

DOI:10.11918/202504011

# 融合多尺度特征的敦煌壁画风格迁移

曹岩, 郭炳森, 冯丹丹, 张焱, 辛子昊

(兰州交通大学 电子与信息工程学院, 兰州 730070)

**摘要:**为解决现有风格迁移技术在处理敦煌壁画时遇到的因高饱和度矿物颜料、精细纹理及层次结构复杂导致的色彩失真、细节模糊、层次结构失调问题,本文提出多尺度敦煌风格迁移网络,改进循环生成对抗网络来实现高质量敦煌壁画风格迁移。通过引入自适应局部膨胀卷积网络,结合可变形卷积动态捕捉细节纹理边缘和膨胀卷积增强纹理远距离关联性,有效恢复深层特征捕捉壁画笔触细节。设计双域网络,通过全局注意力分支建模壁画整体色调协调性,局部分组卷积分支强化笔触细节,解决迁移过程中的信息丢失和色彩层次弱化问题。提出路径融合网络,利用多膨胀率深度可分离卷积并行处理与动态门控融合机制,优化元素间逻辑关系与比例协调。研究表明,本文所提方法在 FID、LPIPS 和 L2 损失指标上分别降低了 5.81%、4.36% 和 5.73%,而 SSIM 提升了 8.12%。用户调研显示,其在内容保真度、风格匹配度和视觉吸引力方面表现突出。本文方法有效解决了迁移敦煌壁画时色彩层次、纹理细节与空间布局的保留难题,可为敦煌壁画艺术的数字化保护与创新传播提供新的思路。

**关键词:**深度学习;风格迁移;生成对抗网络;多尺度特征;敦煌壁画

中图分类号: TP391.41

文献标志码: A

文章编号: 0367-6234(2026)05-0073-10

## Style transfer of Dunhuang murals fusing multi-scale features

CAO Yan, GUO Bingsen, FENG Dandan, ZHANG Yi, XIN Zihao

(School of Electronic & Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China)

**Abstract:** To address color distortion, detail blurring, and structural incoherence in existing style transfer techniques when processing Dunhuang murals—caused by highly saturated mineral pigments, intricate textures, and complex layered structures—this paper proposes a Multi-scale Dunhuang Style Transfer Network based on an improved Cycle-Consistent Generative Adversarial Network for high-quality artistic style transfer. We introduce an adaptive local dilated convolutional-net that dynamically captures detailed texture edges using deformable convolution and enhances long-range texture dependencies through dilated convolution, thereby restoring deep features to preserve brushstroke details. A dual scope net is designed to mitigate information loss and color-layer weakening during style transfer, employing a global attention branch to model overall tonal harmony and a local grouped convolution branch to reinforce stroke details. Additionally, a pathwise fusion net optimizes logical relationships and proportional coordination between elements using multi-dilation-rate depthwise separable convolutions for parallel processing and a dynamic gated fusion mechanism. Experimental results show that the proposed method achieves reductions of 5.81%, 4.36%, and 5.73% in FID, LPIPS, and L2 loss, respectively, and an improvement of 8.12% in SSIM. User studies confirm its superiority in content fidelity, style consistency, and visual appeal. This approach effectively resolves challenges in preserving color layers, texture details, and spatial layouts in Dunhuang murals transfer, offering a novel approach for Dunhuang murals digitization and innovative dissemination.

**Keywords:** deep learning; style transfer; generative adversarial network; multi-scale features; Dunhuang murals

敦煌壁画作为墙壁上的视觉艺术表现形式,构成了世界艺术宝藏中的重要部分,其因悠久的历史、丰富多样的绘画内容以及卓越的艺术价值而著称<sup>[1]</sup>。然而,由于其存在形式的独特性与创作过程的复杂度,使得敦煌壁画艺术相较于油画和水墨画

等其他艺术形式更难以普及。此外,自然因素及人为活动导致的剥落等多种破坏,使其保存和传承变得越来越迫切。因此,实现敦煌壁画艺术形象的自动化创作,对于保护并推动敦煌壁画艺术的发展具有至关重要的意义。

收稿日期: 2025-04-03; 录用日期: 2025-06-06; 网络首发日期: 2025-09-01

网络首发地址: <https://link.cnki.net/urlid/23.1235.T.20250901.1446.005>

基金项目: 甘肃省自然科学基金(23JRRA913, 25JRRA177); 中央引导地方科技发展资金(25ZYJF001)

作者简介: 曹岩(1982—), 男, 副教授

通信作者: 郭炳森, 17335382899@163.com

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纹理与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<sup>[2]</sup>。将敦煌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<sup>[3-4]</sup>。

Gatys 等<sup>[5]</sup>首次将卷积神经网络引入到图像风格转移中,通过最小化内容和风格损失,实现了从一张图像到另一张图像的风格转移。单模型多风格方法<sup>[6]</sup>,通过单一网络实现多样式输出。例如,Dumoulin 等<sup>[7]</sup>采用条件实例归一化技术,通过调整归一化参数实现多风格适配;Zhang 等<sup>[8]</sup>设计的多风格生成网络,利用风格图像的二阶统计特征驱动迁移过程。然而,此类方法因风格表征共享易导致特征混淆,特定风格细节(如笔触独特性)可能被弱化。单模型单风格方法<sup>[9]</sup>,则为每种艺术形式独立构建专用网络。例如,Johnson 等<sup>[10]</sup>通过引入感知损失函数优化网络参数,实现高精度快速迁移;Zhang 等<sup>[11]</sup>针对动漫风格开发残差 U-net 架构。Hu 等<sup>[12]</sup>提出一种基于细节特征提取融合的中国山水画风格迁移网络,该网络主要针对山水画的独特风格,较大程度地保留山水场景的特征。因此,相比之下,在对敦煌壁画进行风格迁移时采用单模型单风格迁移的方法更具说服力。

除基于卷积神经网络的模型外,生成对抗网络<sup>[13]</sup>提供了一种通过对抗性训练实现跨域图像转换的有效范式,使输入图像具有目标域的风格。Cheng 等<sup>[14]</sup>引入潜在扩散模型(latent diffusion model, LDM),在此基础上进行内容反转和概率融合,但对自然图像的影响并不显著,在某些情况下可能很难生成预期的风格化图像。为了解决这一问题,一些方法引入了循环一致性约束,如 CycleGAN 模型<sup>[15]</sup>和 DualGAN 模型<sup>[16]</sup>,通过将生成图像与反向变换后的原始输入进行比较,以保持输入图像的内容不变。Wang 等<sup>[3]</sup>提出了一种基于改进对比学习框架的 DunhuangGAN 模型,该模型通过优化特征对齐和风格判别机制,显著增强了敦煌壁画艺术风格迁移的准确性和视觉保真度。Gui 等<sup>[4]</sup>提出了基于 CycleGAN 的融合 CBAM 注意力机制的敦煌壁画风格迁移方法,生成的敦煌壁画风格图像在视觉效果和艺术性上得到提升。

近年来,随着对比语言-图像预训练和潜在扩散模型的发展,基于示例到图像的生成模型因其卓越的生成能力而受到广泛关注。Zhang 等<sup>[17]</sup>提出了一种基于反转的扩散模型 InST,通过优化文本嵌入空间实现细粒度的示例驱动风格迁移。Wang 等<sup>[18]</sup>提出了 StyleDiffusion 模型,基于扩散模型的特征解耦策略,以增强对图像内容与风格的可控编辑能力。

Huang 等<sup>[19]</sup>提出的 QuantArt 模型提升了风格迁移视觉保真度,但仍面临生成准确性的挑战。针对这一问题,Chung 等<sup>[20]</sup>提出 StyleID,一种基于预训练扩散模型的风格迁移方法,通过替换自注意力机制中的键和值实现风格适配,无需额外优化或文本监督。此外,Huang 等<sup>[21]</sup>提出了一种免训练的注意力驱动风格迁移框架,通过替换自注意力层键值适配参考风格,实现高效的风格特征融合。

本文以 CycleGAN 模型<sup>[15]</sup>为基线架构,对敦煌壁画这一独特艺术形式进行针对性优化。敦煌壁画以矿物颜料为主,色彩饱和度高且具有独特的古朴质感。在进行风格迁移任务时,处理高饱和度颜色可能会导致色彩失真或褪色无法准确复现壁画中矿物颜料的层次感和历史沉淀感。且壁画中的山水、人物和建筑元素具有丰富的细节和纹理,在迁移过程中可能会丢失这些精细的特征导致生成图像显得模糊或失真。敦煌壁画注重层次感和空间布局,山水、人物和建筑元素之间具有明确的比例协调性,模型往往难以理解壁画中元素的层次关系,可能导致生成的山水背景呈现扁平化或杂乱堆砌。针对以上问题,本文提出一种新的多尺度敦煌风格迁移网络(multi-scale Dunhuang style transfer network)来实现从目标图片到敦煌壁画风格的迁移任务。

## 1 多尺度敦煌风格迁移网络

### 1.1 整体网络架构

选择 CycleGAN<sup>[15]</sup>模型作为本文敦煌壁画迁移模型的骨干网络,所提出的迁移模型如图 1 所示,其由两个互为镜像的生成对抗网络构成闭环系统,包括两个改进后的生成器和两个判别器。两个方向的转换路径均配备了改进后的生成器和判别器,形成双向对称的对抗训练机制。

改进的生成器借鉴深度级联设计思想,由编码器、转换器和解码器构成,编码器的结构由初始卷积层、两个下采样块和双域网络(dual scope net, DSNT)组成。解码器结构包括另外一个双域网络、两个上采样块、自适应局部膨胀卷积网络(adaptive local dilated convolutional-net, ALDC-NT)和输出卷积块。在编码器和解码器之间的连接区域是转换器,由残差模块组和相同数量的路径融合网络(pathwise fusion net, PFNT)并联而成,负责恢复和增强数据信息,保留更多输入图像的内容特征。其中,初始卷积层、下采样块与上采样块均采用统一的构建范式,即 Conv2d 和 ConvTranspose2d 后依次接实例归一化(InstanceNorm2d)与 ReLU 激活,分别简记为 CIR 与 DIR;最终输出卷积层单独采用 Tanh 激活。判别器的结构使用 PatchGAN<sup>[15]</sup>的设计,对输入的图像进行卷积、归一化和激活操作。

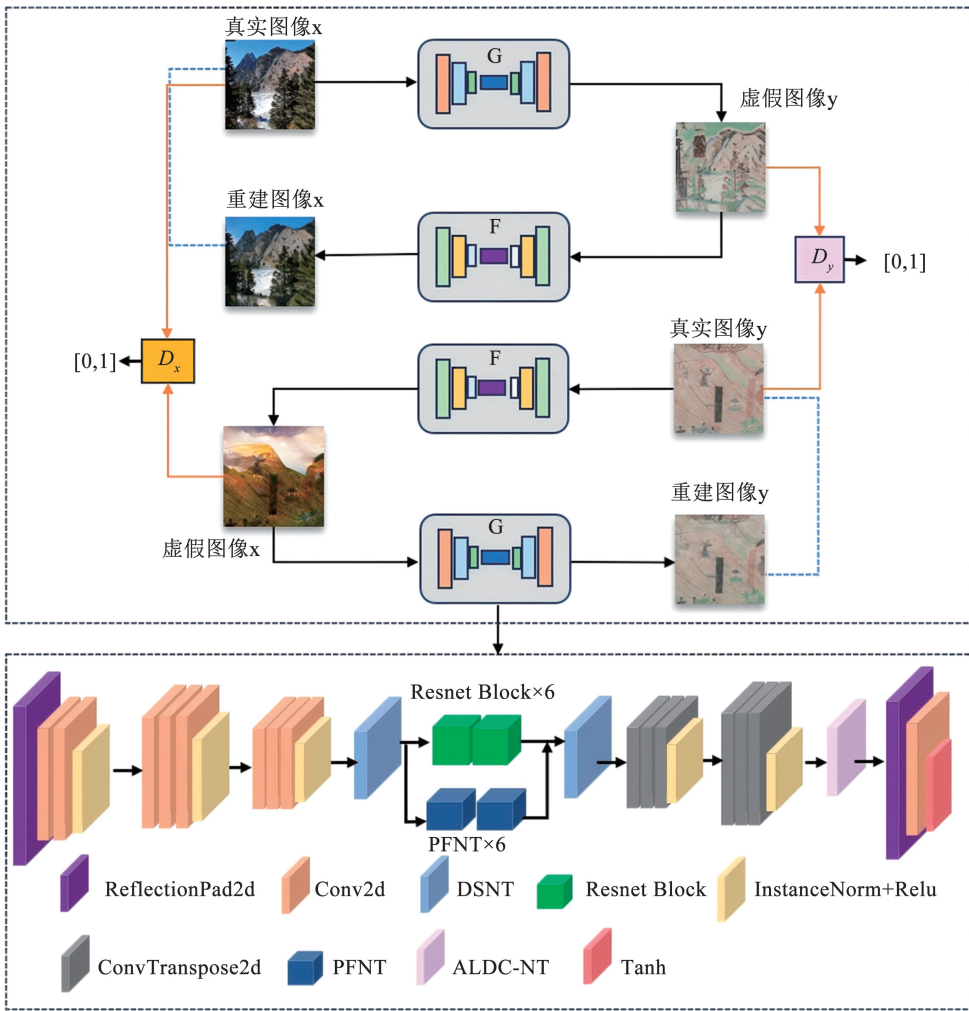


图 1 总体模型框架

Fig. 1 Overall model framework

### 1.2 双域网络 DSNT

针对生成器下采样过程中特征图空间压缩导致的结构信息丢失和色彩层次弱化问题,本文提出双域网络 DSNT,如图 2 所示,其核心思想是通过局部分支与全局分支的协同建模,局部分支通过细粒度特征提取,有效保留了壁画独特的笔触细节和纹理特征,而全局分支则通过长程依赖建模,确保了色彩分布的协调性与风格一致性。

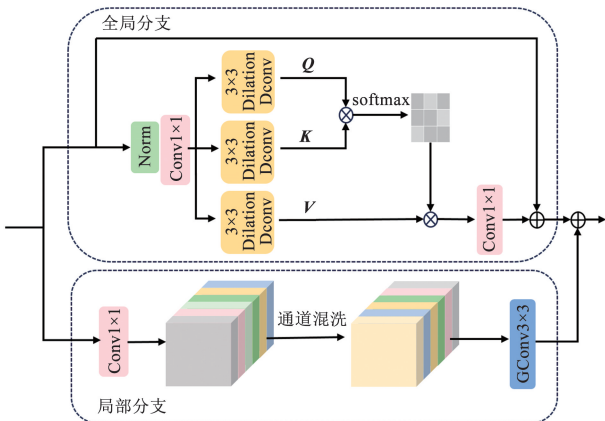


图 2 双域网络

Fig. 2 Dual scope net

在 DSNT 的局部分支中,首先使用  $1 \times 1$  卷积来调整通道维度,将输入特征映射到目标通道数,以减少后续计算冗余。随后执行通道混洗操作以进一步混合通道信息,通道混洗将输入张量沿通道维度划分为组,其中每组采用深度可分离卷积来诱导通道混洗。而后将每组得到的输出张量沿通道维度连接,生成新的输出张量。最后利用分组数为 2 的  $3 \times 3$  分组卷积 (grouped convolution, GC) 增强局部特征提取能力,在减少参数数量的同时增加局部特征多样性,促进特征多样性以及与全局分支互补。局部分支表述为

$$L = G_{C_3 \times 3}(S(C_{1 \times 1}(U))) \quad (1)$$

式中: $L$  为局部通道的输出, $G_{C_3 \times 3}$  为卷积核大小为  $3 \times 3$  的分组卷积, $S$  为通道混洗操作, $C_{1 \times 1}$  为  $1 \times 1$  的卷积, $U$  为输入特征。

在 DSNT 的全局分支中,首先通过层归一化、 $1 \times 1$  卷积和膨胀率为 2 的  $3 \times 3$  的膨胀卷积进行特征规范化与映射,生成查询 ( $Q$ )、键 ( $K$ ) 和值 ( $V$ ) 3 个形状为  $H \times W \times C$  的张量。随后  $Q$  被重塑为

$Q_1 \in \mathbb{R}^{HW \times C}$ ,  $K$  被重塑为  $K_1 \in \mathbb{R}^{C \times HW}$ 。使用  $Q$  和  $K$  的点积计算注意力分数,显著降低了计算复杂度。全局分支表述如下:

$$F = C_{1 \times 1} \text{Attention}(Q_1, K_1, V_1) + U \quad (2)$$

$$\text{Attention}(Q_1, K_1, V_1) = V_1 \text{Softmax}(K_1 Q_1 / \alpha) \quad (3)$$

式中: $F$  为全局通道的输出,  $C_{1 \times 1}$  为卷积核大小为  $1 \times 1$  的卷积,  $\alpha$  为一个可学习的缩放参数,用于在应用 Softmax 函数之前控制  $Q_1$  和  $K_1$  的矩阵乘法的大小。整体模块计算的输出计算为

$$\text{Out} = L + F \quad (4)$$

在生成器的关键下采样模块之后加入该网络,用于补偿空间压缩导致的信息损失,并在上采样模块前提前建模跨区域的色彩关联性,避免高频细节在插值过程中被平滑。这种互补式结构能够在生成过程中精确维持敦煌壁画的艺术特质,包括其独特的色彩层次表现、历史沉淀形成的质感以及整体美学风格。

### 1.3 路径融合网络 PFNT

针对生成器中转换器的残差块在敦煌壁画风格迁移中结构层次建模不足、多尺度特征缺失等问题,本文设计路径融合网络 PFNT,如图 3 所示,通过多膨胀率深度可分离卷积并行处理与动态门控融合机制,实现多层次特征的自适应提取与融合。

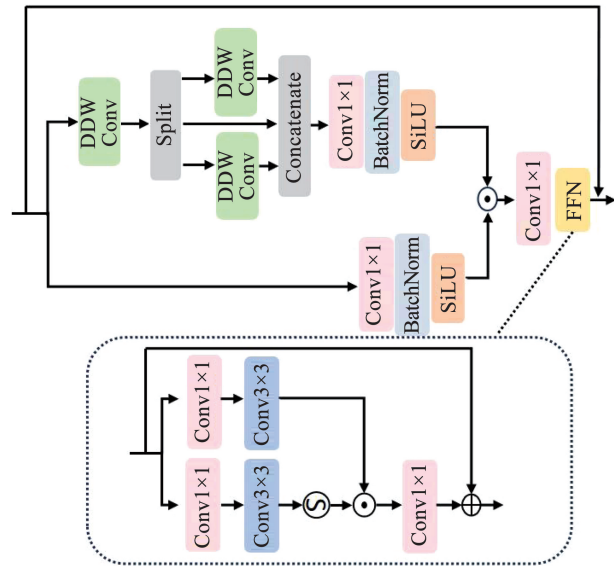


图 3 路径融合网络

Fig. 3 Pathwise fusion net

该网络使用 3 个具有膨胀比  $d \in \{1, 2, 3\}$  的不同 DWConv 层来捕获低、中和高阶间的相互作用。给定输入  $x_e \in \mathbb{R}^{C_e \times HW}$  首先经过  $DW_{5 \times 5, d=1}$  提取低阶特征,并沿通道维度分为 3 组:  $x_a \in \mathbb{R}^{C_a \times HW}$ ,  $x_b \in \mathbb{R}^{C_b \times HW}$ ,  $x_d \in \mathbb{R}^{C_d \times HW}$ 。随后,  $x_a, x_d$  分别分配给  $DW_{5 \times 5, d=2}, DW_{7 \times 7, d=3}$  扩大感受野以捕捉纹理连续性,建模元素间远距离空间关系,而  $x_b$  用作原始信

息通路,保持映射不变。而后将  $x_a, x_b, x_d$  的输出连接形成多阶上下文,表示为  $Y = \text{Concat}(Y_a, Y_b, Y_d)$ ,并通过  $1 \times 1$  逐点卷积融合来自所有分支的多阶特征, BatchNorm 和 SiLU 激活函数用于值分支和门控分支。随后将得到的特征与另一高层次特征分支进行通道间的信息交互。最后将融合后特征通过  $1 \times 1$  卷积和 FFN(前馈网络)进一步非线性变换增强模型的表达能力,加入跳跃连接优化过程中的稳定性和收敛速度。整体模块的输出如式(5)、式(6)所示:

$$M = \text{SiLU}(\text{BN}(\text{Conv}_{1 \times 1}(x_e))) \odot \text{SiLU}(\text{BN}(\text{Conv}_{1 \times 1}(Y))) \quad (5)$$

$$T = \text{FFN}(C_{\text{conv}_{1 \times 1}}(M)) + x_e \quad (6)$$

将多尺度特征提取模块与生成器的残差连接组并联集成,通过层次化特征融合机制增强模型对图像语义的理解能力。显著提升模型在处理敦煌壁画时对层次感和空间布局的理解能力,保留山水、人物和建筑元素之间的明确逻辑关系且比例协调,减少了生成图像出现扁平化或杂乱堆砌的问题。

### 1.4 自适应局部膨胀卷积网络 ALDC-NT

改进后的生成器上采样模块可能无法有效恢复深层特征,如敦煌壁画的细节笔触、纹理,导致生成图像边缘模糊或细节失真。为了更精细地处理这些深层特征,本文设计了 ALDC-NT 网络,如图 4 所示,其中,可变形卷积动态捕捉纹理细节边缘,膨胀卷积则增强纹理的远距离关联性。



图 4 自适应局部膨胀卷积网络

Fig. 4 Adaptive local dilated convolutional-net

ALDC-NT 是一个 U 形结构,由两个线性可变形卷积(linear deformable convolution, LDC)<sup>[22]</sup>层和膨胀率  $R$  为 2、4 和 4 的膨胀卷积层组成。首先对输入  $x$  应用可变形卷积操作进行批量归一化和 ReLU 激活得到  $x_1$ 。然后,将输出经过膨胀率为 2 和 4 的膨胀卷积,并进行批量归一化和 ReLU 激活得到特征  $dx_1, dx_2$ ,其中膨胀率为 2 的分支聚焦局部笔触细节,膨胀率为 4 的分支建模复杂纹理关系。其次,将  $dx_1, dx_2$  在通道维度上拼接,并对该结果再次应用膨胀率为 4 的膨胀卷积,得到  $Dconv_3$  而后将  $x_1$  和  $Dconv_3$  在通道维度上拼接,以防止卷积操作造成的信息丢失。再次,应用可变形卷积操作,动态调整卷积核的位置,以适应不同区域的特征变化,确保壁画

中的细腻线条和复杂纹理被精确再现。最后进行批量归一化和 ReLU 激活输出  $F_{out}$ 。网络的前向传播见式(7)。

$$F_{out} = LD_{Conv2}(\text{Concat}(\mathbf{x}_1, D_{Conv3}(\text{Concat}(\mathbf{dx}_1, \mathbf{dx}_2)))) \quad (7)$$

式中:  $D_{Conv3}$  是将  $\mathbf{dx}_1, \mathbf{dx}_2$  在通道维度上拼接, 并对该结果应用膨胀率为 4 的膨胀卷积操作,  $LD_{Conv2}$  是  $\mathbf{x}_1$  和  $D_{Conv3}$  在通道维度的拼接后应用可变形卷积操作。

将 ALDC-NT 加入模型上采样模块后, 充分利用了网络深层特征的特点, 前 3 个层通过不同扩张率的膨胀卷积层, 使得模型能够捕捉从局部到全局的多尺度特征, 使用跳跃连接机制允许网络直接传递底层的高分辨率信息至输出层, 减少深层次信息丢失, 从而有效地保留壁画中的原始细节, 避免因多次上采样导致的图像模糊或失真现象。

### 1.5 损失函数

本文将对抗性损失 (adversarial loss)<sup>[15]</sup> 应用于两个映射函数。对于生成器  $G$  有如下映射关系  $G: X \rightarrow Y$ , 对应判别器  $D_y$ ,  $X$  和  $Y$  分别为源域和目标域, 目标损失表示为

$$L_{GAN}(G, D_y, X, Y) = E_{y \sim p_{data(y)}}[\log D_y(y)] + E_{x \sim p_{data(x)}}[\log(1 - D_y(G(x)))] \quad (8)$$

式中:  $G$  为生成与  $Y$  域中图像相似的图像;  $D_y$  为判断生成的图像  $G(x)$  与真实图像  $y$  之间的差异性;  $E$  为取得样本的平均值。另外一个生成器  $F$  及其对应的判别器  $D_x$ , 目标函数的构建过程完全镜像对称。

引入循环一致性损失 (cycle consistency loss)<sup>[15]</sup>, 确保生成的图像能够保留原始图像的重要特征。具体来说, 如果将一个图像  $x$  从  $X$  域通过生成器  $G$  转换到  $Y$  域, 然后再通过  $F$  生成器转换回  $X$  域, 那么最终得到的图像应该与原始图像  $x$  非常接近。同样, 这个原则也适用于从  $Y$  域到  $X$  域再到  $Y$  域的转换。循环一致性损失表示为

$$L_{cyc}(G, F) = E_x[\|F(G(x)) - x\|_1] + E_y[\|G(F(y)) - y\|_1] \quad (9)$$

引入本体映射损失 (identity loss), 这一部分的损失确保当输入图像已经是生成器目标域中的图像时, 生成器不会对这幅图像做出不必要的改变或转换。如果输入是目标域  $Y$  中的一个真实样本  $y$ , 则希望  $G(y) \approx y$ , 即生成器  $G$  应当尽量不改变输入图像。这可以通过最小化  $G(y)$  与  $y$  之间的差异来实现。同理, 如果输入是源域  $X$  中的一个真实样本  $x$ , 则期望  $F(x) \approx x$ 。本体映射损失如式(10)所示。

$$L_{identity}(G, F) = E_y[\|G(y) - y\|_1] + E_x[\|F(x) - x\|_1] \quad (10)$$

综上所述, 可以得到如下总体目标损失函数:

$$L(G, F, D_x, D_y) = L_{GAN}(G, D_y, X, Y) + L_{GAN}(F, D_x, Y, X) + \lambda L_{cyc}(G, F) + \gamma L_{identity}(G, F) \quad (11)$$

式中,  $\lambda$  和  $\gamma$  为控制不同损失函数相对重要性的权重系数。  $\lambda = 10, \gamma = 0.5$ , 与 CycleGAN 中的权重设置一致。

## 2 实验概况及结果分析

### 2.1 实验环境和数据集

本文实验环境为 Windows11 操作系统, 硬件环境为 Intel Core i5-13400F 处理器, 16.0 GB 系统内存, 采用 NVIDIA 显存 16GB 的 GeForce RTX4060TI GPU, 对比实验均在相同环境下进行。针对敦煌壁画高分辨率线条与剥落纹理的保留需求, 本文采用 batch size = 1 小批次训练, 为抑制梯度噪声、提升细节保真度并保障实验公平, 基线与改进模型统一使用 AdamW 优化器。训练 epoch 设为 200, 损失连续 10 轮波动小于 0.005 时判定收敛, 全程无过拟合。

敦煌壁画融合了中原汉文化、印度佛教艺术、波斯文化以及西域各民族的艺术风格, 形成了独特的多元文化特征, 其整体色调既华丽又和谐, 展现了高超的色彩运用技巧。线条流畅而有力, 山石的轮廓线清晰, 树木的枝叶描绘细腻。山水布局注重远近层次的表现, 近景的山石、树木刻画细致, 远景的山峰则逐渐虚化, 形成深远的空间感。经许可, 数据集其中的一些图像来自敦煌研究所的公共在线资源和专有数据, 如图 5 所示, 数据集包含各种类别和分辨率的图像, 对于敦煌艺术研究人员, 自然图像的风格迁移任务具有重要意义, 因此本文在内容图中选择了更多山脉和自然风景的内容图像, 并在风格图中尝试选择了自然风景的敦煌壁画, 共 4 100 张内容图和风格图, 其中内容图 2 800 张, 风格图 1 300 张。将输入图像的大小统一调整为  $224 \times 224$ , 使用 8:2 拆分将其分成训练集和测试集。此外, 为了提高模型的训练效率与生成图像效果, 还进行了旋转、裁剪等必要的数据增强处理, 确保输入模型的图像数据均匀和规范化。



图5 敦煌壁画数据集

Fig. 5 Dunhuang mural dataset

### 2.2 风格迁移结果定性分析

为了验证本文方法的有效性, 本文在定性分析中将其与 7 种流行的风格迁移方法进行了比较, 结果如图 6 所示。由图 6 可以看出, CycleGAN<sup>[15]</sup> 能够在没有成对训练数据的情况下学习两个不同域之间的映射关系, 但在图像转换过程中可能会错误地解

读内容图像,导致细节区域出现不自然的纹理和块状结构,色彩失真或褪色,层次和空间布局欠佳。AdaAttN<sup>[23]</sup>利用自适应注意力归一化实现特征分布的一致性,然而在保持内容完整性方面仍有限制,导致生成图像时可能出现结构细节丢失和颜色块匹配不佳的情况。CAST<sup>[24]</sup>通过替代二阶损失的方式保留了大部分的风格特征,使得颜色接近于目标样式,但在内容结构方面存在学习不稳定的问题。AesPA-Net<sup>[25]</sup>专注于增强注意力机制,在内容保真度方面表现良好,但色彩饱和度不足且细节纹理还原较弱。基于扩散模型的方法 InST<sup>[17]</sup>和 IP-Adapter<sup>[26]</sup>能够

创造出生动的图像,但这些方法往往对原始图像的语义和内容进行了大量的修改,难以维持风格与内容之间的平衡。StyleID<sup>[20]</sup>是一种基于预训练扩散模型的风格迁移方法,能够较好地保留内容语义,但其风格迁移结果偏向油画质感,与敦煌壁画的矿物颜料颗粒感存在显著差异。本文所提网络生成的结果图像具有高质量和较为真实的敦煌壁画风格。在风格特征上提取到接近原风格图像的色彩、细节特征和层次关系,并完整地呈现内容图像的语义特征。在针对不同类型的敦煌壁画风格进行风格迁移时,均可输出高质量、未丢失语义且风格鲜明的图像。



图 6 对比实验

Fig. 6 Comparative experiment

### 2.3 风格迁移结果定量分析

为分析模型迁移后图像的质量,本文采用的迁移后图像定量分析的指标有以下 4 种。1) FID。FID 关注图像的全局特征,适用于评估生成图像集的整体质量和多样性。在敦煌壁画风格迁移中,FID 能够反映生成图像的全局整体质量。2) L2 损失。L2 损失主要用来评估生成图像与原始内容图像之间的像素级差异。在敦煌壁画风格迁移中,用于评估图像在多大程度上保留了原始内容图像的内容信息。3) 结构相似性 SSIM。SSIM 用于比较原始图像和目标图像的亮度分布、对比度和结构信息。

首先在各个方面都得到一个相似性得分,然后综合得出整体的相似度评分,证明迁移后壁画的艺术完整性。4) 学习感知图像块相似度 LPIPS。LPIPS 主要关注图像的局部和细节特征,更加符合人类视觉系统对图像质量的感知。在相同参数和实验环境下,针对每种方法选取相同数量的内容图片和风格图片进行分析,从测试集随机选取 200 张图片进行指标测试,结果如表 1 所示。通过指标对比可以发现,相比其他方法本文提出的模型在迁移敦煌壁画风格的实验中,生成的结果图像在全局特征、局部特征细节、整体结构层次、亮度分布和对比度等

方面均有优异表现,较好地还原了敦煌艺术特有的高饱和和矿物颜料、精细纹理和复杂层次结构。为对比不同模型生成图片的时间效率,在相同环境和参数下,对单幅  $224 \times 224$  图像端到端风格迁移平均时长进行比较,结果如表 1 所示。由表 1 可以看到: CycleGAN 和 CAST 基于对抗训练的域映射,相比其他对比实验更轻量化;相对的,AdaAttN 和 AesPA-Net 依赖注意力机制复杂度升高;StyleID、InST 和 IP-Adapter 为扩散模型,其生成过程依赖多步迭代的去噪范式,生成速度通常较慢;本文方法专注多尺度风格解耦,虽相对原模型生成图片时间增加,但与其他对比实验相比更可满足实际需求。

表 1 定量分析指标结果对比

Tab. 1 Comparison of quantitative analysis index results

| 模型         | FID            | LPIPS        | L2           | SSIM         | 生成时长/s       |
|------------|----------------|--------------|--------------|--------------|--------------|
| CycleGAN   | 387.294        | 0.551        | 0.489        | 0.628        | 0.434        |
| AdaAttN    | 354.266        | 0.543        | 0.465        | 0.645        | 0.553        |
| CAST       | 377.450        | 0.665        | 0.578        | 0.396        | <b>0.421</b> |
| StyleID    | <b>353.901</b> | 0.533        | 0.481        | 0.667        | 4.278        |
| AesPA-Net  | 366.284        | 0.547        | 0.492        | 0.621        | 1.432        |
| InST       | 460.316        | 0.672        | 0.504        | 0.351        | 1.899        |
| IP-Adapter | 478.531        | 0.688        | 0.623        | 0.317        | 3.521        |
| Ours       | 364.787        | <b>0.527</b> | <b>0.461</b> | <b>0.679</b> | 0.449        |

本研究针对敦煌壁画艺术风格迁移效果进行了系统性用户调研。实验共招募 329 名具有多元化专

业背景的参与者,其中艺术领域从业者占比 38.6% (含美术教育、文物保护等方向),非艺术领域参与者占比 61.4% (涵盖计算机科学、人文社科等学科),最终回收有效问卷 289 份。评估体系由 3 个核心指标构建,分别是内容保真度、风格匹配度、视觉吸引力。内容保真度:量化生成图像与源图像在主题元素、物体构成及场景布局层面的相似性。风格匹配度:评估生成图像与目标敦煌壁画在笔触特征、色彩体系及艺术表现手法方面的一致性。视觉吸引力:从美学维度衡量图像在构图平衡、视觉舒适度及艺术感染力方面的综合表现。实验采用被试间随机分组评测范式:每位参与者被分层随机分配至单一模型与单一评价指标组合,观看单张生成图像并依据 0~3.00 分连续滑块完成对应维度评分。对收集到的分数进行平均数和标准差计算,结果见表 2 和图 7。

表 2 用户投票评分

Tab. 2 User voting ratings

| 模型         | 内容保真度       | 风格匹配度       | 视觉吸引力       |
|------------|-------------|-------------|-------------|
| CycleGAN   | 2.39 ± 0.33 | 2.48 ± 0.31 | 2.33 ± 0.21 |
| AdaAttN    | 2.54 ± 0.29 | 2.41 ± 0.35 | 2.38 ± 0.24 |
| CAST       | 2.33 ± 0.27 | 2.44 ± 0.25 | 2.27 ± 0.16 |
| StyleID    | 2.59 ± 0.30 | 2.47 ± 0.28 | 2.31 ± 0.19 |
| AesPA-Net  | 2.22 ± 0.32 | 2.53 ± 0.24 | 2.36 ± 0.15 |
| InST       | 1.37 ± 0.43 | 1.45 ± 0.36 | 2.53 ± 0.14 |
| IP-Adapter | 0.85 ± 0.45 | 1.27 ± 0.40 | 2.06 ± 0.22 |
| Ours       | 2.54 ± 0.24 | 2.58 ± 0.23 | 2.71 ± 0.11 |

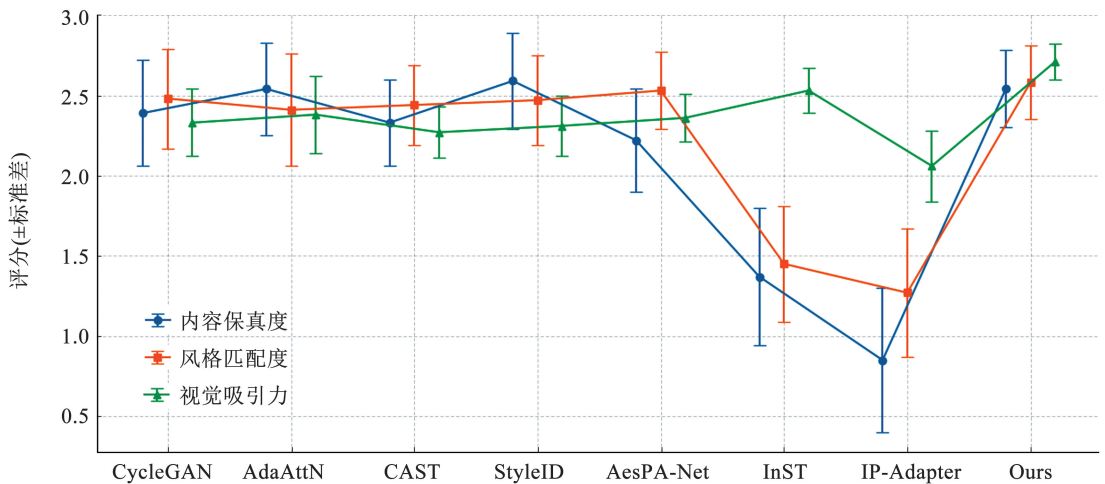


图 7 用户分数均值和标准差统计折线图

Fig. 7 Line plot of user score mean and standard deviation statistics

本文提出的方法在 3 个评估维度均取得显著优势,具体表现为:内容保真度均值 2.54 ( $\sigma = 0.24$ )、风格匹配度均值 2.58 ( $\sigma = 0.23$ )、视觉吸引力均值 2.71 ( $\sigma = 0.11$ )。实验结果验证了本方法在敦煌艺

术风格迁移任务中的有效性,不仅实现了源图像语义特征的高保真度留存,同时也精准捕捉了敦煌壁画特有的矿物颜料色域,说明该方法在内容完整性、风格准确性与审美价值之间达到了优化平衡,符合

大众对数字敦煌艺术的审美期待。

### 2.4 消融实验

对本文模型进行消融实验得到的风格迁移图像如图 8 所示。

图 8 中第 3 列为 CycleGAN 原模型生成的风格图像。(a) 列图为在生成器中加入 PFNT 模块的效果,该模块在保持分辨率的同时提取多尺度的信息,可解决原模型在生成图像时,山体、天空和人物等常见元素之间逻辑关系和比例协调不足和失真问题,确保图像内容的和谐统一,但仍存在局部纹理模糊,色彩协调性欠缺;(b) 列图是在 (a) 列基础上加入 DSNT 模块,该模块可增强图像全局和局部特征的

建模,使得生成图像可以更好地捕捉壁画的整体结构和局部色彩分布,使得树枝、天空及山顶等元素的迁移色彩更加自然且生动,增强视觉表现力,但局部纹理不太清晰;(c) 列图是在 (a)、(b) 列基础上加入 ALDC-NT,可进一步捕捉壁画中山水、枝叶等元素的精细笔触,使得生成图像中山体的轮廓更加清晰,树枝、叶片等细微纹理特征条理清晰,与风格图像特征更为贴合,整体青绿山水的青绿色色调统一协调,图像整体结构、元素逻辑关系也与壁画风格图像有较好的匹配,内容图像中的结构在迁移后的风格图像中均保留了原始细节。

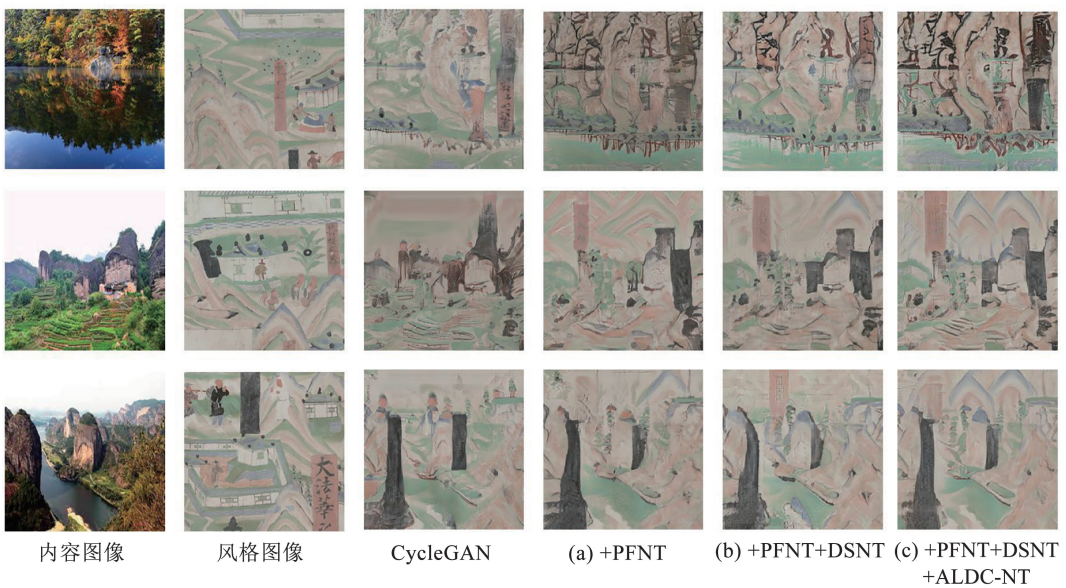


图 8 消融实验

Fig. 8 Ablation experiments

表 3 是加入不同模块后的指标变化,可以看到,本文所提方法在 FID、SSIM、LPIPS 和 L2 损失上均有较大提高,由此体现出生成图像在全局特征、局部特征细节和整体结构层次、亮度分布和对比度等方面均有出色的提升。

表 3 消融实验结果指标对比

Tab. 3 Comparison of indicators of ablation experiment results

| 模型       | FID           | LPIPS        | L2           | SSIM         |
|----------|---------------|--------------|--------------|--------------|
| CycleGAN | 387.29        | 0.551        | 0.489        | 0.628        |
| (a)      | 374.03        | 0.547        | 0.481        | 0.633        |
| (b)      | 368.75        | 0.530        | 0.474        | 0.657        |
| (c)      | <b>364.78</b> | <b>0.527</b> | <b>0.461</b> | <b>0.679</b> |

### 2.5 其他数据集的实验

为验证本文方法在不同艺术风格上的泛化能

力,本文在中国水墨画和梵高油画两种差异显著的艺术风格上进行了迁移实验。实验采用与敦煌壁画相同的网络结构和训练参数,仅替换风格图像数据集,以验证模型对多样化艺术风格的适应能力。中国水墨画以黑白灰为主色调,强调笔触的流动性和画面的留白意境。由图 9 第 3 列结果可以看到:本文方法成功捕捉了水墨画的晕染效果和笔触特征,生成的山水图像呈现出典型的水墨渐变层次;在细节保留方面,模型较好地处理了水墨特有的飞白效果,树枝和山石的轮廓线自然流畅。梵高油画以鲜艳的色彩和独特的笔触著称。由图 9 的第 6 列可以看到:本文方法有效学习了梵高标志性的漩涡状笔触特征与强烈的色彩对比,在严格保持内容图像结构的基础上,成功再现了油画颜料的厚重质感,且色彩过渡自然,无显著色块分割现象。



图9 其他数据集实验

Fig.9 Experiments on other datasets

实验结果验证了本文方法在多个数据集上的有效性和可行性,同时体现了模块的理论价值,并证明了本文模型具有良好的泛化能力。

### 3 结 语

本文针对敦煌壁画风格迁移任务中存在的色彩失真、纹理模糊及层次结构失调等问题,提出了一种改进 CycleGAN 的多尺度敦煌风格迁移网络。设计 ALDC-NT 网络,该网络是线性可变形卷积组合和膨胀卷积组合组成的 U 型结构,加入该网络可进一步捕捉壁画中山水、人物和建筑等元素的精细笔触。提出卷积和注意力组合 DSNT,该网络可增强图像全局和局部特征的建模,使得模型可以更好地捕捉敦煌壁画的整体结构和局部色彩分布。提出 PFNT,该网络在保持分辨率的同时提取多尺度的信息,更好地保留山水、人物和建筑元素之间的层次逻辑关系且比例协调。实验结果表明,本文所提方法在 FID、SSIM、L2 损失、LPIPS 指标和生成图片速度方面表现优异,用户调研显示其在内容保真度、风格匹配度与视觉吸引力方面表现突出,验证了模型在艺术特征保留与生成质量上的优势。本研究可为敦煌壁画的数字化保护与创新传播提供技术支撑,其设计思路可扩展至其他传统艺术风格的自动化创作。但模型引入多尺度网络导致参数量增加,训练时单次迭代耗时增多,DSNT 的全局注意力权重与 PFNT

门控信号缺乏直观语义解释,难以定位错误迁移区域。未来工作将聚焦于轻量化模型设计、跨域风格解耦与可解释性增强,进一步推动传统艺术数字化保护的实用化进程。

### 参考文献

[1] ROMÃO X, BERTOLIN C. Risk protection for cultural heritage and historic centres: Current knowledge and further research needs[J]. International Journal of Disaster Risk Reduction, 2022, 67: 102652. DOI: 10.1016/j.ijdrr.2021.102652

[2] ISOLA P, ZHU Junyuan, ZHOU Tinghui, 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: 1125. DOI: 10.1109/CVPR.2017.632

[3] WANG Weining, LI Yifan, YE Huan, et al. Dunhuang GAN: A generative adversarial network for Dunhuang mural art style transfer[C]//2022 IEEE International Conference on Multimedia and Expo (ICME). Taipei: IEEE, 2022: 1 DOI:10.1109/ICME52920.2022.9859987

[4] 贵向泉, 曹磊, 李立. 融合 CBAM 注意力机制的敦煌壁画风格迁移[J]. 计算机系统应用, 2025, 34(4): 276

GUI Xiangquan, CAO Lei, LI Li. Style transfer of Dunhuang murals with CBAM attention mechanism[J]. Computer System Application, 2025, 34(4): 276. DOI:10.15888/j.cnki.csa.009848

[5] GATYS L A, ECKER A S, BETHGE M. Image style transfer using convolutional neural networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 2414. DOI: 10.1109/CVPR.2016.265

[6] GAO Wei, LI Yijun, YIN Yihang, et al. Fast video multi-style transfer[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Snowmass: IEEE, 2020: 3222.

- DOI: 10.1109/WACV45572.2020.9093420
- [7] DUMOULIN V, SHLENS J, KUDLUR M. A learned representation for artistic style [ PP/OL ]. V1 ( 2016-10-24 ) [ 2025-03-21 ]. <https://doi.org/10.48550/arXiv.1610.07629>
- [8] ZHANG Hang, DANA K. Multi-style generative network for real-time transfer [ C ]//Proceedings of the European Conference on Computer Vision ( ECCV ) Workshops. Munich: ECCV, 2018: 349. DOI: 10.1007/978-3-030-11018-5\_32
- [9] CHEN Dongdong, YUAN Lu, LIAO Jing, et al. Stylebank: An explicit representation for neural image style transfer [ C ]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 1897. DOI: 10.1109/CVPR.2017.296
- [10] JOHNSON J, ALAHI A, LI Feifei. Perceptual losses for real-time style transfer and super-resolution [ C ]//Computer Vision-ECCV 2016. Amsterdam: Springer International Publishing, 2016: 694. DOI: 10.1007/978-3-319-46475-6\_43
- [11] ZHANG Lmin, JI Yi, LIN Xin, et al. Style transfer for anime sketches with enhanced residual U-net and auxiliary classifier GAN [ C ]//2017 4th IAPR Asian Conference on Pattern Recognition ( ACPR ). Nanjing: IEEE, 2017: 506. DOI: 10.1109/ACPR.2017.61
- [12] 胡景浩, 耿国华, 熊美均, 等. 基于细节特征提取融合的中国山水画风格迁移 [ J ]. 工程科学与技术, 2025, 57(1): 98  
HU Jinghao, GENG Guohua, XIONG Meijun, et al. A style transfer method for chinese landscape painting based on detail feature extraction and fusion [ J ]. Advanced Engineering Sciences, 2025, 57(1): 98. DOI: 10.12454/j.jsuese.202300295
- [13] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks [ J ]. Communications of the ACM, 2020, 63(11): 139. DOI: 10.1145/3422622
- [14] CHENG Bin, LIU Zuhao, PENG Yunbo, et al. General image-to-image translation with one-shot image guidance [ C ]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris: IEEE, 2023: 22736. DOI: 10.1109/ICCV51070.2023.02078
- [15] ZHU Junyan, 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: 2223. DOI: 10.1109/ICCV.2017.244
- [16] YI Zili, ZHANG Hao, TAN Ping, et al. Dualgan: Unsupervised dual learning for image-to-image translation [ C ]//Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 2849. DOI: 10.1109/ICCV.2017.310
- [17] ZHANG Yunxin, HUANG Nisha, TANG Fan, et al. Inversion-based style transfer with diffusion models [ C ]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 10146. DOI: 10.1109/CVPR52729.2023.00978
- [18] WANG Zhizhong, ZHAO Lei, XING Wei. StyleDiffusion: Controllable disentangled style transfer via diffusion models [ C ]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris: IEEE, 2023: 7677. DOI: 10.1109/ICCV51070.2023.00706
- [19] HUANG Siyu, AN Jie, WEI Donglai, et al. QuantArt: Quantizing image style transfer towards high visual fidelity [ C ]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 5947. DOI: 10.1109/CVPR52729.2023.00576
- [20] CHUNG J, HYUN S, HEO J P. Style injection in diffusion: A training-free approach for adapting large-scale diffusion models for style transfer [ C ]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024: 8795. DOI: 10.1109/CVPR52733.2024.00840
- [21] HUANG Bo, XU Wenlu, HAN Qizhuo, et al. AttenST: A training-free attention-driven style transfer framework with pre-trained diffusion models [ PP/OL ]. V1 ( 2025-03-11 ) [ 2025-03-21 ]. <https://doi.org/10.48550/arXiv.2503.07307>
- [22] ZHANG Xin, SONG Yingze, SONG Tingting, et al. LDConv: Linear deformable convolution for improving convolutional neural networks [ J ]. Image and Vision Computing, 2024, 149: 105190. DOI: 10.1016/j.imavis.2024.105190
- [23] LIU Songhua, LIN Tianwei, HE Dongliang, et al. Adaattn: Revisit attention mechanism in arbitrary neural style transfer [ C ]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021: 6649. DOI: 10.1109/ICCV48922.2021.00658
- [24] ZHANG Yuxin, TANG Fan, DONG Weiming, et al. Domain enhanced arbitrary image style transfer via contrastive learning [ C ]//ACM SIGGRAPH 2022 Conference Proceedings. New York: ACM, 2022: 1. DOI: 10.1145/3528233.3530736
- [25] HONG K, JEON S, LEE J, et al. AesPA-Net: Aesthetic pattern-aware style transfer networks [ C ]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris: IEEE, 2023: 22758. DOI: 10.1109/ICCV51070.2023.02080
- [26] HU Ye, ZHANG Jun, LIU Sibao, et al. Ip-adapter: Text compatible image prompt adapter for text-to-image diffusion models [ PP/OL ]. V1 ( 2023-08-13 ) [ 2025-03-21 ]. <https://doi.org/10.48550/arXiv.2308.06721>

(编辑 吕雪梅)