

学习纹理一致性与连续性的壁画修复方法^{*}

唐善成 王妍 周彤

(西安科技大学通信与信息工程学院 西安 710600)

摘要:为解决现有图像修复方法在修复破损壁画时存在上下文信息关注不足、图像模糊和纹理不一致等问题,提出了一种学习纹理一致性与连续性的壁画修复方法(LTCC-MIM)。以掩码自编码器理论为基础,首先采用 Transformer 架构的图像修复网络来融合图像的局部特征与全局上下文信息,增强图像各部分之间的相互关系和整体结构的理解能力。其次设计上下文特征增强模块(CFEM),实现区域联合学习和多尺度特征聚合,强化局部纹理连续性并生成精细的图像细节,缓解图像模糊。最后引入结构相似度(SSIM)作为约束有效关注壁画的纹理特性,以此约束学习过程,解决纹理不一致问题,优化修复效果。对真实壁画进行数字化修复的实验结果表明,所提出的方法在主客观评价方面均优于比较方法,其中峰值信噪比(PSNR)值平均提升了 2~8 dB,SSIM 平均提升了 2%~10%,感知相似度(LPIPS)平均下降了 2%~9%,FID 分数平均下降了 1~5。

关键词:壁画修复;Transformer;编码器-解码器;上下文特征增强模块;计算机视觉

中图分类号: TP391.41;TN911 **文献标识码:** A **国家标准学科分类代码:** 510.4050

Learning-based texture coherence and continuity for
mural inpainting method

Tang Shancheng Wang Yan Zhou Tong

(College of Communication and Information Technology, Xi'an University of Science and Technology, Xi'an 710600, China)

Abstract: To address the existing issues in image restoration methods when repairing damaged murals, such as insufficient attention to contextual information, image blurring and texture inconsistency, this paper proposes a learning based texture coherence and continuity for mural inpainting model (LTCC-MIM). Based on the theory of masked autoencoders, this paper first employs an image inpainting network with a Transformer architecture to integrate local features and global contextual information of the image, thereby enhancing the understanding of interrelationships among various parts of the image as well as its overall structure. Next, a contextual feature enhancement module (CFEM) is designed to achieve regional joint learning and multi-scale feature aggregation, strengthening local texture continuity and generating fine image details to alleviate image blurring. Structural similarity metrics are incorporated as constraints to prioritize mural-specific texture patterns, thereby harmonizing inconsistencies and optimizing restoration quality. Experimental results on digital restoration of real murals demonstrate that the proposed method outperforms comparative methods in both subjective and objective evaluations. Specifically, it achieves an average improvement of 2~8 dB in peak signal-to-noise ratio (PSNR), 2%~10% in structural similarity index (SSIM) and a reduction of 2%~9% in learned perceptual image patch similarity (LPIPS). Additionally, the fr chet inception distance (FID) score decreased by an average of 1~5 units.

Keywords: mural inpainting; Transformer; encoder-decoder; contextual feature enhancement module; computer vision

0 引言

壁画作为世界文化遗产的瑰宝,以其千年艺术生命力展现中华文明辉煌,成为窥探古代文明的窗口。然而其物理状

态极为脆弱,面临自然侵蚀与人为损害的双重威胁^[1],存在剥落、裂缝和霉菌滋生等问题,既影响艺术美感,也威胁历史信息的完整性,因此,壁画保护与修复工作尤为迫切。

图像修复技术的不断进步,为壁画修复与保护带来了

新的解决方案,图像修复方法可以分为传统的图像修复方法和基于深度学习的图像修复方法。传统的图像修复方法主要包括基于扩散的方法和基于 patch 的方法。基于扩散的方法利用受损区域的边缘信息,确定扩散的方向,然后将已知区域的信息逐渐向缺失区域扩散。Bertalmio 等^[2]采用基于扩散的方法,通过数学模型模拟自然界的扩散过程,利用图像中已知区域的局部信息,并将这些信息平滑地扩散至受损区域,实现图像修复。基于 patch 的图像修复技术也是传统方法中的一个重要分支。Ding 等^[3]提出了一种非局部纹理匹配和非线性滤波的图像修复方法,该方法能够使用非局部纹理度量再现底层纹理细节,并无缝平滑像素强度,获得自然外观的修复图像。Wang 等^[4]提出了一种线条图引导下的壁画修复方法,使用多个候选 patch 的线性组合来构建目标 patch。乔景慧等^[5]提出了一种基于 L2 归一化三分量双色反射模型的高光去除方法。该方法通过构建包含漫反射、镜面反射与体散射的三分量模型,更准确地描述透明材质表面的反射特性,并利用指数变换与聚类分析实现高光区域的检测与信息恢复。

随着深度学习技术的不断进步,基于深度卷积神经网络(convolutional neural network, CNN)和生成对抗网络(generative adversarial network, GAN)等的图像修复方法在图像修复领域取得了显著成效。CNN 通过学习图像的深层特征,能够自动提取图像中的复杂结构和纹理信息。Wang 等^[6]提出了一种基于多尺度自适应部分卷积和类笔画蒙版的唐卡壁画修复方法,能够以更简单的结构生成逼真的图像内容。Chen 等^[7]提出了一种基于部分卷积和滑动窗口方法的数字敦煌壁画图像修复策略 PCSW,解决了小样本问题,提高网络精度。Cao 等^[8]使用 GAN 修复古代壁画,并引入扩张性卷积来增加网络的接受野。Ciortan 等^[9]提出了一种基于 GAN 的图像修复模型,该模型具有两个生成器,一个用于生成边缘,另一个用于生成颜色,同时通过随机移动掩码的形态变化来模拟艺术品中腔隙的随机性。开志强等^[10]提出了一种融合全局语义学习与显著目标感知的修复网络。该网络通过门控语义学习机制建模远距离上下文依赖,结合多尺度特征聚合模块增强语义重建能力,显著提升了修复区域的语义合理性与视觉连贯性。相比于基于 CNN、GAN 模型,扩散模型具有更好的泛化能力,不易出现模式崩溃的问题,且不需要特定掩码的训练就可以产生高保真的输出。Lugmayr 等^[11]提出了一种基于去噪扩散概率模型的图像修复方法 RePaint,它通过调节反向扩散过程来修复图像中缺失或损坏的区域。但对掩码的分布较为敏感,如果训练时使用的掩码分布与测试时的掩码类型差异较大,模型的修复效果会下降。张绪等^[12]提出了一种基于扩散模型的二阶段细化图像修复模型 Diff-2sIR。该模型通过改进 U-Net 架构与扩散采样策略,对初步修复结果进行二次细化,有效缓解了修复区域与未修复区域之间的不一致性问题,但该方法在修复复杂纹理方面仍

存在一定局限。Transformer 是 Vaswani 等^[13]在 2017 年提出的一种基于注意力机制的神经网络架构,最初用于自然语言处理任务。Wan 等^[14]将 Transformer 应用于图像修复领域,将 Transformer 的外观先验重建与 CNN 的纹理补充结合在一起,利用 Transformer 恢复了多样化的连贯结构以及一些粗糙的纹理。但需先通过 Transformer 生成外观先验,再用 CNN 上采样,增加了推理时间和资源消耗。滕诗宇等^[15]提出了一种融合多尺度分层特征与全局-局部协同 Transformer 的图像修复模型。该方法通过设计多尺度分层特征融合模块增强不同层次特征的整合能力,并引入矩形窗口注意力机制与局部前馈神经网络构建全局-局部协同 Transformer 模块,提升模型对全局结构语义的把握与局部细节特征的恢复质量。但在生成结果的多样性方面仍存在局限。He 等^[16]提出的掩码自编码器(masked autoencoders, MAE)作为一种可扩展的视觉自监督学习方法,通过随机掩码输入图像块并重建缺失像素来学习图像表示,但难以恢复壁画图像中精细的笔触纹理。

虽然上述方法能够修复特定的壁画损伤,但仍存在一些问题:传统的图像修复方法在修复大面积损伤和复杂壁画时精确度不足,基于 CNN 的图像修复方法难以学习语义一致的纹理,易混淆有效与受损像素,导致修复时出现上下文信息关注不足问题。基于 GAN 的修复方法对常规图像效果良好,但用于壁画时,常忽视结构的重要性,导致修复后出现图像模糊和纹理不一致等问题。针对上述问题,本文提出了一种学习纹理一致性与连续性的壁画修复方法(learning-based texture coherence and continuity for mural inpainting model, LTCC-MIM),得到更高质量的修复图像,显著提升壁画修复的可靠性与有效性。

本文的贡献主要包括:

1)提出了一组非对称的编码器-解码器结构,利用编码器中 Transformer 架构提取图像的多尺度特征,并通过解码器中的 Transformer Block 层对图像进行重建,在捕获图像上下文信息的同时提高了壁画图像的全局纹理一致性;

2)设计上下文特征增强模块(contextual feature enhancement module, CFEM),通过模拟长期的空间依赖性来捕捉图像区域间的复杂相互作用,生成更精细的细节,学习了局部的纹理连续性特征,强化局部特征间的联系,更准确地重建壁画图像的纹理和结构,缓解图像模糊;

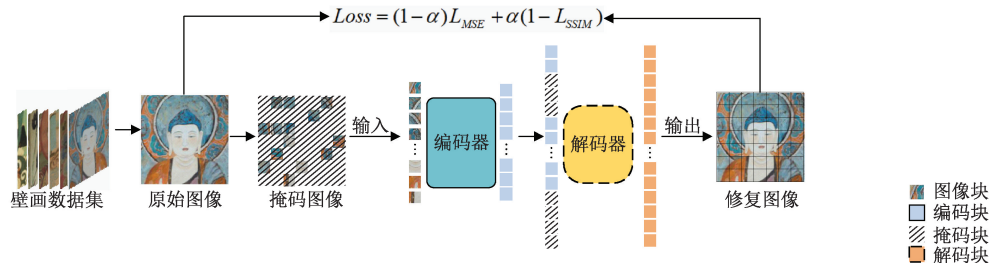
3)提出均方误差和结构相似度作为模型的损失函数,结合像素损失,将传统像素损失转变为基于结构相似度的错误度量,以此约束学习过程,解决纹理不一致问题,优化重建效果。

1 本文算法

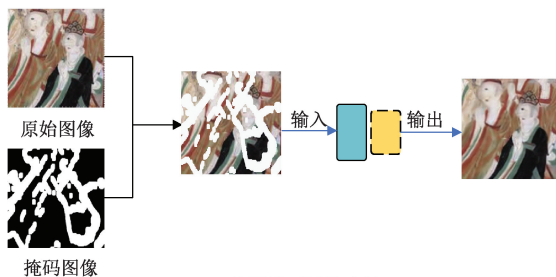
本方法基于掩码自编码器框架,输入为 224×224 的 RGB 壁画图像。如图 1 所示,模型首先将图像划分为 16×16 的非重叠块,并通过线性投影转换为嵌入向量(Patch

Embedding)。随后,采用均匀分布随机采样 75% 的图像块并予以掩码,仅将剩余可见块输入编码器。编码器采用 Vision Transformer 架构,并引入 CFEM 以聚合多尺度特征。解码器接收编码后的可见块与可学习的掩码标记,通

过 Transformer 块重建被掩码区域。最终,通过结合均方误差(mean squared error, MSE)与结构相似度(structural similarity index measure, SSIM)的损失函数优化重建质量。



(a) 壁画表面特征提取
(a) Mural surface feature extraction



(b) 掩码壁画图像修复
(b) Mask mural image inpainting

图 1 学习纹理一致性与连续性的壁画修复模型整体框架

Fig. 1 Overall framework of learning-based texture coherence and continuity for mural inpainting model

1) 图像输入和预处理

首先,将待输入壁画图像加载并调整为统一尺寸 224×224 ,确保所有图像输入到模型中具有相同的分辨率,有助于模型的训练与推理,将输入图像进行归一化处理,如式(1)所示。

$$\text{img}_{\text{norm}} = \frac{\text{img} - \text{imageNet_mean}}{\text{imageNet_std}} \quad (1)$$

式中: imageNet_mean 和 imageNet_std 是 ImageNet 数据集的均值和标准差,以便与预训练模型的输入一致。掩码生成与缺损率计算提供与输入图像同尺寸的掩码图(Mask),掩码图表示壁画的缺损区域,预处理为二值化形式,使缺损区域的像素值为 1,其他部分为 0。利用自定义函数计算图像中缺损区域占比(Mask rate),若缺损率过低,则无需修复,可直接输出完整图像。

2) 壁画特征学习

模型由非对称编码器-解码器结构组成,编码器是一个标准的 Vision Transformer,通过添加位置嵌入的线性投影来嵌入 patches,通过一系列 Transformer Blocks 提取图像特征以学习上下文信息。在编码器中引入 CFEM,实现区域联合学习和多尺度特征聚合,强化局部纹理连续性并生成精细的图像细节。解码器将编码器的输出与 Mask tokens 和 Position Embeddings 相结合,通过 Transformer

块来重建编码器未保留的图像块,生成重建图像,对重建前后的图像计算像素和 SSIM 的损失。特征学习过程首先把壁画图像随机掩码一部分,将掩码图像分割成同等数量的 16×16 大小图像块作为输入。舍弃被掩码块所遮盖的图像块,保留未被遮盖的图像块作为模型输入。并以被遮盖的图像块作为重建目标,迫使模型通过未被遮盖的图像块重建出被遮掩的图像信息,从而完成图像的特征学习任务。

3) 壁画修复

通过自动编码器的结构对图像的缺损区域进行填补,修复后的图像再与掩码图结合,保留非掩码区域的原始信息,保证在修复区域上更好地保留壁画图像的结构和纹理。

输入大小为 224×224 的三维图像,经过 Patch Embedding 转换成一系列 1 维序列。随机采样一部分图像块,加入位置编码以及类别编码输入到编码器中提取图像特征。将输出结果加入掩码和位置编码输入解码器中,归一化后输出重建图像,各模块如图 2 所示。

1.1 编码器

Patch Embedding: 图像 patch 会被展平成 $\mathbf{X} \in \mathbf{R}^{Q \times (P^2 \cdot C)}$ 一维特征序列, $P = HW/Q$ 表示 patch 的个数, P^2 代表序列的像素大小, C 表示通道数,因此 $P^2 \cdot C$ 代表

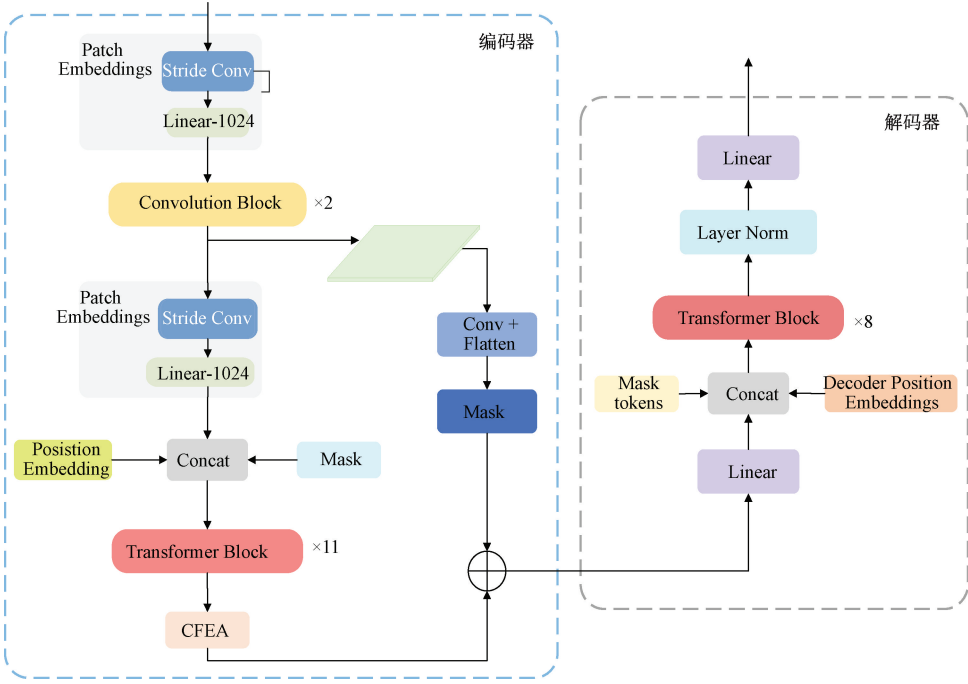


图2 LTCC-MIM 结构示意图

Fig. 2 Schematic diagram of LTCC-MIM

每个序列的维度,随后对 Patch Embeddings 进行线性变换,如式(2)所示。

$$z_0 = [x_p^1 E; x_p^2 E; \dots; x_p^n E] \quad (2)$$

为了保持序列的顺序性,引入位置编码 E_{pos} 来补充被遮掩的 patches,确保未被掩码遮住的 patches 能够保留其位置信息。如式(3)所示。

$$z_1 = [x_p^1 E; x_p^2 E; \dots; x_p^n E] + E_{pos} \quad (3)$$

加入可学习的类别编码用来获得图像的特征,如式(4)所示。

$$z_2 = [x_{class}; x_p^1 E; x_p^2 E; \dots; x_p^n E] + E_{pos} \quad (4)$$

掩码策略:把图像分割成 16×16 大小图像块,使用服从均匀分布的采样策略对这些图像块随机采样一部分,同时遮掩掉余下的另一部分。这种随机掩码的训练方式可以让重建任务更加多样化,有效增强模型的重建能力,本文选用掩码比率为 75%,该掩码比率可获得最佳的模型训练效果。

每个区块由多个交替排列的模块构成,首先是多头自注意力(multi-head self-attention, MSA)模块,它负责捕捉序列内部的依赖关系。紧接着是多层感知器(multi-layer perceptron, MLP)模块,它对序列进行非线性变换,MLP 模块作为一种高级的人工神经网络,能够将输入序列映射成固定长度的向量形式,这些向量随后被送入自注意力模块,以便进行进一步的处理,从而在输入序列中提取出更加细致和深层次的特征。可以用公式表示为:

$$z_l = \text{MSA}(\text{LN}(z_{l-1})) + z_{l-1}, l = 1, \dots, L \quad (5)$$

$$E_2 = \text{MLP}(\text{LN}(z_l)) + z_l, l = 1, \dots, L \quad (6)$$

式中: z_l 是 Transformer Layer 的通用语义表示, $\text{LN}(\ast)$

表示层归一化。

编码器捕获的多尺度特征 E_1 和 E_2 分别展现了图像的细节和概貌信息。为了让解码器更有效地处理这些特征嵌入,在编码器的最后阶段对 E_1 进行 $\text{stride}=2$ 的下采样,并将其与 E_2 结合,最终生成了融合多尺度特征的可见标记(visible tokens) E_2 。

$$E_n = \text{Conv}(E_1, \text{stride} = 2) + E_2 \quad (7)$$

1.2 解码器

解码器的任务是重建图像,它由多个 Transformer 块组成。解码器首先接收两组输入:一组是编码器处理过的可见图像块的特征表示;另一组是用于指示解码器预测区域的 Mask tokens。这些 Mask tokens 是可训练的向量,每个向量代表一个需要预测的图像块。在进入 Transformer 块之前,解码器的线性层会调整它们的维度,并与 Mask tokens 结合。为了使模型能够识别 Mask tokens 在图像中的不同位置,会在输入 Transformer 块之前为整体添加 Decoder Positional Embedding,将输出结果归一化后再通过解码器的 Linear 层来生成重建的图像。

1.3 CFEM

为了更精准地识别出哪些区域对于修复图像破损是关键,开发了 CFEM,它不仅强化了图像中局部特征的相互联系,而且确保了图像整体的连贯性。受 Yu 等^[17]的启发,本研究并未采用固定尺度的补丁匹配策略,而是实施多尺度特征聚合,在不同尺度上捕捉深层次的语义特征。这种方法有效地平衡了处理精确度与复杂度,特别适用于处理包括尺度变化在内的复杂情况。模块结构如图 3 所示。

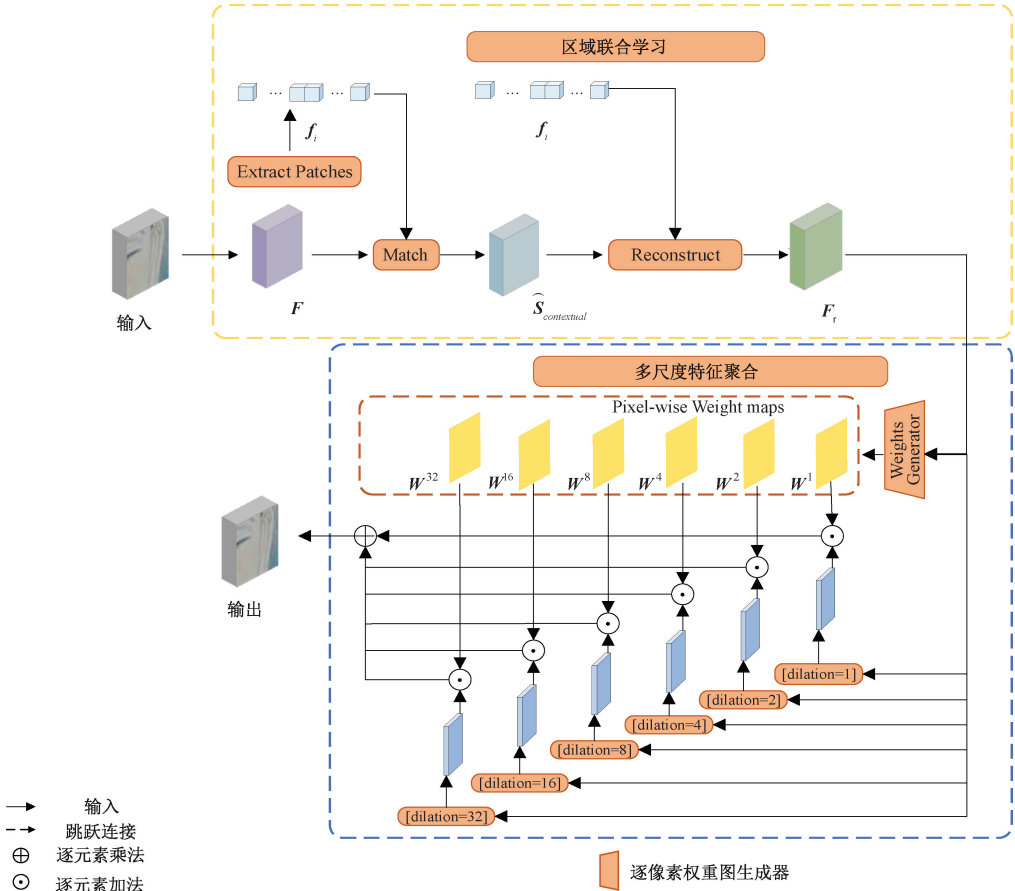


图 3 上下文特征增强模块

Fig. 3 Contextual feature enhancement module

具体操作是,从特征映射 F_{rec} 中提取 3×3 pixels 的补丁,然后计算这些补丁之间的余弦相似度。

$$S_c^{i,j} = \left\langle \frac{f_i}{\|f_i\|_2}, \frac{f_j}{\|f_j\|_2} \right\rangle \quad (8)$$

其中, f_i 和 f_j 分别代表 feature map 的第 i 个 patch 和第 j 个 patch,然后使用 softmax 函数处理这些相似度值,从而得到每个 patch 的注意力分数。

$$\hat{S}_c^{i,j} = \frac{\exp(S_c^{i,j})}{\sum_{j=1}^N \exp(S_c^{i,j})} \quad (9)$$

再利用提取的 patch,基于注意力重建 feature map:

$$\hat{f}_i = \sum_{j=1}^N f_j \cdot \hat{S}_c^{i,j} \quad (10)$$

为了重建特征映射 F_{rec} 的第 i 个 patch,执行的操作包括卷积、信道软最大化和反卷积。在重建特征图的过程中,使用了 4 组具有不同扩张率的扩展卷积层来捕捉图像多尺度语义特征:

$$F_r^k = \text{Conv}_k(F_r) \quad (11)$$

为了更有效地整合多尺度的语义特征,设计了一个像素级权重图生成器 G_w ,目的是预测每个像素的权重。 G_w 由两个卷积层组成,卷积核的大小分别为 3 和 1,这两个卷

积层后面都有 ReLU 非线性激活,并且输出通道的数量被设定为 6。像素级的权重图的计算公式为:

$$W = \text{softmax}(G_w(F_{\text{rec}})) \quad (12)$$

$$W^1, W^2, W^4, W^8, W^{16}, W^{32} = \text{Slice}(W) \quad (13)$$

式中: $\text{Softmax}(\cdot)$ 为通道级 Softmax, $\text{Slice}(\cdot)$ 为通道级 Slice,最后,通过将多尺度语义特征进行元素级别的加权求和,得到了更为精细的特征图 F_c :

$$F_c = (F_r^1 \odot W^1) \oplus (F_r^2 \odot W^2) \oplus (F_r^4 \odot W^4) \oplus (F_r^8 \odot W^8) \oplus (F_r^{16} \odot W^{16}) \oplus (F_r^{32} \odot W^{32}) \quad (14)$$

由于部分卷积层的掩码更新机制被利用,无需像 CNN 区分图像的前景和背景像素,采用了跳过连接来避免 patch-shift 操作可能导致的语义信息丢失,并在网络架构中集成了卷积层和反卷积层,提高了计算效率。

1.4 损失函数

LTCC-MIM 采用 MSE 和 SSIM 作为模型的损失函数,对 Masked patch 重建前后的图像计算像素和 SSIM 的损失,可用公式表示为:

$$L = (1-\alpha)\text{MSE}(X_r, \hat{X}_r) + \alpha(1-\text{SSIM}(X_r, \hat{X}_r)) \quad (15)$$

式中: α 表示平衡像素和 SSIM 之间相对重要性的权重因

子,本文设置 $\alpha = 0.5$, $\mathbf{X}_r$ 表示原始数据, $\hat{\mathbf{X}}_r$ 表示重建数据,SSIM表示结构相似度指数,MSE可用公式表示:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\mathbf{X}_r - \hat{\mathbf{X}}_r)^2 \quad (16)$$

2 实验结果与分析

2.1 数据集

本文采用敦煌壁画作为数据集,该数据集包含5 428幅壁画,对原始壁画图像裁剪,统一调整为 224×224 pixels大小的RGB图像。由于数据集样本数量有限,为提升模型的泛化能力,对壁画图像进行缩放、旋转、镜像翻转、亮度变换等数据增强方法进行扩充,扩充后训练集约为43 000幅,测试集为5 000幅。掩码选择两种类型的掩码,不同比例的随机掩码和中心掩码。不同比例的随机掩码采用Liu等^[18]提供的NVIDIA Irregular Mask数据集中的随机掩码,采用了其中0%~10%、10%~20%、20%~30%、30%~40%、40%~50%、50%~60%的不同比例掩码来模拟不同程度的壁画破损,验证本文模型适用于不同程度的破损。其次,采用正方形中心掩码模拟大面积破损,以验证本文模型对大区域破损修复的有效性。最后采用真实破损壁画进行修复实验,验证本文模型对于真实破损壁画修复的有效性。并与图像修复方法LAMA^[19],DA-CLIP^[20],MISF^[21],RePaint^[11]进行对比分析。其中,LAMA主要采用快速傅里叶卷积作为核心算子使其在浅层即具备图像全局感受野,提升对大规模缺失区域和周期性结构的修复能力,但难以学习局部纹理一致性。DA-CLIP通过训练一个可学习的图像控制器,在冻结原始CLIP模型参数的同时,使固定图像编码器能够从退化输入中提取无退化的高质量内容嵌入,但在复杂、特定的壁画纹理细节上可能泛化不足。MISF是一种基于动态滤波和交互式推理的方法,将图像修复重建为多层级滤波任务,动态核预测机制能够根据壁画不同区域的语义内容自动调整滤波策略。RePaint是基于去噪扩散概率模型,重采样机制能逐步修正生成区域与保留壁画结构之间的不协调性,通过调节反向扩散过程来修复图像中缺失或损坏的区域,但其计算成本高且对掩码分布敏感。

2.2 实验设置和评价指标

模型在64位Windows 11上使用Python 3.9.19环境搭载的PyTorch 1.13.1完成训练,硬件环境为Intel(R) Core(TM) i7-14700KF@3.40 GHz, RAM 32 GB,训练中使用的是NVIDIA GeForce RTX 4080 SUPER GPU,显存为32 GB。Batch Size为10,学习率为 1×10^{-4} ,Epoch为1 000。

为了评估实验的修复结果,选取峰值信噪比(peak signal to noise ratio, PSNR)、SSIM、感知相似度(learned perceptual image patch similarity, LPIPS)和FID分数(Fréchet inception distance, FID)作为评价指标。PSNR

和SSIM值越大,表明修复图像越接近原始图像;LPIPS和FID值越小,表明修复图像与原图像之间的感知差异越小。

2.3 添加不同比例随机掩码修复实验

为了分析本文模型对于随机破损修复的有效性,对测试集中的壁画进行了不同比例掩码的修复实验,图4展示了6幅壁画的修复结果对比,可以看出,LAMA模型修复结果如图4(c)所示,第1幅壁画存在轻微模糊问题,第2幅、第3幅及第5幅壁画存在块状伪影问题,第4幅及6幅壁画存在严重伪影及修复不完整问题。DA-CLIP模型修复结果如图4(d)所示,第2幅壁画脸颊区域的纹理修复不合理,第3幅壁画存在结构断裂和特征缺失问题,第4幅壁画存在结构紊乱问题,第5幅壁画存在面部特征缺失且修复错误问题,第6幅壁画存在细节信息修复错误问题。图4(e)为MISF的修复结果,MISF修复后的壁画较平滑,但整体存在模糊修复现象。对于第1幅壁画,修复后头部区域出现了轻微模糊现象,第2幅和第3幅壁画清晰度不高。RePaint模型修复结果如图4(f)所示,对第1幅壁画修复效果较好,但头部发髻部分存在伪影和模糊,第2幅壁画存在脸颊区域纹理修复不合理问题,第3幅存在色彩失真问题,第4幅存在结构紊乱问题,第5幅存在特征缺失问题,第6幅同样存在修复错误问题。本文提出模型修复结果如图4(g)所示,修复后图像在结构连续,纹理细节都有了进一步提升,线条连贯更自然,获得了更好的视觉效果。为进一步验证修复质量,采用4个指标进行定量分析,如表1所示,本文模型在PSNR、SSIM指标上总体高于其他模型,表明图像失真更少,在LPIPS、FID指标上总体更低,表明修复图像与原图像之间的感知差异更小,修复效果更好,这些定量分析结果验证了本文模型对于不同破损面积壁画修复的有效性。

2.4 添加中心掩码修复实验

为了证明本文模型对破损区域有较好的壁画修复效果,使用中心掩码来模拟壁画破损区域进行修复实验,修复结果如图5所示。LAMA模型修复结果如图5(c)所示,所有图像修复后的壁画存在掩码块状残留,这是由于LAMA是一种简单的、单阶段的方法,难以处理纹理复杂且含有较多语义信息的壁画。

DA-CLIP模型修复结果如图5(d)所示,出现了严重的结构紊乱和特征缺失问题,未完成面部特征的修复。第1幅、第4幅和第6幅的修复结果存在明显的结构紊乱,人脸的眼睛、鼻子等特征结构未得到良好的修复,第2幅、第3幅和第5幅壁画的脸部出现了特征缺失、色彩差异的问题。图5(e)为MISF的修复结果,修复后存在内容模糊和修复伪影问题。第5幅人脸及第6幅眼睛部分的纹理细节信息均不清晰。RePaint模型修复结果如图5(f)所示,出现了结构修复不合理问题,并存在修复残留,未完成面部特征的修复。第1幅、第2幅和第3幅人脸的眼睛、鼻子存

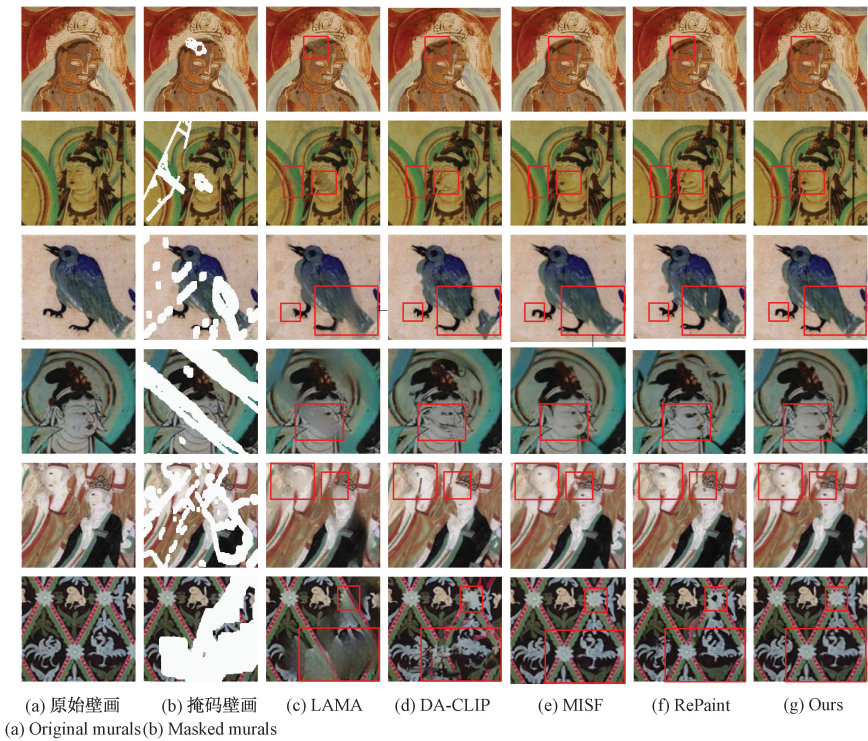


图 4 不同模型对添加不同比例随机掩码壁画修复结果对比

Fig. 4 Comparison of different models in inpainting of murals with added different proportions of random masks

表 1 不同模型对添加不同比例随机掩码壁画修复结果评价指标比较

Table 1 Comparative evaluation metrics of different models in inpainting of murals with added different proportions of random masks

评价指标	模型	[1%,10%)	[10%,20)	[20%,30%)	[30%,40%)	[40%,50%)	[50%,60%)	平均值
PSNR ↑	LAMA	36.157 1	32.805 7	26.842 4	24.093 6	22.731 0	16.858 3	26.581 4
	DA-CLIP	36.048 5	36.714 1	23.775 9	22.210 5	20.979 7	17.705 4	26.239 0
	MISF	39.600 6	33.854 2	31.398 0	28.188 3	29.592 9	25.708 2	31.390 4
	RePaint	36.637 1	36.617 5	23.843 6	24.878 7	24.917 9	21.221 5	28.019 4
	Ours	39.231 9	36.693 6	34.714 1	30.408 9	29.897 0	27.904 8	33.141 7
SSIM ↑	LAMA	0.976 6	0.959 3	0.896 9	0.847 3	0.798 6	0.630 7	0.851 6
	DA-CLIP	0.971 2	0.960 2	0.892 5	0.823 2	0.813 6	0.696 2	0.859 5
	MISF	0.983 9	0.940 7	0.905 4	0.876 1	0.898 6	0.855 0	0.910 0
	RePaint	0.971 9	0.956 5	0.890 7	0.884 9	0.882 6	0.850 5	0.906 2
	Ours	0.986 5	0.964 2	0.943 0	0.920 9	0.912 5	0.903 7	0.938 5
LPIPS ↓	LAMA	0.017 4	0.019 6	0.088 4	0.132 9	0.164 5	0.309 1	0.122 0
	DA-CLIP	0.185 2	0.010 8	0.093 5	0.133 1	0.131 2	0.163 4	0.119 5
	MISF	0.017 2	0.034 3	0.082 8	0.068 0	0.056 1	0.056 6	0.052 5
	RePaint	0.018 5	0.010 9	0.097 2	0.090 6	0.071 7	0.079 9	0.061 5
	Ours	0.010 7	0.021 8	0.037 2	0.040 8	0.051 2	0.034 5	0.032 7
FID ↓	LAMA	3.909 2	6.695 1	8.559 5	13.235 0	16.088 8	15.510 2	10.666 3
	DA-CLIP	5.186 7	4.451 9	8.983 9	9.087 6	14.868 8	9.806 5	8.730 9
	MISF	5.212 3	6.882 0	8.084 4	8.770 6	10.035 0	4.922 0	7.317 7
	RePaint	5.152 0	4.106 7	7.708 4	8.217 1	12.238 1	7.114 8	7.422 9
	Ours	3.696 4	6.013 8	6.993 6	6.792 0	9.785 0	3.731 0	6.168 6

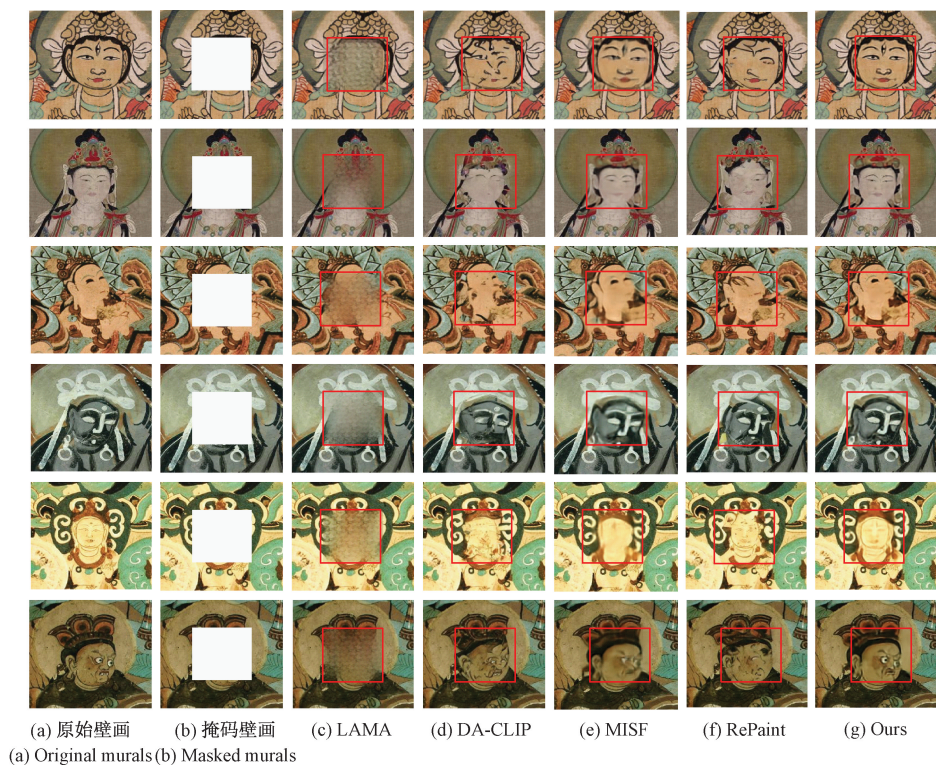


图5 不同模型对添加中心掩码壁画修复结果对比

Fig. 5 Comparison of different models in inpainting of murals with added central masks

在结构断裂和特征缺失的问题,第4幅壁画修复后与周围纹理不和谐,第5幅和第6幅仍存在修复残留,结构紊乱,色彩差异问题。本文提出的模型修复结果如图5(g)所示,整体修复结果相对于其他4种模型更加合理,能够更好地结合壁画图像破损区域周围的上下文信息,整体的重建效果也更好,所生成的壁画在结构、细节和视觉感知等多个方面都更贴近于原始的真实壁画。同样地,实验采用了评价指标对修复结果进行定量分析,从表2中可以看出,对大区域破损壁画修复,本文模型在4种指标上总体优于其他对比模型,验证了本文模型对于大面积破损壁画修复的有效性。

2.5 真实破损修复实验

为进一步验证模型对真实破损敦煌壁画的修复效果,选取6幅壁画进行实验,结果如图6所示。各模型修复效果对比:LAMA与MISF在第1、2幅壁画中修复不彻底,存在残留痕迹。RePaint与DA-CLIP在第1、2幅壁画修复后破损残留明显,第6幅壁画DA-CLIP未完成脸部和身体修复。本文模型对于所有壁画修复更为彻底,视觉一致性更佳。第1幅壁画有效去除了修复痕迹,第2幅壁画结构线条拟合良好,第3、4幅壁画额头和脸颊区域修复彻底,第5幅壁画颜色一致且无修复残留,第6幅壁画面部和眼睛轮廓修复完整度高。实验结果表明,相较于其他模型,本文模型在真实破损敦煌壁画的修复中表现出更高的修复质量和完整性,验证了其在处理实际复杂破损场景中

的有效性。

2.6 消融实验

为更好地分析本文模型中各部分在壁画修复中的作用,进行消融实验。如图7所示,图7(a)为壁画的原始图像,图7(b)为掩码图像,图7(c)为所提方法的基础模型(Baseline)的修复结果,图7(d)为所提方法的基础模型和上下文特征聚合增强模块(Baseline+CFEA)的修复结果;图7(e)为本文所提出模型修复结果(Baseline+CFEA+SSIM),包括基础模型、CFEM、限制结构相似度。从图7的修复结果可以看出,图7(c)图修复结果存在明显的细节丢失和图像模糊的问题。图7(d)与(c)修复结果相比,缓解了图像模糊,但修复后的内容存在语义准确性较差的问题。如7(e)所示,加入结构相似度作为限制后,大致缓解了图像模糊和纹理不一致问题,表3为消融实验的定量评价结果,从中可以看出本文模型具有较好的修复效果。

2.7 损失函数对比实验

为了更好地分析本文损失函数相比其他损失函数在壁画修复中的作用,进行对比实验验证本文损失函数对于壁画修复的有效性。分别与风格损失、边缘损失、对抗损失进行对比。保持了模型主干网络(Baseline+CFEA)不变,仅替换损失函数,在相同的壁画数据集上进行修复实验。其中,风格损失通过比较重建图像与输入图像的Gram矩阵来度量它们的风格差异;边缘损失主要用于保

表 2 不同模型对添加中心掩码壁画修复结果评价指标比较

Table 2 Comparative evaluation metrics of different models in inpainting of murals with added central masks

评价指标	模型	Img1	Img2	Img3	Img4	Img5	Img6	平均值
PSNR ↑	LAMA	20.777 8	21.244 2	21.157 6	19.321 5	18.564 7	22.923 6	20.664 9
	DA-CLIP	21.732 8	25.122 6	23.126 7	23.080 4	21.631 9	26.418 8	23.518 7
	MISF	25.109 1	28.463 0	26.712 7	25.780 0	23.570 1	26.829 2	26.077 4
	RePaint	25.466 5	27.419 3	27.214 7	23.082 0	24.728 9	25.289 4	25.533 5
	Ours	27.885 7	30.180 0	28.177 7	26.850 0	25.387 5	28.064 2	27.757 5
SSIM ↑	LAMA	0.814 3	0.816 0	0.816 7	0.796 4	0.795 6	0.810 3	0.808 2
	DA-CLIP	0.876 5	0.873 7	0.871 3	0.868 9	0.865 6	0.876 4	0.872 1
	MISF	0.874 2	0.881 4	0.864 6	0.876 1	0.875 8	0.849 8	0.870 3
	RePaint	0.893 8	0.897 6	0.883 7	0.881 8	0.875 3	0.869 1	0.883 6
	Ours	0.924 2	0.911 0	0.907 2	0.904 5	0.898 3	0.880 6	0.904 3
LPIPS ↓	LAMA	0.143 3	0.222 7	0.148 3	0.176 7	0.144 6	0.155 2	0.165 1
	DA-CLIP	0.074 7	0.108 6	0.093 8	0.087 3	0.073 5	0.072 3	0.085 0
	MISF	0.093 3	0.152 3	0.125 5	0.129 2	0.074 9	0.130 7	0.117 7
	RePaint	0.064 8	0.199 2	0.090 7	0.087 9	0.065 5	0.081 5	0.098 3
	Ours	0.047 7	0.114 7	0.072 6	0.081 7	0.057 5	0.083 8	0.076 3
FID ↓	LAMA	13.223 1	15.073 9	8.302 1	13.167 3	11.223 5	12.126 4	12.186 1
	DA-CLIP	6.251 7	12.208 8	6.201 1	10.456 3	6.037 7	8.594 3	8.291 7
	MISF	8.235 8	11.765 7	7.542 2	11.768 2	8.822 9	9.168 5	9.550 6
	RePaint	6.326 4	11.839 9	7.149 3	10.653 0	8.243 2	8.829 7	8.840 3
	Ours	5.672 8	9.669 3	5.379 0	9.799 7	7.579 8	8.009 8	7.685 1

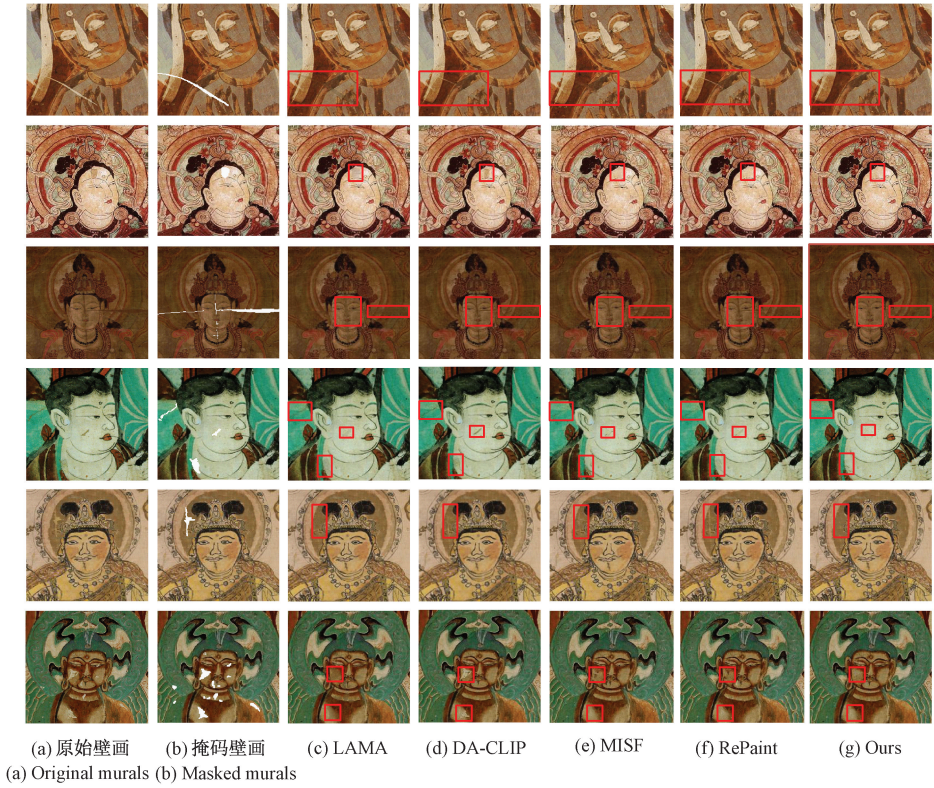


图 6 真实破损壁画修复结果对比

Fig. 6 Comparison of results of inpainting of real damaged mural

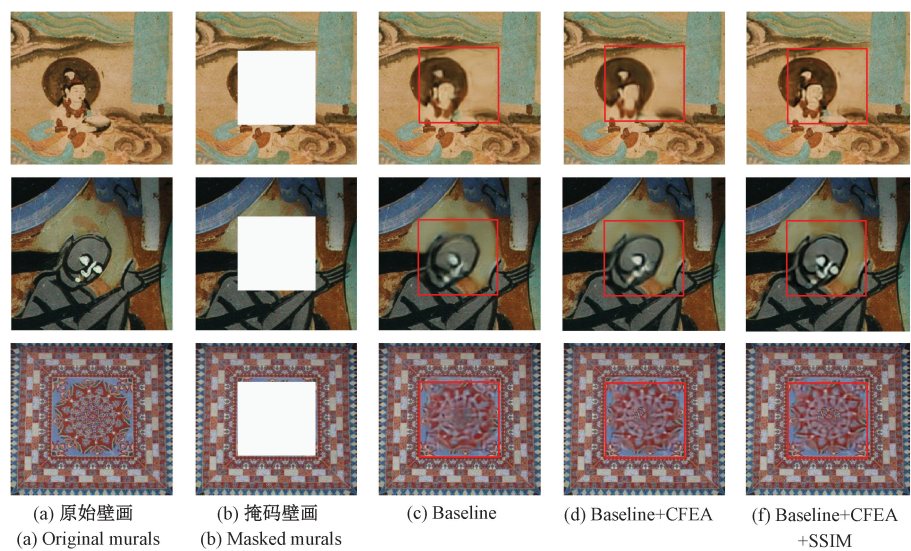


图 7 消融实验壁画修复结果

Fig. 7 Ablation study results for mural inpainting

表 3 消融实验评价指标比较

Table 3 Comparison of evaluation metrics in ablation experiments

评价指标	模型	Img1	Img2	Img3
PSNR ↑	Baseline	29.199 8	26.898 5	25.743 4
	Baseline+CFEA	30.803 3	27.930 0	26.828 7
	Baseline+CFEA+SSIM	31.059 6	29.594 6	27.477 0
SSIM ↑	Baseline	0.892 5	0.878 6	0.839 4
	Baseline+CFEA	0.903 7	0.893 3	0.865 6
	Baseline+CFEA+SSIM	0.909 5	0.903 2	0.879 6
LPIPS ↓	Baseline	0.106 4	0.116 5	0.173 4
	Baseline+CFEA	0.093 1	0.110 7	0.156 9
	Baseline+CFEA+SSIM	0.080 7	0.085 5	0.145 9
FID ↓	Baseline	14.124 0	11.407 1	3.429 1
	Baseline+CFEA	11.219 6	12.358 9	3.065 7
	Baseline+CFEA+SSIM	10.983 9	11.589 0	3.131 1

证修复图像的边缘与输入图像的边缘在结构上的一致性;对抗损失是基于 GAN 的对抗损失。生成器的目标是生成足够逼真的图像,以误导鉴别器。各个损失函数对修复效果的提升如图 8 所示,其中图 8(a)为壁画原始图像,图 8(b)为模拟壁画破损图像,图 8(c)为风格损失修复结果,图 8(d)为边缘损失修复结果,图 8(e)为对抗损失修复结果,图 8(f)为本文损失修复结果。表 4 为不

同损失修复结果评价指标的对比,从表 4 数据可以看出,采用“本文损失”(MSE+SSIM)的模型,在大多数测试图像上取得了最优或接近最优的 SSIM 和 LPIPS 值,并且在平均 FID 分数上显著低于其他损失函数。这表明采用的损失函数组合在促进结构相似度、降低感知差异方面更具优势,从而验证了其针对壁画修复任务的有效性。

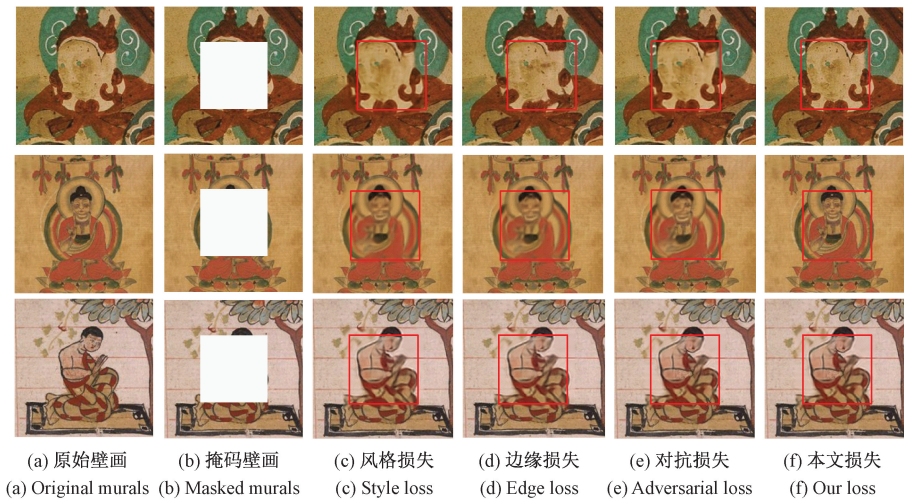


图 8 损失函数对比实验结果

Fig. 8 Results of loss function comparison experiment

表 4 损失函数对比实验评价指标

Table 4 Evaluation metrics for loss function comparison experiment

评价指标	损失函数	Img1	Img2	Img3
PSNR ↑	风格损失	29.283 5	28.913 6	29.159 8
	边缘损失	30.054 8	30.568 2	31.152 6
	对抗损失	30.562 6	31.253 6	31.156 9
	本文损失	30.265 8	32.158 6	32.595 6
SSIM ↑	风格损失	0.905 2	0.896 5	0.901 8
	边缘损失	0.908 9	0.904 5	0.905 7
	对抗损失	0.910 6	0.907 9	0.914 6
	本文损失	0.935 6	0.918 1	0.925 4
LPIPS ↓	风格损失	0.112 6	0.106 9	0.108 5
	边缘损失	0.093 9	0.092 5	0.095 6
	对抗损失	0.095 6	0.092 8	0.100 5
	本文损失	0.075 2	0.078 9	0.065 8
FID ↓	风格损失	11.569 2	11.236 8	10.422 5
	边缘损失	11.326 9	10.822 4	10.265 9
	对抗损失	10.782 8	10.239 7	9.652 3
	本文损失	8.268 4	9.259 4	6.912 6

3 结 论

提出的学习纹理一致性与连续性的壁画修复模型在敦煌壁画修复领域展现了显著的潜力和有效性。模型通过采用 Transformer 架构,能够有效捕获图像的多尺度特征,并在编码器中学习局部纹理特征,同时在解码器中重建图像,强化了局部特征间的联系,提高了图像的全局纹理一致性。引入 CFEM,进一步学习了模型对图像区域间复杂相互作用的捕捉能力,强化局部特征间的联系,更准确地重建图像的纹理和结构,学习纹理连续性特征,缓解

图像模糊。此外,将结构相似度作为主要损失函数,结合像素损失,解决纹理不一致问题,使得修复后的壁画在结构、细节和视觉感知上更贴近原始状态。实验结果表明,在不同程度破损壁画的修复实验中所提模型均优于其他对比模型,验证了其在壁画修复中的有效性和优越性。在未来的工作中,将致力于进一步优化模型,提高学习丰富特征的能力,进一步提升壁画修复质量。

参考文献

[1] WANG H, LI Q Q, JIA S. A global and local feature weighted method for ancient murals inpainting [J]. International Journal of Machine Learning and Cybernetics, 2020, 11(6): 1197-1216.

[2] BERTALMIO M, SAPIRO G, CASELLES V, et al. Image inpainting [C]//27th Annual Conference on Computer Graphics and Interactive Techniques, 2000: 417-424.

[3] DING D, RAM S, RODRIGUEZ J J. Image inpainting using nonlocal texture matching and nonlinear filtering[J]. IEEE Transactions on Image Processing, 2018, 28(4): 1705-1719.

[4] WANG H, LI Q Q, ZOU Q. Inpainting of dunhuang murals by sparsely modeling the texture similarity and structure continuity[J]. Journal on Computing and Cultural Heritage(JOCCH), 2019, 12(3): 1-21.

[5] 乔景慧,苏冠赫,闫书源. 三分量双色反射模型驱动的透明 PET 瓶图像高光去除方法研究[J]. 电子测量与仪器学报, 2025, 39(2): 92-101.

QIAO J H, SU G H, YAN SH Y. Research on highlight removal method driven by three component dichromatic reflection model for transparent PET bottle images[J]. Journal of Electronic Measurement

- and Instrumentation, 2025, 39(2): 92-101.
- [6] WANG N Y, WANG W L, HU W J, et al. Thanka mural inpainting based on multi-scale adaptive partial convolution and stroke-like mask [J]. IEEE Transactions on Image Processing, 2021, 30: 3720-3733.
- [7] CHEN M, ZHAO X D, XU D Q. Image inpainting for digital Dunhuang murals using partial convolutions and sliding window method [J]. Journal of Physics: Conference Series, 2019, 1302(3): 032040.
- [8] CAO J F, ZHANG Z B, ZHAO A D, et al. Ancient mural restoration based on a modified generative adversarial network [J]. Heritage Science, 2020, 8(1): 1-14.
- [9] CIORTAN I M, GEORGE S, HARDEBERG J Y. Colour-balanced edge-guided digital inpainting: Applications on artworks[J]. Sensors, 2021, 21(6): 2091.
- [10] 开志强, 苗锡奎, 马天磊, 等. 基于全局语义学习和显著目标感知的激光干扰图像修复[J]. 仪器仪表学报, 2024, 45(7): 38-51.
- KAI ZH Q, MIAO X K, MA T L, et al. Laser jamming image inpainting based on global semantic learning and salient target awareness [J]. Chinese Journal of Scientific Instrument, 2024, 45(7): 38-51.
- [11] LUGMAYR A, DANELLJAN M, ROMERO A, et al. RePaint: Inpainting using denoising diffusion probabilistic models [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 11461-11471.
- [12] 张绪, 胡峻峰. 基于扩散模型的二阶段细化图像修复模型[J]. 电子测量技术, 2025, 48(17): 142-150.
- ZHANG X, HU J F. Diff-2sIR: Diffusion-based refinement two stage image restoration model [J]. Electric Measurement Technology, 2025, 48 (17): 142-150.
- [13] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [J]. Advances in Neural Information Processing Systems, 2017, 30: 5999-6009.
- [14] WAN Z Y, ZHANG J B, CHEN D D, et al. High-fidelity pluralistic image completion with transformers [C]//IEEE/CVF International Conference on Computer Vision, 2021: 4692-4701.
- [15] 滕诗宇, 何丽君. 融合多特征与全局-局部 Transformer 的图像修复算法 [J]. 电子测量技术, 2025, 48 (6): 121-129.
- TENG SH Y, HE L J. Fusion of multi-features and global-local Transformer for image inpainting [J]. Eletronic Measurement Technology, 2025, 48 (6): 121-129.
- [16] HE K M, CHEN X L, XIE S L, et al. Masked autoencoders are scalable vision learners [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 16000-16009.
- [17] YU J H, LIN Z, YANG J M, et al. Generative image inpainting with contextual attention [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2018: 5505-5514.
- [18] LIU G L, REDA F A, SHIH K J, et al. Image inpainting for irregular holes using partial convolutions [C]//European Conference on Computer Vision (ECCV), 2018: 85-100.
- [19] SUVOROV R, LOGACHEVA E, MASHIKHIN A, et al. Resolution-robust large mask inpainting with fourier convolutions [C]//IEEE/CVF Winter Conference on Applications of Computer Vision, 2022: 2149-2159.
- [20] LUO Z W, GUSTAFSSON F K, ZHAO Z, et al. Controlling vision-language models for multi-task image restoration [J]. ArXiv preprint arXiv:2310.01018, 2023.
- [21] LI X G, GUO Q, LIN D, et al. Misf: Multi-level interactive siamese filtering for high-fidelity image inpainting [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 1869-1878.

作者简介

唐善成, 副教授, 主要研究方向为人工智能、深度学习、自然语言理解。

E-mail: tangshancheng@xust.edu.cn

王妍(通信作者), 硕士研究生, 主要研究方向为机器视觉、图像修复。

E-mail: 23207223121@stu.xust.edu.cn

周彤, 硕士研究生, 主要研究方向为自然语言理解、文字识别。

E-mail: 2713545241@qq.com