



## 论文

# 结构引导的热贡壁画图像大面积缺损修复技术研究

卢嘉钰<sup>1,2,3</sup>, 张效娟<sup>1,2,3\*</sup>

1. 青海师范大学计算机学院, 西宁 810016

2. 藏语智能全国重点实验室, 西宁 810016

3. 青海省非物质文化遗产研究基地, 西宁 810016

\* E-mail: zhxj@qhnu.edu.cn

收稿日期: 2024-12-31; 接受日期: 2025-03-17; 网络版发表日期: 2025-04-21

国家自然科学基金项目(批准号: 62262056)和青海省重点研发与成果转化项目(编号: 2021-GX-111)资助

**摘要** 热贡壁画是人类非物质文化遗产热贡艺术的重要表现形式之一, 具有色彩浓郁艳丽、线条繁复细密等特点. 然而, 长久以来受环境等因素影响, 壁画常出现剥落、龟裂及霉变等问题, 而现有的图像自动修复算法难以重建壁画复杂的结构并生成合理的纹理. 针对热贡壁画结构纹理复杂、破损区域不规则及图像数据集较小等挑战, 提出了一种结构引导的热贡壁画图像大面积缺损修复方法. 首先设计交互式线稿修复模块, 通过引入专家经验对大面积破损的复杂线稿进行修复, 为后续提供较为准确的结构先验; 接着将线稿通过结构引导修复模块避免低层指导信息被削弱的同时, 能够更有效地和纹理信息融合. 该模块将编码的结构信息逐层注入多尺度修复网络中, 同时设计边缘生成判别器进行约束以优化修复结构, 使其更接近真实壁画; 此外, 多尺度特征修复网络引入傅里叶卷积聚合和门控卷积模块以获得更大的感受野来恢复大面积且不规则的缺损区域, 并加入颜色一致性损失以有效降低图像修复过程中颜色等方面的误差. 与AOTGAN, CTSDG, LaMa, ZITS等七种图像修复方法进行定量、定性和主观实验对比分析. 实验表明, 所提方法在缺损比例为30%~50%的大面积缺损图像修复上表现最优, SSIM和PSNR分别提升了37.91%和15.3%, FID和LPIPS则分别下降了94.67%和85.96%.

**关键词** 壁画修复, 结构引导, 傅里叶卷积, 交互式

## 1 引言

热贡艺术发源于13世纪初, 主要包括唐卡、壁画、堆绣、雕塑、建筑彩绘等多种艺术形式, 以佛教中的佛本生故事, 藏族历史人物和神话、传说、史诗等为主要内容<sup>[1]</sup>. 2009年热贡艺术被联合国教科文组

织列入人类非物质文化遗产代表作名录<sup>[2]</sup>. 热贡艺术经过不断提炼、发展, 形成独具一格的艺术特征, 其作品造型准确生动, 工笔精细绝美, 色彩艳丽富于装饰性, 充分发挥了线条的节奏感、运动感和立体感<sup>[1,3,4]</sup>. 留存于热贡当地的壁画大多以寺院中的老壁画为主, 主要保存在隆务寺、郭麻日寺、年都乎寺等

引用格式: 卢嘉钰, 张效娟. 结构引导的热贡壁画图像大面积缺损修复技术研究. 中国科学: 技术科学, 2025, 55: 1422–1434  
Lu J Y, Zhang X J. Research on structure-guided restoration technique for large-area defects in Regong mural images (in Chinese). Sci Sin Tech, 2025, 55: 1422–1434, doi: 10.1360/SST-2024-0343

国家级文物重点保护单位。而由于部分寺院年代久远以及保存不当,寺内壁画普遍存在不同程度的霉变、龟裂起甲、剥落、褪色等迹象,亟待抢救,如图1所示。同时由于壁画的画幅大、破损也较大,对其修复造成了较大的困难。目前,壁画的修复工作多半是以人工修复为主,但该方式修复周期长,且具有不可逆性,可能会由于画师对壁画的理解错误造成效果不佳或者二次破坏的情况发生。

随着数字图像修复技术的不断成熟,衍生出的壁画图像修复技术可以一定程度修复壁画图像,可以低成本且快速地完成大量修复工作。Biradar<sup>[5]</sup>和Sridevi<sup>[6]</sup>等人基于中值滤波、差分曲率驱动等提出传统的修复方法,但传统方法在处理破损区域的大量图像数据时生成的修复图像往往缺乏语义一致性和纹理结构一致性。

Pathak等人<sup>[7]</sup>提出了上下文编码器,将基于上下文的像素预测驱动的无监督特征学习应用于大孔洞图像修复,但其受限于处理固定的低分辨率图像和中心区域破损。Iizuka等人<sup>[8]</sup>提出了一种全局和局部一致的图像修复方法以解决上下文编码器的缺陷,采用双判别器以保证恢复图像的全局语义和局部语义的连贯性,同时在补全网络的中间四层使用空洞卷积<sup>[9]</sup>,增加所提取特征的感受野。

但是使用标准卷积网络修复破损图像时,通常采用有效像素和缺失部分的平均值作为填充,容易使较大的空洞修复区域缺乏纹理信息,产生色差和模糊等伪影,严重影响视觉感官<sup>[10]</sup>。Yu等人<sup>[11]</sup>提出改进门控卷积以取代网络中的普通卷积,更好地解决了将所有输入视为普通卷积中有效像素的问题,为每个空间位置的每个通道提供了一种可学习的动态特征选择机制来泛化部分卷积。

以往的深度学习技术中许多对于结构较复杂的图像无法重建合理的结构,它们的结果通常是过度平滑

或模糊的。Nazeri等人<sup>[12]</sup>提出了一种结合对抗性边缘学习的生成图像修复方法EdgeConnect (generative image inpainting with adversarial edge learning),使用边缘生成器在缺损区域生成一个粗略的轮廓,为第二阶段图像补全网络提供图像结构的先验信息。Xiong等人<sup>[13]</sup>提出并验证了利用结构先验显式指导图像修复任务是一个有效的图像修复方法。Guo等人<sup>[14]</sup>提出以耦合的方式对结构约束的纹理合成和纹理引导的结构重建进行建模,并设计双向门控特征融合模块交换和融合结构和纹理信息。

然而上述许多工作生成整体结构复杂和细节高保真的图像内容仍然是一个巨大的挑战,尤其是当图像缺损区域占比较大时。Suvorov等人<sup>[15]</sup>利用快速傅里叶卷积<sup>[16]</sup>对具有全局感受野的频率场特征进行编码,他们主要针对具有复杂纹理和大面积缺失区域的图像进行鲁棒性修复,论文中实验的掩码比例设置在50%左右,但它们难以保证壁画图像完整的复杂结构。Zheng等人<sup>[17]</sup>提出级联调制生成网络利用傅里叶卷积块的编码器从带有掩码的输入图像中提取多尺度特征,并对解码器每个尺度层设置级联全局空间调制块以更好地合成整体结构和局部细节。Dong等人<sup>[18]</sup>在固定的低分辨率草图空间中使用基于注意力的Transformer模型恢复整体图像结构,通过零初始化的残差加法有效地与其他预训练的修复模型集成。Zeng等人<sup>[19]</sup>提出了聚合上下文变换模块,在10%~60%图像缺失区域进行实验,解决修复图像中可能产生的扭曲结构和模糊纹理,实现了大缺失区域的细粒度合成。曹岩等人<sup>[20]</sup>针对覆盖范围在40%~50%的大规模不规则掩码提出了多元场景下的大面积破损图像修复方法,以解决图像修复过程中图像像素之间语义联系不连贯等问题。

但是以上的自动修复网络往往会生成不合理的结构。因此,研究人员利用手动绘制破损区域纹理结构的方式,来指导网络进行修复。Jo等人<sup>[21]</sup>则利用SC-FE-GAN (face editing generative adversarial network with user's sketch and color)网络通过学习和理解用户输入的草图和颜色信息,自动调整人脸特征,生成高质量、逼真的编辑结果。Zeng等人<sup>[22]</sup>提出SketchEdit (mask-free local image manipulation with partial sketches)基于用户在图像上绘制的草图指导网络进行图像局部编辑。然而将输入图像和结构等低层控制信息结合输入

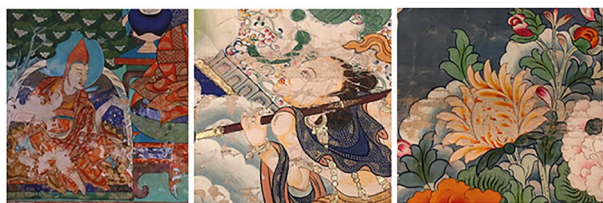


图1 破损壁画(局部)  
Figure 1 Damaged mural (partial).

至网络中会在卷积神经网络空间中逐渐减少, 最终生成的结果不足以传达用户意图甚至产生错误的引导. Liu等人<sup>[23]</sup>在每个跳跃连接中设计一个结构生成块, 将低级的控制信息分别注入到每一个结构生成块中, 并将调制特征与原始解码器特征连接以生成合理的结构. Liu等人<sup>[24]</sup>提出SketchRefiner (towards interactive image inpainting via sketch refinement)通过线稿优化等模块来校准和优化用户提供的草图以解决草图的错位和不连贯的问题, 同时引入快速傅里叶卷积模块进行纹理的修复. 尽管它在自然图像(尤其是结构较为简单的图像)的修复中表现出色, 但当应用于结构复杂的热贡壁画图像时, 往往难以有效利用线稿信息, 进而可能导致修复结果出现结构混乱的情况. 以上这些方法在修复或完成某些特定对象, 如人脸、印刷文字等方面表现出色, 但不适合修复纹理结构复杂的热贡壁画.

相比于自然图像, 热贡壁画图像含有大量的结构、复杂的纹理以及特殊的人物形态, 具有色彩浓艳丽、线条繁多且细密、纹理复杂内涵丰富等特点. 直接应用目前流行的修复算法, 往往出现结构缺失和细节模糊等现象. 然而目前的壁画图像修复也主要针对较小缺损或结构简单的图像进行修复, 较大缺损或结构复杂的缺损图像修复结果同样会出现纹理缺失和线条模糊等现象. 郭开俊等人<sup>[25]</sup>提出多阶段堆叠融合下的壁画修复模型, 设计了粗修复、精细化修复、全局细节修复等三个阶段以解决壁画细节修复的难题; 丁元等人<sup>[26]</sup>针对敦煌壁画数字化修复过程中存在的掩膜区域语义不合理的问题提出频域扩散信息引导下的敦煌壁画修复模型, 提出一种信息对比学习策略来实现对语义相关性的监督. 上述论文的工作重心均聚焦于敦煌壁画图像的修复, 值得注意的是, 相较于敦煌壁画数据集, 热贡壁画数据集展现出更为复杂的结构特征和的颜色分布. 廖年鸿等人<sup>[27]</sup>提出的轻量化运算修复网络, 专注于针对与热贡壁画图像特征相仿的唐卡图像进行修复. 尽管该方法在处理缺损较小或结构相对简单的唐卡图像时表现出色, 但在面对结构更为复杂, 缺损面积更大的图像时, 其修复结果仍存在一定的不足.

综上所述, 目前的图像修复算法已经取得了良好的效果. 但应用到结构纹理复杂、同时破损较大且数据集较小的热贡壁画数据集时, 容易出现结构不合理, 纹理模糊等现象. 因此本文基于生成对抗架构, 针对热

贡壁画图像提出一种结构引导的热贡壁画图像大面积缺损修复技术(structure-guided restoration technique for large-area defects in Regong mural images, SGRT), 首先设计半交互的方式为修复网络提供更接近原图的结构先验信息, 这一步骤显著提升了网络对壁画复杂结构的理解能力. 之后通过结构引导修复模块将多尺度的结构先验信息注入纹理修复模块中引导修复, 从而在修复过程中提供强有力的结构引导. 同时, 在多尺度纹理修复模块引入门控卷积、跳跃连接和多尺度傅里叶卷积模块等使得网络能够更为针对性地对于不规则且大面积破损的热贡壁画图像进行结构引导下的修复. 最后设计了整体图像和边缘细节相结合的判别器, 确保在保持图像整体一致性的同时, 精细地恢复出复杂的图像结构和细节.

本文的主要创新和贡献如下.

(1) 提出了一种基于结构引导的热贡壁画图像多尺度修复网络, 通过半交互的方式修复缺损的线稿图, 为具有复杂结构的壁画图像提供结构先验信息.

(2) 设计了结构引导修复模块和边缘判别器, 边缘判别器确保了生成器恢复出结构一致的图像, 结构引导修复模块将多尺度的结构先验信息引导纹理修复, 避免结构先验信息在网络传输中减弱结构紊乱的同时实现与纹理信息的深度融合.

(3) 通过引入门控卷积、跳跃连接和多尺度傅里叶卷积模块(multi-scale Fourier convolution module, MFC)等提高网络对于信息的有效提取, 有效地修复不规则且破损较大的热贡壁画图像, 能够融合强引导的结构先验信息, 保留了原壁画艺术风格的同时, 修复出结构纹理清晰, 颜色均匀的热贡壁画图像.

## 2 本文模型

### 2.1 网络整体框架

热贡壁画图像结构繁复、细节繁多, 尤其面临数据集稀缺的挑战下, 为了实现对较大面积破损的热贡壁画图像进行修复, 使之尽可能接近原图, 本文设计了一个结构引导的热贡壁画图像大面积缺损修复网络, 网络的整体结构如图2所示. 网络整体采用生成对抗架构, 生成器部分包括线稿交互修复模块、结构引导修复模块和多尺度纹理修复模块. 判别器则以Isola等人<sup>[28]</sup>提出的基于patch的生成对抗网络的判别器部

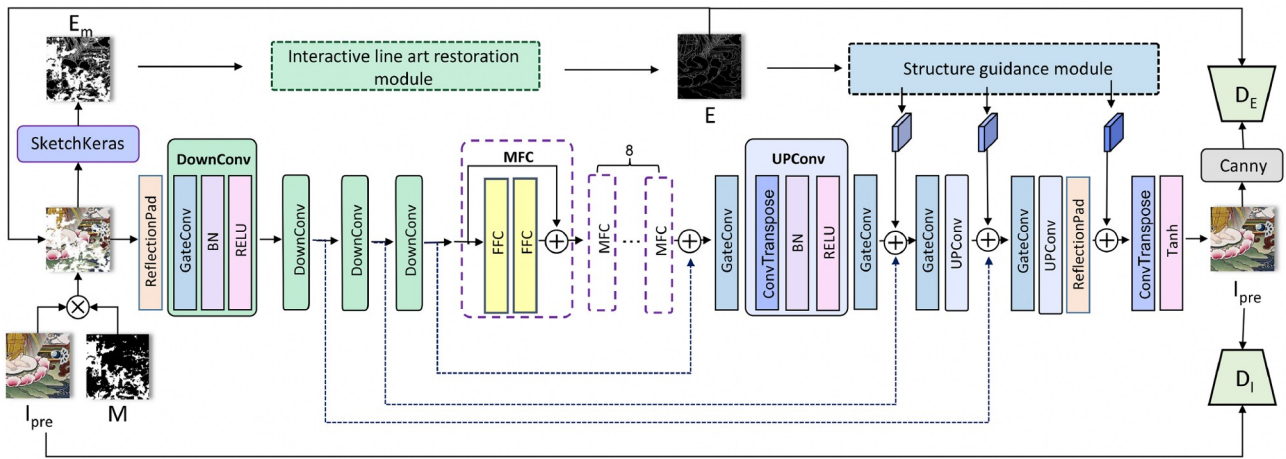


图 2 (网络版彩图)结构引导的热贡壁画图像大面积缺损修复网络

Figure 2 (Color online) Structure-guided repair network of large area defect in Regong mural images.

分(PatchGAN)为基础, 分别设计了图像判别器 $D_I$ 和边缘判别器 $D_E$ .

本文热贡壁画图像修复网络的工作流程如下. 首先使用SketchKeras线稿提取算法<sup>1)</sup>从输入的缺损壁画图像中提取破损线稿图, 热贡壁画非遗传承人等艺术家的指导下进行线稿的交互修复, 旨在为之后的修复提供更精确的结构先验信息, 确保修复后的线稿尽可能贴近原壁画的结构特征; 随后将补全的线稿图像通过结构引导修复模块进行特征提取, 并将提取的多尺度结构特征逐层注入到多尺度纹理修复模块的解码器中, 从而更有效地引导图像的修复过程. 同时, 将破损的壁画图像输入至多尺度纹理修复模块中, 首先经过反射填充层和下卷积模块, 在保持图像信息完整的同时, 通过门控卷积机制有选择地处理输入像素, 从而更有效地应对不规则的缺损区域, 并显著减少修复结果中的伪影现象; 接下来, 提取到的特征被送入多尺度傅里叶卷积模块中, 该模块巧妙地将傅里叶卷积模块与普通卷积相结合, 不仅擅长处理较大破损的图像, 还能细致入微地捕捉和修复局部纹理信息, 确保修复后的图像在整体和细节上都达到更好的效果; 接着将输出的特征输入至之后的门控卷积模块和上卷积模块, 同时结合从下卷积经过跳跃连接得到的不同尺度信息能够更有效地捕捉图像的上下文信息和细节特征, 也减少了结构引导过强导致的细节生硬. 最后, 两

个判别器分别专注于图像整体内容和图像边缘特征的判别, 保持图像整体一致性的同时, 精细地恢复出复杂的图像结构和细节, 从而显著提升结构复杂图像的修复效果.

## 2.2 交互式修复模块

自动的图像修复算法修复热贡壁画图像这类线条繁多且细密、纹理复杂内涵丰富的图像效果较差, 往往会出现结构紊乱、纹理模糊等现象. 同时, 壁画作为独特艺术形式, 需遵循特定规范和标准, 自动修复方法可能导致修复结果与壁画内容不符. 针对此问题, 本文提出线稿交互修复的思想, 借助热贡艺术壁画传承人的经验确保图像整体结构的准确性从而能够为后续的修复工作提供准确的结构先验信息并生成更接近原图的壁画图像. 交互式修复模块采用SketchKeras算法(表示为SK)提取破损图像的线稿, 该算法能够更为模拟壁画图像的真实线稿. 用 $I_{gt}$ 表示原始图像,  $E_{gt}$ 和 $E_m$ 分别表示完整图像和破损图像的线稿图, 则 $E_m$ 表示为

$$E_m = SK(I_{gt} \times M), \quad (1)$$

式中,  $M$ 表示模拟得到的mask(缺失区域为1, 未缺失区域表示0). 在专业人员的指导下, 对线稿的破损区域进行准确修复得到 $E_i$ , 确保线稿的完整性和准确性. 最终

1) <https://github.com/higumax/sketchKeras-pytorch>.

得到修补完整的线稿图 $E$ :

$$E = E_m + E_i. \quad (2)$$

### 2.3 结构引导修复模块

现有的基于先验信息引导的图像修复方法, 主要聚焦于在图像层面将输入图像与先验信息进行简单融合, 以此作为神经网络的初始输入. 然而这些低层输入的指导信息在特征空间的传递过程中会逐渐减弱其影响力, 导致生成的图像内容可能出现结构上的混乱、细节信息的丢失或逻辑上的不一致. 为了克服这一局限性, 本文设计了一个结构引导修复模块, 该模块旨在有效保留低层次的指导信息, 并同时实现与纹理信息的深度融合, 从而优化基于深度学习的图像修复流程. 不同于将低级控制信息直接附加到输入图像上的做法, 而是将这些控制信息嵌入到修复网络解码器的每一层级中, 以便控制信息能够在卷积神经网络的特征空间内进行更为精确和细致的结构引导处理. 在此过程中, 采用了相加而非拼接的连接方式, 以避免过强的指导信息可能导致的生成结果不合理的问题, 从而确保图像修复过程的高质量完成. 该模块的具体结构如图3所示.

### 2.4 多尺度纹理修复网络

热贡壁画数据集具有元素复杂、风格多样、色彩艳丽等特点, 然而其数据集规模相对较小, 这为大面积缺损的图像修复任务带来了挑战. 当前主流的大面积缺损修复算法在处理此类数据集时, 往往难以以充分学习图像特征而导致修复结果不尽人意, 具体表现为修复区域与原图在纹理、结构和风格上的不一致, 并且可能产生伪影或模糊现象. 本文设计了结合门控卷积和多尺度傅里叶卷积模块(MFC)的多尺度纹理修复网络, 如图2所示. 该网络利用门控卷积机制来动态调整特征流, 增强模型对关键信息的捕捉能力. 随后, 多尺度傅里叶卷积模块凭借结合频域分析和局部卷积模块, 以有效捕捉图像的全局信息和局部信息, 残差连接进一步增强了模型对全局信息的整合能力, 其具体结构如图4所示. 傅里叶卷积通过将图像转换到频域进行处理, 能够更有效地捕捉图像的全局结构信息. 特别是当数据集较小时, 模型往往难以从有限的训练样本中学习到足够的特征. 傅里叶卷积通过引入频域特征, 增加了模型的输入维度和特征表示能力, 使得模

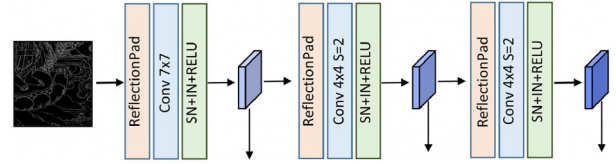


图3 (网络版彩图)结构引导模块

Figure 3 (Color online) Structure guidance module.

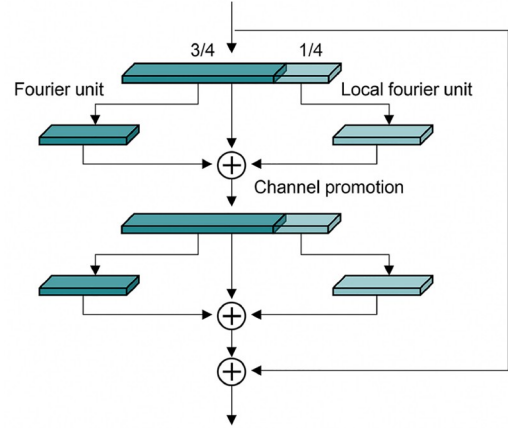


图4 (网络版彩图)多尺度傅里叶卷积模块

Figure 4 (Color online) Multi-scale Fourier convolution module.

型能够更好地学习到图像中的潜在规律和特征. 为了弥补傅里叶卷积在精细纹理合成方面的不足, 快速傅里叶卷积(fast Fourier convolution, FFC)模块将四分之一维度的数据通过局部傅里叶卷积模块精准提取图像的局部特征. 其中采用SpectralTransform, 通过将输入数据转换到频域并提取频谱特征, 实现了频域和空域特征的有效融合.

为了更好地修复合成细节纹理, 在整个纹理修复网络中引入跳跃连接, 将下卷积模块得到的特征直接传递至上卷积模块. 同时, 结构引导模块得到的多尺度结构特征也逐层注入其中, 作为解码器的一部分, 以增强模型对局部细节的恢复能力. 保留全局结构信息的同时确保局部特征的精确重建, 在处理大面积破损时, 能够保持图像的整体一致性和连贯性, 生成与原图更为一致的修复结果.

### 2.5 判别器

本文采用了两个基于PatchGAN<sup>[28]</sup>架构的判别器, 分别用于区分原图和生成图的整体质量以及它们的边缘细节. PatchGAN是一种特殊的卷积神经网络, 它通

过对图像中的局部区域patches进行分类, 来判断整个图像的真实性. 减少计算量的同时, 还使得判别器能够捕捉到图像中的精细纹理和局部特征. 这种设计使得PatchGAN能够关注图像的局部细节, 使得模型在训练过程中能够更细致地捕捉图像的细节特征.

本文在传统的一个判别器判别整体图像质量的基础上增加了另一个用于边缘细节的判别器, 有助于准确恢复边缘信息, 从而保持图像的连贯性和视觉质量. 其输入是经过Canny算法得到的真实图像和预测图像的边缘图像, 与生成器进行对抗性训练, 促使生成器更精细地控制图像结构的生成, 使得预测图像的边缘更加锐利和准确, 接近真实图像.

### 3 损失函数

#### 3.1 $L_1$ 损失函数

在计算 $L_1$ 损失时, 仅考虑掩码 $\mathbf{M}$ 中标记为无效(即需要被修复)的像素, 使得模型能够更加专注于需要修复的区域, 而忽略已知的背景或不需要修复的区域, 从而提高了模型对这些区域修复质量的关注度, 并进而提升了整体的修复效果.

$$L_1 = (1 - \mathbf{M}) \|\mathbf{I}_{\text{pre}} - \mathbf{I}_{\text{gt}}\|_1, \quad (3)$$

式中,  $\mathbf{I}_{\text{pre}}$ 和 $\mathbf{I}_{\text{gt}}$ 分别代表预测图像和真实图像. 其中,  $\mathbf{M}$ 是一个0-1掩码, 其中1表示掩码区域.

#### 3.2 对抗损失

对抗性损失由判别器损失 $L_D$ 和生成器损失 $L_G$ 组成. 基于PatchGAN的判别器被表示为 $D$ , 而整个结构引导的热贡壁画图像大面积缺损修复方法则被视为生成器 $G$ . 因此, 对抗损失可以表示为

$$L_D = -E_{\mathbf{I}_{\text{gt}}}[\log D(\mathbf{I}_{\text{gt}})] - E_{\mathbf{I}_{\text{pre}}, \mathbf{M}}[\log D(\mathbf{I}_{\text{pre}}) \odot (1 - \mathbf{M})] - E_{\mathbf{I}_{\text{pre}}, \mathbf{M}}[\log(1 - D(\mathbf{I}_{\text{pre}})) \odot \mathbf{M}], \quad (4)$$

$$L_G = -E_{\mathbf{I}_{\text{pre}}}[\log D(\mathbf{I}_{\text{pre}})], \quad (5)$$

$$L_{\text{GP}} = E_{\mathbf{I}_{\text{gt}}} \|\nabla_{\mathbf{I}_{\text{gt}}} D(\mathbf{I}_{\text{gt}})\|^2, \quad (6)$$

$$L_{\text{adv}} = L_D + L_G + \lambda_{\text{GP}} \times L_{\text{GP}}, \quad (7)$$

式中,  $D(x)$ 表示判别器对图像 $x$ 的预测概率;  $E$ 表示期望,  $L_{\text{GP}}$ 表示梯度惩罚.  $\lambda_{\text{GP}} = 1 \times 10^{-3}$ .

#### 3.3 感受野匹配损失

传统的感知损失通过预先训练的基础网络来评估从预测图像和目标图像中提取的特征之间的距离. 它不需要精确的重建, 允许重建图像中的变化. 大掩码修复的重点需要转移到对全局结构的理解上. 因此, 引用了一个基于快速增长感受野的基础网络的高感受野感知损失<sup>[15]</sup> $L_{\text{hrf}}$ . 与传统感知损失不同的是 $L_{\text{hrf}}$ 使用扩张卷积来实现, 以此提高大掩码修复的质量:

$$L_{\text{hrf}} = E \left( \left[ \phi_{\text{hrf}}(\mathbf{I}_{\text{gt}}) - \phi_{\text{hrf}}(\mathbf{I}_{\text{pre}}) \right]^2 \right), \quad (8)$$

式中,  $\phi_{\text{hrf}}$ 表示具有扩张卷积的预训练分割ResNet50<sup>[29]</sup>所训练得到的高感受野感知损失.

#### 3.4 颜色一致性损失

提出的颜色一致性损失函数是对 $L_2$ 损失函数进行了修改, 使其更加关注考虑颜色一致性. 首先将图像分为红色(R)、绿色(G)、蓝色(B)三个通道, 并分别计算每个通道上预测图像与目标图像之间的平方差异. 这一步确保了颜色一致性损失能够考虑到图像中的颜色信息, 而不仅仅是亮度或灰度值. 使用掩码来区分已知区域和缺失区域, 并且应用不同的权重. 可以调整允许损失函数在训练过程中更加灵活地关注图像的需要修复的部分. 计算所有颜色通道上差异的平均值, 然后乘以像素权重, 最后取平均值作为最终的损失值. 颜色一致性损失函数(color consistency loss function, CCL)可以表示为

$$L_{\text{CCL}} = \frac{1}{n} \sum_{i \in \text{pixels}} \left[ \omega_{\text{known}}(1 - m_i) + \omega_{\text{missing}} m_i \right] \times \left[ (p_i^{\text{R}} - t_i^{\text{R}})^2 + (p_i^{\text{G}} - t_i^{\text{G}})^2 + (p_i^{\text{B}} - t_i^{\text{B}})^2 \right], \quad (9)$$

式中,  $n$ 是参与计算的像素总数.  $m_i$ 是掩码在像素 $i$ 上的值(0表示已知, 1表示缺失).  $\omega_{\text{known}}$ 和 $\omega_{\text{missing}}$ 分别是已知区域和缺失区域的权重, 分别为10和5.  $p_i^{\text{R}}, p_i^{\text{G}}, p_i^{\text{B}}$ 分别是预测图像在像素 $i$ 上的红色、绿色、蓝色通道值.  $t_i^{\text{R}}, t_i^{\text{G}}, t_i^{\text{B}}$ 分别是目标图像在像素 $i$ 上的红色、绿色、蓝色通道值.

#### 3.5 总损失

综上所述, 总损失函数如下:

$$L_{\text{final}} = \lambda_{L_1} L_1 + \lambda_{\text{adv}} L_{\text{adv}} + \lambda_{\text{hrf}} L_{\text{hrf}} + \lambda_{\text{CCL}} L_{\text{CCL}}. \quad (10)$$

本文的总损失 $L_{\text{final}}$ 如公式(10)所示, 受文献[18]启发, 结合实验调整, 最终权重设置为 $\lambda_{L1}=10$ ,  $\lambda_{\text{adv}}=10$ ,  $\lambda_{\text{hrf}}=30$ ,  $\lambda_{\text{CCL}}=15$ .

## 4 实验

### 4.1 数据集

为了验证本文所提出的热贡壁画图像修复网络的有效性和泛化性, 本文构建了热贡壁画图像数据集和CelebA<sup>[30]</sup>图像数据集. 本文使用的壁画图像均采集于青海省黄南藏族自治州同仁地区的佛教寺院和画院. 由于文物保护环境与数字采集设备的限制, 原始采集图像存在几何形变与光照干扰问题, 需要对数据集进行预处理. 本文采用透视变换、亮度均衡化等操作对图像进行调整, 初步得到547张热贡壁画图像. 鉴于原始采集的图像具有较高的分辨率(约2500×3500)且直接降采样后过于模糊的图像无法用于训练, 本文以256×256像素为基准单位, 沿水平和垂直方向进行图像裁剪, 同时剔除模糊、残缺的子图像, 最终构建包含19570张的符合训练条件的热贡壁画图像数据集. 之后, 按照8:1:1的比例将数据集划分为训练集、验证集和测试集. 在训练阶段, 由于数据量较大, 若逐个绘制线稿图, 工作量巨大, 且可能会出现偏差, 因此采用SketchKeras算法对裁剪后的图片进行线稿提取, 得到对应的线稿数据集, 在实际应用阶段, 则由热贡壁画艺术家进行指导绘制.

与普通掩码数据集不同, 热贡壁画图像的损坏类型包括但不限于颜料沿折痕脱落、部分颜料块状或点状脱落、不规则方向裂纹等. 这些缺损既包括逐行分布的零散缺失, 也涵盖连续的大块缺失, 甚至还存在这两种缺损模式的复合情况, 这无疑增加了图像修复的难度与挑战性. 因此本文采用Photoshop模拟这几种破损形状, 针对性地制作掩码数据集以针对性地解决壁画缺损的问题. 经过裁剪处理后, 图像的缺损比例主要集中于30%至50%的范围内, 且缺损形态颇为复杂多样, 符合壁画真实的破损情况. 如图5所示, 裁剪成256×256大小的掩码图, 共得到5320张掩码图. 按5:1随机分为训练集和测试集.

为验证本文方法的泛化性, 使用公共数据集CelebA<sup>[30]</sup>进行实验, 本文随机选取25000张, 按8:1:1的比例将这些图像分为训练集、测试集和验证集.

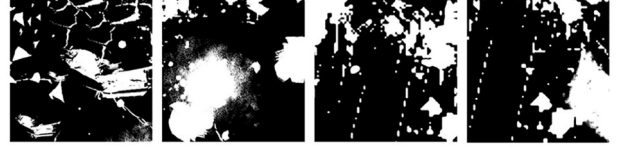


图5 掩码数据集  
Figure 5 Mask dataset.

### 4.2 实验设置

本试验采用有监督训练模式, 生成器学习率为 $1 \times 10^{-3}$ , 判别器学习率为 $1 \times 10^{-4}$ , Batchsize为8, 采用Adam优化器. 本实验均在Ubuntu20.04操作系统上完成, 硬件环境: CPU Intel(R) Xeon(R) Platinum 8362, GPU NVIDIA RTX3090 24GB; 软件环境: Python=3.8, Pytorch=1.10.1, CUDA=11.1.

### 4.3 评价指标

在定量对比实验中, 实验使用峰值信噪比(peak signal-to-noise ratio, PSNR)、结构相似度(structural similarity index, SSIM)、弗雷歇距离分数(frèchet inception distance, FID)、学习感知图像块相似度(learned perceptual image patch similarity, LPIPS). PSNR能够较好地反映图像的客观质量, 主要衡量图像的整体失真程度. SSIM用于衡量两幅图像的相似性, 包括亮度、对比度和结构三个方面, SSIM值越接近1, 表示修复的图像结构信息越相似于原始图像. FID通过计算修复后图像与真实图像在Inception网络中的特征向量的距离来衡量图像质量, FID越小, 表示修复的图像分布越接近原始图像分布. LPIPS是基于深度学习的方法, 能够更好地捕捉到人眼的视觉感知特性, 更符合人类对图像质量的评价, 对细节和纹理的失真更为敏感.

### 4.4 实验结果

采用壁画数据集和CelebA<sup>[30]</sup>数据集进行图像修复对比实验, 对比方法包括近年来几种先进的图像修复方法: CTSDG<sup>[14]</sup> (image inpainting via conditional texture and structure dual generation), AOTGAN<sup>[19]</sup> (aggregated contextual transformations for high-resolution image inpainting), LaMa<sup>[15]</sup> (resolution-robust large mask inpainting with Fourier convolutions), ZITS<sup>[18]</sup> (incremental transformer structure enhanced image

inpainting with masking positional encoding), Deflocnet<sup>[23]</sup> (deep image editing via flexible low-level controls), SketchEdit及SketchRefiner, 以验证本文方法在壁画图像修复上的可靠性和优越性. CTSDG以耦合的方式对结构约束纹理合成和纹理引导的结构重建进行建模, 以便两个生成任务更好地相互促进, 以获得更准确的结果. AOTGAN提出用于高分辨率图像修复的聚合上下文变换生成对抗网络, 捕获信息丰富的远程上下文, 以进行上下文推理, 模型有助于生成器合成细粒度纹理. LaMa针对大面积缺失和高分辨率图像修复提出基于快速傅里叶卷积模块的网络并提出高感受野感知损失, 能够实现修复结果的结构和纹理的一致性. ZITS提出新的增强Transformer结构, 在固定的低分辨率草图空间中使用强大的基于注意力的Transformer模型恢复整体图像结构, 并结合掩码位置编码、傅里叶CNN纹理恢复等模块修复出纹理清晰、结构合理的图像.

同时对比了一些利用交互信息的对比实验. Deflocnet, SketchEdit和SketchRefiner在修复过程中, 都利用了用户提供的交互信息. Deflocnet将用户提供的颜色和结构控制信息直接注入到每个结构生成块中在CNN特征空间中进行草图线条细化和颜色传播, 有效传递用户意图, 并得到想要的编辑结果. SketchEdit提出仅需要用户简单的草图输入, 就能实现图像的局部修改的图像编辑网络, 使用互相关损失函数来鲁棒地校准和细化草图, 并在第二阶段从特征空间中的抽象草图中提取信息特征, 并调整修复过程.

4.4.1 定量分析

表1列出了不同修复算法在热贡壁画图像数据集和CelebA<sup>[30]</sup>数据集上的PSNR, SSIM, FID和LPIPS. 结果表明, 本文提出方法表现较好, 证明了本文方法的优越性. 相比其他修复算法的评价指标得分, 热贡壁画数据集上本文方法的PSNR, SSIM, FID和LPIPS提升明显, 分别提升8.385, 0.129, 73.688和0.098; CelebA<sup>[30]</sup>数据集上本文方法的PSNR, SSIM, FID和LPIPS分别提升7.862, 0.082, 17.115和0.068. 表明本文方法的修复结果在图像像素结构、分布相似性和感知相似性方面与真实图像更为接近.

4.4.2 定性分析

热贡壁画图像数据集上的修复实验可视化结果如图6(a)所示, 可以看出本文方法相较其他七类修复算法有更好的表现. CTSDG, AOTGAN, LaMa, ZITS虽然提出了提升大面积缺损修复或者高分辨率图像修复结果的算法, 但对于结构复杂, 元素众多且数据集较小的热贡壁画数据集上, 在没有线稿指导的情况下无法恢复出缺损区域的结构. 例如第一行的佛脚和花叶的部分出现了结构模糊的现象; Deflocnet虽然结合了用户给定的草图和颜色信息, 但是不能很好地利用这些信息, 可以从图中看出, 仍会出现结构纹理扭曲的情况; 相比之下, 本文的修复算法能够得到结构更加合理清晰、纹理质量和视觉质量更高的结果. 可以看出, 对于元素众多且构成复杂的图像, SketchRefiner所提出的草图优化网络反而会导致部分关键结构信息缺失

表 1 热贡壁画数据集和CelebA数据集上定量指标评估结果<sup>a)</sup>  
Table 1 Quantitative evaluation results on the Regong mural and CelebA datasets<sup>a)</sup>

| Methods                 | Regong mural dataset |              |               |              | CelebA <sup>[30]</sup> dataset |              |              |              |
|-------------------------|----------------------|--------------|---------------|--------------|--------------------------------|--------------|--------------|--------------|
|                         | PSNR↑                | SSIM↑        | FID↓          | LPIPS↓       | PSNR↑                          | SSIM↑        | FID↓         | LPIPS↓       |
| AOTGAN (TVCG2023)       | 23.075               | 0.871        | 27.465        | 0.093        | 28.193                         | 0.907        | 14.160       | 0.089        |
| CTSDG (ICCV2021)        | 22.948               | 0.876        | 43.637        | 0.088        | 30.650                         | 0.945        | 17.122       | 0.064        |
| LaMa (WACV2022)         | 23.912               | 0.877        | <u>13.901</u> | 0.060        | 30.492                         | 0.926        | 7.350        | 0.043        |
| ZITS (CVPR2022)         | 23.673               | 0.869        | 20.904        | 0.074        | <u>30.900</u>                  | 0.933        | <u>4.898</u> | <u>0.032</u> |
| Deflocnet (CVPR2021)    | 22.114               | 0.843        | 77.834        | 0.114        | 30.693                         | <u>0.949</u> | 13.364       | <u>0.032</u> |
| SketchEdit (ICVF2022)   | <u>28.035</u>        | <u>0.945</u> | 40.413        | <u>0.045</u> | 28.212                         | 0.937        | 19.545       | 0.039        |
| SketchRefiner (TMM2024) | 22.983               | 0.870        | 26.945        | 0.065        | 30.601                         | 0.939        | 16.927       | 0.040        |
| SGRT                    | <b>30.499</b>        | <b>0.972</b> | <b>4.146</b>  | <b>0.016</b> | <b>34.055</b>                  | <b>0.989</b> | <b>2.430</b> | <b>0.021</b> |

a) 加粗字体表示最优结果, 下划线字体表示次优结果

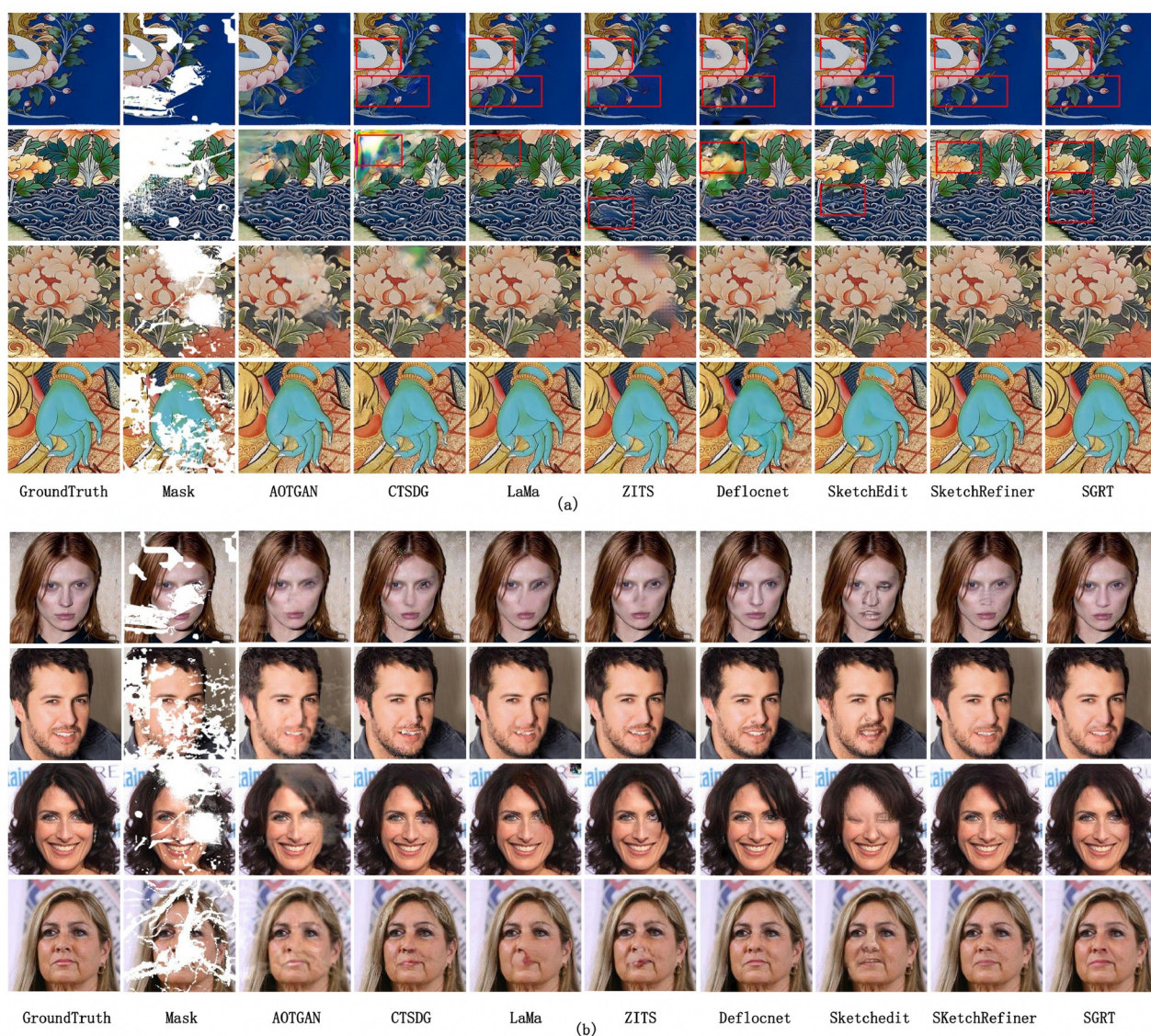


图 6 在不同数据集上的修复结果对比图. (a) 热贡壁画数据集; (b) CelebA<sup>[30]</sup>数据集  
Figure 6 Comparison of restoration results on different datasets. (a) Regong mural dataset; (b) CelebA<sup>[30]</sup> dataset.

和噪音干扰的情况同时存在.

CelebA<sup>[30]</sup>数据集上的修复实验可视化结果如图6(b)所示, 本文方法相较于其他七类修复算法也有更好的表现. AOTGAN和LaMa在重建结构方面比较欠缺, 如第一行的鼻子几乎无法修复, 同时AOTGAN出现了较严重的伪影; CTSDG和ZITS相较于在结构恢复方面有更好的结果, 但对于大面积缺失, 恢复的结构仍是不合理的; Deflocnet在结构和元素较为简单的人脸数据集上能够较好地恢复出合理的结构和纹理, 但是仍会因为没能很好利用结构和颜色交互信息出现部分扭曲的纹

理等, 如第一行和第二行的鼻子处; SketchEdit的修复结果出现了严重的扭曲, 没能完美融合结构和纹理信息; SketchRefiner的修复结果存在局部结果模糊与细节失真问题, 如第一行和第四行的鼻子处最为明显; 本文方法能够融合结构引导信息和纹理特征, 恢复出视觉质量更好的结果.

#### 4.4.3 主观分析

由于热贡壁画图像的复杂性和特殊性, 只使用修复自然图像的通用评价指标对热贡壁画图像的修复结

果对修复结果的质量进行评价是不完整不准确的, 为了全面评估修复结果, 本文针对热贡壁画图像的修复结果设计了主观评价实验环节。

本实验邀请15名参与者对热贡壁画图像的修复结果进行主观评价。分别按照未接触过热贡壁画、对热贡壁画有一定了解和热贡壁画修复专家或热贡壁画艺术传承人的类别分为三组, 每组五人, 每组分别记为实验组1、实验组2和实验组3。分别向三组参与者匿名展示本文方法和对比方法, 要求参与者们从结构连贯性、色彩还原度、细节保留、文化元素准确性和整体视觉美感五个方面对修复结果进行打分。满分为5分, 三组按照1:3:6的权重对初步打分结果进行加权计算, 得到最终的主观评价结果。最终结果如表2所示, 分数越高代表主观评价越好。实验结果表明, 本文方法的评分明显高于其他方法, 这表明我们提出的热贡壁画图像修复方法是可行有效的, 并且相对符合人类的审美。

4.5 消融实验

为了验证交互式线稿修复、边缘判别器、跳跃链

接与卷积层设计以及颜色一致性损失等四个模块的有效性, 本文进行了消融实验, 分别在去除交互式线稿修复, 移除边缘判别器, 将跳跃连接和卷积层替换为普通卷积层, 忽略颜色一致性损失, 以及不去除任何模块的五种情况下进行实验, 并确保该消融实验均在同一参数、硬件和软件条件下进行。表3显示了五种情况下在热贡壁画数据集上的PSNR, SSIM, FID和LPIPS结果。消融实验的视觉对比结果如图7所示, 在去除交互式线稿修复模块时, 大面积缺损修复结构中复杂的区域(如图中的花朵和云朵)出现了明显的结构模糊现象, 这证明交互式线稿修复模块(interactive line art restoration, ILAR)通过引入用户提供的线稿信息, 为修复网络提供了明确的结构指导, 从而有效提升了修复结果的准确性和结构清晰度。在移除边缘判别器(edge discriminator, ED)之后, 修复图像的边缘的锐度和连贯性有所下降, 整体修复结果的视觉质量降低。进一步地, 将跳跃连接和门控卷积(skip connections and gated convolutions, SC+GC)等替换为普通卷积时, 修复结果的纹理质量和细节表现均有所下降, 这一现象在细节

表 2 热贡壁画数据集上不同修复方法的主观评价结果<sup>a)</sup>

Table 2 Subjective evaluation results of different restoration methods on the Regong mural dataset<sup>a)</sup>

| Methods                  | Experimental group 1 | Experimental group 2 | Experimental group 3 | Results (weighted average) |
|--------------------------|----------------------|----------------------|----------------------|----------------------------|
| AOTGAN (TVCG2023)        | 3.11                 | 2.98                 | 3.05                 | 3.047                      |
| CTSDG (ICCV2021)         | 3.12                 | 2.90                 | 3.10                 | 3.040                      |
| LaMa (WACV2022)          | 3.55                 | <u>3.58</u>          | 3.30                 | 3.397                      |
| ZITS (CVPR2022)          | 3.64                 | 3.15                 | 3.35                 | 3.380                      |
| Deflocnet (CVPR2021)     | 3.15                 | 2.26                 | 2.72                 | 2.710                      |
| SketchEdit (ICVF2022)    | <u>3.82</u>          | 3.57                 | <u>3.63</u>          | <u>3.673</u>               |
| SketchRefiner (TMM 2024) | 3.20                 | 2.95                 | 3.26                 | 3.137                      |
| SGRT                     | <b>4.32</b>          | <b>4.21</b>          | <b>4.28</b>          | <b>4.270</b>               |

a) 加粗字体表示最优结果, 下划线字体表示次优结果

表 3 消融实验定量结果<sup>a)</sup>

Table 3 Quantitative results of ablation experiments<sup>a)</sup>

| Methods   | PSNR↑         | SSIM↑        | FID↓         | LPIPS↓       |
|-----------|---------------|--------------|--------------|--------------|
| w/o ILAR  | 24.182        | 0.903        | 14.297       | 0.057        |
| w/o ED    | 28.991        | 0.965        | 5.761        | 0.022        |
| w/o SC+GC | 28.260        | 0.960        | 6.027        | 0.024        |
| w/o CCL   | 28.178        | 0.961        | 10.417       | 0.035        |
| SGRT      | <b>30.499</b> | <b>0.972</b> | <b>4.146</b> | <b>0.016</b> |

a) 加粗字体表示最优结果

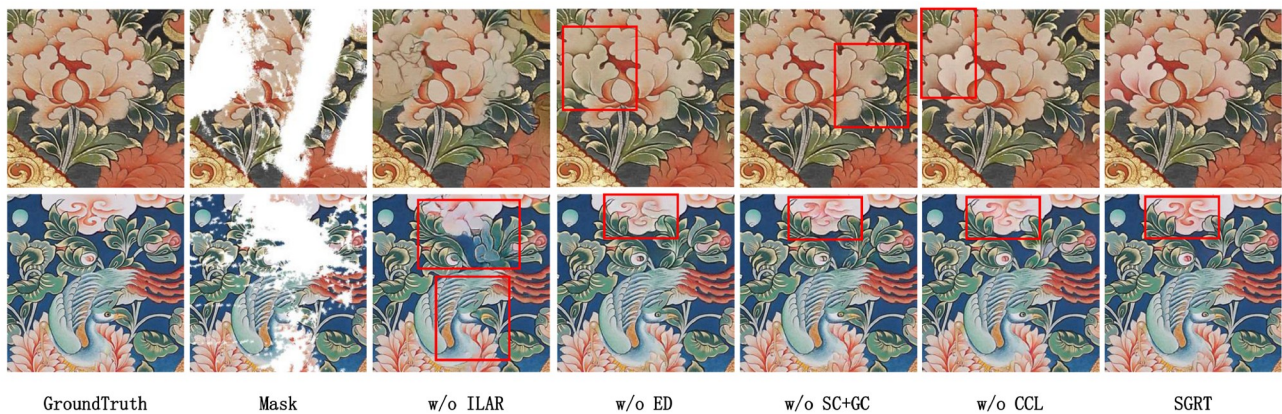


图 7 热贡壁画图像数据集消融实验可视化结果  
Figure 7 Visualization results of ablation experiments on the Regong mural image dataset.

丰富的区域中更为明显. 移除了颜色一致性损失项后, 修复结果在颜色分布和一致性方面出现了偏差, 破坏了整体视觉效果的和谐性. 综上所述, 从实验指标及视觉效果的双重维度来看, 各模块的引入之后能够使得修复后图像的质量明显提升.

## 5 结论

本文针对结构复杂、元素众多且数据集较小的热贡壁画提出了一种结构引导的热贡壁画图像大面积缺损修复算法. 交互式线稿修复模块提供的更为准确的结构先验信息能够有效帮助网络在修复小而复杂的数据集时克服结构模糊和纹理扭曲等难题. 通过引入边缘判别器, 修复网络能够更准确地捕捉并复原图像的边缘细节, 更好地保持修复结果的结构完整性和

视觉连贯性. 跳跃连接和门控卷积等模块的设计确保了修复过程中的特征信息完整性和细节表达能力. 此外, 颜色一致性损失的引入进一步提升了修复结果的视觉和谐度和整体美感, 有助于修复网络在恢复缺损区域时保持与周围区域的颜色一致性, 从而避免了颜色分布和一致性方面的偏差. 实验结果表明, 本文方法在热贡壁画的大面积缺损修复任务中表现出色, 表现优于比较的七个现有算法, 并且具有较好的泛化性. 能够为热贡壁画的保护与传承提供了有力的技术支持, 也为其他具有类似特点的文化遗产图像修复提供了新的思路和方法. 但是虽然本文方法考虑到了使用交互线稿提取来避免修复结果产生不符合文化的不合理结果, 但是尚未考虑热贡壁画不同艺术风格等内容, 后续研究将结合相关知识进行更深一步的改进.

## 参考文献

- 1 Liao N H. A Deep-learning research and application on Thangka inpainting based on params and computational methods (in Chinese). Dissertation of Masteral Degree. Xining: Qinghai Normal University, 2022. 1–9 [廖年鸿. 基于参数与运算轻量化的深度学习唐卡图像修复研究与应用. 硕士学位论文. 西宁: 青海师范大学, 2022. 1–9]
- 2 Li X. Let the ancient traditional culture “come alive”: Summary of the construction of Qinghai cultural ecological protection experimental zone (in Chinese). Qinghai Party Life, 2017, 9: 64–66 [李欣. 让古老的传统文化“活色生香”——青海文化生态保护实验区建设综述. 青海党的生活, 2017, 9: 64–66]
- 3 Hu J Y. Research on traditional mural animal patterns in Regong area of Qinghai (in Chinese). Orient Collect, 2024, 2: 45–47 [胡景源. 青海热贡地区传统壁画动物纹样研究. 东方收藏, 2024, 2: 45–47]
- 4 Bo G. Stylistic characteristics of early Regong Buddhist painting art from the perspective of Qing Dynasty murals in Nian duhu Temple (in Chinese). Qinghai Minzu Yanjiu, 2008, 3: 136–138 [伯果. 从年都乎寺的清代壁画看早期热贡佛画艺术的风格特征. 青海民族研究, 2008, 3: 136–138]

- 5 Biradar R L, Kohir V V. A novel image inpainting technique based on median diffusion. *Sadhana*, 2013, 38: 621–644
- 6 Sridevi G, Srinivas Kumar S. Image inpainting based on fractional-order nonlinear diffusion for image reconstruction. *Circuits Syst Signal Process*, 2019, 38: 3802–3817
- 7 Pathak D, Krahenbuhl P, Donahue J, et al. Context encoders: Feature learning by inpainting. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, 2016. 2536–2544
- 8 Iizuka S, Simo-Serra E, Ishikawa H. Globally and locally consistent image completion. *ACM Trans Graph*, 2017, 36: 1–14
- 9 Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. 2016. arXiv: [1511.07122](https://arxiv.org/abs/1511.07122)
- 10 Elharrouss O, Almaadeed N, Al-Maadeed S, et al. Image inpainting: A review. *Neural Process Lett*, 2020, 51: 2007–2028
- 11 Yu J, Lin Z, Yang J, et al. Free-form image inpainting with gated convolution. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul, 2019. 4470–4479
- 12 Nazeri K, Ng E, Joseph T, et al. Edgeconnect: Structure guided image inpainting using edge prediction. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. Seoul, 2019. 3265–3274
- 13 Xiong W, Yu J, Lin Z, et al. Foreground-aware image inpainting. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, 2019. 5833–5841
- 14 Guo X, Yang H, Huang D. Image inpainting via conditional texture and structure dual generation. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. Montreal, 2021. 14114–14123
- 15 Suvorov R, Logacheva E, Mashikhin A, et al. Resolution-robust large mask inpainting with Fourier convolutions. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. Waikoloa, 2022. 3172–3182
- 16 Chi L, Jiang B, Mu Y. Fast Fourier convolution. In: *Proceedings of the 34th International Conference on Neural Information Processing Systems (NIPS)*. Vancouver, 2020, 33: 4479–4488
- 17 Zheng H, Lin Z, Lu J, et al. Image inpainting with cascaded modulation GAN and object-aware training. In: *Computer Vision—ECCV 2022: 17th European Conference*. Tel Aviv, 2022, Part XVI. 277–296
- 18 Dong Q, Cao C, Fu Y. Incremental transformer structure enhanced image inpainting with masking positional encoding. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New Orleans, 2022. 11348–11358
- 19 Zeng Y, Fu J, Chao H, et al. Aggregated contextual transformations for high-resolution image inpainting. *IEEE Trans Visual Comput Graphics*, 2023, 29: 3266–3280
- 20 Cao Y, Xin Z H, Wu K J, et al. BMTA: Inpainting of large area damaged images in multiple scenarios. *J Front Comput Sci Technol (in Chinese)*, 2024. doi: 10.3778/j.issn.1673-9418.2406095 [曹岩, 辛子昊, 邬开俊, 等. BMTA: 多元场景下的大面积破损图像修复. 计算机科学与探索, 2024. doi: 10.3778/j.issn.1673-9418.2406095]
- 21 Jo Y, Park J. Sc-Fegan: Face editing generative adversarial network with user's sketch and color. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul, 2019. 1745–1753
- 22 Zeng Y, Lin Z, Patel VM. Sketchedit: Mask-free local image manipulation with partial sketches. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New Orleans, 2022. 5941–5951
- 23 Liu H, Wan Z, Huang W, et al. Deflocnet: Deep image editing via flexible low-level controls. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville, 2021. 10760–10769
- 24 Liu C, Xu S, Peng J, et al. Toward interactive image inpainting via robust sketch refinement. *IEEE Trans Multimedia*, 2024, 26: 9973–9987
- 25 Wu K J, Shan H Q, Tian B, et al. Multi-stage stacked fusion for Dunhuang mural restoration (in Chinese). *J Univ Electron Sci Technol China*, 2024, 53: 551–559 [邬开俊, 单宏全, 田彬, 等. 多阶段堆叠融合下的敦煌壁画修复. 电子科技大学学报, 2024, 53: 551–559]
- 26 Ding Y, Wu K J. Inpainting model of Dunhuang Murals guided by frequency domain diffusion information (in Chinese). *J Beijing Univ Aeronaut Astronaut*. 2024. doi: 10.13700/j.bh.1001-5965.2024.0701 [丁元, 邬开俊. 频域扩散信息引导下的敦煌壁画修复模型. 北京航空航天大学学报, 2024. doi: 10.13700/j.bh.1001-5965.2024.0701]
- 27 Liao N H, Zhang X J, Peng C Y, et al. LRGAN: A computational lightweight image inpainting neural network. *Sci Sin Tech*, 2022, 52: 447–457
- 28 Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, 2017. 5967–5976
- 29 He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, 2016. 770–778

- 30 Liu Z, Luo P, Wang X, et al. Deep learning face attributes in the wild. In: Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, 2015. 3730–3738

## Research on structure-guided restoration technique for large-area defects in Regong mural images

LU JiaYu<sup>1,2,3</sup> & ZHANG XiaoJuan<sup>1,2,3\*</sup>

<sup>1</sup> School of Computer, Qinghai Normal University, Xining 810016, China

<sup>2</sup> The State Key Laboratory of Tibetan Intelligence, Xining 810016, China

<sup>3</sup> Qinghai Province Intangible Cultural Heritage Research Base, Xining 810016, China

The Regong mural is a significant manifestation of the Regong art, a human intangible cultural heritage, characterized by its rich and vivid colors as well as intricate and dense lines. However, prolonged environmental exposure has caused severe degradation, including peeling, cracking, and mildew. Existing automatic image inpainting methods fail to reconstruct their complex structures and generate plausible textures. Addressing the challenges posed by the complex structure and texture of Regong murals, irregular damaged areas, and a limited image dataset, this paper proposes a structure-guided method for restoring large-area defects in Regong mural images. Firstly, an interactive line art restoration module is designed to leverage expert experience in repairing the complex line art of large damaged areas, providing a more accurate structural prior to subsequent steps. Subsequently, the line art is processed through a structure-guided restoration module to avoid weakening low-level guidance information while more effectively integrating with texture information. This module injects encoded structural information into a multi-scale restoration network layer by layer, and an edge adversarial discriminator is designed to constrain and optimize the restoration structure, making it closer to the real mural. Furthermore, the network incorporates Fourier convolution blocks and gated convolution modules to expand receptive fields for irregular large-area damage recovery, and a color consistency loss is introduced to effectively mitigate color discrepancies. Extensive experiments against seven state-of-the-art methods demonstrate that our approach achieves superior performance on murals with 30%–50% damage ratios. Quantitative evaluations reveal improvements of 37.91% in the structural similarity index (SSIM) and 15.3% in the peak signal-to-noise ratio (PSNR), alongside reductions of 94.67% in the Fréchet inception distance (FID) and 85.96% in the learned perceptual image patch similarity (LPIPS).

**mural restoration, structure guidance, Fourier convolution, interactive**

doi: [10.1360/SST-2024-0343](https://doi.org/10.1360/SST-2024-0343)