基于特征对比的循环生成对抗网络图像风格转换研究

闫娟[a],康鹏帅[b],王士斌[b,c],梅学术[b],李燕[b],刘栋[b]

(河南师范大学 a.信息化建设与管理办公室;b.计算机与信息工程学院;
c."教育人工智能与个性化学习"河南省重点实验室,河南 新乡 453007)

摘 要:无监督图像到图像转换任务是在非配对训练数据的情况下学习源域图像到目标域图像的转换.但是,图像风格转换任务依然面临着图像内容丢失、模型坍塌等现象.为了解决上述问题,提出了一种局部特征对比来保持图像内容,通过特征提取器获得多层图像深层特征,使得图像编码器学习到高级语义信息,获得信息更加丰富的图像特征.同时,增加局部特征对比损失来引导特征提取器学习到有利于图像内容生成的特征.实验结果表明,在大多数情况下,所提方法在 FID 和 KID 分数方面优于之前的方法,图像生成质量有一定的提升.

关键词:特征对比;图像风格转换;对比损失

中图分类号:TP399 **文献标志码**:A **文章编号**:1000-2367(2024)06-0073-07

不同图像之间的转换是通过学习某种映射来完成的,这种转换不仅可以用于图像风格之间的转换,例如将真实照片转换成梵高类型的油画,还可以用于图像内容和结构方面的修改,比如猫与狗、斑马与马之间的转换.得益其出色的表现,图像风格转换任务也被推广到众多领域,例如图像修复[1]、图像去雾[2]、图像编辑[3]、图像高分辨率生成[4]等等.因此,无监督图像之间的转换受到了众多计算机视觉领域研究者的关注.

早期的图像风格转换任务通过对源域图像建立数学模型进行分析,在与目标域不断地对比当中,不断调整转换模型,然后将图像输入到模型输入中,完成图像风格的转换,但也因此无法提取和学习到图像的特征,转换效果较为粗糙.随着深度神经网络(DNN[5])的不断发展,其也被应用在图像转换领域,通过反向传播来更新权重系数,达到与目标域图像近似.基于深度神经网络的图像转换模型在面对复杂图像和大量数据时,参数空间指数上升,泛化能力弱,对数据要求较为苛刻,无法实现大量无监督图像风格转换任务.

当前无监督图像转换任务通常都是基于生成对抗网络(GAN)来实现的,传统的 GAN 模型[6]通过训练一组生成器和鉴别器来完成图像转换任务.但是由于 GAN 的复杂性和模型训练的困难,导致很难获得一个良好的图像转换模型.像循环生成对抗网络(CycleGAN[7])通过采用一对生成器和辨别器实现两个域之间的转换.但其在一些图像结构差别较大的领域,比如猫狗转换上表现不佳.为了提高复杂图像风格转换的质量,通常还引入其他模块,如使用注意力机制[8]的 U-GAT-IT[9],但同时也增加了模型的冗余度.NICE-GAN[10]通过重用鉴别器的编码器对网络进行了简化,取得了令人瞩目的结果.然而,简化后的网络在图像生成结果中也产生了一些新的问题,如翻译图像的结构不平衡和图像部分模糊.

为了解决上述问题,提出了局部特征对比模块.在本文的工作中,目标是希望转换过来的图像可以展现出

收稿日期:2023-07-12;**修回日期**:2024-04-07.

基金项目:国家自然科学基金(62072160);河南省科技攻关计划项目(222102210187);河南省普通本科高等学校智慧教学专项研究项目(202111).

作者简介:闫娟(1982—),女,河南周口人,河南师范大学高级工程师,研究方向为图像处理、人工智能、大数据分析,E-mail:48279674@qq.com.

通信作者:王士斌,E-mail:wangshibin@htu.edu.cn.

引用本文:闫娟,康鹏帅,王士斌,等.基于特征对比的循环生成对抗网络图像风格转换研究[J].河南师范大学学报(自然科学版),2024,52(6):73-79.(Yan Juan,Kang Pengshuai,Wang Shibin,et al.Research on image style transformation of cyclic generation adversarial network based on feature contrast[J].Journal of Henan Normal University(Natural Science Edition),2024,52(6):73-79.DOI:10.16366/j.cnki.1000-2367.2023.07.12.0003.)

目标域的外观,同时保留住输入图像的结构或内容,而不是使用原始像素或特征.具体来说,通过在循环生成对抗网络模型中引入局部特征对比模块,将源域图像和输出图像通过同一个编码器来提取局部特征向量,为了简化网络和减少参数,通过重复使用鉴别器中前半部分的编码器作为图像局部特征的提取器,随后在输出图像内容丰富区域选取锚点补丁[11],在源域图像相同位置选取正样本补丁,在源域图像的其他部分随机抽取 N 个负样本补丁.同时,加入局部特征对比损失来减少锚点与正样本之间的差距,拉大与负样本之间的差距,以此来提高生成器的编码性能和约束模型学习.这样,编码器就可以学习到两个不同领域之间的共性,如物体的形状,同时对差异保持不变,如物体的纹理.实验结果表明,本文的模型取得了更好的图像转换效果.

1 图像风格转换相关工作

1.1 图像风格转换

图像风格转换旨在将一幅图像的风格转换为另一幅图像的风格,并尽可能保留源域图像的内容特征.传统的图像风格转换方法最早由 HERTZMANN 等[12]提出,他们在单个输入输出训练图像上使用非参数纹理模型.随着深度学习的不断探索,GATYS 等[13]首次提出了基于卷积神经网络的风格转移方法,他们通过 VGG 网络[14]来表示图像的语义风格信息和内容纹理特征信息,并通过迭代的方式不断地更新网络参数,从而使输出图像不断接近目标域图像.但是,这些方法在风格转移算法方面建模困难,耗时长,效果不佳.

1.2 无监督图像风格转换

图像到图像的转换技术一般需要大量的成对数据,而获取这些数据非常耗时耗力,而无监督图像风格转换是一种不需要成对数据集的转换方法.代表性的有文献[7]提出的 CycleGAN 模型,可以将其看成是一个循环生成的网络,利用对偶学习的思路将源域图像生成目标图像之后再转换为源域图像,需要要求输入的图像域和目标域之间具有双射关系,其通过循环一致性损失来保证原始图像的结构不变,使用对抗损失强化输出图像的外观特征,提出身份损失去控制生成图像整体的颜色变化,具备强大的数据生成能力.

最新的研究方面,文献[9]提出的 U-GAT-IT 模型,通过使用类激活映射并引入自适应层实例归一化,构建了一个端到端的弱监督跨域转换模型.文献[10]提出了 NICE-GAN 网络模型,将判别器赋予双重价值,同时进行编码和判别,通过复用判别器的编码器来替代目标域图像的编码器,不再需要额外的编码组件,网络结构更加紧凑,减少了网络复杂度和网络参数.

1.3 对比学习

对比学习[15]广泛应用于无监督表示学习,其核心思想是通过最大化相关样本之间的相似性,最小化不相关样本之间的相似性来学习数据表示.文献[11]将对比学习应用到图像转换领域,提出了 CUT 模型,该模型通过最大化互信息的方法学习一个输入输出图像块之间的相似性函数,首次将 InfoNCE loss 应用到了条件图像生成领域,可以实现在单张图像上完成图像转换.随后 HAN 等[16]提出了双重对比方法,他们通过使用两个不同的编码器用于推断未配对数据之间的有效映射,提高了一致性和训练的稳定性.

2 本文方法

针对非配对图像转换后图像内容和结构丢失问题,提出了一种局部特征对比模块,使其注重于图像中物体的内容和外观.该模块由多层特征提取器,特征块对比损失函数组成.下面对整个模块框架,局部特征提取器和损失函数进行介绍.

2.1 模型框架

局部特征对比模型主要包括被重复使用的局部特征提取器 $E_{x\to y}$,两个生成器 $G_{x\to y}$ 和 $G_{y\to x}$,两个判别器 D_x 和 D_y,如图 1 所示.其中 X 代表源域图像的数据分布,Y 代表目标图像的数据分布.局部特征提取器同时作为生成器 $G_{x\to y}$ 和鉴别器 D_x 的编码器,在训练模型时,采用解耦的训练方式,仅在最大化对抗损失的时候对该编码器进行训练.首先对于一张来自 X 域的图像 x,局部特征提取器 $E_{x\to y}$ 首先进行特征提取,得到的特征向量同时反馈给生成器 $G_{x\to y}$ 和鉴别器 D_x,通过生成器 $G_{x\to y}$ 得到目标图像,多尺度辨别器 D_x 判断图像的真假程度.随后,将生成的 Y 域图像分别传递给编码器 $E_{y\to x}$ 和 $E_{x\to y}$,由 $E_{y\to x}$ 得到的特征向量同时也反馈

给生成器 $G_{y\to x}$ 和鉴别器 D_y.最后通过计算对抗损失、循环一致性损失、重构损失和局部特征对比损失更新网络参数.将 Y 域图像转换为 X 域图像与上述过程相同，这里不再赘述.

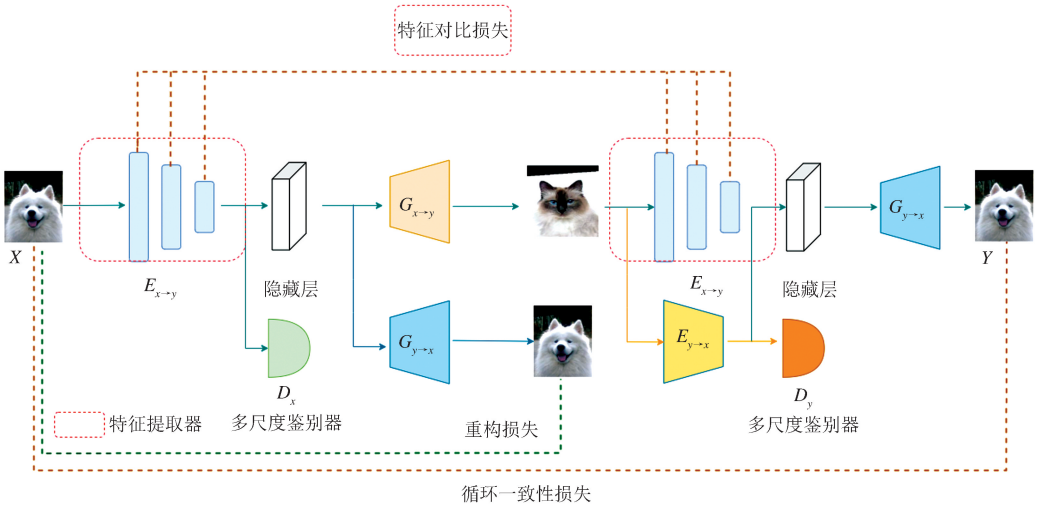

图1　局部特征对比模型框架图

Fig.1　Framework diagram of local feature comparison model

2.2　局部特征提取器

图 2 和图 3 说明了局部特征提取和对比特征采样的实现过程以及特征对比损失的计算，局部特征提取器使用卷积神经网络，可以高效提取特征.对于 X 域和 Y 域的图像都是通过同一个编码器进行两次下采样，为了加快模型的收敛速度，在每次卷积操作之前增加 Spectral _ norm[17]，在每次卷积操作之后加入 LeakyReLU 激活函数，上述正则化和激活函数可以提高神经网络的稳定性和泛化能力.随后在输出图像 Y 中采样一个锚点(z)，也就是查询样本，对于输入图像 X，在锚点相同位置采样一个正样本($z+$)，在除此之外的其他位置随机采样负样本($z-$)，所有的采样都是在网络的空间维度上进行的.同时将它们送入特征对比模块计算特征对比损失，即将其以对抗性的方式对锚点、正样本和生成的负样本进行对比学习，即扩大查询样本与负样本之间的距离，缩小与正样本的差距，达到输出图像近似于目标图像的效果.

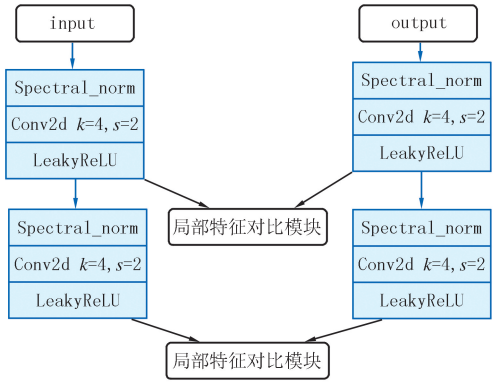

图2　局部特征提取

Fig.2　Local feature extraction

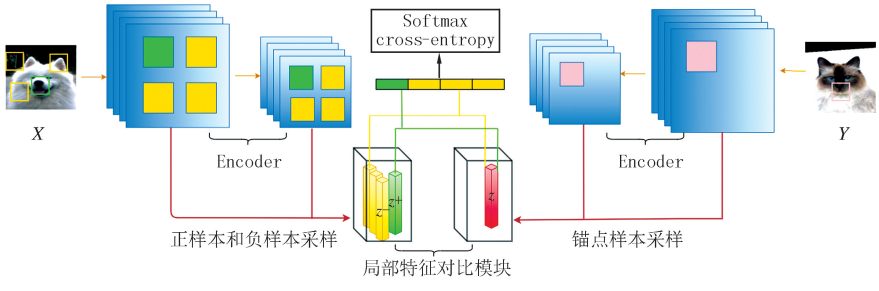

图3　特征对比损失

Fig.3　Feature contrast loss

2.3　损失函数

2.3.1　对抗损失

没有使用原始的 GAN 损失，而是采用了 LSGAN 中提出的最小二乘损失代替交叉熵损失从而让训练更加稳定，图像生成质量更高，与目标图像更加相似.目标函数如下：

$$\min_{G_{x\to y}}\max_{D_y=(C_y\circ E_y)} L_{\text{gan}}^{x\to y} := E_{y\to Y}[(D_y(y))^2]+E_{x\sim X}[(1-D_y(G_{x\to y}(E_x(x))))^2], \tag{1}$$

当最大化 $L_{\text{gan}}^{x\to y}$ 时，固定 E_x，训练 E_y；当最小化 $L_{\text{gan}}^{x\to y}$ 时，同时固定 E_x 和 E_y.

2.3.2 重构损失

使用重构损失来确保可以通过源域图像特征和源域生成器去恢复源域特征,其计算公式如下:

$$\min_{G_{y\to x}} L_{\text{recon}}^{x\to y} := E_{x\sim X}[\,|\,x - G_{y\to x}(E_x(x))\,|_1], \tag{2}$$

其中,$|\cdot|_1$ 计算 L_1 范数,E_x 保持不变.同样,也可以定义 $L_{\text{recon}}^{y\to x}$.

2.3.3 循环一致性损失

单纯地使用对抗损失会使目标域生成器只倾向于改变图像风格,从而导致模式崩塌问题.因而使用了 CycleGAN 中的 L_1 损失来计算循环一致性损失,可以很好地保留图像内容.其计算公式如下:

$$\min_{G_{y\to x}G_{x\to y}} L_{\text{cycle}}^{x\to y} := E_{x\sim X}[\,|\,x - G_{y\to x}(E_y(G_{x\to y}(E_x(x))))\,|_1]. \tag{3}$$

2.3.4 特征对比损失

通过编码器提取的图像特征包含丰富的图像表示信息.为了使生成的目标域图像与源域图像在图像结构上和图像内容更加接近,将生成图像和源域图像作为输入,通过重复使用同一个编码器的 L 层计算图像深层特征.其中 $s\in\{1,2,\cdots,S_l\}$,S_l 表示每一层中选取样本的数量.在生成图像内容丰富的区域选取锚点,在源域图像中的同一位置选取正样本,并在源域图像的其他位置选取 N 个负样本.目标是在图像特征向量表征空间中将正样本与锚点样本(z)之间的特征距离拉近,将负样本与锚点样本之间的特征距离拉远,其计算公式如下:

$$L_{\text{Feature-PatchNCE}}(G,H,X) = E_{x\sim X}\sum_{l=1}^{L}\sum_{s=1}^{S_l}\ell(\hat{z}_l^{\,s}, z_l^{\,s}, z_l^{S/s}). \tag{4}$$

2.3.5 总损失

判别器的总损失为:

$$\max_{E_x,C_x,E_y,E_y} \lambda_1 L_{\text{gan}} + \lambda_{2-1} L_{\text{Feature-PatchNCE}}. \tag{5}$$

生成器的总损失为:

$$\min_{G_{x\to y}G_{y\to x}} \lambda_1 L_{\text{gan}} + \lambda_{2-2} L_{\text{cycle}} + \lambda_3 L_{\text{recon}}. \tag{6}$$

在实验中,λ_1,λ_{2-1},λ_{2-2},λ_3 分别被固定为 $\lambda_1=1$,$\lambda_{2-1}=1$,$\lambda_{2-2}=10$,$\lambda_3=10$.

3 实验分析

3.1 数据集

实验中使用了 4 种常见的无配对基准数据集,分别为 horse↔zebra、summer↔winter、vangogh↔photo 和 cat↔dog.其中 horse↔zebra 来源于 CycleGAN,它包含 2 401 张训练图像,260 张测试图像,分别为 1 067/120(horse),1 334/140(zebra),这些图像都是从 ImageNet[18] 中收集的;summer↔winter 是从 Flickr API 上下载的,剪掉了黑白照片,其中夏天和冬天的训练集和测试集分别为 1 231/309(summer),962/238(winter);vangogh↔photo 来自于 CycleGAN,它包含 400 张梵高画,7 038 张照片,重复使用梵高画的训练集作为测试集,将照片分为 6 287 张训练集,751 张作为测试集;cat↔dog 在 DRIT[19] 中被介绍,该数据集是从谷歌图像中截取的,其中猫和狗的训练集和测试集分别为 771/100(cat),1 264/100(dog).在实验中,将所有数据集进行裁剪并调整大小为 256×256.

3.2 实验设置

所有实验均在 Pytorch 框架上进行,遵循了 NICEGAN 的框架设定,增加了局部特征对比损失,并相应地提取编码器均匀分布点的特征.在生成器中使用 ReLU 作为激活函数,在鉴别器中使用斜率为 0.2 的 LeakyReLU.使用学习率为 0.000 1 的 Adam 优化器,在 NVIDIA A100 显卡上训练所有模型.对于数据增强,以 0.5 的概率水平翻转图像,将其大小调整为 286×286,并随机裁剪为 256×256.所有实验的 BatchSize 设置为 1.设置权重衰减为 0.000 1.所有模型都经过了 300 k 次迭代训练.

3.3 评价指标

在本文中,采用图像风格转换领域常用的评价指标 FID 和 KID 来评估图像生成质量.FID 对每个比较图像集的 InceptionNet 隐藏激活函数进行高斯分布拟合,然后计算这些高斯之间的 Frechet 距离.当 FID 分

数越低，生成的图像越接近真实图像，也就意味着生成的网络结果越好.计算公式如下：

$$\mathrm{FID}=\|\mu_r-\mu_g\|^2+\mathrm{Tr}\left(\sum_r+\sum_g-2\left(\sum_r\sum_g\right)^{11/2}\right). \tag{7}$$

KID 与 FID 类似，KID 通过计算特征之间最大平均差的平方去测量两组样本之间的差异.此外，与 FID 不同的是，KID 有一个简单的无偏估计器，这使得它更可靠，特别是当初始特征通道的数量远大于图像的数量时.较小的 KID 分数表示真实图像与生成图像之间具有良好的视觉相似性.

3.4　对比分析

与其他 5 种优秀的图像转换模型进行了可视化分析，如图 4 所示.从生成的图像上来看，UNIT[20] 和 MUNIT[21] 也都成功转换了图像的语义信息，但在图像清晰度上还存在比较大的差异，尤其是在 vangogh→photo 任务上，这是因为原来梵高的图像没有锐利的边缘，在模型上效果会比较差.U-GAT-IT-light 的模型结构与 NICEGAN 大致相同，它们在图像转换方面取得了更好的效果，但生成的图像仍会出现部分模糊，在图像的结构和细节方面仍然存在不足.与其他图像转换方法相比，本文提出的局部特征对比很好地处理了图像细节和结构部分，更好地还原了目标图像的特征，如图 4第 2 列所示.

图4　不同图像风格转换模型的可视化结果

Fig.4　Visualization results of different image style transfer models

表 1 展示了不同图像转换模型的定量评估结果，如表 1 所示，本文的方法在大多数情况下都可以获得最低的 FID 和 KID 分数，这说明本文的方法在不同的任务上都有良好的转换能力.相比之下，其他方法仅在某些数据集上表现良好.例如 CycleGAN 只擅长修改了图像的低级特征(例如颜色和纹理)，因此在 cat→dog 数据集上取得了最差的 FID=125.30.但在 horse→zebra 结果上较差的原因在于图像内容和风格相差较大，并且图像背景占据图像内容较多部分，模型在提取源图像特征与目标域图像特征进行对比时会增大误差，因此很难控制图像的整体风格.相比之下，本文算法在 cat↔dog 数据集上的表现较为突出，dog→cat 得到了最好 FID=41.67 和 KID=0.86，cat→dog 得到了最好 FID=37.01 和 KID=0.56.

3.5　消融实验

为了验证本论文提出的局部特征对比的有效性，在 cat↔dog 数据集和 summer↔winter 数据集上进行了消融实验.进行了两次实验，第一次实验不添加特征对比模块，第二次实验添加特征对比模块.“√”表示在原有模型的基础上增加了此方法，所有的消融实验都遵循了相同的参数设置.

表 2 展示了分离特征对比模块前后实验的对比，根据表 2 的结果，增加特征对比之后，大部分数据集得到提升，特别在 cat→dog 的实验中，FID 和 KID 分数分别提升了 7.66 和 0.64，这证明了特征对比模块产生了好于之前模型的实验效果，在图像转换方面具有实效性.

表 1 不同图像风格转换模型的客观指标评价结果

Tab. 1 Evaluation results of objective metrics for different image style transfer models

模型	dog→cat		winter→summer		photo→vangogh		zebra→horse	
	FID	KID×100	FID	KID×100	FID	KID×100	FID	KID×100
MUNIT[21]	53.25	1.26	99.14	4.66	130.55	4.50	193.43	7.25
UNIT[20]	59.56	1.94	95.93	4.63	136.80	5.17	170.76	6.30
CycleGAN[7]	119.32	4.93	79.58	1.36	136.97	4.75	156.19	5.54
U-GAT-IT-light[9]	80.75	3.22	80.33	1.82	137.70	6.03	145.47	3.39
NICE-GAN[10]	48.79	1.50	76.44	1.22	**122.27**	**3.71**	149.48	4.29
Ours	**41.67**	**0.86**	**70.99**	**0.86**	122.31	4.23	**128.20**	**2.34**
模型	cat→dog		summer→winter		vangogh→photo		horse→zebra	
	FID	KID×100	FID	KID×100	FID	KID×100	FID	KID×100
MUNIT[21]	60.84	2.42	114.08	5.27	138.86	6.19	128.70	6.92
UNIT[20]	63.78	1.94	112.07	5.36	143.96	7.44	131.04	7.19
CycleGAN[7]	125.30	6.93	78.76	0.78	135.01	4.71	95.98	3.24
U-GAT-IT-light[9]	64.36	2.49	88.41	1.43	123.57	4.91	113.44	5.13
NICE-GAN[10]	44.67	1.20	**76.03**	**0.67**	112.00	2.79	**65.93**	**2.09**
Ours	**37.01**	**0.56**	76.57	0.96	**95.03**	**2.38**	79.47	2.11

表 2 增加特征对比模块前后对比

Tab. 2 Add the feature comparison module before and after comparison

分组	数据集	方法	FID	KID	分组	数据集	方法	FID	KID
		特征对比模块					特征对比模块		
实验一	cat→dog		44.67	1.20	实验二	cat→dog	√	**37.01**	**0.56**
	dog→cat		48.79	1.50		dog→cat	√	**41.67**	**0.86**
	winter→summer		76.44	1.22		winter→summer	√	**70.99**	**0.86**
	summer→winter		**76.03**	**0.67**		summer→winter	√	76.57	0.96

4 总 结

本文在循环生成对抗网络上提出了一种局部特征对比模块进行图像风格转换.局部特征对比模块从输入和输出图像上获取图像内容丰富的特征,使用特征对比损失更好的维持图像内容,使其与目标域图像特征接近,从而提升生成图像的效果.与现有的 5 种优秀图像风格转换模型相比,本文的图像转换方法在 4 种常用的数据集上取得了良好的效果,消融实验表明本文提出方法的可靠性.

致谢:本论文数值计算得到了河南师范大学高性能计算中心的计算支持.

参 考 文 献

[1] ZHENG H T,LIN Z,LU J W,et al.Image inpainting with cascaded modulation GAN and object-aware training[C]//European Conference on Computer Vision.Cham:Springer,2022.

[2] SONG Y D,HE Z Q,QIAN H,et al.Vision transformers for single image dehazing[J].IEEE Transactions on Image Processing:a Publication of the IEEE Signal Processing Society,2023,32:1927-1941.

[3] DEKEL T,GAN C,KRISHNAN D,et al.Sparse,smart contours to represent and edit images[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.Salt Lake City:IEEE,2018.

[4] ROMBACH R,BLATTMANN A,LORENZ D,et al.High-resolution image synthesis with latent diffusion models[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR).New Orleans:IEEE,2022.

[5] KRIZHEVSKY A，SUTSKEVER I，HINTON G E.ImageNet classification with deep convolutional neural networks[J].Communications of the ACM，2017，60(6)：84-90.

[6] GOODFELLOW I，POUGET-ABADIE J，MIRZA M，et al.Generative adversarial networks[J].Communications of the ACM，2020，63(11)：139-144.

[7] ZHU J Y，PARK T，ISOLA P，et al.Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//2017 IEEE International Conference on Computer Vision(ICCV).Italy：IEEE，2017.

[8] VASWANI A，SHAZEER N M，PARMAR N，et al.Attention is all you need[EB/OL].[2024-04-06].http://arxiv.org/pdf/1706.03762.

[9] LEE H Y，LI Y H，LEE T H，et al.Progressively unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation[J].Sensors，2023，23(15)：6858.

[10] CHEN R F，HUANG W B，HUANG B H，et al.Reusing discriminators for encoding：towards unsupervised image-to-image translation[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR).Seattle：IEEE，2020.

[11] PARK T，EFROS A A，ZHANG R，et al.Contrastive learning for unpaired image-to-image translation[C]// European Conference on Computer Vision.Cham：Springer，2020.

[12] HERTZMANN A，JACOBS C E，OLIVER N，et al.Image analogies[C]//Proceedings of the 28th annual conference on Computer graphics and interactive techniques.New York：ACM，2001.

[13] GATYS L A，ECKER A S，BETHGE M.Image style transfer using convolutional neural networks[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR).Las Vegas：IEEE，2016.

[14] SZEGEDY C，LIU W，JIA Y Q，et al.Going deeper with convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition(CVPR).Boston：IEEE，2015.

[15] HE K M，FAN H Q，WU Y X，et al.Momentum contrast for unsupervised visual representation learning[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR).Seattle：IEEE，2020：9726-9735.

[16] HAN J L，SHOEIBY M，PETERSSON L，et al.Dual contrastive learning for unsupervised image-to-image translation[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops(CVPRW).Nashville：IEEE，2021.

[17] MIYATO T，KATAOKA T，KOYAMA M，et al.Spectral normalization for generative adversarial networks[EB/OL].[2024-04-06].http://arxiv.org/abs/1802.05957v1.

[18] DENG J，DONG W，SOCHER R，et al.ImageNet：a large-scale hierarchical image database[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition.Miami：IEEE，2009.

[19] LEE H Y，TSENG H Y，HUANG J B，et al.Diverse image-to-image translation via disentangled representations[C]//Computer Vision-ECCV 2018：15th European Conference，Munich：ACM，2018.

[20] LIU M Y，BREUEL T，KAUTZ J.Unsupervised image-to-image translation networks[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems.California：ACM，2017.

[21] HUANG X，LIU M Y，BELONGIE S，et al.Multimodal unsupervised image-to-image translation[C]//Computer Vision-ECCV 2018：15th European Conference.Munich：ACM，2018.

Research on image style transformation of cyclic generation adversarial network based on feature contrast

Yan Juan[a]，Kang Pengshuai[b]，Wang Shibin[b,c]，Mei Xueshu[b]，Li Yan[b]，Liu Dong[b]

(a. Information Construction and Management Office；b. School of Computer and Information Engineering；c. Key Lab of "Artificial Intelligence and Personalized Learning in Education" in Henan Province，Henan Normal University，Xinxiang 453007，China)

Abstract：The unsupervised image-to-image translation task is to learn the transformation of source domain images to target domain images in the case of unpaired training data. However，the image style conversion task still faces phenomena such as image content loss and model collapse. In order to solve the above problems，we propose a local feature comparison to preserve image content，and obtain multi-layer image deep features through a feature extractor，allowing the image encoder to learn high-level semantic information and obtain more informative image features. At the same time，local feature contrast loss is added to guide the feature extractor to learn features that are beneficial to image content generation. Experimental results show that in most cases，our method outperforms previous methods in terms of FID and KID scores，and the quality of image generation is improved to a certain extent.

Keywords：feature comparison；image style conversion；contrast loss

[责任编校　陈留院　赵晓华]