计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
[doi: 10.15888/j.cnki.csa.009848] [CSTR: 32024.14.csa.009848]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

融合 CBAM 注意力机制的敦煌壁画风格迁移①

贵向泉[1], 曹　磊[1], 李　立[2]

[1](兰州理工大学 计算机与通信学院, 兰州 730050)
[2](兰州大学 信息科学与工程学院, 兰州 730000)
通信作者: 曹　磊, E-mail: 222085404095@lut.edu.cn

摘　要: 敦煌壁画是人类世界文明史中耀眼的瑰宝. 然而, 现有对敦煌壁画的算法研究主要集中在壁画修复方面, 很少有针对敦煌壁画的色彩风格迁移研究. 因此, 提出一种基于循环生成对抗网络的融合 CBAM 注意力机制的敦煌壁画风格迁移方法. 通过提取输入图像的特征, 将其输入到添加 CBAM 注意力机制的生成器中, 应用注意力机制提升重点区域的风格迁移效果, 抑制边界伪影的产生; 为了更好地保留图像内容的结构信息, 在下采样区和上采样区之间添加了残差网络模块; 并且在损失函数中加入色彩损失, 约束模型提高生成图像的风格化效果. 通过自建的敦煌壁画数据集上进行的实验验证, 所提出的模型在敦煌壁画艺术风格迁移任务中展现出了相较于现有方法的优越性. 该模型能够生成视觉效果更为卓越、艺术韵味更为浓厚的敦煌壁画风格化图像, 为敦煌壁画的创新研究提供了新思路.

关键词: 风格迁移; 循环生成对抗网络; CBAM 注意力机制; 敦煌壁画

引用格式: 贵向泉,曹磊,李立.融合 CBAM 注意力机制的敦煌壁画风格迁移.计算机系统应用. http://www.c-s-a.org.cn/1003-3254/9848.html

Style Transfer of Dunhuang Murals with CBAM Attention Mechanism

GUI Xiang-Quan[1], CAO Lei[1], LI Li[2]

[1](School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China)
[2](School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China)

Abstract: Dunhuang murals are dazzling treasures in the history of human world civilization. However, existing algorithmic studies on Dunhuang murals mainly focus on mural restoration, seldom concentrating on color style transfer. Therefore, a style transfer method for Dunhuang murals which incorporates the CBAM attention mechanism based on recurrent generative adversarial network is proposed in this study. By extracting the features of the input image and feeding them into the generator which is added with the CBAM attention mechanism, the attention mechanism is applied to improve the style transfer effect of the focus area and suppress the generation of boundary artifacts. To better retain the structural information of the image content, a residual network module is added between the down-sampling region and the up-sampling region. In addition, a color loss is added to the loss function to improve the stylization effect of the generated image by constraining the model. Experiments conducted on the self-constructed Dunhuang mural dataset validate the superiority over existing methods of the proposed model in the task of Dunhuang mural art style transfer. This model can generate stylized images of Dunhuang murals with more excellent visual effects and stronger artistic flavor, providing a new idea for innovative research on Dunhuang murals.

Key words: style transfer; cycle generative adversarial network (CycleGAN); CBAM attention mechanism; Dunhuang murals

① 基金项目: 甘肃省重点研发计划-工业类项目 (22YF7GA159); 甘肃省教育厅产业支撑计划 (2023CYZC-25); 甘肃省基础研究计划-软科学专项 (22JR4ZA084)
收稿时间: 2024-09-12; 修改时间: 2024-10-10, 2024-11-29; 采用时间: 2024-12-06; csa 在线出版时间: 2025-03-04

敦煌壁画作为中国古代艺术宝库中的璀璨明珠, 其卓越之处不仅体现在绘画、雕塑与建筑等多重艺术形式的精妙融合, 更在于其色彩艺术的非凡造诣. 这些壁画以其无与伦比的绘画技艺、栩栩如生的人物刻画和对色彩运用的炉火纯青而著称于世, 深刻映射出各个历史阶段独特的审美追求与艺术风貌. 尤为引人注目的是敦煌壁画的色彩艺术, 以其强烈的视觉冲击力与鲜明的个性风格, 成为研究其色彩美学与艺术创新的重要窗口. 因此, 聚焦于敦煌壁画的色彩研究, 不仅是对这一人类文化遗产的深情守护, 更推动了敦煌壁画艺术传承与创新. 随着计算机技术的不断发展, 使用人工智能技术对敦煌壁画图像进行特征提取, 将壁画艺术风格与其他图像相结合, 为艺术作品的二次创作提供了新思路, 也促进了中华优秀传统文化的创造性转化与创新性发展.

图像风格迁移技术是一种计算机视觉技术, 其目的是将一幅风格图像的风格纹理色彩等特征应用于另一幅内容图像上, 同时保持内容图像的主要结构和语义内容不变. 图像风格迁移技术已广泛应用于艺术创作、图像视频编辑和时尚设计等领域, 它赋予了用户独特的创造力, 使他们能够轻松创作出既独特又引人入胜的视觉艺术作品. 最早的图像风格迁移方法主要是使用纹理建模的技术, 这些技术更多的是对像素级别的底层图像特征进行操作, 在处理简单纹理时有一定的效果, 但对于复杂的风格图像迁移任务则显得力不从心. 近年来随着计算机水平的不断发展, 图像风格迁移技术也日新月异. 特别是在卷积神经网络 (convolutional neural network, CNN) 被提出之后, 越来越多的人利用基于深度学习的风格迁移方法在一些领域进行创作, 例如油画[1]、山水画[2]和水墨画[3], 使其成为图像风格迁移任务中的主流方法. Gatys 等人[4]提出了一种基于优化的图像风格迁移研究算法, 利用预训练的 VGG 网络能够有效地提取图像纹理信息和语义信息, 并基于这些不同层次上的信息构建了内容损失和风格损失. 但是这种方法计算成本高且迭代优化过程复杂, 十分耗时. 基于此, Johnson 等人[5]提出了一个快速风格迁移网络模型, 大大提升了风格迁移速度, 但是生成的风格化图像质量不是很好. Liu 等人[6]提出了一种新的自适应注意力归一化模块, 在每个像素点的基础上自适应地执行注意力归一化, 提高了风格化图像的质量. 这些模型在图像风格迁移任务中都取得了很好的效果, 但是在对敦煌壁画进行风格迁移时, 发现得到的最终效果并不理想. 主要存在以下问题: (1) 生成的图片中存在不稳定的因素, 会影响图片质量, 例如会在图片的天空、湖面部分出现不必要敦煌壁画艺术元素; (2) 对敦煌壁画的风格艺术学习能力有限, 不能将敦煌壁画艺术中的色彩元素很好的学习到, 生成的图片效果只是相当于加了一层滤镜. 出现这些问题的原因是敦煌壁画与西方的绘画艺术有着本质的区别. 不像西方艺术中的写实风格, 敦煌壁画的风格更加抽象, 它的色彩厚重且浓烈, 有很强的视觉冲击力, 相比于中国传统绘画, 它的色彩又更加突出, 这些原因导致很难使用现有的方法将敦煌壁画的特征直接应用到目标图片上. 为了实现高质量的敦煌壁画艺术风格迁移任务, 本文提出了一种新的基于循环生成对抗网络 (cycle generative adversarial network, CycleGAN)[7]的融合卷积注意力 (convolutional block attention module, CBAM)[8]的敦煌壁画风格迁移模型来实现从目标图片到敦煌壁画风格的迁移任务. 首先, 为有效减少在天空等空白区域生成不恰当地敦煌壁画风格元素的问题, 引入了 CBAM 注意力机制. 这一机制的引入会优化生成图片过程中的区域处理策略, 通过增强模型对关键区域的关注度, 改善这些区域的视觉效果, 从而确保生成的图像既符合敦煌壁画的艺术特色, 又能在细节处理上更加精准与自然. 此外, 由于敦煌壁画内容过于抽象, 色彩厚重, 整体具有高度概括性, 与现代图片有很大差异, 据此提出了色彩损失, 以缓解域间风格差异过大带来的负面影响. 最后, 通过改进的残差网络结构来更好的保留图像内容信息. 本文的主要贡献可以概括为以下几点: (1) 在生成器网络结构上引入了 CBAM 注意力机制用于引导生成器对主要区域进行特征提取和风格迁移, 以防止在次要区域生成太多的敦煌壁画元素; (2) 提出了针对敦煌壁画风格迁移任务的色彩损失函数, 以缓解域间风格差异过大给风格迁移任务带来的负面影响; (3) 在原有生成器网络的基础上添加残差网络模块, 从而避免梯度消失问题并增加多尺度不变特性. (4) 针对敦煌壁画风格迁移任务缺乏公共数据集的问题, 创建了一个敦煌壁画数据集 DHdata, 该数据集包含了 3 000 张从晋朝到清朝风格的 jpg 格式图片和 5 000 张现实世界图片. 在此数据集上进行了大量实验, 实验结果表明, 所提出的方法在风格迁移过程中不论是风格一致性还是图片内容细节保留方面均优于对照模型, 生成

图像的视觉效果和艺术性更好.

1 相关工作

1.1 无配对图像域的风格迁移

无配对图像域的风格迁移方法适用于图像集合, 它通过将输入映射到基于生成对抗网络 (generative adversarial network, GAN) 的目标域来转移图像风格. Pix2Pix[9]首先将 GAN 应用于图像迁移任务, 它需要配对的训练数据, 但是在某些风格迁移任务中很难获得配对的数据集, 这限制了它的发展. CycleGAN 以 Pix2Pix 的工作为基础, 通过循环一致性网络实现在非配对数据集上的风格迁移, 并取得了很好效果. StyTr^2[10]第一次将视觉 Transformer 应用于图像的风格迁移任务, 利用自注意力机制捕获图像的长程依赖关系, 避免了内容泄露问题. 为了提高风格迁移网络的灵活性, MUNIT[11]和 StarGAN[12]实现了多模态和多域图像翻译. CUT[13]考虑到如果追求过于严格的循环一致性损失可能会限制图像风格迁移的能力, 所以引入了补丁级的对比学习来保留图像内容, 并且还可以只使用一张图片就能进行风格迁移. 然而, 直接使用从同一编码器中提取的特征来计算 CUT 中的对比度损失可能会限制其性能, 这主要受域间风格差异的影响. 为了解决这个问题, LseSim[14]使用空间相关图来计算对比度损失, 捕获图像中的空间关系, 有效克服了域间特异性带来的负面影响. 在壁画的风格迁移领域, Fang 等人[15]提出了一种渐进式风格注意网络 PSANet, 采用多层次损失函数策略, 得到了满意的唐卡壁画图像风格化效果. 由于对敦煌壁画的研究主要集中壁画修复方面, Wang 等人[16]提出的 DunhuangGAN 是目前已知唯一在敦煌壁画风格迁移领域的工作, DunhuangGAN 在基于改进的对比学习框架下, 利用多重损失函数进行优化以得到视觉效果较好的风格化图像.

1.2 注意力机制

注意力机制自提出以来, 就广泛应用于基于 RNN、CNN 等神经网络模型的各种自然语言处理任务中, 在许多计算机视觉任务中均取得了较好的效果, 包括图像分割[17]、目标检测[18]、语义分割[19]、图像生成[20]、三维重建与理解[21]、视频分析[22]和视觉问答[23]等. Vaswani 等人[24]开创性地提出了自注意力机制, 由于其能更好的处理长程依赖关系, 成为在自然语言处理领域最具影响力的模型之一. Wang 等人[25]首次自注意力机制引入计算机视觉领域, 构建了一个通用的捕获长程依赖关系的网络结构, 在视频分类和静态图像识别方面取得了显著成效. 与以往通过多尺度特征融合来捕获上下文的方法不同, Fu 等人[26]提出了一种双重注意力网络, 通过上下文依赖的自注意力机制来获得更精确的语义分割结果. He 等人[27]提出一种基于注意力的自适应金字塔上下文网络模块进行语义分割. 在风格迁移领域, Zhang 等人[28]提出用于风格迁移的分层视觉 Transformer, 在不同的窗口形状中捕获短期和长期依赖, 得到了理想的迁移效果. 聂雄锋等人[29]提出了一种融合注意力机制的多模态动漫风格迁移方法, 在图像和视频风格迁移任务中均取得了较好的效果. Yu 等人[30]提出了协方差注意力网络 (CovAttN) 的特征融合方法, 从通道相关性和空间分布两个方面对齐内容特征和风格特征, 使生成的风格化图像质量更高.

2 本文方法

2.1 敦煌壁画风格迁移模型架构

本文提出的基于循环生成对抗网络的敦煌壁画迁移模型如图 1 所示. 其由两个生成器$G_{\text{attention}}$、G_2和两个判别器D_y、D_c组成, 目的是学习从目标图片域X到敦煌壁画艺术图像域Y的风格映射. 由预训练的 VGG16 网络提取输入图像的特征, 再输入到生成器中. $G_{\text{attention}}$是添加了 CBAM 注意力机制的生成器, 用于实现从目标图片C_1到敦煌壁画艺术图像S_1的风格迁移. 判别器D_y是一个 70×70 的 PatchGAN, 用于判断生成的风格化图像的真实性, 判别器D_c是用于判断 G_2生成的内容图像是否与输入的内容图像一致. 其中, 使用色彩损失让生成图像的风格化效果更好, 并组合使用多个损失函数, 对模型进行优化, 约束模型的训练过程, 以得到高质量的风格化图像. 因为本文研究重点是目标图片到敦煌壁画风格化的迁移, 所以将 CBAM 注意力机制添加在能将目标图片转换为敦煌壁画风格图像的生成器$G_{\text{attention}}$中.

2.2 基于 CBAM 注意力机制的生成器网络

注意力机制源于对人类视觉的研究, 模拟人类在处理复杂信息时能够选择性的关注重要部位的能力, 通过让模型处理输入数据时能够聚焦于关键信息, 同时忽略其他可见信息, 从而提高处理效率和准确性.

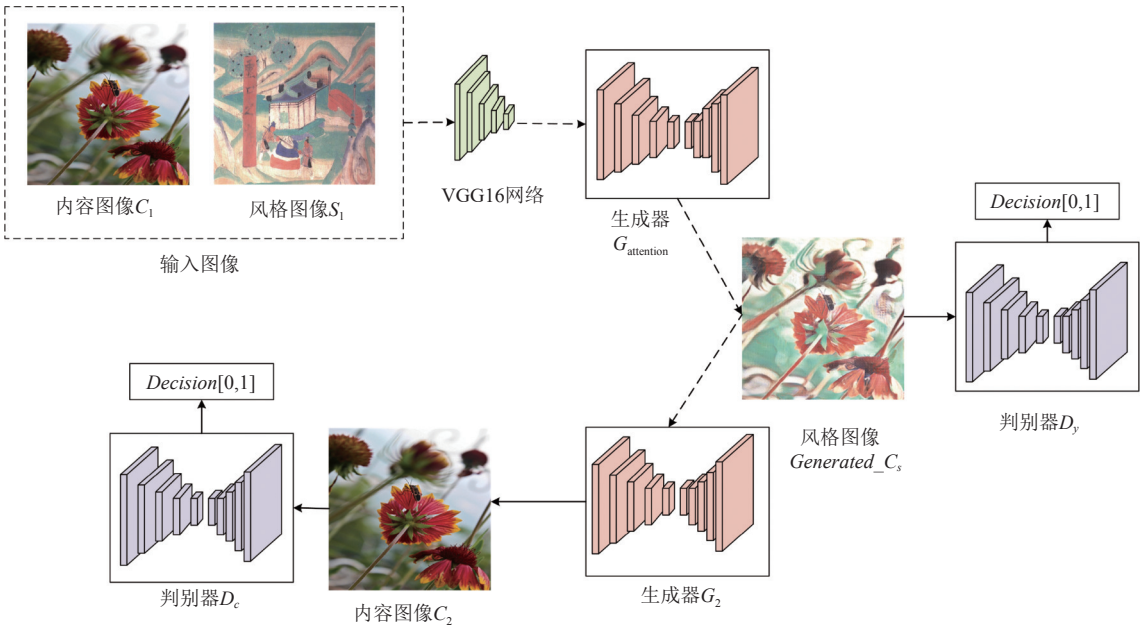

图 1　敦煌壁画风格迁移网络结构

所提出的加入 CBAM 注意力机制的生成器网络结构如图 2 所示. 主要的网络结构是: (1) 3 个下采样模块主要用于提取输入图片的特征, 1 个 CBAM 注意力模块捕捉内容图像和风格图像之间的关系, 并确保在风格迁移过程中保持图像的内容不变, 确保在减小特征图尺寸的同时, 保留全局信息, 并且对每个通道都进行了加权, 以突出重要特征; (2) 中间区 9 个残差网络模块用于特征学习以生成更真实的数据样本; (3) 4 个上采样模块通过反卷积还原图像低级特征, 生成风格化图像.

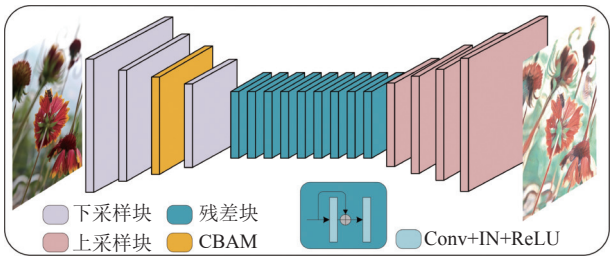

图 2　添加 CBAM 注意力机制的生成器模型

2.2.1　下采样区

在生成器$G_{attention}$网络的浅层设置下进行卷积操作, 采用较大的卷积核尺寸以及较大的步长, 优化特征图的维度处理, 旨在提升计算效率并保留特征的多样性, 为支持后续残差网络模块的高效特征提取. 此下采样区设计了 4 层卷积结构: 首层卷积采用步长为 1, 后 3 层则统一设计步长为 2, 通过一系列的卷积操作, 使输入图像的特征空间被细化缩减至原来的 1/8, 有效压缩了的特征映射的范围.

2.2.2　中间区

尽管中间层的残差网络通过跳跃连接实现了层与层之间的信息通道, 但原图像中的部分关键特征信息仍有可能在传递过程中丢失. 为了更有效地学习图像内容的结构信息, 在下采样的最后和上采样的起始中间嵌入了残差网络模块. 如图 3 所示, 此模块内包含 3 个 3×3 的卷积层, 其中首个卷积层的输出经过权重调整之后与第 2 个卷积层的输出一起作为第 3 个卷积层的输入, 强化了特征信息的传递与保留. η为权重, 具体可表示为:

$$L = \eta L_1 + L_2 \tag{1}$$

其中, L_1代表残差网络模块第 1 层的输出, L_2代表残差网络模块第 2 层的输出, η是权重参数.

2.2.3　上采样区

在生成器$G_{attention}$的上采样区第 2 层引入 CBAM 注意力机制, 整合全局和局部空间信息, 提高生成图像的协调性和质量. CBAM 注意力结构如图 4 所示, 首先将通道数为C、尺寸大小为N的$F_1 \in \mathbb{R}^{C\times H\times W}$作为输入特征映射, 然后将中间状态$F_2$和输出状态$F_3$表示为:

$$F_2 = M_C(F_1) \otimes F_1 \tag{2}$$

$$F_3 = M_S(F_2) \otimes F_2 \tag{3}$$

其中, M_C 和 M_S 分别为通道和空间注意力图, ⊗表示元素的乘法. 通过减少信息弥散和放大全局交互来提高深度神经网络性能, 能更加全面准确地提取到敦煌壁画独特的艺术风格特征.

通道注意力模块通过挖掘特征图各通道之间的关联性, 聚焦于图像中的关键特征区域. 其结构展示如图 5 所示. 为了更有效地计算通道注意力, 首先对输入特征采用最大池化和平均池化来整合其通道信息. 然后将这些信息送入一个含有隐含层的多层感知机 (multi-layer perceptron, MLP) 并行处理, 处理后的特征经过将元素求和操作, 并使用激活函数, 最后生成了通道注意力特征图.

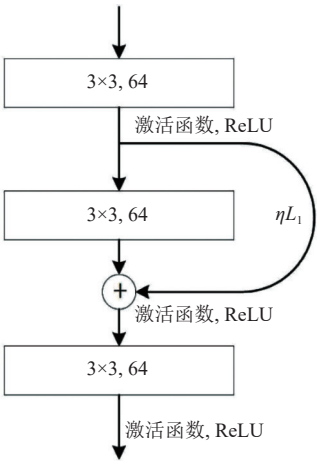

图 3 残差网络模块结构

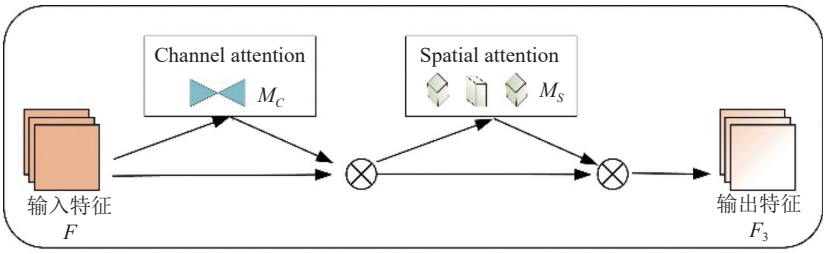

图 4 CBAM 注意力层

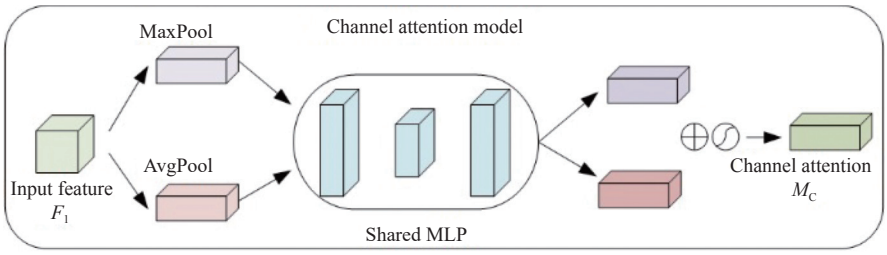

图 5 通道注意力模块

空间注意力模块通过使用两种不同的池化操作来整合特征图的全局信息, 如图 6 所示, 该过程使用两个池化层来压缩通道信息, 分别得到表示整个通道的平均池化特征和最大池化特征两个映射: $F^s_{\text{avg}} \in \mathbb{R}^{1\times H\times W}$ 和 $F^s_{\max} \in \mathbb{R}^{1\times H\times W}$. 然后将这两个特征映射通过卷积层进行融合, 并通过激活函数, 最终生成空间注意力特征图.

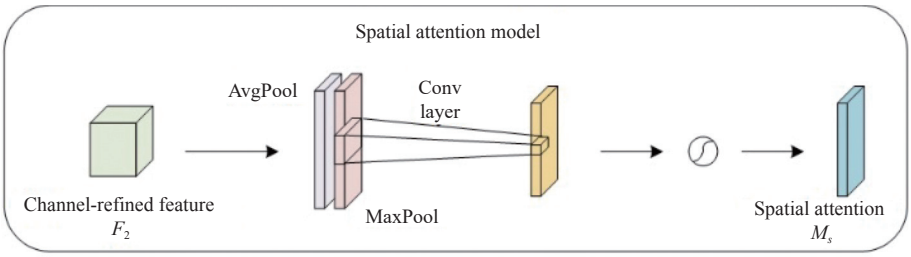

图 6 空间注意力模块

2.3 损失函数

损失函数 L 包含对抗损失 L_1、循环一致性损失 L_2 和色彩损失 L_3 这 3 个部分.

2.3.1 对抗损失

CycleGAN 的对抗损失是 GAN 训练过程的一个关键组成部分, 旨在使生成器能够生成逼真的图像, 对

抗损失L_1如式 (4):

$$\begin{aligned}L_1(G_{\text{attention}},D,X,Y)=&E_{y\sim pdata(y)}\left[\log D_y(y)\right]\\&+E_{x\sim pdata(x)}\left[\log D_c(G_2(x))\right]\\&+E_{x\sim pdata(x)}\left[\log\left(1-D_y(G_{\text{attention}}(x))\right)\right]\\&+E_{y\sim pdata(y)}\left[\log(1-D_c(G_2(y)))\right]\end{aligned}\tag{4}$$

其中, 生成器$G_{\text{attention}}$试图将目标图像生成具有敦煌壁画风格的图像$G_{\text{attention}}(x)$, D_y为尽可能区分生成图像$G_{\text{attention}}(x)$和真实图像Y的判别器. 生成器G_2是将风格化之后的图像重新转换为目标域的图像.

2.3.2 循环一致性损失

对于输入的图像x, 经过生成器$G_{\text{attention}}$生成的风格化图像为$G_{\text{attention}}(x)$, 再通过生成器G_2逆向转换得到$G_2(G_{\text{attention}}(x))$, 这个结果应该与$x$近似相等, 对于输入图像$y$也是相似的操作. 通过最小化循环一致性损失来确保图像在经过转换之后还能回到原始状态, 具体如式 (5):

$$\begin{aligned}L_2(G_{\text{attention}},G_2)=&E_{x\sim pdata(x)}\left[\|G_2(G_{\text{attention}}(x))-x\|\right]\\&+E_{y\sim pdata(y)}\left[\|G_{\text{attention}}(G_2(y))-y\|\right]\end{aligned}\tag{5}$$

2.3.3 色彩损失

敦煌壁画是一种注重彩色的艺术, 其色彩具有很强的装饰性. 用基于 HSV (hue, saturation, value) 色相通道的色彩损失模拟敦煌壁画艺术的色彩, 首先提取能够代表图像颜色特征的色相通道直方图向量, 然后计算生成图像的色相直方图向量与敦煌壁画的余弦相似度, 得到色彩损失L_3如式 (6):

$$L_3(G_{\text{attention}},X,Y)=E_{x\sim pdata(x),y\sim pdata(y)}\left[\frac{Hue(G_{\text{attention}}(x))\times Hue(y)}{\|Hue(G_{\text{attention}}(x))\|\times\|Hue(y)\|}\right]\tag{6}$$

其中, Hue是从图像中提取的色彩通道的 8 分区直方图矢量.

由以上 3 个损失可以得到本文的损失函数L如式 (7):

$$\begin{aligned}L\left(G_{\text{attention}},G_2,D_y,D_c\right)=&L_1(G_{\text{attention}},D,X,Y)\\&+L_2(G_{\text{attention}},G_2)+L_3(G_{\text{attention}},X,Y)\end{aligned}\tag{7}$$

3 实验及结果分析

3.1 实验数据准备

通过对《中国敦煌壁画全集》电子版中的壁画插图进行裁剪收集和扩充, 本文建立了适用于敦煌壁画风格迁移任务的数据集 DHdata, 包括 3 000 张具有典型的敦煌壁画艺术风格的图片和 5 000 张现实世界的图片. 敦煌壁画跨越时间长, 从北魏一直到元朝都有创作, 每个朝代都有其独特的创作手法和艺术风格, 本文在实验中主要使用唐朝壁画进行风格迁移任务.

3.2 实验设置

本文使用 PyTorch 框架搭建网络模型, 在 Linux 环境下使用 NVIDIA A100 GPU 来训练和测试网络模型. 实验中的自建数据集图像大小均为 256×256 px, 批处理参数 batch_size 设置为 4, 训练轮次 epoch 为 200, 学习率为 0.0001.

3.3 定量分析

本文使用结构相似性 (*SSIM*)、*FID* (Frechet inception distance) 和 *IS* (inception score) 作为衡量生成的风格化图像质量好坏的评价指标. *SSIM* 主要通过亮度、对比度、结构 3 个方面评估原图与生成图像相似度的指标, 常用于衡量模型生成图像的真实性, 公式如式 (8):

$$SSIM(x,y)=\frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)}\tag{8}$$

其中, μ_x和μ_y分别是两幅图像块的平均亮度, σ_x^2和σ_y^2分别是两幅图像的对比度, σ_{xy}是两幅图像的协方差, 用以衡量它们的结构相似性, C_1和C_2是避免分母为 0 的常数. *SSIM* 是模拟人眼对图像结构的评估, 更符合人类视觉系统, 能够反映感知上的图像质量. *SSIM* 值越高则代表图像的质量越好, 通过表 1 可以看出, 本文网络模型生成的风格化图像与其他方法生成的图像对比最具有真实性, 在 *SSIM* 指标下取得了最好的效果.

表 1 定量分析

模型	*SSIM*	*FID* ↓	*IS*
文献[4]	0.50	264.19	2.05
ArtFlow	0.63	200.57	3.01
CUT	0.54	288.34	2.62
LseSim	0.61	207.07	3.22
StyTr^2	0.62	195.18	3.12
Ours	0.70	189.52	3.27

FID 是一种用于评估生成模型和真实数据分布之间差异的指标, 广泛用于生成模型的训练和评估中, 是一个客观的评估指标, 用以避免人为主观因素对评估结果的影响, 其计算公式如式 (9):

$$FID(P,G)=\|\mu_P-\mu_G\|^2+\text{tr}\left(\Sigma_P+\Sigma_G-2(\Sigma_P\Sigma_G)^{1/2}\right)\tag{9}$$

其中, P是真实图像分布的特征向量集合, G是生成图

像分布的特征向量集合, μ_P 和 μ_G 分别表示 P 和 G 的特征向量集合的均值, Σ_P 和 Σ_G 分别是 P 和 G 的特征向量集合的协方差矩阵, $\mathrm{tr}\left(\Sigma_P+\Sigma_G-2(\Sigma_P\Sigma_G)^{1/2}\right)$ 是协方差矩阵的迹的平方根. 主要评价生成图片的质量和多样性, *FID* 值越低代表生成图像越接近真实的图像分布. 本文方法在 *FID* 指标上表现同样优秀.

IS 是一个衡量图像生成模型综合质量的评价指标, 将生成器生成的图像模型输入到 Inception V3 Network 中, 再对该网络的输出值做统计分析, 其计算公式如式 (10):

$$IS(G)=\exp(\mathbb{E}_{x\sim p_g}D_{\mathrm{KL}}(p(y|x)\|p(y))) \tag{10}$$

其中, $x\sim p_g$ 表示 x 是从 p_g 中生成的图像样本, $D_{\mathrm{KL}}(p\|q)$ 是分布 p 和 q 间的 KL 散度, $p(y|x)$ 是给定图像 x 分类为 y 的概率, $p(y)=\int_x p(y|x)p_g(x)$ 表示类别的边缘分布, exp 是便于比较最终计算的 *IS* 值.

IS 用 KL 散度综合图像质量和图像多样性两个指标来评价模型的好坏, *IS* 值越大说明模型效果越好, 在综合考虑生成图片清晰度和多样性指标 *IS* 下, 注重保留内容的 StyTr^2 和本文所提出方法几乎不分上下, 即使它在 *FID* 评价指标上表现并不突出. 综合比较而言, 本文提出方法在风格质量和内容保持方面均表现出色.

3.4 定性评价

为验证所本文模型的有效性, 使用几个主流模型所做实验效果如图 7 所示, 包括 Gatys 等人[4]提出的基于图像迭代优化的方法、可逆网络风格迁移 ArtFlow、引入了对比学习的 CUT 网络、基于图像空间特征图的 LseSim 和使用 Transformer 网络架构的 StyTr^2. 文献[4]虽然能迁移部分颜色和纹理, 但是迁移后内容丢失严重图片较模糊. ArtFlow 虽然避免了对通用迁移过程中的内容泄露, 但是对某些物体迁移过程中的细节还是有所欠缺, 例如第 1 行塔尖的缺失以及第 5 行的整体背景变色等. 相比其他方法, CUT 和 LseSim 可以学习到更抽象的艺术风格, 但是在某些区域还是会被模糊处理, 尤其 CUT 方法的第 1 行天空区域出现了明显不必要的纹理, 最后 1 行房屋底部的台阶缺失. 而 LseSim 则没有很好的学习到敦煌壁画色彩特征, 整体风格偏暗, 偏向于冷色调. StyTr^2 在壁画迁移任务中取得了较好的效果, 但是经过迁移后, 图片中的细节部分, 例如第 1 行和第 2 行的行人部分被模糊处理. 相比之下, 本文所提出的方法不仅能够保留内容的显著特征和细节纹理, 还能通过学习内容与风格之间的关系, 有效整合二者. 这样, 生成的结果既保留了原始的语义信息, 又融入了风格图像的丰富色彩.

为了更加直观地看到不同模型间的差异及本文模型添加 CBAM 注意力机制的有效性, 对于不同模型的实验结果如图 8 所示. 从第 1 行方框部分中可以看到本文所提出的模型对原图的细节部分有很大程度的保留, 而文献[4]整体内容保留的不是很好, ArtFlow、CUT 和 StyTr^2 等模型在细节部分也有所失真. 相比较而言, LesSim 和本文模型则能够很好地将内容图像的细节保留. 从图 8 的第 2 行可以看出, 在图片的顶端部分, CUT 和 LseSim 模型出现了很多不规则的纹理, 极大降低了风格化图片的质量和美感, 而文献[4]和 StyTr^2 模型所生成的风格化图像则有明显的阴影产生. 对比其他模型, 本文所提出的模型能有效抑制不必要纹理的生成. 从图 8 的第 3 行中, 也可以直观地看到本文所提出的模型, 在风格化之后的图像空白区域有更少的伪影产生, 说明了添加注意力机制的有效性和必要性.

同时, 本文设计了一份针对生图片的色彩、清晰度等方面的调查问卷, 因为对生成的风格化图片存在个人的主观感受, 所以用户的直观感受也是评价图片风格迁移效果的一项关键性指标. 随机选取 6 张内容图片与其风格化之后的图片, 共有 72 人进行投票, 与 5 种对比模型生成的风格化图片进行比较, 参与者选出自己认为生成效果最好的一张, 结果以百分比表示. 从表 2 可以看出, 用户对本文所提出的模型生成的敦煌艺术风格化图片有更高的关注度.

表 2 用户调查 (%)

模型	风格一致性	内容保留	整体质量
文献[4]	4.8	5.2	4.3
ArtFlow	10.1	8.9	7.8
CUT	18.7	12.4	15.4
LseSim	17.3	25.4	20.1
StyTr^2	19.6	22.5	17.2
Ours	29.5	25.6	35.2

3.5 消融实验

3.5.1 CBAM 注意力模块消融实验

为了验证添加 CBAM 注意力模块的有效性, 对是否添加注意力模块分别做了实验, 实验结果如图 9 所示. 图 9(b) 是没有添加 CBAM 注意力模块风格迁移的结果, 图 9(c) 是添加 CBAM 注意力模块之后风格迁移的结果.

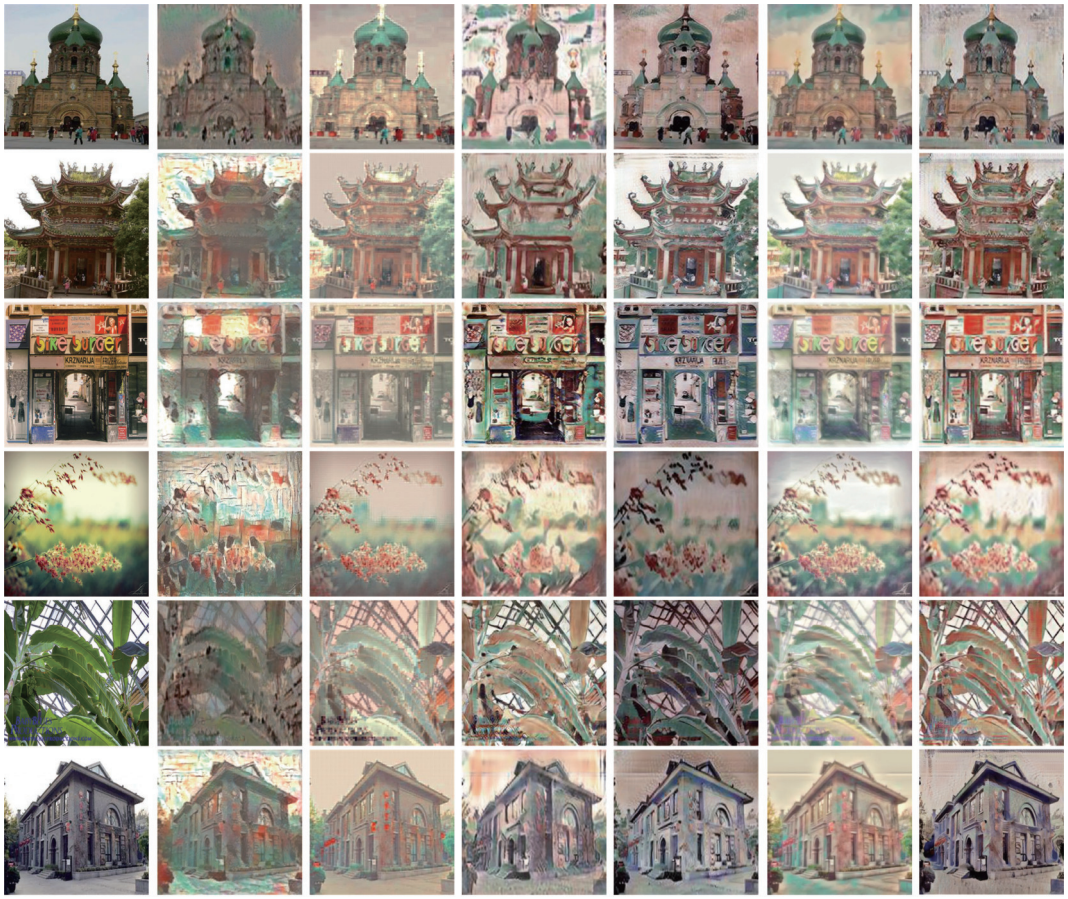

图 7　对比试验

图 8　不同模型实验结果差异

实验结果显示, 未融入 CBAM 注意力机制的模型在风格迁移之后生成的图片质量有明显不足, 具体表现在图 9(b) 这一列, 明显生成了不必要的风格特征, 尤其第 2、3、4 行在空白区域产生了突兀的斑点, 影响了图片的整体观感. 但是在添加了 CBAM 注意力机制后, 可以看到在生成图像原来有斑点的位置, 都变得更自然, 整体图片风格更协调, 有效提升了的生成图像的质量, 证明了所添加模块的有效性.

(a) 内容图像 (b) 没有添加CBAM (c) 添加CBAM

图 9 CBAM 注意力消融实验

3.5.2 色彩损失函数消融实验

为了验证所提出的色彩损失函数的有效性, 表 3 提供了对色彩损失函数消融实验的量化结果, 直观地反映了该损失函数对整个网络的影响. 从表 3 的数据可以清晰地看出, 去掉色彩损失函数之后生成的风格化图像无法很好地学习到敦煌壁画的风格特征, 添加损失函数之后不管是生成图像的结构相似性还是与真实图像的相似度均优于基准实验. 从表 3 可以看出, 色彩损失函数在内容图像的风格化过程中是比较重要的.

表 3 消融实验的 *FID* 对比

模型	*SSIM*	*FID* ↓
w/o Lcolor	0.65	199.47
Full net	0.70	189.52

注: w/o表示没有使用

4 结论与展望

本文提出了一种基于 CycleGAN 的融合 CBAM 注意力机制的敦煌壁画风格迁移方法. 为了提高对敦煌壁画艺术风格的特征提取能力, 利用 CBAM 注意力机制中的通道注意力和空间注意力得到风格图像的通道特征和聚合之后的空间特征; 添加的残差跳跃连接更有效地学习图像内容的结构信息, 高效整合全局信息; 添加色彩损失函数可以约束风格迁移之后的生成图像与真实的风格域图像更接近. 实验结果表明, 所提方法生成的风格化图像质量相较于基准模型不管在主观还是客观方面都有一定的提升. 由于敦煌壁画具有鲜明的各个朝代独特的风格, 下一步将专注于研究在特定的洞窟或朝代中生成风格化图像.

参考文献

1 Liu Y. Improved generative adversarial network and its application in image oil painting style transfer. Image and Vision Computing, 2021, 105: 104087. [doi: 10.1016/j.imavis.2020.104087]

2 Gui XQ, Zhang BX, Li L, *et al*. DLP-GAN: Learning to draw modern Chinese landscape photos with generative adversarial network. Neural Computing and Applications, 2024, 36(10): 5267–5284. [doi: 10.1007/s00521-023-09345-8]

3 He B, Gao F, Ma DQ, *et al*. ChipGAN: A generative adversarial network for Chinese ink wash painting style transfer. Proceedings of the 26th ACM International Conference on Multimedia. Seoul: ACM, 2018. 1172–1180.

4 Gatys LA, Ecker AS, Bethge M. Image style transfer using convolutional neural networks. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 2414–2423.

5 Johnson J, Alahi A, Li FF. Perceptual losses for real-time style transfer and super-resolution. Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 694–711.

6 Liu SH, Lin TW, He DL, *et al*. AdaAttN: Revisit attention mechanism in arbitrary neural style transfer. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 6629–6638.

7 Zhu JY, Park T, Isola P, *et al*. Unpaired image-to-image translation using cycle-consistent adversarial networks. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 2242–2251.

8 Woo S, Park J, Lee JY, *et al*. CBAM: Convolutional block attention module. Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich: Springer, 2018. 3–19.

9 Isola P, Zhu JY, Zhou TH, *et al*. Image-to-image translation with conditional adversarial networks. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 5967–5976.

10 Deng YY, Tang F, Dong WM, *et al*. StyTr^2: Image style transfer with Transformers. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 11316–11326.

11 Huang X, Liu MY, Belongie S, *et al*. Multimodal unsupervised image-to-image translation. Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich: Springer, 2018. 179–196.

12 Choi Y, Choi M, Kim M, *et al*. StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 8789–8797.

13 Park T, Efros AA, Zhang R, *et al*. Contrastive learning for unpaired image-to-image translation. Proceedings of the 16th European Conference on Computer Vision. Glasgow: Springer, 2020. 319–345.

14 Zheng CX, Cham TJ, Cai JF. The spatially-correlative loss for various image translation tasks. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 16402–16412.

15 Fang J, Li H, Jia Y, *et al*. Thangka mural style transfer based on progressive style-attentional network and multi-level loss function. Journal of Electronic Imaging, 2023, 32(4): 043007.

16 Wang WN, Li YF, Ye H, *et al*. DunhuangGAN: A generative adversarial network for dunhuang mural art style transfer. Proceedings of the 2022 IEEE International Conference on Multimedia and Expo (ICME). Taipei: IEEE, 2022. 1–6.

17 Rahman MM, Marculescu R. Medical image segmentation via cascaded attention decoding. Proceedings of the 2023 IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2023. 6211–6220.

18 Zhou Q, Shi HM, Xiang WK, *et al*. DPNet: Dual-path network for real-time object detection with lightweight attention. IEEE Transactions on Neural Networks and Learning Systems, 2024: 1–15.

19 Li X, Xu F, Liu F, *et al*. A synergistical attention model for semantic segmentation of remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 5400916.

20 Cao MD, Wang XT, Qi ZG, *et al*. MasaCtrl: Tuning-free mutual self-attention control for consistent image synthesis and editing. Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision. Paris: IEEE, 2023. 22503–22513.

21 Zhu ZY, Ma XJ, Chen YX, *et al*. 3D-VisTA: Pre-trained Transformer for 3D vision and text alignment. Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision. Paris: IEEE, 2023. 2899–2909.

22 Pan YC, Shang YY, Liu T, *et al*. Spatial-temporal attention network for depression recognition from facial videos. Expert Systems with Applications, 2024, 237: 121410. [doi: 10.1016/j.eswa.2023.121410]

23 Lu SY, Liu MZ, Yin LR, *et al*. The multi-modal fusion in visual question answering: A review of attention mechanisms. PeerJ Computer Science, 2023, 9: e1400. [doi: 10.7717/peerj-cs.1400]

24 Vaswani A, Shazeer N, Parmar N, *et al*. Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.

25 Wang XL, Girshick R, Gupta A, *et al*. Non-local neural networks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 7794–7803.

26 Fu J, Liu J, Tian HJ, *et al*. Dual attention network for scene segmentation. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 3141–3149.

27 He JJ, Deng ZY, Zhou L, *et al*. Adaptive pyramid context network for semantic segmentation. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 7511–7520.

28 Zhang CY, Xu XG, Wang L, *et al*. S2WAT: Image style transfer via hierarchical vision Transformer using strips window attention. Proceedings of the 38th AAAI Conference on Artificial Intelligence. 2024. 781.

29 聂雄锋, 王俊英, 董方敏, 等. 融合注意力机制的多模态动漫风格迁移方法. 计算机工程与应用, 2023, 59(15): 223–234. [doi: 10.3778/j.issn.1002-8331.2204-0338]

30 Yu XM, Zhou G. Arbitrary style transfer via content consistency and style consistency. The Visual Computer, 2024, 40(3): 1369–1382. [doi: 10.1007/s00371-023-02855-5]

(校对责编: 王欣欣)