关注(如过程⑥、⑦所示). 与另外 2 维注意力关注方法不同的是,为了提取和整合输入特征图更多特征,此处应用了卷积操作. $W\in\mathbf{R}^{k\times k\times c_{\mathrm{in}}\times c_{\mathrm{out}}}$ 代表卷积核随机初始化的权重,通过过程⑥得到融入空间注意力关注值 $\boldsymbol{\alpha}_s$ 的卷积核参数 W',再与 Y_1 进行卷积操作,如过程⑦所示,输出结果为

$$Y_2=Y_1W', \quad (3)$$

其中 $W'=W\odot\boldsymbol{\alpha}_s$. 最后进行输出通道维度的注意力关注(如过程⑧所示). 上述输出结果 Y_2 与输出通

道的注意力值 $\boldsymbol{\alpha}_0$ 按通道进行对应相乘,最终输出结果为

$$\boldsymbol{Y} = \boldsymbol{Y}_2 \odot \boldsymbol{\alpha}_0. \tag{4}$$

MDConv 模块运用了动态卷积,对输入特征图分别在输入通道维度、卷积核空间维度、输出通道维度上生成了对应注意力,使得注意力关注值根据输入图像的变化而灵活变化,提升了模型的特征表达能力. 更重要的是,使用了较少卷积核,在增加较小的运算前提下,提供了更优越的性能以捕获多维的注意力关注信息.

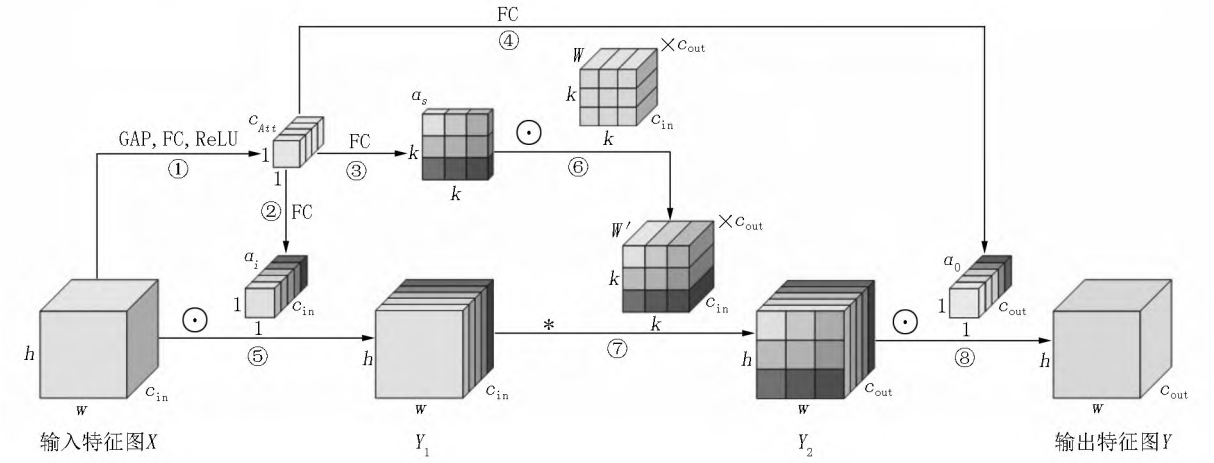

图 6　MDConv 模块内部结构

5　实验与分析

5.1　数据集

本文选择 CelebA-HQ 高清人脸数据集进行实验,在数据集内包含了 30 000 幅人脸图像,其中 28 000幅作为训练集,剩余 2 000 幅作为测试集. 将人脸图像分成 2 类,分别为男性人脸和女性人脸.

5.2　实验细节

本文实验环境是基于版本号为 1.7.0 的 Pytorch 框架实现的,采用 RTX3090 进行 GPU 加速,CPU 型号为 E5-2678 v3,CUDA 版本为 11.0.

在训练过程中,将输入图像和输出图像分辨率均设为 256×256; batchsize 设为 8; 超参数 λ_{sty}、λ_{ds} 和 λ_{cyc} 都设为 1; λ_{ds} 线性衰减为 0. 采用 Adam 优化器进行梯度下降优化,其中生成器 G、判别器 D 和风格编码器 E 的学习率设为 10^{-4},映射网络 F 的学习率设为 10^{-6},b_1 设为 0,b_2 设为 0.99,权值衰减值设为 10^{-6}.

5.3　对比实验效果

5.3.1　参考图像引导转换　图 7 为参考图像引导转换方式实验对比结果. 在此方式下,生成图像在保留输入人脸身份的同时,具有与对应参考图像类似的风格. 从风格转换效果来看,本文方法改进效果显著,转换的人脸发型细节更为丰富,与参考图像更相似,并且能较好地区分参考图像的头发和背景阴影(如图 7 第 4 列和第 5 列所示).

从人脸纹理细节来看,本文方法生成的人脸纹理细节也更加自然、美观(如图 7 第 2、3 列所示),生成的女性人脸淡化了输入的男性人脸轮廓特征,更贴近于真实的女性人脸特征.

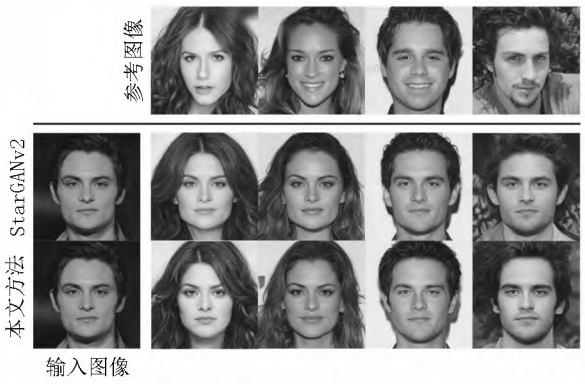

图 7　参考图像引导转换对比结果

5.3.2　潜码引导转换　图 8 为潜码引导转换方式的实验对比结果. 在此方式下,生成图像在保留输入人脸身份的同时,具有随机的风格外观. 可看出本文方法生成的图像,人脸纹理细节和风格转换效果均有着较高的质量. 尤其在图 8 第 3 列和第 4 列中,StarGANv2 方法生成的图像脸部存在输入人脸头发的残留痕迹,而本文方法较好地改善了这一问题.

图 8 潜码引导转换对比结果

5.4 评价指标

在生成对抗网络生成图像领域中,评价图像质量具有较强的主观性,因此本文采用主、客观相结合的方式进行实验评价. 采用学习感知图像相似度 LPIPS [30] 作为客观评价指标,调查用户选图占比结果作为主观视觉评价.

5.4.1 LPIPS LPIPS 用于衡量图像间的多样性,是通过从文献 [31] 提出的 AlexNet 提取的特征之间的 L_1 距离来测量生成图像的多样性,其值越高代表图像多样性越高. 在本文实验中,对于参考图像引导转换方式,测试集的每幅输入图像会抽取 10 幅参考图像进行引导转换. 在生成的 10 幅图像中,计算每 2 幅图像的 LPIPS 值,然后取平均值,最后对所有测试图像计算出的 LPIPS 值取平均值; 而对于潜码引导转换方式,在测试集中每幅输入图像通过 10 个随机潜码进行引导转换,LPIPS 值计算方式与前者一致. 不同方式对比结果如表 1 所示.

表 1 LPIPS 值对比结果

	参考图像引导转换	潜码引导转换
StarGANv2	0.382	0.441
本文方法	0.401	0.451

5.4.2 主观视觉评价 为了进一步评价生成图像的改进效果,邀请 20 名用户根据自身偏好进行选图,统计占比结果作为主观视觉评价结果,并对 2 种生成图像方式分别进行比较.

1) 对于参考图像引导转换方式,保持 2 幅输入图像和 5 幅参考图像一致,StarGANv2 方法和本文方法分别生成 10 幅图像. 让 20 名调查用户根据自身喜好在一共生成的 20 幅打乱的图像中分别选出人脸纹理细节最好的 5 幅图像和风格转换效果最好的 5 幅图像. 分别汇总后各包含 100 幅图像,选图占比结果如表 2 所示.

表 2 参考图像引导转换选图占比结果

方法	人脸纹理细节/%	风格转换效果/%
StarGANv2	32.0	29.0
本文方法	68.0	71.0

2) 对于潜码引导转换方式,保持 2 幅输入图像一致,分别输入 5 个随机潜码. StarGANv2 方法和本文方法分别生成 10 幅图像. 选图方式和 1) 一致,分别汇总后各包含 100 幅图像,选图占比结果如表 3 所示.

表 3 潜码引导转换选图占比结果

方法	人脸纹理细节/%	风格转换效果/%
StarGANv2	27.0	34.0
本文方法	73.0	66.0

5.5 消融实验

为了验证本文方法的有效性,进行了消融实验,对比 PSConv 模块和 MDConv 模块对模型的单独作用以及它们的共同作用,实验结果如图 9 所示. 从图 9 可看出: 仅在生成器内加入 PSConv 模块或者仅在判别器内加入 MDConv 模块,效果都不够理想,如发型变形、缭乱,人脸细节不够自然. 而本文方法生成的图像效果改进显著,较大程度地避免了这些问题. 由此可看出,生成器和判别器是彼此制约权衡的,若只改进一方,则有可能造成不良影响.

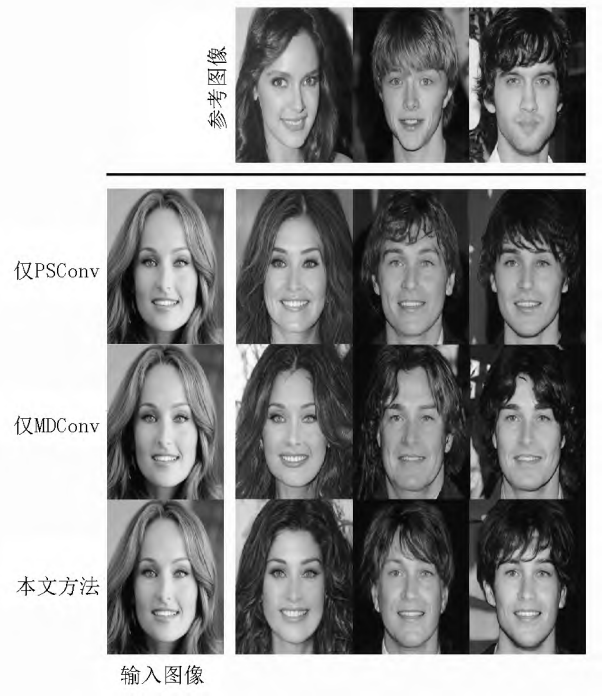

图 9 消融实验结果

6　结束语

本文提出了结合多尺度特征和多维注意力的人脸风格转换模型(MFMA-StarGANv2) . 通过在 StarGANv2 生成器中加入 PSConv 模块,从而提高生成器融合图像不同尺度特征的能力; 为了建立对输入图像特征的多维注意力,本文提出了 MDConv 模块,并将它融入 StarGANv2 判别器中,从而增强判别器对真假人脸的判别能力. 实验结果表明本文方法相比 StarGANv2 方法,转换的人脸风格更加美观,生成的人脸纹理细节更加自然,LPIPS 的值也得到了提升,并进行了消融实验进一步说明了本文方法的有效性. 本文提出的方法不仅适用于人脸风格转换领域,未来将会尝试将改进思路应用在图像生成的其他领域,不断改进模型,提高模型的泛化能力.

7　参考文献

[1] 费建伟,夏志华,余佩鹏,等. 人脸合成技术综述 [J] . 计算机科学与探索,2021,15(11) : 2025-2047.

[2] FU Yun,GUO Guodong,HUANG T S. Age synthesis and estimation via faces: a survey [J] . IEEE Transactions on Pattern Analysis and Machine Intelligence,2010,32(11) : 1955-1976.

[3] GOODFELLOW I,POUGET-ABADIE J,MIRZA M,et al. Generative adversarial nets [EB/OL] . [2022-06-16] . https: //arxiv. org/pdf/1406. 2661. pdf.

[4] LI Mu,ZUO Wangmeng,ZHANG D. Deep identity-aware transfer offacial attributes [EB/OL] . [2022-09-06] . https: //a-rxiv. org/pdf/1610. 05586. pdf.

[5] LIU Rujie,SHEN Wei. Learning residual images for face attribute manipulation [EB/OL] . [2022-09-08] . https: //doc. taixueshu. com/foreign/arXiv161205363. html.

[6] HE Kaiming,ZHANG Xiangyu,REN Shaoqing,et al. Deep residual learning for image recognition [EB/OL] . [2022-09-02] . https: //zhuanlan. zhihu. com/p/370863670.

[7] CHOI Y,CHOI M,KIM M,et al. Stargan: unified generative adversarial networks for multi-domain image-to-image translation [EB/OL] . [2022-09-02] . https: //arxiv. org/pdf/1711. 09020. pdf.

[8] LIU Ming,DING Yukang,XIA Ming,et al. Stgan: a unified selective transfer network for arbitrary image attribute editing [EB/OL] . [2022-09-08] . https: //blog. csdn. net/WhaleAndAnt/article/details/104677489.

[9] SHEN Yujun,GU Jinjin,TANG Xiaoou,et al. Interpreting the latent space of gans for semantic face editing [EB/OL] . [2022-09-08] . https: //arxiv. org/abs/1907. 10786v3.

[10] YANG Guoxing,FEI Nanyi,DING Mingyu,et al. L2m-gan: learning to manipulate latent space semantics for facial attribute editing [EB/OL] . [2022-09-08] . https: //www. xueshufan. com/publication/3182270175.

[11] WANG Huipo,YU Ning,FRITZ M. Hijack-gan: unintended use of pretrained,black-box gans [EB/OL] . [2022-09-08] . https: //arxiv. org/abs/2011. 14107v1.

[12] KHODADADEH S,GHADAR S,MOTIIAN S,et al. Latent to latent: a learned mapper for identity preserving editing of multiple face attributes in StyleGAN-generated images [EB/OL] . [2022-09-08] . https: //blog. csdn. net/xjm850552586/article/details/123656232.

[13] HUANG Xun,BELONGIE S. Arbitrary style transfer in real-time with adaptive instance normalization [EB/OL] . [2022-09-08] . https: //blog. csdn. net/a19990412/article/details/84729453/.

[14] KARRAS T,LAINE S,AILA T. A style-based generator architecture for generative adversarial networks [EB/OL] . [2022-09-08] . https: //blog. csdn. net/NGUever15/article/details/122299290.

[15] KARRAS T,LAINE S,AITTALA M,et al. Analyzing and improving the image quality of stylegan [EB/OL] . [2022-09-08] . https: //blog. csdn. net/lynlindasy/article/details/104495583.

[16] KARRAS T,AITTALA M,LAINE S,et al. Alias-free generative adversarial networks [EB/OL] . [2022-09-06] . https: //arxiv. org/pdf/2106. 12423. pdf.

[17] CHOI Y,UH Y,YOO J,et al. Stargan v2: diverse image synthesis for multiple domains [EB/OL] . [2022-09-08] . https: //blog. csdn. net/weixin_43135178/article/details/126828444.

[18] KARRAS T,AILA T,LAINE S,et al. Progressive growing-of gans for improved quality,stability,and variation [EB/OL] . [2022-09-07] . https: //arxiv. org/pdf/1710. 10196. pdf.

[19] KINGMA D P,WELLING M. Auto-encoding Variational-Bayes [EB/OL] . [2022-09-07] . https: //arxiv. org/pdf/1312. 6114. pdf.

[20] KINGMA D P,DHARIWAL P. Glow: generative flow with-invertible 1x1 convolutions [EB/OL] . [2022-09-09] . https: //arxiv. org/pdf/1807. 03039. pdf.

[21] LONG J,SHELHAMER E,DARRELL T. Fully convolutional networks for semantic segmentation [EB/OL] . [2022-09-08] . https: //ieeexplore. ieee. org/document/7478072.

[22] RONNEBERGER O,FISCHER P,BROX T. U-NET: convolutional networks for biomedical image segmentation [EB/OL]. [2022-09-08]. https://blog.csdn.net/weixin_36670529/article/details/102809431.

[23] SZEGEDY C,LIU Wei,JIA Yangqing,et al. Going deeper with convolutions [EB/OL]. [2022-09-08]. https://zhuanlan.zhihu.com/p/158914902.

[24] TAN Mingxing,LE Q V. Mixconv: mixed depthwise convolu-tional kernels [EB/OL]. [2022-09-09]. https://arxiv.org/pdf/1907.09595.pdf.

[25] LI Duo,YAO Anbang,CHEN Qifeng. Psconv: squeezing feature pyramid into one compact poly-scale convolutional layer [EB/OL]. [2022-09-08]. https://arxiv.org/abs/2007.06191.

[26] YANG B,BENDER G,LE Q V,et al. CondConv: conditionally parameterized convolutions for efficient inference [EB/OL]. [2022-09-10]. https://arxiv.org/pdf/1904.04971.pdf.

[27] CHEN Yinpeng,DAI Xiyang,LIU Mengchen,et al. Dynamic convolution: attention over convolution kernels [EB/OL]. [2022-09-08]. https://blog.csdn.net/m0_47180208/article/details/118570067.

[28] LI Chao,ZHOU Aojun,YAO Anbang. Omni-dimensional dynamic convolution [EB/OL]. [2022-09-11]. https://arxiv.org/pdf/2209.07947.pdf.

[29] YU Fisher,KOLTUN V. Multi-scale context aggregation by dilated convolutions [EB/OL]. [2022-09-11]. https://arxiv.org/pdf/1511.07122.pdf.

[30] ZHANG R,ISOLA P,EFROS A A,et al. The unreasonable effectiveness of deep features as a perceptual metric [EB/OL]. [2022-09-08]. https://arxiv.org/pdf/1801.03924.pdf.

[31] KRIZHEVSKY A,SUTSKEVER I,HINTON G E. Imagenet classification with deep convolutional neural networks [J]. Advances in Neural Information Processing Systems,2017,60(6):84-90.

The Face Style Conversion Combining Multiscale Feature Fusion and Multi – Dimensional Attention

LIU He,ZHOU Yong*,PAN Yi,ZHANG Jintao

(School of Computer and Information Engineering,Jiangxi Normal University,Nanchang Jiangxi 330022,China)

Absrtact: According to the phenomenon that the face images generated by StarGANv2 exist poor style reconstruction effect and unnatural face texture details,a face style conversion model called MFMA-StaGANv2 (multiscale feature and multi-dimensional attention StarGANv2) combining multiscale features and multi-dimensional attention is proposed. In order to improve the ability of the model to extract image features,a multiscale feature fusion module is embedded in the generator of StarGANv2. In order to improve the ability of the model to distinguish true and false face images,a multi-dimensional attention module called MDConv is proposed and embed in the StarGANv2 discriminator. Compared with StarGANv2 on CelebA-HQ dataset,the results show that the style of the face images generated by our method is more beautiful,the details of face texture are more natural,and the values of LPIPS (learned perceptual image patch similarity) are improved.

Key words: face style conversion; face attributes synthesis; multiscale feature fusion; multi-dimensional attention

(责任编辑:冉小晓)