DOI:10.11918/202410044

级联迭代精细化的敦煌壁画图像修复

曹 岩,辛子昊,马 宁,朱崇凯,邬开俊

(兰州交通大学 电子与信息工程学院,兰州 730070)

摘 要:为解决破损敦煌壁画图像修复中存在的结构修复不合理及图像特征纹理不够细化的问题,提出一种级联迭代精细化壁画图像修复模型。该模型通过多层残差连接网络(multi residual-net,MR-Net)结合空间和通道注意力,利用迭代层归一化残差模块(layernorm residual module,LRM)及普通卷积进行先修复,逐步提取壁画图像的全局和局部特征信息;结合深度可分离卷积下的多头转置注意(dsconv layernorm transposed-attention,DSLT-A)和门控部分卷积特征聚合传播模块(feature aggregation propagation module,FAPM),对图像的全局信息和局部特征纹理信息进行重构;通过部分卷积多尺度门控模块增加修复过程中缺失的原始图像特征信息,对壁画图像的纹理信息进行细化。进行人为不规则掩码修复实验、模拟霉斑掩码修复实验以及真实破损壁画修复实验。实验结果表明:与4种对比模型中表现最好的数据相比,本文所提模型修复图像的峰值信噪比提升了1.178 dB,结构相似性指数提升了0.010 2,平均平方损失函数降低了0.000 9,3种指标均优于对比模型。所提模型能够修复出更高质量的壁画图像,为提高敦煌壁画图像修复水平以及后续研究数字化敦煌壁画图像修复提供技术支撑和参考。

关键词:深度学习;壁画修复;门控模块;部分卷积;特征细化

中图分类号:TP391.41 **文献标志码**:A **文章编号**:0367-6234(2026)09-0125-10

Cascade iterative refinement of Dunhuang mural image inpainting

Cao Yan, Xin Zihao, Ma Ning, Zhu Chongkai, Wu Kaijun

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

Abstract: To address the issues of unreasonable structural inpainting and insufficiently refined image texture features in the inpainting of damaged Dunhuang mural images, this paper proposes a cascaded iterative fine-grained mural image inpainting model. This model utilizes a multi-layer residual-net (MR-Net) combined with spatial and channel attention mechanisms,and employs an iterative layernorm residual module (LRM) together with standard convolution for initial inpainting, progressively extracting both global and local feature information of the mural images. Subsequently, the dsconv layernorm teansposed-attention (DSLT-A) and the gated feature aggregation propagation module (FAPM) are combined to reconstruct the global and local texture features. A partial convolution multi-scale gating module is used to enhance missing original image feature information during the inpainting process, refining the texture information of the mural images. Experiments were conducted using artificially irregular mask inapinting, simulated mold spot mask restoration, and real damaged mural inpainting. The experiment results show that, under three objective evaluation metrics, compared with the best-performing data among four comparative models, the model proposed in this paper improves the peak signal-to-noise ratio of the inpainted images by 1.178 dB, structural similarity index by 0.010 2, and reduces mean squared loss by 0.000 9, all outperforming the other comparative models. The proposed model is capable of inpainting higher-quality mural images, providing technical support and reference for improving the inpainting level of Dunhuang mural images and promoting future research in digital inpainting of Dunhuang murals.

Keywords: deep learning; mural inpainting; gating module; partial convolution; feature refineme

敦煌壁画以其精湛的绘画技艺和多样化的题材内容闻名,展示了中国古代艺术和宗教信仰的深厚积淀。作为教育工具,其对帮助人们理解古代艺术、宗教和文化有着重要意义。然而,由于时间、天气和环境污染因素,壁画面临着自然和人为破坏的风险。除了对壁画进行现场修复以外,数字化的方式也是当今文物保护的重要手段。数字图像修复技术的出现为壁画修复带来新的保障,其通过不断学习图像

收稿日期:2024-10-21;**录用日期**:2024-12-30;**网络首发日期**:2025-10-31
网络首发地址:https://link.cnki.net/urlid/23.1235.T.20251030.1854.004
基金项目:甘肃省自然科学基金(23JRRA913);内蒙古自治区重点研发与成果转化计划项目(2023YFSH0043,2023YFDZ0054)
作者简介:曹 岩(1982—),男,副教授,硕士生导师;邬开俊(1978—),男,教授,博士生导师
通信作者:邬开俊,wkj@mail.lzjtu.cn

的特征信息,促进生成符合壁画间语义联系的像素,令修复后的壁画在视觉上更生动逼真[1-2]。因此,数字化图像修复技术对保护敦煌壁画具有重要意义。

数字化图像修复方法可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。其中,传统的图像修复方法又分为基于偏微分的图像修复方法和基于样本的图像修复方法。基于偏微分的图像修复是利用数学中的偏微分方程将图像中已知的像素平滑地扩散至破损图像中,实现对图像的修复[3-4]。Bertalmio 等[5]提出梯度和曲率信息共同调节扩散模型(Bertalmio-Sapiro-Caselles Ballester, BSCB),该模型利用已知信息沿等光线方向平滑传播来修复图像。而基于样本的图像修复方法是通过查询并使用与缺失区域相似度最高的区域来修复破损图像的。Efros 等[6]提出基于马尔科夫随机场的非参数纹理合成修复方法,该方法通过大量时间计算得出与已知区域近似的相似度得分实现修复。但以上两种传统的图像修复方法无法获得较强的自适应能力,缺乏对破损图像语义信息的理解。

为进一步加强对破损壁画图像的修复,研究人员运用基于深度学习的图像修复方法实现对图像的修复。Yang 等[7]利用残差块替代卷积层,设计了一个空间上下文编码器联合结构以提供先验信息,同时训练一个共享生成器以修复图像纹理信息,从而补全图像。在此基础上,Zeng 等[8]基于 U-Net 结构,引入了多尺度编解码器,采用逐层注意力转移网络填充图像的缺失区域,通过跳跃连接将注意力转移网络学习到的重构特征与潜在特征进行解码,取得了良好的视觉和语义修复效果。为解决修复过程中图像精细部分内容缺失的问题,Wang 等[9]在 U-Net 的编码器中引入分层金字塔卷积和动态归一化机制,用于获取具有不同感受野的图像特征;在解码器中,采用金字塔注意力机制学习更加精细的特征信息。Chen 等[10]集成 4 组空洞卷积层的密集多尺度融合模块,扩大模型的感受野范围,更有效地捕捉了图像中的深层次信息。此外,为进一步提升生成对抗网络(generative adversarial network, GAN)在生成图像时的语义准确性和结构一致性,引入了视觉几何组(visual geometry group, VGG)特征匹配损失与自导回归损失相结合的训练策略。但上述方法在尝试修复大面积受损区域时,容易出现结构上的混乱以及边界模糊不清等问题。为解决上述问题,结合 Transformer 的图像修复方法逐渐兴起。Xu 等[11]的两阶段盲人脸修复方法通过频率引导 Transformer 的精准定位与自上而下细化编码器–解码器架构的精细修复相结合,为盲人脸修复领域提供了一种高效且高质量的解决方案。在此基础上,Zhan 等[12]结合基于 Transformer 的编码器和自适应注意力感知层的细化阶段,实现了一种从低分辨率图像到高质量、高分辨率图像的转换策略。在修复过程中,为充分利用掩码信息,Quan 等[13]提出了掩码感知 Transformer 模型,其独特的多头上下文注意力机制和样式操作模块为高分辨率图像的缺失区域修复问题提供了高效、精准的解决方案,同时也为图像修复领域的研究开辟了新的思路。但以上方法受限于对图像深层语义特征的有效学习与理解,也会造成修复结构不合理与纹理修复不够细化的问题,进而影响图像整体的一致性和真实感。

针对以上问题,本文提出一种级联迭代精细化的敦煌壁画修复模型。第 1 阶段,提出多层残差连接网络(multi residual-net, MR-Net),结合空间和通道注意力,利用迭代层归一化残差模块(layernorm residual module, LRM)进行先修复,逐步提取图像中的全局和局部信息。第 2 阶段,提出多头转置注意(dsconv layernorm transposed-attention, DSLT-A)和门控部分卷积特征聚合传播模块(feature aggregation propagation module, FAPM),加强对壁画结构及纹理信息修复的准确性和质量;设计由整体到细节的精修复网络(DF-Net),更深层次理解壁画像素间的语义关系,细化图像特征边缘;引入 multi-scale gating module[14]并改进为部分卷积多尺度门控模块(pconv scale gating module, PSGM),提高不同尺度信息间的交互,进一步解决壁画图像修复过程中结构修复不合理及原始特征信息丢失的问题。

1 级联迭代精细化敦煌壁画修复模型

1.1 整体模型

敦煌壁画具有独特的艺术风格和精湛的绘画技巧,因此其特征细节丰富,结构形式多样。为将破损壁画的结构和纹理细节进行精确修复,获得更加真实的视觉效果[15],本文提出一种级联迭代精细化图像修复网络模型,其框架如图 1 所示。该模型的第 1 阶段由结合空间和通道注意力的 LRM 组成,破损壁画图像经过 4 层残差块后获得初步修复后的图像。模型通过多次堆叠残差块,逐步增强对破损壁画内容的理解。第 2 阶段主要由基于 7 层深度可分离卷积的 DSLT-A 和 FAPM 组成,该组合用于细化和增强第 1 阶段得到的修复图像。在第 2 阶段中,输入特征大小为 $H \times W \times C$ 的图像。对特征经过第 1 层 Transformer Block 模块后的下采样(down)操作,将输入图像的高度与宽度进行减半处理,同时扩张壁画图像的通道数,图像大小变为 $H/2 \times W/2 \times 2C$。

以同样方式经过 3 层 Transformer Block 结构，逐步改变壁画尺寸，图像大小变为 $H/8 \times W/8 \times 8C$，使模型能够感受图像的细节信息，增强不同通道信息间的联系。最后，以同样的方式将图像进行上采样（up），恢复壁画图像的原始大小，并逐层增加图像的原始信息。

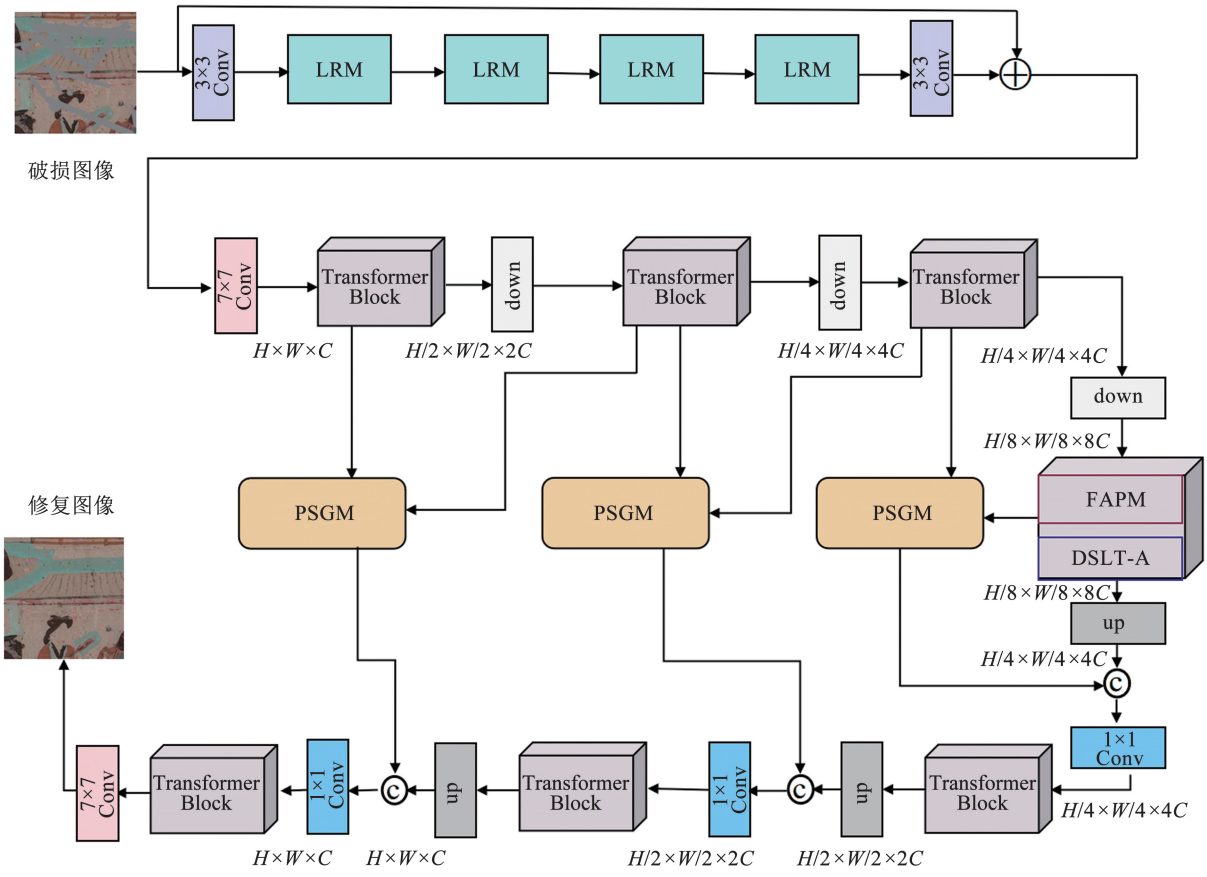


图 1　模型总体框架

Fig. 1　Overall framework of the model

1.2　MR-Net

由于模型对破损壁画图像细节信息掌握程度较低，破损区域语义联系不连贯，因此，采用结合空间和通道注意力并行的迭代层归一化残差连接网络 MR-Net。该网络采用 3 × 3 Conv 将带有掩码的壁画图像特征进行初步提取，然后将提取到的信息传入 LRM。LRM 的结构如图 2 所示。

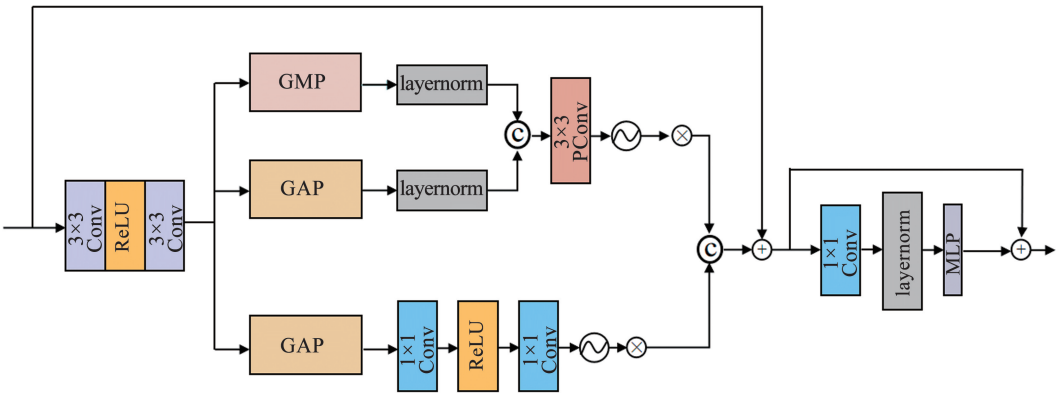


图 2　LRM 结构

Fig. 2　LRM structure

首先，采用 ReLU 激活函数引入非线性因素增强学习能力，利用全局最大池化（golbal max pool，GMP）和全局平局池化（golbal average pool，GAP）降低过拟合风险，利用归一化层（layernorm）将每个维度上的特征进行独立的归一化处理，防止由于残差块迭代次数的增加而令模型难以应对输入数据分布

变化。其次,空间注意力和通道注意力两个模块并行处理,得到修复过程所需要的相应特征。空间注意力模块将输入的特征图在通道维度上进行压缩与卷积处理,生成空间注意力特征,进一步提高模型对壁画结构信息的感知能力,促进壁画生成结构的合理性[16];通道注意力模块将特征在空间维度上进行压缩,得到每个通道的平均值,然后经过 Sigmoid 函数激活,得到每个通道的重要性信息,这有助于模型在破损壁画图像重建时关注对修复纹理细节更有用的特征信息。最后,将提取到的特征进行拼接处理,通过归一化层和多层感知机层(multilayer perceptron,MLP)加强特征的融合,提取深度特征信息以促进修复效果的生成。其中,以双注意力模块为主的注意力计算单元的表达式为

$$\boldsymbol{F}_{\text{out}} = \boldsymbol{T}_{\text{in}} + \text{Conv}(\text{Concat}([S(\boldsymbol{x}), C_{\text{n}}(\boldsymbol{x})])) \quad (1)$$

式中:$\boldsymbol{F}_{\text{out}}$为经过注意力单元后得到的特征图,$\boldsymbol{T}_{\text{in}} \in \mathbf{R}^{H\times W\times C}$为初始输入特征图,$\boldsymbol{x}$ 为输入张量经过两层 3×3 卷积后的特征图,S 为空间注意力,C_{n} 为通道注意力,Conv(·)为卷积层,Concat(·)为沿通道维度上的连接操作。

1.3 精修复网络 DF-Net

敦煌壁画受储存环境及时间的影响,出现了表面破损、色彩特征不明显的问题。这从根本上导致了壁画局部特征信息的缺失和像素之间语义信息的匮乏,为修复带来了困难[17]。

为使修复出的结果纹理信息更丰富,结构信息更合理,提出一种结合深度可分离卷积的 DSLT-A 和 FAPM 以及 PSGM 的 DF-Net。该网络采用逐层提取特征的方式对第 1 阶段的粗修复结果进行细节处理和纹理修复,通过堆叠 7 层 DSLT-A 与 FAPM 组成的 Transformer Block 增强模型对特征的深层次提取,进而加深模型对破损壁画结构信息和纹理细节的理解。最后利用 PSGM,通过跳跃连接完成不同尺度特征信息的融合,结合深度可分离卷积实现多通道破损壁画特征信息的保留,有效解决修复后图像特征纹理不够细化的问题。

1.3.1 DSLT-A

在修复过程中,破损壁画会因纹理细节信息丢失而导致修复结果出现视觉紊乱及纹理信息不够丰富的问题,为增强破损壁画不同通道间的特征联系,提出了 DSLT-A,其结构如图 3 所示。首先,模型将特征图输入至并联 1×1 Conv 与 3×3 深度可分离卷积(depthwise separable, DWConv)的 DSLT-A 中,以提取每个通道的局部特征并捕获图像中的空间信息,降低模型的计算复杂度和参数量[18],减少过拟合的风险。随后,DSLT-A 利用其多个注意力头捕获多种特征,计算通道间的交叉协方差,实现生成破损壁画全局间的注意力特征图。DF-Net 的表达式为:

$$\hat{\boldsymbol{X}} = \boldsymbol{W}\boldsymbol{F}_{\text{at}}(\hat{\boldsymbol{Q}}, \hat{\boldsymbol{K}}, \hat{\boldsymbol{V}}) + \boldsymbol{X} \quad (2)$$

$$\boldsymbol{F}_{\text{at}}(\hat{\boldsymbol{Q}}, \hat{\boldsymbol{K}}, \hat{\boldsymbol{V}}) = \hat{\boldsymbol{V}} \times \text{Softmax}(\hat{\boldsymbol{K}} \times \hat{\boldsymbol{Q}}/\alpha) \quad (3)$$

式中:$\boldsymbol{X}$ 为原始输入的特征图,$\hat{\boldsymbol{X}}$ 为经过 DSLT-A 后得到的特征图,$\boldsymbol{W}$ 为经过卷积层之后的特征,α 为可学习的放缩参数,$\hat{\boldsymbol{Q}}$、$\hat{\boldsymbol{K}}$、$\hat{\boldsymbol{V}}$ 分别为查询向量、键向量、值向量,Softmax(·)为激活函数,$\boldsymbol{F}_{\text{at}}$为 DSLT-A 输出的注意力特征图。

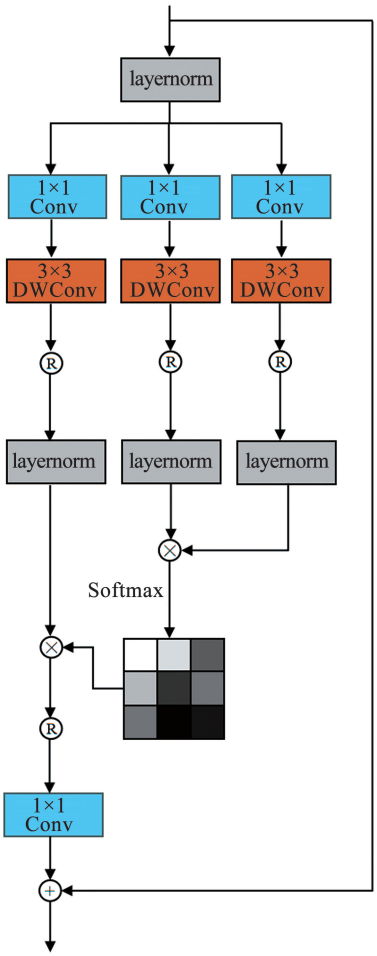


图 3 DSLT-A 结构

Fig. 3 DSLT-A structure

1.3.2 FAPM

在修复破损图像时,有用像素和无关像素均会被输入至模型中,只利用普通卷积无法区分出有用像素,进而提出 FAPM,其结构如图 4 所示。首先,对输入的特征图经沿不同维度分别执行独立归一化操作。其次,通过聚焦线性注意(foused linear attention)模块加快模型对缺失壁画图像的推理过程,增强其捕捉全局信息的能力并提高前向传播的速度。将注意力处理后的特征信息交由 1×1 Conv 与 3×3 部分卷积(PConv)进行处理后经过 GeLU 激活函数激活,使模块能够更好地填充缺失区域的信息,增加模型的泛化能力。在门控机制作用下,

FAPM 可以过滤掉修复过程中的杂乱信息，动态的选择出对填补缺失区域更有用的像素。该模块在前向传播的过程中能捕捉到更丰富的上下文信息，对于提取底层语义信息具有更好的表现性，也大大减少了杂乱信息对模型学习深层特征的干扰。最后，FAPM 对筛选出的特征信息进行聚合，在保持整体结构与局部信息一致性的同时，提升修复纹理部分的真实性。对于聚焦线性注意力模块，采用文献[19]的计算方式增强该结构对破损壁画中重要信息的集聚能力以及特征多样性的表达能力，具体计算公式为：

$$\boldsymbol{O}=S_{\mathrm{im}}(\boldsymbol{Q},\boldsymbol{K})\boldsymbol{V}=\phi_{\mathrm{p}}(\boldsymbol{Q})\phi_{\mathrm{p}}(\boldsymbol{K}^{\mathrm{T}})\boldsymbol{V}+\mathrm{DWConv}(\boldsymbol{V}) \tag{4}$$

$$S_{\mathrm{im}}(\boldsymbol{Q}_i,\boldsymbol{K}_j)=\phi_{\mathrm{p}}(\boldsymbol{Q}_i)\phi_{\mathrm{p}}(\boldsymbol{K}_j^{\mathrm{T}}) \tag{5}$$

$$\phi_{\mathrm{p}}(\boldsymbol{x})=\boldsymbol{f}_{\mathrm{p}}(\mathrm{ReLU}(\boldsymbol{x})) \tag{6}$$

$$\boldsymbol{f}_{\mathrm{p}}(\boldsymbol{x})=\frac{\|\boldsymbol{x}\|}{\|\boldsymbol{x}^{**p}\|}\boldsymbol{x}^{**p} \tag{7}$$

式中：$\boldsymbol{O}$ 为注意力的输出，$S_{\mathrm{im}}(\cdot)$ 为相似度函数，$\boldsymbol{x}^{**p}$ 为 $\boldsymbol{x}$ 的逐元素 p 次幂，$\phi_{\mathrm{p}}(\cdot)$ 为特征映射函数，ReLU(·)为激活函数，*DWConv* 为在聚焦线性注意模块中的深度可分离卷积，$\boldsymbol{f}_{\mathrm{p}}(\cdot)$ 为注意力作用下的特征图。

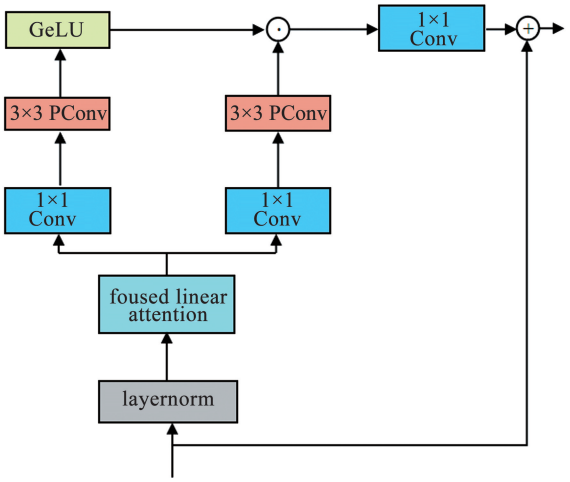


图 4　FAPM 结构

Fig. 4　FAPM structure

1.3.3　PSGM

在修复壁画图像的过程中，往往因模型层数深度和模块数量的增加而造成某些原始图像特征丢失，从而导致模型对壁画局部区域特征信息理解失准。为使模型在修复图像的过程中能兼顾某些重要的原始特征[20]，引入了 PSGM 以降低修复图像与原始图像结构信息不合理的问题，并通过增加特征信息以补充局部纹理信息的缺失，丰富模型对上下文信息的理解，使模型更好地重建缺失区域的结构特征[21]。PSGM 结构如图 5 所示。

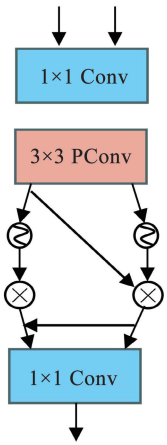


图 5　PSGM 结构

Fig. 5　PSGM structure

将 DF-Net 中的第 1 层 Transformer 模块与第 2 层 Transformer 模块、第 2 层 Transformer 模块与第 3 层 Transformer 模块、第 3 层 Transformer 模块与第 4 层 Transformer 模块提取到的特征分别传入 PSGM 中，利用不同大小的缩放因子将其结合在一起并调整为适合不同解码器所需要的尺度，在通道维度上将特征信息进行融合；利用 1×1 Conv 调整尺度大小，利用 3×3 PConv 提取相应的特征；利用 Sigmoid 函数激活得到两个特征，将不同尺度的两部分特征融合在一起，并转换为增强型特征。PSGM 计算过程表达式为：

$$\boldsymbol{F}=\mathrm{Conv}(\mathrm{Concat}([(\boldsymbol{T}_1^{\mathrm{out}}),(\boldsymbol{T}_2^{\mathrm{out}})^{\uparrow}])) \tag{8}$$

$$\boldsymbol{F}_1,\boldsymbol{F}_2=\mathrm{Split}(\mathrm{PConv}(\boldsymbol{F})) \tag{9}$$

$$\boldsymbol{F}_3=\mathrm{Sigmoid}(\boldsymbol{F}_1)\times\boldsymbol{F}_2 \tag{10}$$

$$\boldsymbol{F}_4=\mathrm{Sigmoid}(\boldsymbol{F}_2)\times\boldsymbol{F}_1 \tag{11}$$

$$\boldsymbol{F}_{\mathrm{e}}=\mathrm{Conv}(\mathrm{Concat}([\boldsymbol{F}_3,\boldsymbol{F}_4])) \tag{12}$$

式中：$\boldsymbol{T}_i^{\mathrm{out}}(i=1,2,3,\cdots,n)$ 为 n 个 Transformer 模块的输出，$\boldsymbol{F}$ 为输入 PSGM 中的图像特征，↑为上采样操作，PConv(·)为部分卷积操作，Split(·)为分割成块操作，$\boldsymbol{F}_1$、$\boldsymbol{F}_2$、$\boldsymbol{F}_3$、$\boldsymbol{F}_4$ 为通道分割后的特征图，$\boldsymbol{F}_{\mathrm{e}}$ 为增强型特征。

1.4　损失函数[22]

1.4.1　重建损失

修复后的图像与原始壁画图像之间存在像素差异，从而引入重建损失进行度量。其表达式为

$$L_{\mathrm{re}}=\|\boldsymbol{I}_{\mathrm{out}}-\boldsymbol{I}_{\mathrm{g}}\|_1 \tag{13}$$

式中：L_{re} 为重建损失，$\boldsymbol{I}_{\mathrm{out}}$ 为输出修复后的壁画图像，$\boldsymbol{I}_{\mathrm{g}}$ 为真实壁画图像。

1.4.2　感知损失

感知损失通常利用卷积神经网络提取图像的高级特征，这些特征反映了图像的内容和结构，而非单纯的像素值。采用感知损失的目的是提高生成图像在结构和语义上与真实目标图像的匹配程度。其表

达式为

$$L_{\text{perc}} = E\left[\sum_i \frac{1}{N_i} \| L_i(\boldsymbol{I}_{\text{out}}) - L_i(\boldsymbol{I}_{\text{g}}) \|_1\right] \quad (14)$$

式中：L_{perc}为感知损失，$E[\cdot]$为取期望操作，$L_i(\cdot)$为在 ImageNet 上预训练 VGG19 的第 i 层的激活函数，N 为特征层数。

1.4.3 风格损失

引入风格损失的目的是将原壁画的风格应用于修复后的图像上。风格损失的计算方式为从原风格图像和生成图像中提取多个层的特征图，并计算相应的 Gram 矩阵。然后，比较真实风格图像和生成图像的 Gram 矩阵之间的差异，若假设输入图像的大小为 $C \times H \times W$，那么风格损失的表达式为

$$L_{\text{style}} = E[\| \boldsymbol{G}_j^{\Phi}(\boldsymbol{I}_{\text{out}}) - \boldsymbol{G}_j^{\Phi}(\boldsymbol{I}_{\text{g}}) \|_1] \quad (15)$$

式中：L_{style}为风格损失，$\boldsymbol{G}_j^{\Phi}$ 为输入图像相应激活图构建的 L_i 的 Gram 矩阵。

1.4.4 对抗性损失

引入对抗性损失，是为了通过判别器和生成器之间的对抗作用，生成高质量的壁画图像，并增强修复壁画的视觉效果，其表达式为

$$L_{\text{adv}} = E[\log D(\boldsymbol{I}_{\text{g}})] + E\{\log[1 - D(\boldsymbol{I}_{\text{out}})]\} \quad (16)$$

式中：L_{adv}为对抗性损失，D 为具有光谱归一化的补丁 GAN 判别器。

1.4.5 总体损失

将重建损失、感知损失、风格损失以及对抗性损失结合，得到总体损失，其公式为

$$L = \lambda_{\text{r}} L_{\text{re}} + \lambda_{\text{p}} L_{\text{perc}} + \lambda_{\text{s}} L_{\text{style}} + \lambda_{\text{a}} L_{\text{adv}} \quad (17)$$

式中：L 为总体损失，λ_{r}、λ_{p}、λ_{s}、λ_{a} 分别为重建损失、感知损失、风格损失、对抗性损失对应的超参数，分别取 1、1、250、0.1。$\lambda_{\text{r}} = 1$ 说明该损失项的重要性与其他损失项相当，确保生成图像在像素级别上尽可能接近目标图像。$\lambda_{\text{p}} = 1$ 强调高层次特征的重要性，有助于生成更自然、更真实的图像。$\lambda_{\text{s}} = 250$ 意味着风格损失在整体损失函数中占有非常高的权重，强调风格的一致性。$\lambda_{\text{a}} = 0.1$ 意味着对抗性损失的贡献相对较小，但仍然不可忽视，较小的权重可以避免对抗性损失对生成过程的过度影响。

2 实验结果及分析

2.1 数据集

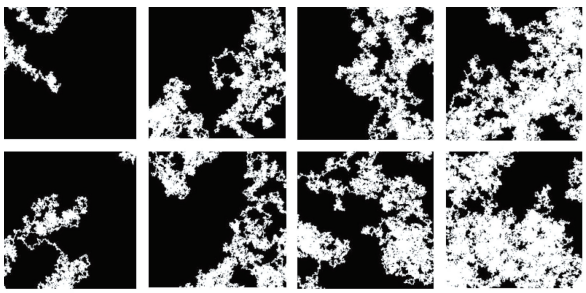
采用敦煌壁画图像数据集验证本文模型的有效性和准确性。由于数据集中图像均为未经修补的原始图像，因此，局部壁画信息损失较为严重，纹理信息相对模糊。壁画数据集包含佛像、莲台等种类及大小均不等的图像，采用裁剪、变形等方法随机选取 8 000张图像作为训练集，1 000 张作为测试集，如图 6所示。为训练模型绘制了模拟霉斑掩码，掩码率为 0% ~80%，如图 7 所示。为进一步验证模型的有效性，在绘制掩码数据集训练的同时也采用掩码率为 40% ~60% 的公开不规则掩码数据集验证本文模型应对不同数据的泛化能力。



(a) 佛像 (b) 侍女 (c) 莲台 (d) 建筑 (e) 服饰

图 6 敦煌壁画数据集

Fig. 6 Dunhuang mural datasets



(a) 0%~<20% (b) 20%~<40% (c) 40%~<60% (d) 60%~80%

图 7 不同范围掩码数据集

Fig. 7 Datasets with different range masks

2.2 实验环境及实验设置

实验运行环境为 Windows 11 操作系统，硬件配置为 Intel(R) Core i7-13650H CPU @ 3.40 GHz，16.0 GB RAM，NVIDIA GeForce RTX4060。对比实验均在相同配置下进行。模型训练时，学习率设为 10^{-4}，训练周期为 200 轮。

2.3 评价指标

考虑像素值与视觉感知两方面对修复结果的影响，选用平均平方损失函数(mean squared loss，L_2)、结构相似性指数(structural similarity index，I_{SSM})、峰值信噪比(peak signal-to-noise ratio，R_{PSN})作为评价

指标。

2.4　对比实验模型

为全面评估本文模型的有效性，在敦煌壁画数据集上进行修复实验和对比实验。选用近年来较流行的 LGNet[13]、Shift-Net[23]、CSWT[24]、RFR[25] 作为对比模型。LGNet[13] 是一种从局部到全局的细化三阶段图像修复模型，引入了跳跃连接与高效注意力增强修复精度；Shift-Net[23] 是一种基于 U-Net 的模型结构，采用了一种新颖的位移机制，能够在不同的空间位置之间进行信息交流；CSWT[24] 采用条纹窗口机制，将输入图像划分为不同的区域，通过变换器结构捕获长距离依赖关系；RFR[25] 是一种结合递归推理特征和知识一致性注意力模块构建的模型，能够从不同尺度提取特征，使其能够捕捉图像中的细节和上下文信息。

2.5　人为添加随机不规则掩码及模拟霉斑掩码修复实验

采用随机不规则且掩码率为 40% ~60% 的掩码及霉斑掩码，模拟图像的不同程度损坏状态，开展效果验证实验。抽样选取 5 张敦煌壁画图像作为样本，与 4 种模型进行对比，结果如图 8 所示。通过图 8可以发现，本文模型与 4 种对比模型均可以修复结构较为简单的壁画图像，差别在于处理结构和纹理信息相对复杂的图像。Shift-Net 模型修复后的图像虽然语义信息较为符合原图像，但色彩分布不均匀，人脸结构扭曲且伴有纹理细节模糊；CSWT 模型修复后的图像虽未出现色彩分层不均匀的问题，但仍存在纹理细节模糊的缺陷，人体结构修复图像局部区域出现了结构修复不合理以及明显修复痕迹的现象（图 8(c) 中最后 1 行）；RFR 模型修复后的图像在色彩分层和人体修复方面均出现了部分结构断裂和纹理模糊的问题，且许多区域的结构被过度修复；LGNet 模型修复后的图像可以清楚地实现色彩分层，清晰地显示人物整体结构与合理的纹理填充，但仍存在纹理细节信息不够细化的问题；而本文模型在局部细节以及整体结构上表现出更优的视觉效果，在人物头像、佛本生故事场景等方面的修复均有良好的表现。

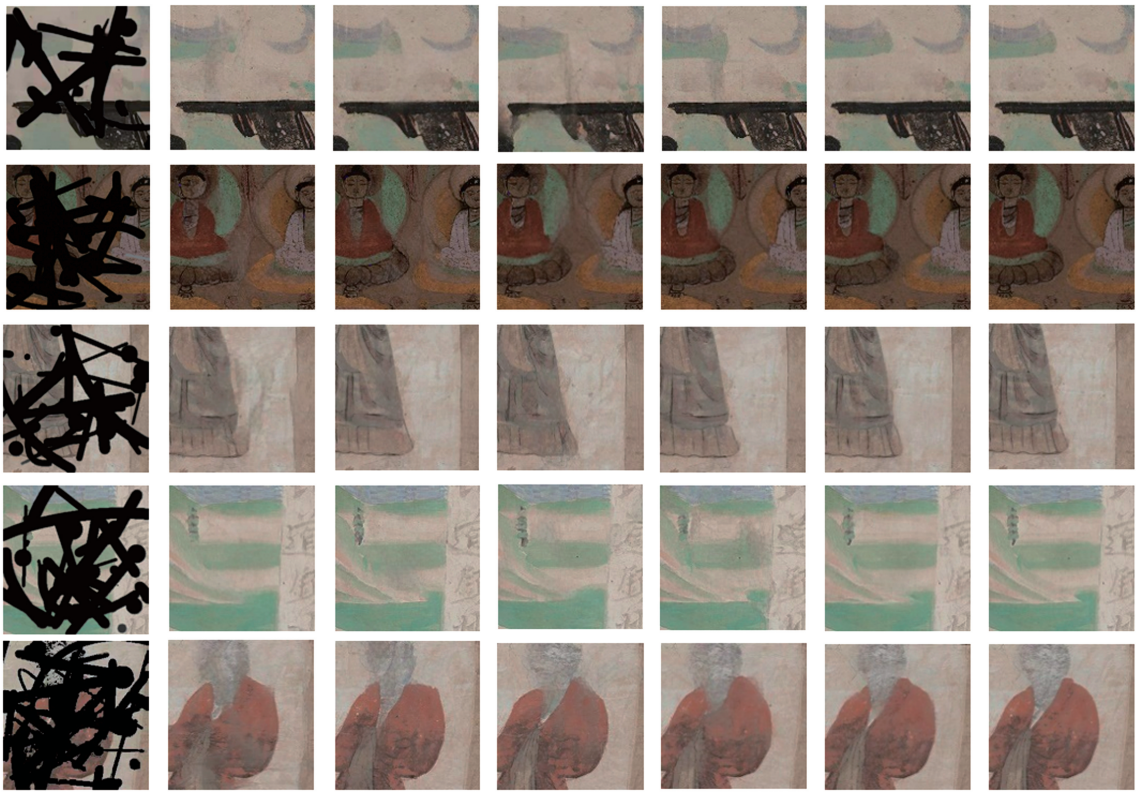


图 8　不同模型在随机掩码下的修复结果对比

Fig. 8　Comparison of inpainting results of different models under random masks

表 1 为各种模型在随机不规则掩码下定量结果对比。由表 1 可以看出，本文模型的 R_{PSN}、I_{SSM} 均高于对比模型，L_2 均低于对比模型，说明本文模型在客观和主观上均取得了更好的修复效果。

表 1 随机不规则掩码下敦煌数据集定量结果

Tab. 1 Quantitative results of the Dunhuang dataset under random irregular masks

模型	R_{PSN}/dB	I_{SSM}	L_2
Shift-Net	27.133	0.924 0	0.022 4
CSWT	30.286	0.943 6	0.019 8
RFR	28.505	0.930 0	0.021 0
LGNet	34.397	0.957 8	0.021 9
本文模型	**34.988**	**0.962 1**	**0.013 8**

模拟霉斑掩码是模拟受天气、人为因素影响下破损壁画的真实场景,从而使模型适应真实的壁画修复环境。本文模型与其他 4 种对比模型修复结果如图 9 所示。由图 9 可以看出:由于选取图像的纹理与结构信息较为复杂,Shift-Net 模型出现了色彩重合及图像信息模糊;CSWT 模型修复后,图像局部区域出现了结构不合理的现象,有的人物头像的明显结构未能修复出;RFR 模型修复后,出现局部色彩混杂现象且纹理特征不够细化;LGNet 模型能够对缺失区域进行填充,但修复后壁画图像的结构仍出现了断裂、衔接不合理的现象。

为进一步验证模拟真实壁画损坏场景下的修复效果,各模型在霉斑掩码修复实验中的定量结果对比见表 2。由表 2 可以看出,与对比模型中修复效果最好的模型相比,本文模型修复出的图像的 R_{PSN} 提升了 1.178 dB,I_{SSM} 提升了 0.010 2,L_2 降低了 0.000 9,在模拟真实壁画破损场景后,本文模型仍具有有效性,修复出的图像也更接近原图。

相较于 4 种对比模型,本文模型在人为添加随机不规则掩码及模拟霉斑掩码修复实验中能够获得更清晰、更连贯的结构信息以及更合理、更丰富的纹理细节信息。



(a) Mask (b) Shift-Net (c) CSWT (d) RFR (e) LGNet (f) 本文模型 (g) 原始图像

图 9 不同模型在霉斑掩码下的修复实验对比

Fig. 9 Comparison of inpainting experiments of different models under mildew masks

表 2 霉斑掩码下敦煌数据集定量结果

Tab. 2 Quantitative results of Dunhuang dataset under mildew masks

模型	R_{PSN}/dB	I_{SSM}	L_2
Shift-Net	29.355	0.952 1	0.021 1
CSWT	31.424	0.964 4	0.014 7
RFR	30.569	0.955 9	0.015 9
LGNet	34.601	0.964 3	0.022 1
本文模型	**35.779**	**0.974 6**	**0.013 8**

2.6 真实破损壁画图像修复实验

为进一步验证本文模型对真实破损壁画图像修复结果的有效性,选取真实破损壁画图像进行修复实验,结果如图 10 所示。由图 10 可以看出:Shift-Net 模型不能理解像素间的语义关系,修复不彻底,如未能对佛像衣服边缘划痕处作出良好的判断,也未能作出有效的修复;CSWT 模型修复的结果存在

残留的阴影，如图 10(d)中第 1 行图像，模型可以对残缺位置进行填充，但其未能建立起像素之间的连贯性；RFR 模型修复的结果出现了小部分伪影及像素模糊的问题，如第 2 行佛像，模型在进行残缺位置修复后出现了局部特征模糊的问题；LGNet 模型能够相对较好地重构出颜色及纹理信息，但仍存在白色掩码的残影；相对来说，本文模型修复结构较为流畅，能够有效克服重构结构不合理以及颜色杂乱的问题，其更加注重纹理细节的修复，在视觉上也更具流畅感且修复的完整度也较高。

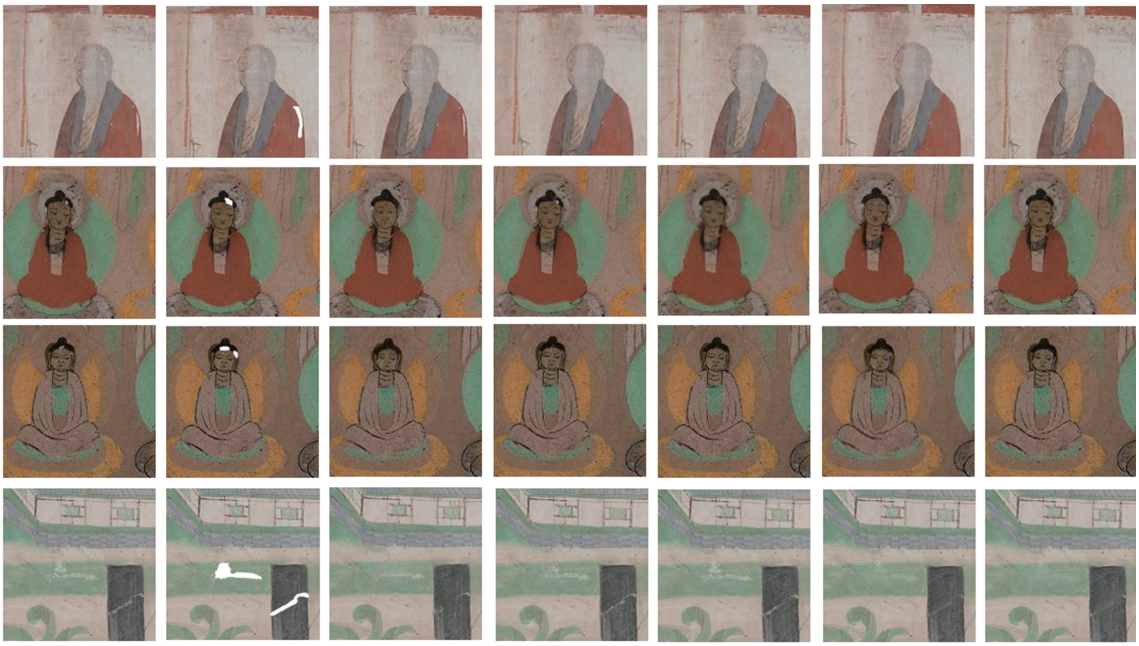


图 10　不同模型在真实破损壁画图像下的修复对比

Fig. 10　Comparison of inpainting of realistic damaged mural images

2.7　消融实验

为验证本文构建的模型对破损壁画图像修复的逐级提升效果，对测试集进行逐模块消融实验，定性分析结果如图 11 所示。其中，图 11(c)为加入迭代 LRM 模块的实验结果，图 11(d)为加入 LRM 模块与结合 DSLT-A 和 FAPM 模块的 Transformer 实验结果，图 11(e)为加入 LRM 模块、Transformer 模块与 FAPM 模块(本文模型)的实验结果。消融实验定量结果见表 3。

由图 11 可以看出：只加入 LRM 模块的修复结果可以大体修复出图像的整体结构，但无法构建出局部的结构轮廓，并且无法做到局部颜色分离；加入 LRM 模块与结合 DSLT-A 和 FAPM 模块的 Transformer 模块的修复结果能够对局部结构轮廓做出合理重建，但无法填充更多的细节信息；本文所提整体模型对壁画图像局部结构的合理性重建以及局部特征的纹理细化均具有有效性，修复的清晰度显著提升，修复后的图像更接近原图。由表 3 可以看出，随着模型深度的递增，修复后图像的客观评价指标也逐渐变优。

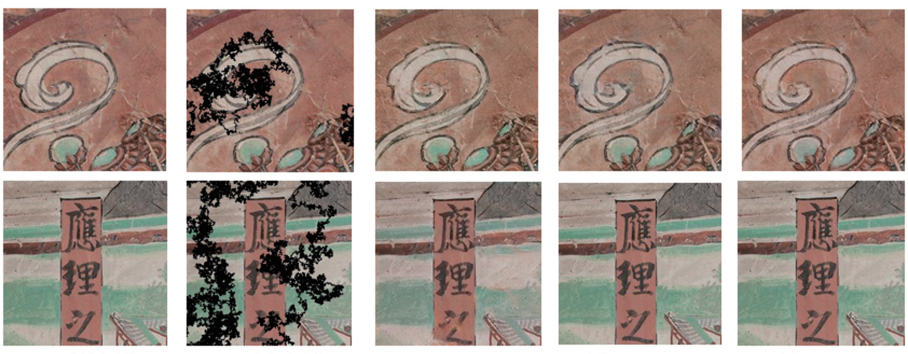


图 11　消融实验定性结果

Fig. 11　Qualitative results of the ablation experiment

表 3 消融实验定量结果

Tab. 3 Quantitative results of the ablation experiment

模块	R_{PSN}/dB	I_{SSM}	L_2
+ LRM	32.421	0.957 8	0.014 8
+ LRM + Transformer	33.672	0.963 7	0.014 2
+ LRM + Transformer + FAPM	34.989	0.970 1	0.013 9

3 结 论

1)提出了一种级联迭代精细化修复网络模型,以应对图像修复方法在修复破损壁画图像时存在的结构不合理及特征纹理细节不够细化的问题。

2)通过融合 LRM 模块,将 DSLT-A、FAPM 与 Transformer 架构进行整合,对破损图像作出先修复,并加深对图像特征的深度提取,有利于实现壁画图像的精细修复并增强修复的一致性。

3)在修复过程中,部分卷积多尺度门控模块的应用有效避免了原始图像中重要部分特征的丢失。

4)在敦煌壁画数据集上的实验结果表明,本文所提模型在面对人为掩码修复以及不规则掩码修复任务时,能够精准完成修复工作。与 4 种对比模型中各项指标表现结果最好的数据相比,本文模型修复后的图像的 R_{PSN} 提升了 1.178 dB,I_{SSM} 提升了 0.010 2,L_2 降低了 0.000 9。从主、客观评价指标来看,本文模型均优于其他对比模型,能够获得更具完整性和清晰化的修复图像。

参考文献

[1] Wang Huan, Li Qingquan, Jia Sen. A global and local feature weighted method for ancient murals inpainting[J]. International Journal of Machine Learning and Cybernetics, 2020, 11(6): 1199. DOI:10.1007/s13042-019-01032-2

[2] 罗海银, 郑钰辉. 图像修复方法研究综述[J]. 计算机科学与探索, 2022, 16(10): 2193
Luo Haiyin, Zheng Yuhui. Survey of research on image inpainting methods [J]. Journal of Frontiers of Computer Science and Technology, 2022, 16(10): 2193. DOI:10.3778/j.issn.1673-9418.2204101

[3] Bertalmio M, Sapiro G, Caselles V, et al. Image inpainting[C]// Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques. New Orleans: ACM, 2000: 418. DOI: 10.1145/344779.344972

[4] Chan T F, Shen Jianhong. Nontexture inpainting by curvature-driven diffusions [J]. Journal of Visual Communication and Image Representation, 2001, 12(4): 438. DOI:10.1006/jvci.2001.0487

[5] Bertalmio M, Vese L, Sapiro G, et al. Simultaneous structure and texture image inpainting [J]. IEEE Transactions on Image Processing, 2003, 12(8): 885. DOI:10.1109/TIP.2003.815261

[6] Efros A A, Leung T K. Texture synthesis by non-parametric sampling[C]//Proceedings of the 7th IEEE International Conference on Computer Vision. Kerkyra: IEEE, 1999: 1035. DOI:10.1109/ICCV.1999.790383

[7] Yang Jie, Qi Zhiquan, Shi Yong. Learning to incorporate structure knowledge for image inpainting[C]//Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York: AAAI, 2020: 12610

[8] Zeng Yanhong, Fu Jianlong, Chao Hongyang, et al. Learning pyramid-context encoder network for high-quality image inpainting [C]//Proceedings of the 32nd IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 1488. DOI: 10.1109/CVPR.2019.00158

[9] Wang Chao, Shao Mingwen, Meng Deyu, et al. Dual-pyramidal image inpainting with dynamic normalization[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(9): 5979. DOI: 10.1109/TCSVT.2022.3165587

[10] Chen Min, Zang Shengrui, Ai Zhenghua, et al. RFA-Net: residual feature attention network for fine-grained image inpainting[J]. Engineering Applications of Artificial Intelligence, 2023, 119: 105814. DOI:10.1016/j.engappai.2022.105814

[11] Xu Zishan, Zhang Xiaofeng, Chen Wei, et al. A review of image inpainting methods based on deep learning[J]. Applied Sciences-Basel, 2023, 13(20): 11189. DOI:10.3390/app132011189

[12] Zhan Xiaobo, Zhai Donghai, Li Tianrui, et al. Image inpainting based on deep learning: a review[J]. Information Fusion, 2023, 90: 74. DOI:10.1016/J.INFFUS.2022.08.033

[13] Quan Weize, Zhang Ruisong, Zhang Yong, et al. Image inpainting with local and global refinement[J]. IEEE Transactions on Image Processing, 2022, 31(20): 2405. DOI: 10.1109/TIP.2022.3152624

[14] Chen Yuantao, Liu Linwu, Tao Jiajun, et al. The improved image inpainting algorithm via encoder and similarity constraint [J]. Visual Computer, 2021, 37(7): 1691. DOI:10.1007/s00371-020-01932-3

[15] 邬开俊, 单宏全, 田彬, 等. 多阶段堆叠融合下的敦煌壁画修复[J]. 电子科技大学学报, 2024, 53(4): 551
Wu Kaijun, Shan Hongquan, Tian Bin, et al. Inpainting of Dunhuang murals under multi-stage stacking and fusion[J]. Journal of University of Electronic Science and Technology of China, 2024, 53(4): 551. DOI:10.12178/1001-0548.2023173

[16] 徐志刚, 杨欣宇. 结合 CSWin-Transformer 和门卷积的壁画图像修复方法[J]. 计算机工程与应用, 2024, 60(21): 215
Xu Zhigang, Yang Xinyu. Mural image inpainting method based on CSWin-Transformer and gate convolution [J]. Computer Engineering and Applications, 2024, 60(21): 215. DOI: 10.3778/j.issn.1002-8331.2307-0238

[17] Zhao Lei, Mo Qihang, Lin Sihuan, et al. UCTGAN: diverse image inpainting based on unsupervised cross-space translation[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 5744. DOI: 10.1109/CVPR42600.2020.00578

[18] Jam J, Kendrick C, Walker K, et al. A comprehensive review of past and present image inpainting methods[J]. Computer Vision and Image Understanding, 2021, 203: 103147. DOI:10.1016/j.cviu.2020.103147

[19] Zeng Yanhong, Fu Jianlong, Chao Hongyang, et al. Aggregated contextual transformations for high-resolution image inpainting[J]. IEEE Transactions on Visualization and Computer Graphics, 2023, 29(7): 3266. DOI:10.1109/TVCG.2022.3156949

[20] Long Chaojun, Li Xinghua, Jing Yinghong, et al. Bishift networks for thick cloud removal with multitemporal remote sensing images [J]. International Journal of Intelligent Systems, 2023, 2023: 9953198. DOI:10.1155/2023/9953198

[21] 吕建峰, 邵立珍, 雷雪梅. 基于深度神经网络的图像修复算法综述[J]. 计算机工程与应用, 2023, 59(20): 1
Lü Jianfeng, Shao Lizhen, Lei Xuemei. Image inpainting algorithm based on deep neural networks [J]. Computer Engineering and Applications, 2023, 59(20): 1. DOI:10.3778/j.issn.1002-8331.2303-0111

[22] Deng Ye, Hui Siqi, Zhou Sanping, et al. T-former: an efficient transformer for image inpainting[C]//Proceedings of the 30th ACM International Conference on Multimedia. Lisbon: ACM, 2022: 6560. DOI: 10.1145/3503161.3548446

[23] Yan Zaoyi, Li Xiaoming, Li Mu, et al. Shift-Net: image inpainting via deep feature rearrangement [C]//Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018: 8. DOI:10.1007/978-3-030-01264-9_1

[24] Chen B W, Liu T J, Liu K H. Lightweight image inpainting by stripe window transformer with joint attention to CNN[C]//33rd IEEE International Workshop on Machine Learning for Signal Processing. Rome: IEEE, 2023: 3. DOI: 10.1109/MLSP55844.2023.10285983

[25] Li Jingyuan, Wang Ning, Zhang Lefei, et al. Recurrent feature reasoning for image inpainting[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE 2020: 7763. DOI:10.1109/CVPR 42600.2020.00778