

文章编号: 1671-7449(2026)03-0352-10

J. Test Measur. Technol., 2026, 40(3): 352-361.

## 融合频域卷积与双向注意力的壁画图像修复算法

王宁<sup>1</sup>, 白瑞君<sup>2</sup>, 李众<sup>1\*</sup>, 谭文博<sup>1</sup>, 罗佳宁<sup>1</sup>

(1. 中北大学 软件学院, 山西 太原 030051; 2. 中北大学 信息与通信工程学院, 山西 太原 030051)

**摘要:** 针对壁画图像修复中存在的长程纹理依赖建模不足、结构连贯性缺失及跨尺度特征融合粗糙等问题, 提出一种基于两阶段生成对抗网络的壁画修复算法。首先, 设计多尺度频域卷积模块, 通过空域-频域并联架构与自适应膨胀率机制动态平衡高低频特征, 解决传统卷积感受野受限导致的矿物颜料渐变失真问题; 其次, 引入轻量级双向注意力单元, 在1/4分辨率层融合方向梯度直方图与稀疏注意力, 增强大尺度结构的连贯性; 最后, 构建混合注意力融合模块, 结合坐标注意力与动态通道加权优化多尺度特征上采样。在敦煌壁画数据集上的实验结果显示, 与当前最新EGMF算法相比, 在[30%, 40%)的掩码比例下, 峰值信噪比指标提升0.13 dB, 结构相似性指标提升1.06%, 且在其他3种掩码比例中, 结果均优于对比算法, 有效提升了修复后图像在纹理、结构和细节上的视觉一致性。

**关键词:** 壁画修复; 频域卷积; 双向注意力; 多尺度融合

中图分类号: TP391

文献标识码: A

DOI: 10.62756/csjs.1671-7449.2026048

**引用格式:** 王宁, 白瑞君, 李众, 等. 融合频域卷积与双向注意力的壁画图像修复算法[J]. 测试技术学报, 2026, 40(3): 352-361.

WANG Ning, BAI Ruijun, LI Zhong, et al. Mural image restoration algorithm integrating frequency domain convolution and bidirectional attention[J]. Journal of Test and Measurement Technology, 2026, 40(3): 352-361.

## Mural Image Restoration Algorithm Integrating Frequency Domain Convolution and Bidirectional Attention

WANG Ning<sup>1</sup>, BAI Ruijun<sup>2</sup>, LI Zhong<sup>1\*</sup>, TAN Wenbo<sup>1</sup>, LUO Jianing<sup>1</sup>

(1. School of Software, North University of China, Taiyuan 030051, China;

2. School of Information and Communication Engineering, North University of China, Taiyuan 030051, China)

**Abstract:** To address the issues of insufficient long-range texture dependency modeling, lack of structural coherence, and rough cross-scale feature fusion in mural image restoration, a mural restoration algorithm is proposed based on a two-stage generative adversarial network. Firstly, a multi-scale frequency domain convolution module is designed, using a spatial-frequency parallel architecture and an adaptive expansion rate mechanism to dynamically balance high and low-frequency features and solve the problem of gradual distortion of mineral pigments caused by the limited receptive field of traditional convolution. Secondly, a lightweight bidirectional attention unit is introduced to fuse the direction gradient histogram and sparse attention at the 1/4 resolution layer, enhancing the coherence of large-scale structures. Further, a hybrid attention fusion module is constructed, combining coordinate attention and dynamic channel weighting to optimize multi-scale feature upsampling. Experimental results on the Dunhuang mural dataset show that, compared with the current latest EGMF algorithm, at a mask ratio of [30%, 40%), the peak signal-to-noise ratio (PSNR) index, and the

【收稿日期】2025-07-24

【作者简介】王宁(1998—), 女, 硕士生, 主要从事图像处理研究。E-mail: 1282315761@qq.com。

\*【通信作者】李众(1974—), 男, 教授, 博士, 主要从事计算机图形图像处理、嵌入式系统安全开发等研究。E-mail: lizhong@nuc.edu.cn。

structural similarity (SSIM) index are improved by 0.13 dB and 1.06%, respectively. And in other 3 kinds of mask ratios, this algorithm outperforms the comparison algorithm, effectively enhancing the visual consistency of the restored image in terms of texture, structure, and details.

**Key words:** mural restoration; frequency domain convolution; bidirectional attention; multiscale fusion

## 0 引言

图像修复旨在对缺失或损坏的像素区域进行合理重建,以恢复图像的视觉真实性和结构完整性,在数字文化遗产保护、影视修复等领域具有重要应用价值。壁画作为不可再生的文化遗产,长期受到自然风化、人为破坏等因素的影响,导致局部剥落、褪色或结构性缺损<sup>[1-2]</sup>。传统修复主要依赖人工干预和化学手段,存在效率低、主观性强等问题。数字图像修复技术为壁画保护提供了新的解决方案<sup>[3]</sup>。

以往的图像修复方法主要基于偏微分方程的扩散模型<sup>[4-5]</sup>和基于样本的块匹配<sup>[6-7]</sup>。这类方法的核心局限在于强假设和局部操作特性:扩散方法通过建立变分模型并求解相应的偏微分方程来逐步扩散已知区域信息至缺损区域,其本质是局部平滑,因而仅适用于平滑或低纹理区域,对壁画中常见的复杂纹理重建能力有限;块匹配方法(如Criminisi算法<sup>[6]</sup>)通过搜索并复制已知区域中的最佳匹配纹理块进行填充,在规则纹理修复上效果显著,但其性能高度依赖已知区域中存在足够且相似的纹理信息,对于大范围缺损或结构复杂、纹理非周期的壁画区域,极易出现纹理重复、结构断裂等不合理现象<sup>[8]</sup>。随着深度学习的发展,基于卷积神经网络<sup>[9]</sup>和生成对抗网络(Generative Adversarial Networks, GAN)<sup>[10]</sup>的方法显著提升了图像修复的质量。单阶段方法(如PCNet<sup>[11]</sup>)采用部分卷积动态更新掩码,计算高效,但因网络结构简单,对复杂纹理的生成能力和全局语义一致性建模不足;两阶段方法(如DeepFillv1/v2<sup>[12-13]</sup>)通过“粗修复+精修复”的策略提升了大范围缺损的修复效果,但两个阶段间的信息传递损失以及卷积操作的局部性仍可能导致修复区域与全局上下文语义不一致,产生边缘模糊或结构扭曲。Zeng等<sup>[14]</sup>使用多尺度空洞卷积模块聚合上下文信息,在高分辨率复杂场景下取得了较高修复能力,却忽略了计算成本的增加。Chen等<sup>[15]</sup>为了提升边缘和纹理的保真度,结合门控机制和局部判别器进行精细化局部小范围修复,但在大区域修复易出现结构错误。

Yu等<sup>[16]</sup>提出了一种语境注意力层来从距离较远

的区域提取近似待修复区域的特征,但修复图像通常会产生扭曲的结构或模糊的纹理。Dosovitskiy等<sup>[17]</sup>在计算机视觉领域引入Transformer,通过自注意力机制实现全局操作,为建模所有像素间的关系提供了新思路,并与卷积形成有效互补。Li等<sup>[18]</sup>使用Transformer来提升大面积破损的修复效果。刘仲民等<sup>[19]</sup>提出了一种基于动态特征选择和像素级通道注意力的壁画修复模型,通过有效可迁移卷积和区域综合归一化提升修复效果,但其动态卷积与注意力机制主要作用于局部特征交互,对大面积破损和复杂长程纹理细节的恢复能力仍显不足。Wei等<sup>[20]</sup>针对低光环境下的壁画修复,提出两阶段增强-修复框架,采用多感受野策略,结合大、小感受野网络进行全局和局部修复,但是未引入线描或边缘引导,在大范围缺损时难以保证生成内容的几何合理性与结构连贯性。LRDiff算法<sup>[21]</sup>将低秩结构先验融入扩散模型,有效提升了壁画修复中全局结构的完整性与纹理一致性,克服了传统方法忽略长程依赖的问题。然而,其对大规模损伤的修复仍存在细节失真,且低秩约束在复杂纹理区域可能过度平滑,导致艺术风格局部失配,影响视觉自然性。EGMF算法<sup>[22]</sup>融合边缘引导与多尺度空间特征,提升了壁画修复的边缘与结构质量,但复杂纹理生成不足,边缘增强模块在深层网络中可能引入噪声导致在大尺度缺损区域修复时出现语义不连贯与细节模糊问题,影响整体修复自然性。

综上所述,当前壁画修复算法普遍面临三大核心挑战,其根源在于现有方法的内在缺陷:1)复杂纹理生成不自然:源于传统卷积感受野有限,难以捕获壁画中矿物颜料渐变等非局部纹理特征;2)长程结构连续性差:源于现有注意力或Transformer机制计算成本高,或缺乏对壁画笔触方向性等几何先验的利用,导致大尺度结构断裂;3)跨尺度特征融合粗糙:源于上采样和多尺度特征拼接过程中,缺乏对空间位置和通道重要性的精细化调控,导致边缘伪影和细节丢失。

针对上述问题与现有方法的局限性,本文基于两阶段生成对抗网络EdgeConnect,提出一种面向壁画图像的修复算法,具体创新包括:1)设计

多尺度频域卷积(Multi-Scale Frequency Convolution, MSF-Conv)模块,通过空域-频域并联架构突破空间卷积的局限,动态平衡高低频信息,从根本上改善长程纹理依赖建模;2)引入轻量级双向注意力单元(Lightweight Bidirectional Attention Unit, LBAU),在低分辨率特征层融合方向梯度直方图(Histogram of Oriented Gradient, HOG)与稀疏注意力,以可接受的计算代价增强大尺度结构的几何一致性;3)构建混合注意力融合(Hybrid Attention Fusion, HAF)模块,结合坐标注意力与动态通道加权,在特征上采样过程中实现空间与通道维度的精准调控,从而实现精细化的跨尺度特征融合。

MSF-Conv 为系统提供了具备全局感知的特征基座, LBAU 在此基础上保证了宏观结构的几何合理性,而HAF 模块则确保了从特征空间到像素空间

重建的精细度。三者协同工作,可使敦煌壁画得到视觉真实、结构连贯且细节丰富的修复结果。

## 1 相关理论

EdgeConnect 模型由两个阶段组成:边缘生成器和图像补全器。G1 和 D1 分别是边缘生成器的生成器和判别器,边缘生成仅关注缺失区域中的假想边缘。G2 和 D2 分别是图像补全器的生成器和判别器。图像补全网络使用假想边缘并估计缺失区域的 RGB 像素强度。这种分阶段对抗训练框架虽然在结构化图像修复中表现出色,能够生成视觉连贯的结果,但其对非规则缺失或复杂纹理修复能力有限,对于大面积缺失的处理也稍显不足。因此,本文针对 EdgeConnect 模型(结构见图 1)进行了改进,提出了一种两阶段生成对抗网络,用于解决上述模型存在的不足。

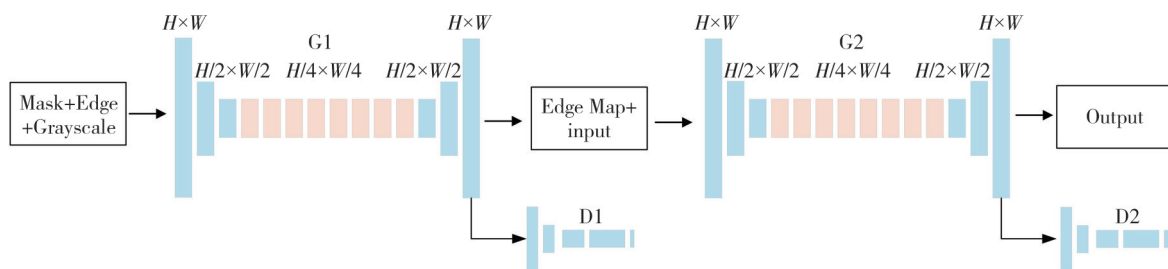


图 1 EdgeConnect 模型结构图

Fig. 1 EdgeConnect model structure diagram

## 2 本文方法

基于 EdgeConnect 框架,本文提出一种专用于壁画修复的两阶段生成对抗网络,如图 2 所示。本方法保留了原框架边缘生成阶段的边缘生成器 G1,用于预测壁画缺失区域的结构信息。针对图

像补全阶段的 G2 生成器,本方法提出 MSF-Conv 模块,通过空域-频域并联处理捕获矿物颜料的非局部相关性。改进的 LBAU 在 1/4 特征图分辨率层实现稀疏化跨区域建模,可有效保持壁画中的全局结构一致性。构建 HAF 模块,将坐标注意力与挤压激励思想相结合,通过轴向特征分解和动态通道加权来优化多尺度特征上采样的过程。

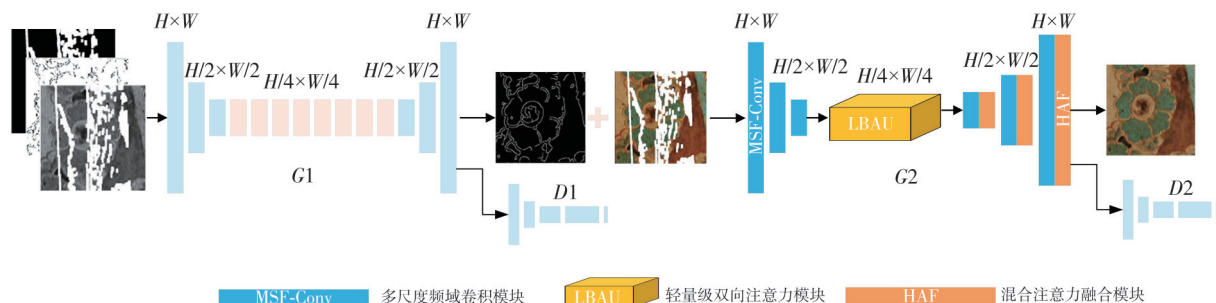


图 2 改进的 EdgeConnect

Fig. 2 Improved EdgeConnect

### 2.1 MSF-Conv 模块

传统卷积神经网络在壁画图像修复中有较大局

限性,有限感受野难以建模全局纹理依赖,网络在重建过程中无法保持壁画特有的连续性笔触和矿物

颜料渐变效果;空间域操作无法有效捕获频域特征,导致周期性纹理重建失真。为此,本文提出采用MSF-Conv模块,其结构如图3所示。

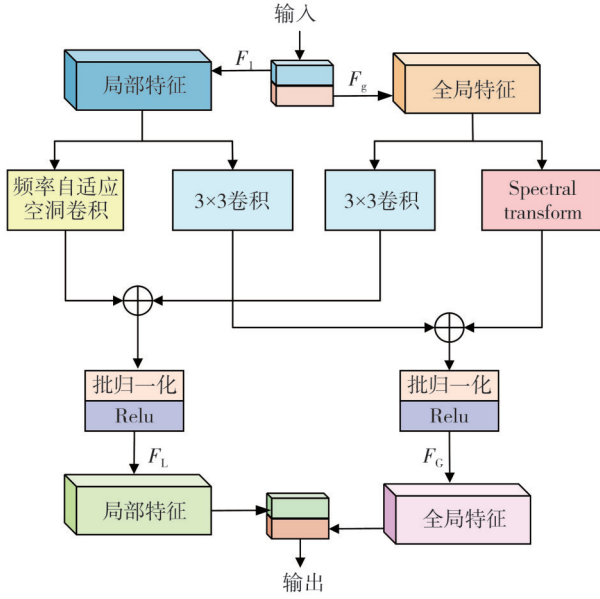


图3 MSF-Conv模块结构图

Fig. 3 Structure diagram of MSF-Conv module

本模块采用双路并行架构,其中,局部分支通过频率自适应空洞卷积与全局标准 $3 \times 3$ 卷积的特征融合实现多尺度局部建模,将二者特征进行融合,感受野与局部纹理尺度匹配的同时,抑制频域操作导致的边缘振荡。局部分支的计算过程为

$$F_L = \text{Relu}\left(\text{BN}\left(\text{FADC}(F_l) + \text{Conv}_{3 \times 3}(F_g)\right)\right), \quad (1)$$

式中: $F_l$ 为局部分支的输入; $F_g$ 为全局分支的输入; $\text{FADC}(\cdot)$ 表示频率自适应空洞卷积。

全局分支也采用双路并行架构,保留局部分支的 $3 \times 3$ 卷积,通过局部感受野提取底层特征,保证对壁画基础几何轮廓(如建筑线条,人物姿态)的稳定性建模。频域路径通过改进的模块捕捉全局频域特征,该模块允许G2生成器从早期开始考虑全局上下文,对于全局周期性的壁画结构更为敏感。全局分支的计算过程为

$$F_G = \text{Relu}\left[\text{BN}\left(\text{Conv}_{3 \times 3}(F_l) + \text{FFT}(F_g)\right)\right]. \quad (2)$$

MSF-Conv模块通过双路并行架构,同时提取包含局部细节和全局周期性纹理的鲁棒性特征,传入后续LBAU进行更深层次的结构推理。

## 2.2 LBAU模块

G2中低分辨率特征层( $H/4 \times W/4$ )承担着全局结构推理与跨区域语义关联的关键功能。传统残差

块在此分辨率下存在长距离依赖建模缺陷,导致对壁画中典型的大尺度结构元素的恢复能力不足。本文在 $1/4$ 特征图分辨率层引入改进的双向路由注意力机制,将方向梯度直方图HOG引入Transformer路由机制,解决了长程依赖缺失,也解决了壁画笔触的各向异性问题。具体来说,在路由分数计算层将原始窗口平均特征 $Q_i$ 、 $K_j$ 替换为HOG特征 $\text{HOG}(Q_i)$ 和 $\text{HOG}(K_j)$ ;利用HOG的主方向直方图构建几何掩码 $M_{\text{geo}}$ 。具体结构如图4所示。

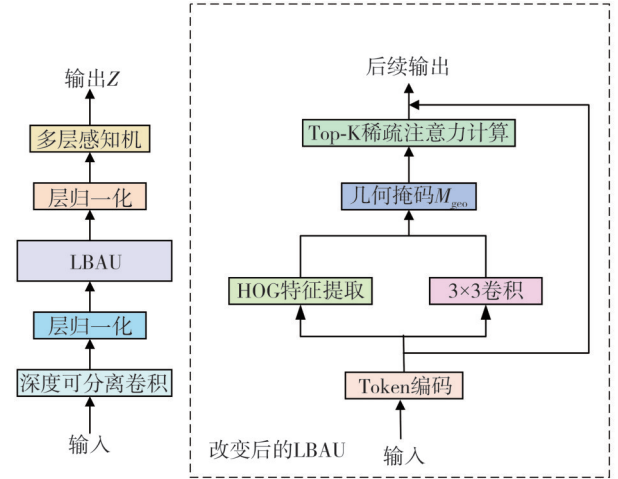


图4 改进的LBAU模块图

Fig. 4 Improved LBAU module diagram

动态路由筛选阶段:在传统BiFormer的动态路由筛选阶段,区域相关性通过平均池化特征计算,缺乏对壁画笔触方向性的感知,本文将其替换为多尺度HOG描述算子。对于输入的特征图 $X$ ,使用Sobel算子获取水平和垂直梯度 $G_x$ 和 $G_y$ ,

$$G_x = X * \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \quad (3)$$

$$G_y = X * \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}. \quad (4)$$

将梯度方向 $\theta = \arctan\left(\frac{G_y}{G_x}\right)$ 在 $[0^\circ, 180^\circ)$ 范围内

均匀划分为8个方向区间,每个区间宽度为 $22.5^\circ$ ,并对梯度幅值 $M$ 进行加权 $M = \sqrt{G_x^2 + G_y^2}$ 。将特征图划分为8个单元格,对每个单元格根据方向 $\theta$ 将像素分配到8个方向区间,用幅值 $M$ 加权统计直方图,得到HOG特征 $\text{HOG}(X) \in \mathbb{R}^{(H/8) \times (W/8) \times 8}$ ,同时生成几何掩码 $M_{\text{geo}}$ ,

$$M_{\text{geo}}(i, j) = \begin{cases} 1, & \text{if } |\theta_{\text{main}}^i - \theta_{\text{main}}^j| \leq 30^\circ, \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

将HOG直方图中幅值加权和最大的方向区间所对应的方向作为该区域的主方向 $\theta_{\text{main}}$ ，若两区域的主方向之差大于 $30^\circ$ ，认为其结构方向不一致，则禁止它们之间的注意力交互，用于增强结构一致性约束。最终得到的路由计算分数过程为

$$R_{ij} = \text{Softmax} \left( \frac{W_q \text{HOG}(Q_i) \cdot (W_k \text{HOG}(K_j))^T}{\sqrt{d}} + \ln M_{\text{geo}} \right), \quad (6)$$

式中： $W_q, W_k \in \mathbb{R}^{8 \times d}$ 为查询和键值的投影矩阵； $\text{HOG}(Q_i)$ 和 $\text{HOG}(K_j)$ 分别为查询区域 $Q_i$ 和键值区域 $K_j$ 的8维梯度直方图； $M_{\text{geo}}$ 为基于HOG主方向一致性的几何掩码。最后，通过Top-K筛选出每个查询区域 $Q_i$ 最相关的4个区域进行稀疏注意力计算。

稀疏注意力计算保证空间注意力建模全局区域关系，通道注意力捕捉特征级联。具体注意力公式为

$$\text{Attn}(Q, K, V) = \text{softmax} \left( \frac{Q_{\text{pix}} K_{\text{sed}}^T}{\sqrt{d}} \right) V_{\text{sed}}, \quad (7)$$

式中： $Q_{\text{pix}}$ 为像素级查询； $K_{\text{sed}}, V_{\text{sed}}$ 分别为筛选的键值对。然后，通过引入门控深度卷积保留细节特征，进行特征增强。具体过程为

$$Y = \text{LayerNorm}(A_{\text{sparse}} + X). \quad (8)$$

最终使用多层感知机(Multilayer Perception, MLP)作为前馈网络的整体输出过程为

$$Z = \text{MLP}(Y) + A_{\text{sparse}} + X. \quad (9)$$

经过LBAU增强后的特征 $Z$ ，保留了壁画的大尺度几何结构信息，将此特征通过后续的HAF模块进行多尺度特征融合与最终图像的重建。

### 2.3 HAF模块

为了将前序模块LBAU处理得到的多尺度特

征进行有效融合，并重建出高质量的修复图像，本文构建了HAF模块。常见修复模型中的金字塔生成结构缺乏跨尺度的融合机制，各层级间主要通过简单的上采样连接，导致壁画笔触方向、矿物颜料分布等空间信息丢失，而直接拼接不同尺度特征会引入通道间干扰。本文通过坐标注意力(Coordinate Attention, CA)模块引导多尺度特征重组，同时采用多分支全连接层挤压激励(Squeeze-and-Excitation, SE)特征，通过多分支结构进一步提升特征表达的精细度和全局信息的整合能力。融合模块的整体结构如图5所示。

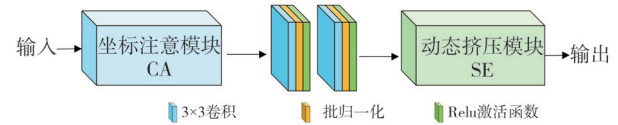


图5 融合模块整体结构图

Fig. 5 Overall structure diagram of fusion module

HAF模块的计算过程为

$$X_{\text{ca}} = x + \text{CA}(x), \quad (10)$$

$$X_{\text{conv}} = X_{\text{ca}} + \text{RELU}(\text{BN}(\text{Conv}_{3 \times 3}(X_{\text{ca}}))), \quad (11)$$

$$Y = \text{SE}(X_{\text{conv}}) = X_{\text{conv}} \otimes S, \quad (12)$$

式中：CA为坐标注意力机制的过程； $x$ 为前序模块的输出； $X_{\text{conv}}$ 为将 $X_{\text{ca}}$ 经过 $3 \times 3$ 卷积批归一化和ReLU激活函数后，与 $X_{\text{ca}}$ 本身进行残差连接所得的结果； $Y$ 为经过动态挤压后的最终输出。

#### 2.3.1 CA机制

壁画边缘往往具有明确的方向性，CA机制通过拼接 $H$ 和 $W$ 方向特征，建立空间长程依赖，对于壁画剥落区域结构连贯性进行了有效改善，同时消除了上采样过程给壁画修复图像带来的伪影，增强色块的连续性。CA机制结构如图6所示。

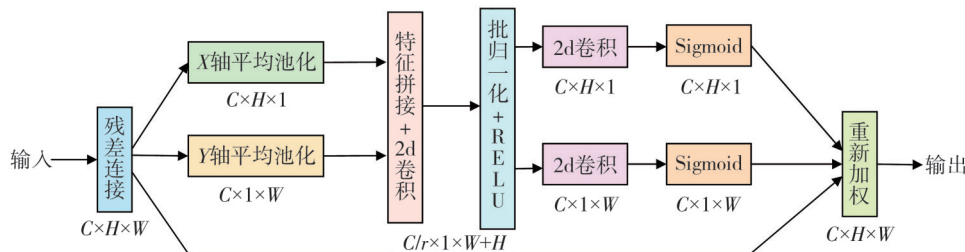


图6 CA机制结构图

Fig. 6 Structure diagram of CA mechanism

#### 2.3.2 动态挤压模块

为了更好地获得融合特征，本节对CA机制后

的壁画特征进行动态压缩。采用基数为4的并行结构替代单一全局压缩，通过独立的全连接层(Fully Connected Layer, FCL)分别捕捉矿物颜料

的颗粒纹理、大色块分布和低频细节等不同粒度特征;其次,通过基数间特征拼接与非线性融合,增强对朱砂、金箔等特殊壁画材料的通道敏感度。该模块结构如图 7 所示。

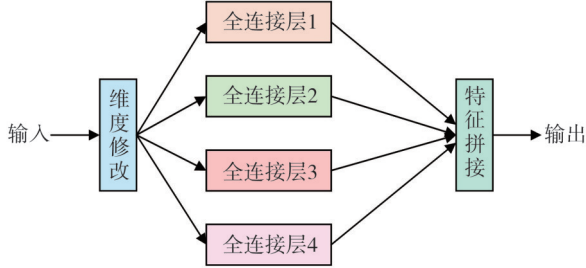


图 7 动态挤压模块结构图

Fig. 7 Structure diagram of dynamic squeeze module

总的来说,该模块通过多路径特征分解与动态加权,显著改善了传统 SE 模块在壁画多尺度纹理和特殊光学特性上的表征能力,同时维持较低的计算开销。

#### 2.4 损失函数

本文算法沿用 EdgeConnect 联合损失函数框架,针对壁画修复任务优化权重分配。总损失函数由四部分组成:边缘生成器损失  $L_{\text{edge}}$ 、图像生成器损失  $L_{\text{image}}$ 、判别器 D1 和 D2 的对抗损失。具体来说,边缘生成器 G1 的损失包含对抗损失  $L_{\text{adv}}^e$  和像素级 L1 损失  $L_{\text{pixel}}^e$ , 计算过程为

$$L_{\text{edge}} = \lambda_{\text{adv}} L_{\text{adv}}^e + \lambda_{\text{pixel}} L_{\text{pixel}}^e \quad (13)$$

式中:  $\lambda_{\text{adv}}$  为对抗损失权重系数;  $\lambda_{\text{pixel}}$  为像素损失权重系数。参考 EdgeConnect 的联合损失函数框架,  $\lambda_{\text{adv}}$  设置为 0.1,  $\lambda_{\text{pixel}}$  设置为 1.0, 以此来强调边缘精度。

图像生成器 G2 的损失包含对抗损失、像素损失、感知损失和风格损失, 具体计算为

$$L_{\text{image}} = \lambda_{\text{adv}} L_{\text{adv}}^i + \lambda_{\text{pixel}} L_{\text{pixel}}^i + \lambda_{\text{perc}} L_{\text{perc}}^i + \lambda_{\text{style}} L_{\text{style}}^i \quad (14)$$

式中:  $\lambda_{\text{perc}}$  为感知损失权重系数;  $\lambda_{\text{style}}$  为风格损失权重系数;  $L_{\text{adv}}^i$  为对抗损失, 用来判别器 D2 判断生成图像的真实性;  $L_{\text{pixel}}^i$  为像素损失, 通过 L1 损失约束低频结构重建;  $L_{\text{perc}}^i$  为基于 VGG-16 的感知损失;  $L_{\text{style}}^i$  为风格损失, 通过 Gram 矩阵保留壁画艺术风格。

式(14)中的 4 个损失权重系数分别设置为 0.1、1.0、0.05 和 0.01。像素损失作为基础监督项, 其权重设置为 1.0, 用于约束生成图像在像素层面的重建精度, 保证低频结构与颜色信息的准确恢复。对抗损失旨在提升图像的视觉逼真度, 但其梯度通常不稳定且量级较大。为了防止其主导整个训练过程并导致模式崩溃, 将  $\lambda_{\text{adv}}$  设置为一个较小的值(0.1)。 $\lambda_{\text{perc}}$  和  $\lambda_{\text{pixel}}$  通过控制变量法进行调整, 以在验证集上取得视觉质量和指标得分的最佳平衡, 先设置  $\lambda_{\text{style}}$  为 0, 在掩膜为 [1%, 10%) 下进行权重控制实验, 得出  $\lambda_{\text{perc}}$  为 0.05 取得最优效果。具体如图 8 所示。

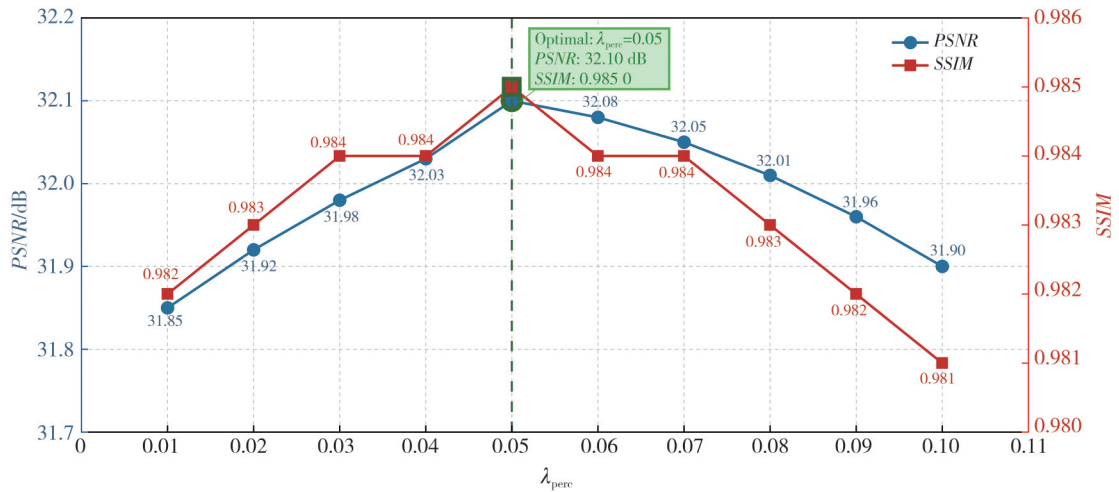


图 8 感知损失权重控制变量实验分析( $\lambda_{\text{style}}=0$ , 掩膜比例 [1%, 10%))

Fig. 8 Controlled variable analysis of perceptual loss weight ( $\lambda_{\text{style}}=0$ , Mask Ratio: [1%, 10%))

最后, 通过固定的像素损失、对抗损失和感知损失权重, 确定风格损失的  $\lambda_{\text{style}}$  数值。图 9 为在掩膜为 [1%, 10%) 下进行权重控制实验, 由该图得出  $\lambda_{\text{style}}$  为 0.05 取得最优效果。

判别器损失包括 D1 和 D2 的对抗损失, 计算过程分别为

$$L_{D1} = E_{I_{\text{edge}}, I_{\text{gt}}} [\ln D1(I_{\text{edge}}, I_{\text{gt}})] + E_{I_{\text{edge}}, I_{\text{fake1}}} [\ln (1 - D1(I_{\text{edge}}, I_{\text{fake1}}))], \quad (15)$$

$$L_{D_2} = E_{I_{\text{comp}}, I_{\text{gt}}} [\ln D1(I_{\text{comp}}, I_{\text{gt}})] + E_{I_{\text{comp}}, I_{\text{fake2}}} [\ln(1 - D1(I_{\text{comp}}, I_{\text{fake2}}))], \quad (16)$$

式中:  $I_{\text{edge}}$  为输入的真实边缘图(由边缘生成器G1生

成);  $I_{\text{gt}}$  为真实完整的图像;  $I_{\text{fake1}}$  为由G1生成的伪造边缘图, 用于判断输入边缘是否真实;  $I_{\text{comp}}$  为由G2生成的修复图像;  $I_{\text{fake2}}$  为由G2生成的伪造修复图像;  $D2$  为图像判别器, 用于判断修复图像是否真实。

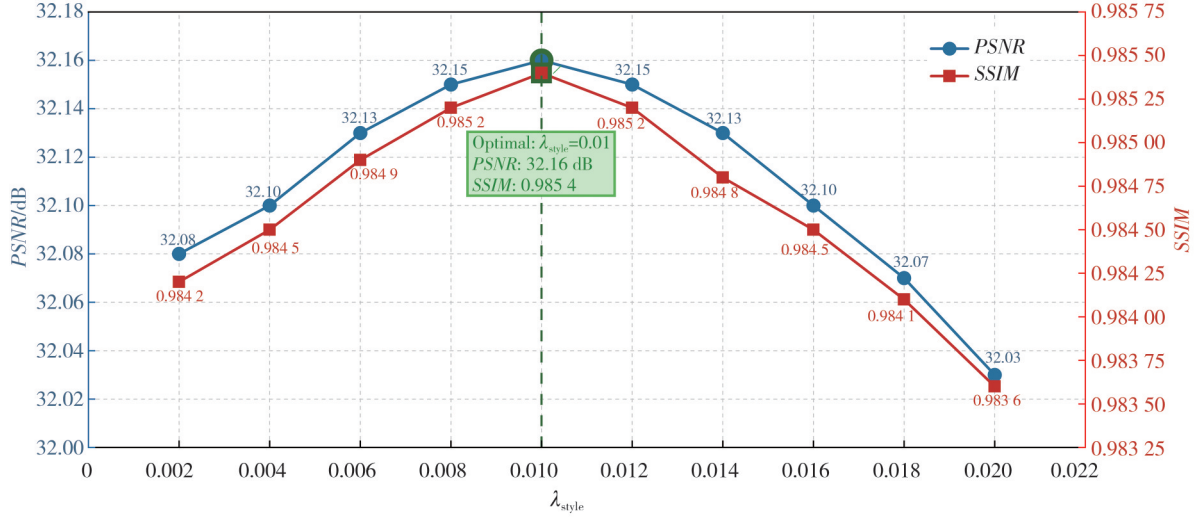


图9 风格损失权重控制变量实验分析( $\lambda_{\text{perc}}=0.05$ , 掩膜比例[1%, 10%])

Fig. 9 Controlled variable analysis of style loss weight ( $\lambda_{\text{perc}}=0.05$ , mask ratio: [1%, 10%])

整体损失函数为

$$L_{\text{total}} = L_{\text{edge}} + L_{\text{image}} + L_{D1} + L_{D2\circ} \quad (17)$$

### 3 实验结果与分析

#### 3.1 实验设置与数据集

本文模型在 Intel(R) Xeon(R) Platinum 8375CPU @ 3.5 GHz 处理器与 NVIDIA GeForce RTX3090 GPU 的硬件环境下训练, 软件环境为 Python 3.9.1 和 CUDA11.2, 并采用 Adam 优化算法进行参数更新。学习率的设置采用了动态调整策略(Warmup), 其中最小学习率设置为  $10^{-7}$ , 最大学习率设置为  $10^{-4}$ , 迭代次数为 500, 线程数设置为  $4 \times 4$ 。

本文采用公开的敦煌壁画数据集, 该数据集共有 18 800 张图片, 通过图像随机旋转、裁剪、镜像翻转等算法对数据集扩充至 44 900 幅图片, 将其按照 8:2 的比例划分为训练集和测试集。

#### 3.2 评价指标

在评估壁画图像修复算法的性能时, 主观评价依赖于人眼对修复图像的直观感受, 而客观评价是通过具体的量化标准来衡量图像的重建质量。这些量化标准涵盖了峰值信噪比(Peak Signal-to-Noise Ratio, PSNR)、结构相似性指数(Mean Structural Similarity Index Measure, SSIM)等。

1) PSNR 反映两幅图像对应像素之间的误差, 是基于均方误差(Mean Squared Error, MSE)定义的, 用于评估原始图像与失真图像之间的质量差异。该值越高, 输出图像中的失真就越少, 从而获得更好的图像重建质量。PSNR 计算公式为

$$MSE(I, K) = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \|I(i, j) - K(i, j)\|^2, \quad (18)$$

$$PSNR(I, K) = 20 \lg \frac{\text{MAX}_I}{\sqrt{MSE}}, \quad (19)$$

式中:  $I$  和  $K$  为需要比较的图像;  $i$  和  $j$  为像素的索引;  $m$  为图像的宽度;  $n$  为图像的高度;  $\text{MAX}_I$  为  $I$  的最大值, 设置为 255。

2) SSIM 是量化评估两个图像之间相似性的指标, 取值范围通常在 0~1 之间。当 SSIM 值越接近 1 时, 表示两个图像在视觉上的相似性越高。SSIM 计算公式为

$$SSIM(I, K) = \frac{(2\mu_I\mu_K + c_1)(2\sigma_{IK} + c_2)}{(\mu_I^2 + \mu_K^2 + c_1)(\sigma_I^2 + \sigma_K^2 + c_2)}, \quad (20)$$

式中:  $\mu_I$  和  $\mu_K$  分别为  $I$  和  $K$  的平均值;  $\sigma_I$  和  $\sigma_K$  分别为  $I$  和  $K$  的方差;  $\sigma_{IK}$  为  $I$  和  $K$  的斜方差;  $c_1 = (k_1 L)^2$  和  $c_2 = (k_2 L)^2$  为鲁棒常数, 其中  $L$  为图像的范围, 设置

为 255, 超参数  $k_1$  和  $k_2$  分别设置为 0.01 和 0.03。

### 3.3 对比试验

#### 3.3.1 定量分析

为了进一步验证所提算法对壁画图像修复的有效性, 对测试集中 1 000 张图像进行不同掩膜区域比例的修复测试, 将所提算法与 AOT-GAN<sup>[14]</sup>、Deepfillv1<sup>[12]</sup>、Deepfillv2<sup>[13]</sup>、GLNet<sup>[15]</sup>、PCNet<sup>[11]</sup>、MIFDFSP<sup>[19]</sup>、MER<sup>[20]</sup>、LRDiff<sup>[21]</sup> 和 EGMF<sup>[22]</sup> 算法对比分析。PSNR 的数值如表 1 所

示, SSIM 的数值如表 2 所示。在 [1%, 10%)、[10%, 20%)、[20%, 30%) 和 [30%, 40%) 不同掩膜区域上, 本文算法较经典的 Deepfillv2 算法, 在 PSNR 指标上分别提升 4.03、1.90、0.91 和 1.47 dB; 在 SSIM 指标上分别提升 1.29%、5.39%、5.93% 和 22.18%。与最新的算法 LRDiff 比较, PSNR 指标分别提升 0.67、0.79、0.38 和 0.21 dB; SSIM 指标分别提升 0.12%、1.10%、4%、4.3%。由此可见, 所提算法较其他方法在多个评价指标下都有较好的提升效果, 尤其是在大比例掩膜区域的修复效果更好。

表 1 敦煌壁画数据集修复结果 PSNR 对比

Tab. 1 Comparison of PSNR results for dunhuang mural dataset restoration

| 掩膜区域<br>比例 | PSNR/dB  |            |            |          |          |          |          |          |          |          |
|------------|----------|------------|------------|----------|----------|----------|----------|----------|----------|----------|
|            | AOT-GAN  | Deepfillv1 | Deepfillv2 | GLNet    | PCNet    | MIFDFSP  | MER      | LRDiff   | EGMF     | 本文方法     |
| [1%, 10%)  | 28.642 4 | 27.635 0   | 28.071 4   | 29.294 3 | 27.882 0 | 30.382 5 | 30.575 4 | 31.434 6 | 31.726 0 | 32.105 6 |
| [10%, 20%) | 26.797 7 | 26.786 6   | 27.226 3   | 26.447 5 | 25.038 9 | 27.535 8 | 27.697 2 | 28.334 8 | 28.875 3 | 29.122 8 |
| [20%, 30%) | 23.328 6 | 23.497 5   | 24.825 6   | 24.046 6 | 21.637 8 | 25.138 5 | 25.296 7 | 25.358 5 | 25.500 4 | 25.736 9 |
| [30%, 40%) | 18.441 8 | 19.350 0   | 20.790 0   | 19.091 8 | 19.650 0 | 22.100 0 | 22.260 0 | 22.320 1 | 22.401 0 | 22.530 1 |

表 2 敦煌壁画数据集修复结果 SSIM 对比

Tab. 2 Comparison of SSIM results for dunhuang mural dataset restoration

| 掩膜区域<br>比例 | SSIM    |            |            |         |         |         |         |         |         |         |
|------------|---------|------------|------------|---------|---------|---------|---------|---------|---------|---------|
|            | AOT-GAN | Deepfillv1 | Deepfillv2 | GLNet   | PCNet   | MIFDFSP | MER     | LRDiff  | EGMF    | 本文方法    |
| [1%, 10%)  | 0.972 0 | 0.975 0    | 0.973 0    | 0.961 0 | 0.972 5 | 0.979 0 | 0.984 0 | 0.984 4 | 0.985 0 | 0.985 6 |
| [10%, 20%) | 0.891 9 | 0.915 7    | 0.918 0    | 0.907 4 | 0.880 4 | 0.942 9 | 0.926 5 | 0.956 5 | 0.959 8 | 0.967 5 |
| [20%, 30%) | 0.858 8 | 0.846 5    | 0.888 0    | 0.869 8 | 0.775 9 | 0.880 8 | 0.867 9 | 0.900 7 | 0.929 7 | 0.940 7 |
| [30%, 40%) | 0.795 7 | 0.777 2    | 0.747 4    | 0.812 2 | 0.671 3 | 0.798 6 | 0.809 4 | 0.870 0 | 0.902 6 | 0.913 2 |

#### 3.3.2 定性分析

通过与经典和先进方法进行视觉对比, 可以直观地看出本文算法在感知质量上的显著优越性, 尤其是对于修复后壁画的全局周期性纹理以及局部裂纹等方面较为突出。

图 10 分别展示了多种图像修复算法在不同掩膜比例下的修复效果对比。可以看出, 在掩膜比例较小的情况下, 各算法均能对图像进行一定程度的修复。AOT-GAN、Deepfillv1 和 Deepfillv2 能够恢复出大致的纹理和颜色, 但细节部分略显模糊; GLNet 和 PCNet 在纹理的细腻度上仍有提升空间; MIFDFSP 和 MER 的修复结果在颜色和纹理的连贯性上表现尚可; LRDiff 和 EGMF 作为较新方法, 在低掩膜率下已展现出良好的修复能力; 而本文算法在细节恢复上更为出色, 能够更好地还原图像的原始纹理和色彩, 视觉效果更加自然。图像在 [10%, 20%) 掩膜区域上, AOT-GAN、GLNet 和 PCNet 的

修复结果开始出现明显的模糊和伪影; Deepfillv2 在部分区域能够保持较好的结构, 但整体细节丢失较多; LRDiff 和 EGMF 在此阶段仍能保持较好的修复质量。图像在 [20%, 30%) 掩膜区域上, AOT-GAN、GLNet 和 PCNet 处理后的结构保持能力有所下降, 修复结果较差; MIFDFSP 和 MER 在颜色和纹理的融合上存在一定问题, 修复区域与周围区域的边界较为明显, LRDiff 和 EGMF 的修复质量已出现下降。图像在 [30%, 40%) 掩膜区域上, Deepfillv2 虽然能够大致恢复图像的轮廓, 但纹理和颜色的准确性大幅下降; MIFDFSP 和 MER 出现了明显的颜色失真, 修复效果不佳; LRDiff 和 EGMF 虽优于传统方法, 但也出现明显质量下降, 有明显的修复痕迹; 而本文算法在面对大面积掩膜时, 由于对频域特征的全局学习以及在融合模块中通过注意力机制的挤压策略, 使得模型在保证全局一致性的前提下, 仍能获得最逼真的细节区域纹理特征, 取得了较优的修复效果。

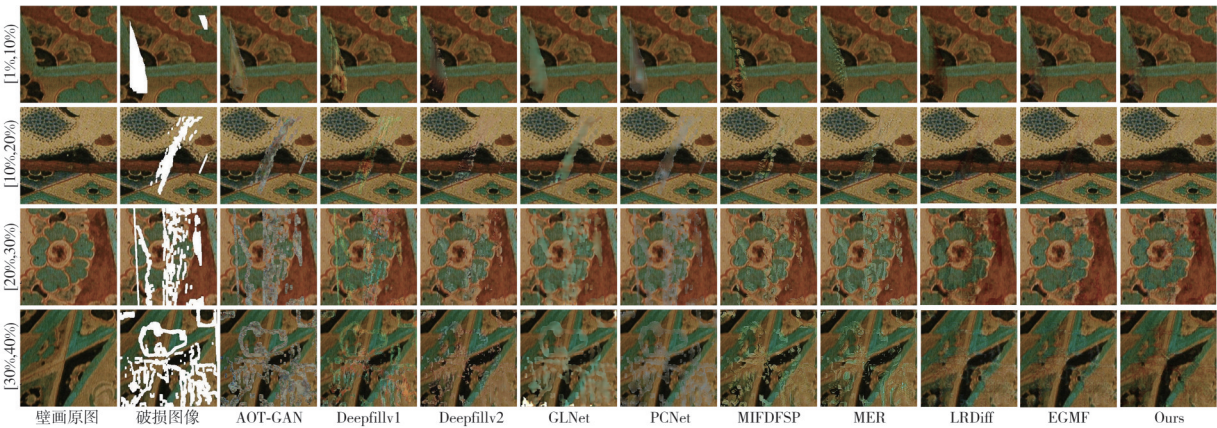


图 10 不同掩码率下各修复算法的修复结果

Fig. 10 Inpainting results of different algorithms under various mask ratios

3.4 消融实验

为了验证所提算法中各个模块的有效性，将MSF-Conv、LBAU、HAF模块依次单独引入至模型中，随机选择在[10%，20%)掩码率下进行壁画修复实验。实验结果如表3所示。

表 3 模块消融实验  
Tab. 3 Module experiment study

| Base | MSF-Conv | LBAU | HAF | PSNR/dB  | SSIM    |
|------|----------|------|-----|----------|---------|
| ✓    |          |      |     | 28.711 9 | 0.966 6 |
| ✓    | ✓        |      |     | 28.949 8 | 0.966 9 |
| ✓    |          | ✓    |     | 28.854 8 | 0.967 2 |
| ✓    |          |      | ✓   | 28.918 7 | 0.966 5 |
| ✓    | ✓        | ✓    | ✓   | 29.122 8 | 0.967 5 |

由表3可知，本文提出的模型较基础模型PSNR和SSIM分别提升了1.43%和0.09%，在引入模块后性能指标均有所提升，将所有模块引入后模型性能达到最佳。

为了评估上述各模块的视觉修复效果，本文通过随机选取[10%，20%)掩码率的图像进行定性分析，结果如图11所示。

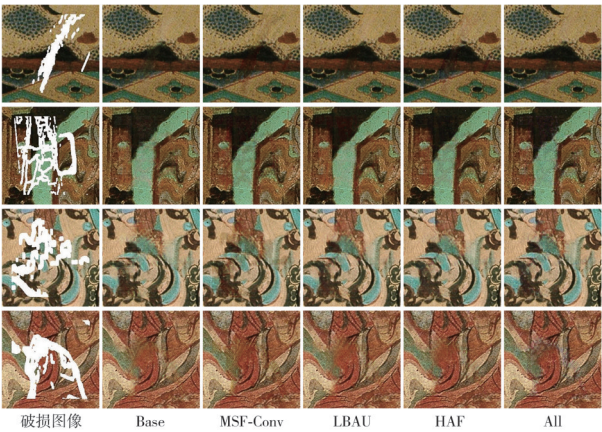


图 11 消融实验结果图

Fig. 11 Ablation experiment results

MSF-Conv模块通过空间域和频域特征联合提取，修复后的图像可有效捕捉壁画全局纹理特征，如周期性棱形纹，但是带来了修复边缘的伪影特征。LBAU模块则显著提升了修复区域和整理的纹理结构连贯性。HAF模块虽然在指标提升方面不明显，但是通过挤压激励策略弥补了MSF-Conv模块带来的修复伪影，3个模块协同使复杂的破损区域能够保持局部一致性的同时重建更好的细节特征。

4 结 论

本文围绕壁画修复中的长程纹理建模、结构连贯性保持与跨尺度特征融合三大挑战，提出了一种创新性两阶段生成对抗网络框架。通过MSF-Conv模块实现了频域全局建模与局部细节修复的协同优化，LBAU模块利用HOG几何先验增强了长距离依赖感知，HAF模块通过动态通道加权抑制了多尺度伪影。消融实验验证了各模块的有效性。与其它方法对比，本算法在PSNR和SSIM指标上均达到最优，视觉上，其在矿物颜料渐变自然度、长程结构连贯性与边缘细节保真度方面均有显著提升，验证了本方法在应对敦煌壁画图像修复核心问题时的有效性。未来工作可聚焦探索轻量化部署与多模态先验的融合，以进一步提升复杂风化场景的修复鲁棒性。

利益冲突声明/Conflict of Interests

所有作者声明不存在利益冲突。  
All authors declare no relevant conflict of interests.

参考文献

[1] 慕天琪. 智能化虚拟复原技术与应用研究[D]. 北

- 京: 北京邮电大学, 2020.
- [2] CAO J, ZHANG Z, ZHAO A, et al. Ancient mural restoration based on a modified generative adversarial network[J]. *Heritage Science*, 2020, 8(1): 1-14.
- [3] 吴余芳. 图像修复算法在国画中的应用[D]. 杭州: 浙江大学, 2008.
- [4] SRIDEVI G, SRINIVAS KUMAR S. Image inpainting based on fractional-order nonlinear diffusion for image reconstruction[J]. *Circuits, Systems, and Signal Processing*, 2019, 38(8): 3802-3817.
- [5] 曹浩杰, 李郁峰, 张权. 一种改进的Criminisi眼底图像修复算法[J]. *中北大学学报(自然科学版)*, 2023, 44(2): 176-181.  
CAO Haojie, LI Yufeng, ZHANG Quan. An improved criminisi fundus image restoration algorithm[J]. *Journal of North University of China (Natural Science Edition)*, 2023, 44(2): 176-181. (in Chinese)
- [6] YAO F. Damaged region filling by improved criminisi image inpainting algorithm for thangka [J]. *Cluster Computing*, 2019, 22(6): 13683-13691.
- [7] YANG S, LIANG H, WANG Y, et al. Image inpainting based on multi-patch match with adaptive size[J]. *Applied Sciences*, 2020, 10(14): 4921.
- [8] CRIMINISI A, PEREZ P, TOYAMA K. Region filling and object removal by exemplar-based image inpainting[J]. *IEEE Transactions on Image Processing*, 2004, 13(9): 1200-1212.
- [9] KÖHLER R, SCHULER C, SCHÖLKOPF B, et al. Mask-specific inpainting with deep neural networks[M]// *Cham: Springer International Publishing*, 2014: 523-534.
- [10] GAO L, ZHOU X. A super-resolution reconstruction algorithm for mural images based on improved generative adversarial network[J]. *Journal of Measurement Science and Instrumentation*, 2024, 15(4): 499-508.
- [11] CHEN Y, GAO Y, ZHU L, et al. PCNet: prior category network for CT universal segmentation model[J]. *IEEE Transactions on Medical Imaging*, 2024, 43(9): 3319-3330.
- [12] YU J, LIN Z, YANG J, et al. Generative image inpainting with contextual attention[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 5505-5514.
- [13] YU J, LIN Z, YANG J, et al. Free-form image inpainting with gated convolution [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 4470-4479.
- [14] ZENG Y, FU J, CHAO H, et al. Aggregated contextual transformations for high-resolution image inpainting[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2023, 29(7): 3266-3280.
- [15] CHEN W, JIANG Z, WANG Z, et al. Collaborative global-local networks for memory-efficient segmentation of ultra-high resolution images[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 8916-8925.
- [16] YU J, LIN Z, YANG J, et al. Generative image inpainting with contextual attention[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 5505-5514.
- [17] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words; transformers for image recognition at scale [EB/OL]. [2025-11-03]. <https://doi.org/10.48550/arXiv.2010.11929>.
- [18] LI W, LIN Z, ZHOU K, et al. MAT: mask-aware transformer for large hole image inpainting[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022: 10748-10758.
- [19] 刘仲民, 严梁. 融合动态特征与注意力的敦煌壁画修复模型[J]. *计算机工程*, 2024, 50(5): 342-353.  
LIU Zhongmin, YAN Liang. Inpainting model of Dunhuang mural fusing dynamic feature and attention[J]. *Computer Engineering*, 2024, 50(5): 342-353. (in Chinese)
- [20] WEI X, FAN B, WANG Y, et al. Progressive enhancement and restoration for mural images under low-light and defective conditions based on multi-receptive field strategy[J]. *NPJ Heritage Science*, 2025, 13: 63.
- [21] LEI T, LI Y, CHENG Z, et al. Low-rank structure guided diffusion for Shaanxi temple mural restoration [J]. *NPJ Heritage Science*, 2025, 13: 295.
- [22] LIU Z, LIU Y, HU W. Mural restoration *via* the fusion of edge-guided and multi-scale spatial features [J]. *NPJ Heritage Science*, 2025, 13: 491.