文章编号:1674-2974(2025)04-0001-15 DOI:10.16339/j.cnki.hdxbzkb.2025261

双向自回归Transformer与快速傅里叶卷积增强的壁画修复

陈永[1,2†],张世龙[1],杜婉君[1]
(1. 兰州交通大学 电子与信息工程学院,甘肃 兰州 730070;
2. 甘肃省人工智能与图形图像处理工程研究中心,甘肃 兰州 730070)

摘 要:针对现有深度学习算法在壁画修复时,存在全局语义一致性约束不足及局部特征提取不充分,导致修复后的壁画易出现边界效应和细节模糊等问题,提出一种双向自回归Transformer与快速傅里叶卷积增强的壁画修复方法.首先,设计基于Transformer结构的全局语义特征修复模块,利用双向自回归机制与掩码语言模型(masked language modeling, MLM),提出改进的多头注意力全局语义壁画修复模块,提高对全局语义特征的修复能力.然后,构建了由门控卷积和残差模块组成的全局语义增强模块,增强全局语义特征一致性约束.最后,设计局部细节修复模块,采用大核注意力机制(large kernel attention, LKA)与快速傅里叶卷积提高细节特征的捕获能力,同时减少局部细节信息的丢失,提升修复壁画局部和整体特征的一致性.通过对敦煌壁画数字化修复实验,结果表明,所提算法修复性能更优,客观评价指标均优于比较算法.

关键词:壁画修复;双向自回归Transformer;掩码语言模型;快速傅里叶卷积;语义增强
中图分类号:TP391.41 **文献标志码**:A

Bidirectional Autoregressive Transformer and Fast Fourier Convolution Enhanced Mural Inpainting

CHEN Yong[1,2†], ZHANG Shilong[1], DU Wanjun[1]
(1. School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China;
2. Gansu Provincial Engineering Research Center for Artificial Intelligence and Graphics & Image Processing, Lanzhou730070, China)

Abstract: Aiming at the lack of global semantic consistency constraints and insufficient acquisition of local features of the current deep learning algorithms in the process of image restoration of broken murals, resulting in the restored murals being prone to boundary effects and blurring of details, this paper proposes a bidirectional autoregressive Transformer with fast Fourier convolutional enhancement of murals restoration method. First, a global semantic feature repair module based on the Transformer structure is designed, and an improved multi-head

* **收稿日期**:2024-04-20
基金项目:教育部人文社会科学研究青年基金资助项目(19YJC760012), Humanities and Social Sciences Youth Foundation of Ministry of Education (19YJC760012);兰州交通大学基础研究拔尖人才项目(2023JC36), Lanzhou Jiaotong University Basic Top-Notch Personnel Project (2023JC36);兰州交通大学重点研发项目(ZDYF2304), Key Research and Development Project of Lanzhou Jiaotong University (ZDYF2304)
作者简介:陈永(1979—),男,甘肃武威人,兰州交通大学教授,博士
† 通信联系人,E-mail:edukeylab@126.com

attention global semantic mural repair module is proposed using the bidirectional autoregressive mechanism with masked language modeling (MLM) to improve the repair capability of global semantic features. Then, a global semantic enhancement module consisting of gated convolution and a residual module is constructed to enhance the global semantic consistency constraint. Finally, the local detail repair module is designed, which adopts large kernel attention (LKA) and fast Fourier convolution (FFC) to improve the ability of capturing detailed features while reducing the loss of local detail information, so as to enhance the consistency of the local and overall features of the repaired murals. The experimental results of the digital restoration of real Dunhuang murals show that the proposed algorithm can effectively restore the structure and texture of the murals, and the subjective visual effect and objective evaluation indexes are better than the comparative algorithms.

Key words: mural inpainting; bidirectional autoregressive Transformer; masked language modeling; fast Fourier convolution; semantic enhancement

敦煌莫高窟内珍藏着数量众多、内容丰富的石窟壁画资源,其以卓越的历史价值和艺术魅力而享誉世界.然而,受地理、气候环境和人为因素等多重影响,莫高窟内的壁画普遍受到龟裂、褪色等病害的侵袭,亟待保护[1].数字化修复由于可以更好地保护古壁画文化遗产,已成为当前的研究热点[2].

目前数字化壁画修复方法主要分为传统方法和深度学习方法,其中传统方法主要包括:①像素扩散[3-5]、②块匹配[6-7]和③稀疏表示[8-11]等方法.传统方法主要适合于修复小范围破损壁画,如裂缝及小面积脱落等.其中基于像素扩散的方法,基本思想是通过迭代方式将破损区域周围的完好区域的像素值扩散到待修复区域,逐步扩散完成破损区域修复.如文献[3]提出了一种扩散修复方法,利用扩散操作来平滑图像的损坏部分.文献[4]提出了改进曲率驱动模型的敦煌壁画修复算法,通过引入自适应控制策略,使得壁画破损区域曲率不同时能够选择不同的修复模型进行动态自适应控制修复,最后利用光滑函数进行角点扩散完成壁画裂缝的修复.文献[5]提出了基于Shannon-Cosine小波精细积分法的壁画修复方法,运用小波数值方法对偏微分方程PDE进行离散处理,完成了对壁画图像的降噪修复.上述像素扩散类修复方法一般利用梯度或曲率完成像素扩散修复,其修复能力有限,无法完成较大破损区域修复.

第二种是基于块匹配的传统修复方法,该类方法主要通过在壁画完好区域搜索匹配块,然后采用平移复制的方式完成破损区域的修复.如文献[6]提出了一种改进的Criminisi图像修复算法,采用优先级函数选择非零元素最少的块进行重构,以及通过查找表进行匹配搜索来提高修复性能.文献[7]提出了改进块匹配的古壁画修复算法,通过引入分散度项来降低块匹配的错误填充率.但块匹配类方法因仅采用块复制粘贴的方式来填充壁画破损区域,易造成修复后壁画结构信息的错位.

第三种是基于稀疏表示的传统修复方法,该类方法利用图像信息稀疏的特点,借助过完备变换的稀疏组合来表示图像信号,然后通过信号重构的方式实现图像缺失部分的恢复[8-9].如文献[10]提出了一种结合全局一致性和局部连续性的壁画修复方法,通过构建具有全局过完备相似块字典来提高壁画的修复性能.文献[11]提出了一种改进多重字典联合自适应学习的稀疏敦煌壁画修复方法,采用非下采样剪切波分解的方法,通过构建多重稀疏子类字典,完成对壁画低频纹理的修复.但稀疏表示类修复方法在修复时,一般在字典系数求解后采用插值的方法进行恢复,插值操作对于较大面积破损修复效果欠佳,无法完成有效修复.

随着深度学习技术的发展,基于深度学习的壁画修复已经成为一种主流方法.深度学习图像修复方法通过深度神经网络模型来自动学习壁画图像特征以填补缺失部分,从而获得视觉上更加逼真的修复图像.目前,国内外学者开展了大量的研究工作,根据深度学习网络架构的不同,主要深度学习修复方法包括:①基于自编码器修复方法[12-18]、②基于生成对抗网络GAN的修复方法[19-24]和③基于Transformer架构修复方法[25-29].

其中,深度学习第一种基于自编码器修复方法,

其结构一般由编码器、解码器组成,主要利用编码器提取壁画图像特征,将其压缩到低维特征空间学习其语义特征,最终通过解码器将特征进行重构并输出修复结果.如文献[12]提出了一种基于混合自编码模型的壁画修复方法,首先将壁画分割成规则大小的图像块,然后采用CNN-VIT网络进行修复,提升对细节的修复能力.文献[13]提出一种渐进式多级特征古画修复算法,设计特征提取网络对不同尺度特征进行提取,并通过解码网络还原特征并修复,但渐进式修复易造成误差累积影响修复结果.文献[14]提出了一种基于超图卷积的双阶段修复模型,利用超图卷积增强特征捕获能力,但该方法未考虑全局语义特征一致性约束,修复结果易出现边界效应与内容模糊等现象.文献[15]提出了一种双流图像修复网络,以耦合的方式增强纹理与结构的一致性,但其采用普通卷积进行特征提取,存在特征提取不充分的问题.文献[16]提出了多尺度图像上下文注意力学习的修复模型,通过设计多尺度注意力模块计算缺失区域与已知区域的余弦相似度并进行注意力传播,并且采用像素卷积将特征图压缩到和输入相同的尺寸完成破损区域修复,但该方法缺乏对整体语义的感知,导致修复结果易出现纹理模糊现象.文献[17]提出了一种基于注意力机制学习的知识辅助修复模型,但是该方法仅将局部信息作为约束条件,缺乏全局语义一致性约束,导致修复结果出现局部细节丢失的问题.文献[18]提出多模态引导图像修复的网络模型,通过预测辅助先验边缘和轮廓,以多阶段的方式辅助修复图像,但该方法在多模态混合时,易引入混合误差,修复易出现细节丢失的问题.上述基于自编码器修复方法在特征处理过程中,一般因缺乏全局语义约束,且存在频繁的编解码操作,在修复过程中易造成特征信息丢失的问题.

深度学习第二种基于生成对抗网络GAN的图像修复方法,主要通过生成器学习原始壁画图像的先验分布,再通过鉴别器和生成器的博弈对抗来完成修复.如文献[19]提出了多阶段渐进推理壁画修复方法,设计多阶段渐进推理网络修复破损边界并逐渐增强区域纹理约束,采用多尺度特征聚合模块融合不同尺度壁画特征,提升对壁画语义的修复能力.文献[20]提出了一种两阶段边缘连接引导的修复方法,首先对缺失区域边缘信息进行预测,然后将其作为先验填充缺失区域,该方法能较好地拟合壁画边缘结构,但修复过程中因缺乏语义一致性约束,修复结果易出现纹理模糊的问题.文献[21]提出了结构引导渐进式壁画修复方法,通过结构重建网络修复壁画受损结构后引导完成纹理恢复,该方法提升了修复结果的整体一致性,但易产生累积误差,修复结果存在边界效应问题.文献[22]提出了一种循环特征推理的图像修复方法,通过区域识别模块识别推断区域并设计循环特征推理模块实现破损区域推理,但是该过程缺乏对像素信息的有效预测和控制,易引入无效像素参与修复过程,导致出现修复模糊的问题.文献[23]提出了基于多路编码器和双重注意力的古壁画修复算法,设计多路编码器来学习古壁画不同尺度的语义特征,并通过解码器完成修复.文献[24]提出一个辅助上下文重建修复网络,通过学习完好区域和破损区域之间图像块的相似性并对破坏区域进行替换修复,但在该过程中由于未对图像全局特征进行学习,导致修复结果出现边界效应的问题.上述基于生成对抗网络GAN修复方法,生成器通过学习壁画图像的先验分布后与鉴别器博弈对抗完成修复,但生成器因缺乏有效的语义引导,修复结果易出现细节丢失问题.

深度学习第三种基于Transformer架构修复方法,其主要原理是利用编码器对输入图像块嵌入向量进行处理,提取图像块之间关系和特征并生成高维特征表示,然后通过解码器生成修复后的图像.如文献[25]提出一种基于掩码感知的Transformer图像修复模型,通过掩码的动态更新,提高修复性能,但该方法只对完好区域的Token序列进行计算,忽略了图像局部与全局的一致性,其结果存在纹理细节丢失的问题.文献[26]设计了一种基于线性注意力的Transformer图像修复网络,但是线性注意力无法很好地捕捉图像全局上下文语义信息,修复后易出现纹理模糊的问题.文献[27]提出了一种基于生成对抗网络的高分辨率修复模型,设计堆叠的聚合上下文块来捕获特征信息,并通过多层转置卷积来解码得到修复图像,但该方法采用标准卷积进行编解码,由于标准卷积感受野受限导致其对全局特征存在提取不充分的问题.文献[28]提出了基于Transformer的修复方法,利用Transformer修复壁画图像的整体结构,而后采用上采样网络恢复缺失图像的纹理细节,但修复过程中上采样采用普通卷积对全局特征信息的捕获能力有限,导致修复结果易出现边界效

应.文献[29]提出了基于Transformer的图像修复方法,利用自编码器对输入图像进行分块编码并形成特征码本,并通过解码器得到修复图像,但该方法在细节修复过程中因缺乏全局语义约束,修复结果易出现纹理模糊现象.上述基于Transformer架构的方法虽然能够有效捕捉图像中长距离的依赖关系,但对于局部特征信息关注不足,易出现纹理模糊与细节丢失的问题.

综上所述,针对现有深度学习算法在壁画修复时,存在全局语义一致性约束不足及局部特征提取不充分,导致修复后的壁画易出现边界效应和细节模糊等问题,提出一种双向自回归Transformer与快速傅里叶卷积增强的壁画修复方法.主要工作包括:首先,设计基于Transformer结构的壁画全局语义特征修复模块,通过双向自回归多头注意力与掩码语言模型机制,对缺失像素进行推理,提升对壁画全局语义特征的修复能力.然后,构建由门控卷积和扩张卷积残差模块组成的全局语义增强模块,增强修复过程中对于壁画全局语义的一致性约束.最后,使用大核注意力机制与快速傅里叶卷积构建局部细节修复模块,通过大核注意力扩大感受野范围学习局部特征表示,并利用快速傅里叶卷积进行细节修复,再通过上采样阶段输出修复后的壁画图像.

1 本文算法

1.1 网络整体结构

壁画图像通常具有复杂的全局语义与局部细节信息,大多数深度学习模型采用卷积神经网络(CNN)提取局部特征,但是难以捕捉壁画全局特征表示.原始Transformer深度学习模型具有较强的全局特征信息提取能力,但其对局部特征学习有限,易造成局部细节丢失[30].因此,为了更好地实现对破损壁画的修复,本文在原始Transformer深度学习模型的基础上,提出了一种双向自回归Transformer与快速傅里叶卷积增强的壁画修复模型,其主要由①全局语义特征修复模块、②全局语义增强模块及③局部细节修复模块三部分构成,整体框架如图1所示.

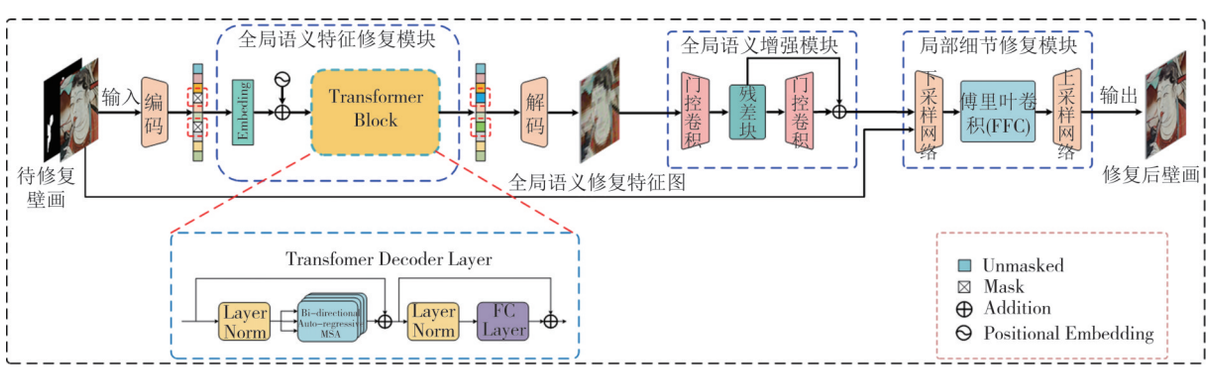


图1 整体框架示意图

Fig.1 Diagram of the overall framework of the model

模型工作时,首先利用编码模块将待修复壁画编码为像素序列,其次通过本文提出的全局语义特征修复模块,提高对全局语义特征的修复能力.然后,进一步设计全局语义增强网络,使用残差块和门控卷积模块提取不同尺度的全局语义特征,增强壁画修复过程中全局语义一致性约束.最后,使用大核注意力机制与快速傅里叶卷积完成局部细节修复,并通过上采样得到修复后的壁画.

1.2 全局语义特征修复模块

在壁画修复过程中,由于普通卷积操作是一种基于局部区域的操作,仅具有局部相关性,其全局特征捕获能力较弱[31],导致壁画修复时易出现边界效应的问题.为了克服上述不足,本文设计了全局语义特征修复模块,利用Transformer结构及双向自回归[32]多头注意力机制建立了壁画像素序列Token之间的依赖关系,并采用掩码语言模型(masked language modeling, MLM)[33]实现对壁画缺失像素的推理,然后通过解码像素序列得到全局语义修复特征图.通过全局语义特征修复模块,提高对全局语义特征的修复能力.

在全局语义特征修复模块中,设计了一种双向自回归机制改进的Transformer全局修复模型,利用其建立序列中像素Token之间的依赖关系[34],并对壁画的全局上下文信息进行特征学习.

在建立壁画像素Token之间的依赖关系时,首先将输入的壁画图像Token序列化为$x=(x_1,x_2,\cdots,x_T)$,

其中T表示序列长度.通过序列化可以将Token像素x_t与前序Token像素$x_{<t}$和后序Token像素$x_{>t}$建立两个方向的依赖关系，从而通过捕获双向的特征信息实现壁画图像的上下文依赖关系.壁画图像序列化后缺失Token像素x_t两个方向的概率分布分别为：$\Pi_{\phi}^{\mathrm{pre}}\left(x_t|x_{<t}\right)$和$\Pi_{\phi}^{\mathrm{pre}}\left(x_t|x_{>t}\right)$.接着，将上述两个概率分布以自回归的方式连接起来，作为缺失像素Token的推理条件，其过程可以用如下公式表示：

$$J^{\mathrm{pre}}(\phi)=E_{x\sim D^{\mathrm{pre}}}\left[\sum_{t=1}^{T}\log\Pi_{\phi}^{\mathrm{pre}}\left(x_t|x_{<t}\right)+\sum_{t=1}^{T}\log\Pi_{\phi}^{\mathrm{pre}}\left(x_t|x_{>t}\right)\right] \tag{1}$$

式中：$J^{\mathrm{pre}}(\cdot)$表示整个序列的概率分布；D^{pre}表示预训练数据的分布；E表示整个序列x从D^{pre}进行采样的期望值；ϕ表示训练过程中模型学习到的参数；t表示Token序号数；x_t表示需要建立依赖关系像素Token；$x_{>t}$表示序列中第t个Token之后的特征信息；$x_{<t}$表示序列中第t个Token之前的特征信息.

由于普通注意力在捕获壁画语义特征信息过程中，仅关注缺失像素Token之前的完好像素特征信息，如图2(a)中完好壁画像素点1、2、3和4的先验信息，忽略了缺失像素Token的后序完好像素5、6和7的特征信息，导致缺失像素Token推理时局部先验信息未能充分利用，易出现推理误差的问题.为了提高对于缺失像素Token的推理性能，本文采用双向注意力进行缺失像素推理.双向注意力与普通注意力不同，如图2(b)所示，双向注意力从两个方向进行注意力计算，使模型不仅考虑到壁画缺失像素Token之前存在的语义特征信息，同时也捕获缺失像素Token之后的特征信息，壁画修复时通过双向注意力可以提升完好区域与受损区域之间的语义一致性.

在图2所示双向注意力过程中，计算如下：

$$F^n=E^n+\mathrm{MSA}\left(\mathrm{LN}\left(E^n\right)\right) \tag{2}$$

$$E^{n+1}=F^n+\mathrm{MLP}\left(\mathrm{LN}\left(F^n\right)\right) \tag{3}$$

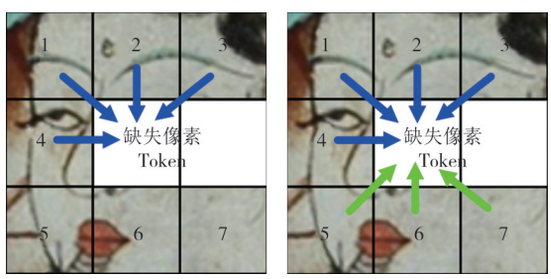
式中：F^n表示第n层多头自注意力层的输出；E^{n+1}表示第n层全连接层的输出；n表示Transformer结构的层数；MSA为多头自注意力层；LN为归一化层；MLP为全连接层.其中，MH计算过程如下：

$$\mathrm{Attention}\left(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}\right)=\mathrm{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{4}$$

$$\mathrm{head}=\mathrm{Attention}\left(\boldsymbol{Q}W_I^Q,\boldsymbol{K}W_I^K,\boldsymbol{K}W_I^V\right)K \tag{5}$$

$$\mathrm{MSA}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\mathrm{Concat}\left(\mathrm{head}_1,\cdots,\mathrm{head}_h\right)W^O \tag{6}$$

式中：Attention表示自注意力分数；head表示独立计算每个头部的注意力分数；Concat表示对不同注意力头之间分数进行合并；h代表注意力头的个数；d_k表示$\boldsymbol{K}$的特征维度；$\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}$是三个可学习的参数矩阵；$W^O$是一个可学习的全连接层，其主要作用是融合来自不同head的输出.



(a)单向注意力　　(b)双向注意力

图2　双向与普通注意力机制

Fig.2　Bidirectional and general attention mechanisms

在建立完像素Token之间依赖关系后，进一步通过掩码语言模型MLM实现破损像素Token的推理.掩码语言模型MLM是一种自然语言处理预测模型，其主要特点是通过捕获完好部分的先验信息来预测缺失的信息[35].本文利用掩码语言模型MLM上述特点，通过学习完好区域壁画特征信息，实现对缺失像素的预测，其过程如图3所示.对于已经序列化的壁画Token序列，利用MLM预测出破损壁画像素Token. MLM预测是指将破损像素$\tilde{x}$的输出概率分布与完好像素的概率分布相接近，实现缺失像素的补全.推理过程可以被表示为如下公式：

$$\log P\left(\tilde{x}|\hat{x};\ \theta\right)=E_x\sum_{\tilde{x}_i\in\tilde{x}}\log P\left(x|\hat{x};\ \theta\right) \tag{7}$$

式中：P表示对破损像素Token推理的概率分布；E_x表示期望值；x表示壁画像素Token序列；$\hat{x}$表示确定破损像素Token的序列；$\tilde{x}$表示破损像素对应的Token；i表示序列中Token序号数；θ表示模型训练过程学习到的参数.

1.3　全局语义增强模块

在完成上述全局语义特征修复后，本文进一步设计了全局语义增强模块来提高全局语义一致性.全局语义增强模块由门控卷积[36](gated convolution，GC)和扩张卷积残差块[37]构成的编解码器构成，结构如图4所示.

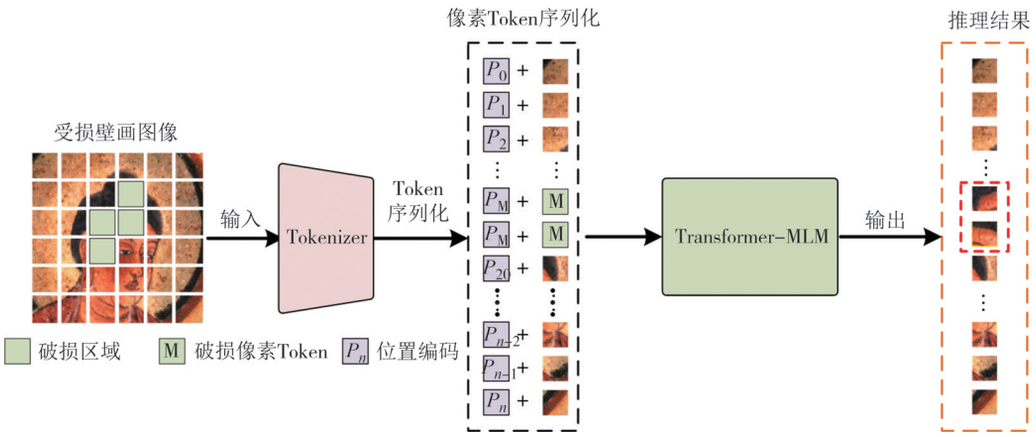


图3 MLM建模推理过程

Fig.3 Masked language modeling inference process

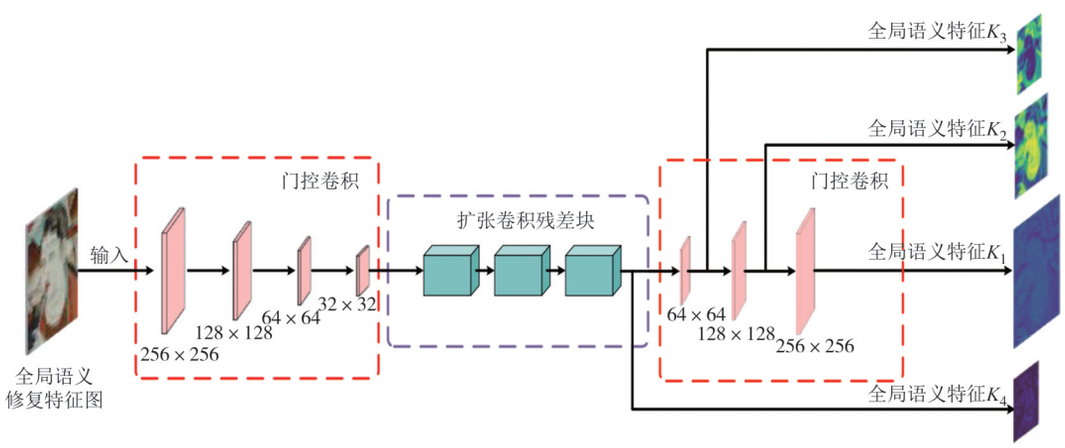


图4 全局语义增强模块结构图

Fig.4 Global semantic enhancement module structure diagram

图4所示全局语义增强模块中,设计了门控卷积语义增强的编码器与解码器结构,实现对全局语义修复结果进行特征提取.门控卷积结合了卷积操作和门控单元,可捕捉壁画图像中不同尺度和位置的上下文信息,获取壁画图像浅层特征,其过程如下:

$$E=\phi\left(\sum\sum W\cdot O\right) \tag{8}$$

$$S=\sigma\left(\sum\sum W\cdot O\right) \tag{9}$$

$$G=(E)\odot(S) \tag{10}$$

式中:E为特征值;S为门控值;W为卷积滤波器;O表示全局语义修复特征图;$\phi(\cdot)$表示ReLU激活函数;$\sigma(\cdot)$为Sigmoid激活函数;G表示门控卷积得到的特征图;$\odot$为逐像素相乘操作.

然后,设计融合扩张卷积的残差块来扩展感受野范围,以便更好地获取壁画上下文信息.在该过程中,扩张卷积根据扩张因子r在水平与垂直两个方向分别选取壁画特征图F上的像素点进行卷积运算,其计算过程如下:

$$y(m,n)=\sum_{i=1}^{M}\sum_{j=1}^{N}F\left(m+r\times i,n+r\times j\right)w(i,j) \tag{11}$$

式中:y为输出特征;F为输入特征图;(m,n)为特征图F上点坐标;r为扩张因子;$w(i,j)$为滤波器;(M,N)表示滤波器大小.

接着将壁画图像进行解码操作,得到不同尺度的全局语义特征.

1.4 局部细节修复模块

在完成全局语义增强后,为了进一步提高壁画修复结果的完整性与细节协调性,减少修复后局部细节丢失,设计了局部细节修复模块,其结构如图5所示.该模块由基于大核注意力机制LKA的采样部分和快速傅里叶卷积FFC网络组成.

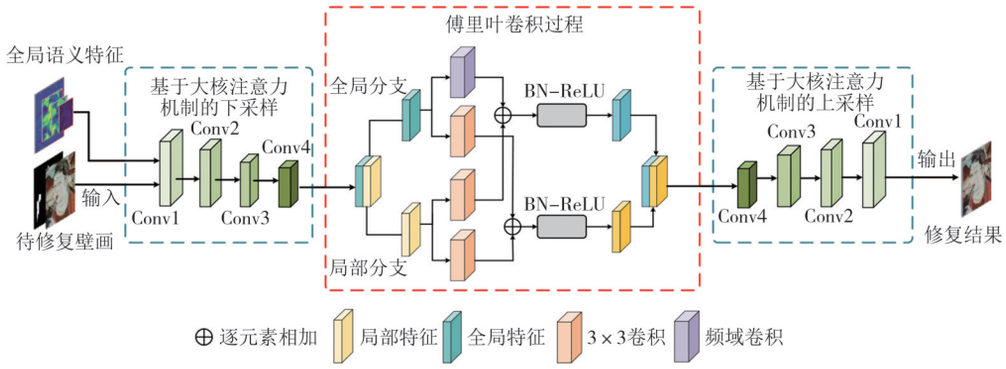


图5　局部细节修复模块图

Fig.5　Partial detail repair module diagram

1.4.1　大核注意力机制

图5所示局部细节修复模块中,引入了大核卷积注意力[38] LKA下采样实现对壁画细节特征的提取,其结构由归一化层(BatchNorm, BN)、注意力层(Attention)、前馈神经网络层(feedforward neural network ,FNN)三部分构成,如图6所示.

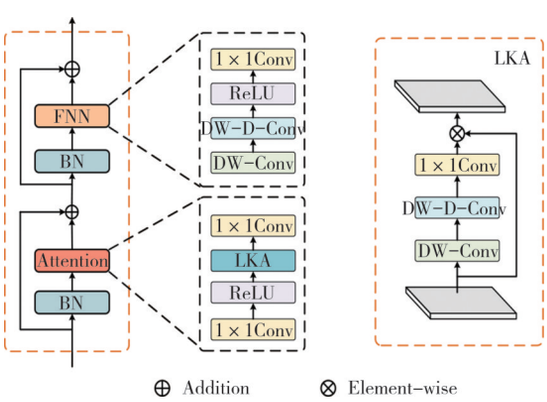


图6　大核注意力机制

Fig.6　Large kernel attention mechanism

大核卷积注意力机制LKA采用深度卷积(depth wise-Conv, DW-Conv)和扩张深度卷积(depth wise-dilated-Conv, DW-D-Conv)来获取壁画细节特征信息.扩张深度卷积输出后采用1×1卷积,然后将输出的注意力图与输入特征相乘,计算过程如下:

$$A_t=\mathrm{Conv}_{1\times1}\left(\mathrm{DW-D-Conv}\left(\mathrm{DW-Conv}\left(\mathrm{Input}\right)\right)\right) \tag{12}$$

$$\mathrm{Output}=A_t\otimes\mathrm{Input} \tag{13}$$

式中:Input为输入特征图;A_t为注意力值;DW－Conv为深度卷积操作;DW－D－Conv表示对特征图进行扩张深度卷积的操作;⊗为逐像素相乘.

1.4.2　快速傅里叶卷积网络

在完成下采样阶段的特征提取后,本文采用快速傅里叶卷积[39]对壁画局部细节进行修复.快速傅里叶卷积基于快速通道级傅里叶变换(fast Fourier transform, FFT),能够将壁画图像从空间域向频域变换,利用傅里叶卷积中通道维度上的频域转换可以处理细节纹理特征,从而实现对壁画频域特征的全局学习的目的,继而完成破损区域的细节修复.快速傅里叶卷积网络将经过下采样得到的壁画特征图按照其通道分为全局特征与局部特征两部分.在局部分支处理过程中,采用3×3卷积提取局部纹理特征,然后与经过全局分支中的3×3卷积处理的特征进行逐元素相加,接着通过归一化层与ReLU激活函数层后输出.快速傅里叶全局分支中频域卷积的具体过程如图7所示.

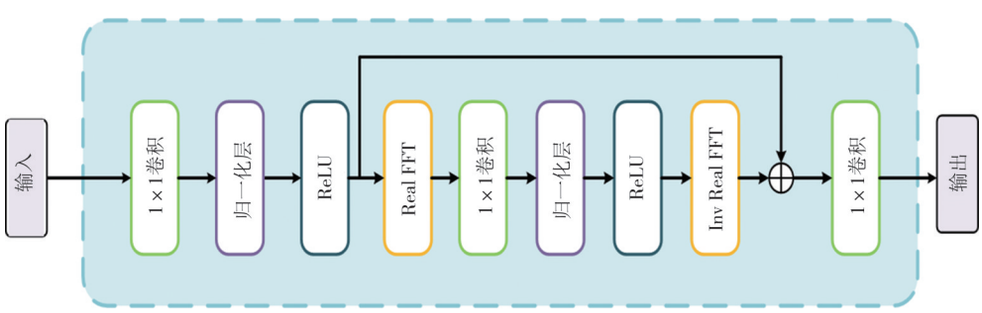


图7　积全局分支中频域卷积

Fig.7　Frequency domain convolution in global branches

在图7中,快速傅里叶卷积FFC对于壁画图像进行频域卷积处理,采用实数快速傅里叶卷积操作,计算过程如下:

$$\mathbb{R}^{H\times W\times C}\rightarrow\mathbb{C}^{H\times\frac{W}{2}\times C} \tag{14}$$

在式(14)的基础上,将张量的实部与虚部连接,该过程表示如下:

$$\mathbb{C}^{H\times\frac{W}{2}\times C}\rightarrow\mathbb{R}^{H\times\frac{W}{2}\times 2C} \tag{15}$$

接着,在频域使用ReLU激活函数,并通过归一化层和1×1卷积完成频域卷积操作,该过程表示如下:

$$\mathbb{R}^{H\times\frac{W}{2}\times 2C}\rightarrow\mathbb{R}^{H\times\frac{W}{2}\times 2C} \tag{16}$$

然后,应用逆变换恢复空间结构,该过程表示如下:

$$\mathbb{R}^{H\times\frac{W}{2}\times 2C}\rightarrow\mathbb{C}^{H\times\frac{W}{2}\times C} \tag{17}$$

$$\mathbb{C}^{H\times\frac{W}{2}\times C}\rightarrow\mathbb{R}^{H\times W\times C} \tag{18}$$

式中:$\mathbb{R}$表示张量的实部;$\mathbb{C}$表示张量的虚部;H为特征图的高度;W为宽度;C为通道数.

完成逆傅里叶变换后,将全局分支修复结果与局部分支中普通卷积后的结果融合,并通过归一化层与RelU激活函数层,最后利用大核卷积上采样后输出修复结果,从而完成破损壁画修复.

2 损失函数

本文损失函数由L_1损失、对抗损失和感知损失构成.其中L_1损失函数$\mathcal{L}_{L1}$仅在壁画图像的未破损区域进行计算,公式如下:

$$\mathcal{L}_{L1}=(1-M)\odot\left|\hat{I}-\tilde{I}\right|_1 \tag{19}$$

式中:M表示0-1掩码,其中1表示壁画破损区域即掩码区域,0表示壁画完好区域;$\odot$表示逐元素相乘;$\hat{I}$与$\tilde{I}$分别表示真实壁画图像与修复后的壁画图像.

对抗损失$\mathcal{L}_{\text{adv}}$的引入有助于生成器生成性能,提高生成壁画图像的真实性以及其结构与纹理的一致性.对抗损失$\mathcal{L}_{\text{adv}}$由鉴别器损失$\mathcal{L}_{\text{D}}$与生成器损失$\mathcal{L}_{\text{G}}$构成,其定义为:

$$\mathcal{L}_D=-\mathbb{E}_{\hat{I}}\left[\log D\left(\hat{I}\right)\right]-\mathbb{E}_{\tilde{I},M}\left[\log D\left(\tilde{I}\right)\odot(1-M)\right]-\mathbb{E}_{\tilde{I},M}\left[\log\left(1-D\left(\tilde{I}\right)\right)\odot M\right] \tag{20}$$

$$\mathcal{L}_G=-\mathbb{E}_{\tilde{I}}\left[\log D\left(\tilde{I}\right)\right] \tag{21}$$

$$\mathcal{L}_{\text{adv}}=\mathcal{L}_{\text{G}}+\mathcal{L}_{\text{D}}+\lambda_{\text{GP}}\mathcal{L}_{\text{GP}} \tag{22}$$

式中:$\mathcal{L}_{\text{GP}}=\mathbb{E}_{\hat{I}}\left\|\nabla_{\hat{I}}D\left(\hat{I}\right)\right\|^2$表示惩罚梯度;$\lambda_{\text{GP}}=10^{-3}$;$\odot$表示逐元素相乘.

感知损失$\mathcal{L}_{\text{HRFPL}}$通过基本预训练网络评估从修复的壁画图像中提取的特征与目标壁画图像之间的距离,通过感知损失来衡量两者之间的相似性,表示为:

$$\mathcal{L}_{\text{HRFPL}}=\mathbb{E}\left(\left[\phi_{\text{HRF}}\left(\hat{I}\right)-\phi_{\text{HRF}}\left(\tilde{I}\right)\right]^2\right) \tag{23}$$

式中:$[\cdot-\cdot]^2$表示逐元素运算;$\phi_{\text{HRF}}\left(\hat{I}\right)$与$\phi_{\text{HRF}}\left(\tilde{I}\right)$可以通过傅里叶卷积实现.

因此,本文模型的总损失函数计算公式可以表示为:

$$\mathcal{L}=\lambda_{L1}\mathcal{L}_{L1}+\lambda_{\text{adv}}\mathcal{L}_{\text{adv}}+\lambda_{\text{HRFPL}}\mathcal{L}_{\text{HRFPL}} \tag{24}$$

式中:$\mathcal{L}_{L1}$、$\mathcal{L}_{\text{adv}}$、$\mathcal{L}_{\text{HRFPL}}$分别为$L_1$损失、对抗损失和感知损失的对应权重.

3 实验结果与分析

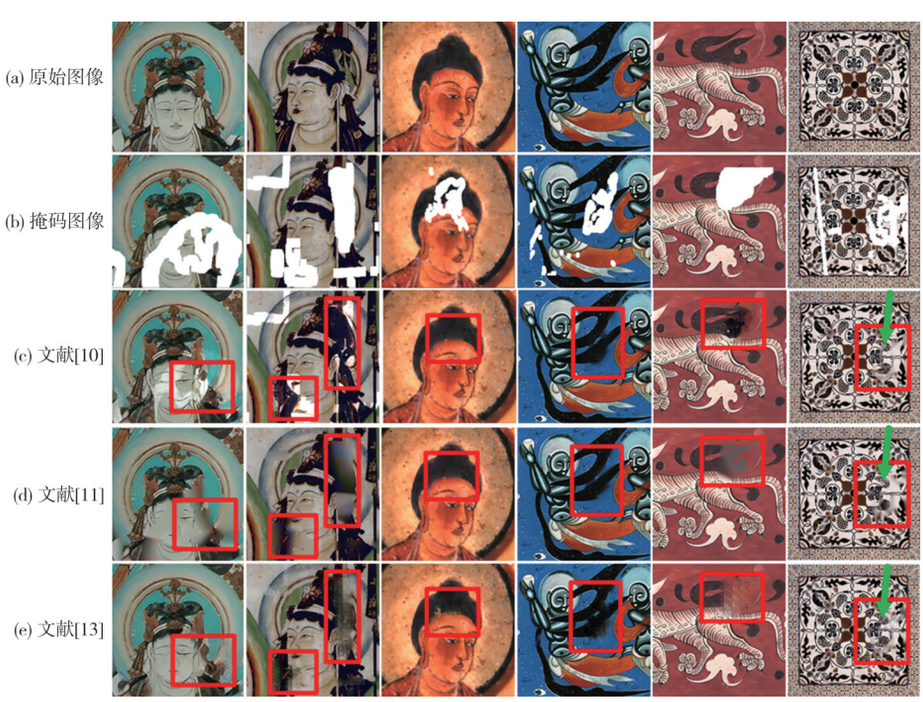
3.1 实验数据集及实验参数设置

为了验证所提方法的有效性,下面进行对比实验.采用自制敦煌壁画数据集,选取高清敦煌壁画图像作为数据集来源,并对其进行扩展后形成21 000张壁画数据集,其中训练数据集含壁画图像14 600张,测试数据集含壁画图像6 400张.同时与文献[10]全局一致性和局部连续性联合修复方法、文献[11]多重字典联合自适应修复方法、文献[13]渐进式修复方法、文献[14]超图卷积双阶段修复方法、文献[15]结构双生成方法、文献[20]边缘连接修复方法、文献[22]循环特征推理方法、文献[40]结构与纹理联合修复方法进行对比.客观评价指标采用峰值信噪比PSNR、结构相似性SSIM和梯度相似性GMSD.实验硬件环境为Intel i9-12900K CPU,24.0G RAM,NVIDIA RTX A5000,对比实验均在相同配置下进行.其中全局语义特征修复阶段采用AdamW[41]优化器,参数$\beta_1=0.9$,$\beta_2=0.95$,学习率为3×10^{-4}并在迭代过程中进行衰减,同时采用学习率分别为1×10^{-3}与1×10^{-4}的Adam优化器训练局部细节修复模块的生成器与鉴别器.

3.2 人为添加随机掩码修复实验

首先进行人为添加随机掩码修复实验,其修复结果如图8所示.其中,图8(a)为原始壁画,图8(b)为添加随机掩码图像.图8(c)为文献[10]结合全局一致性与局部连续性修复方法的结果,该方法有明

显修复残留，如第一幅与第二幅佛像出现白色残留的问题，这是因为该方法通过线性关系建立图像相似块集合字典，无法捕捉全局上下文特征信息，对于大面积破损修复效果较差．图8(d)为文献[11]多重字典联合自适应学习的稀疏壁画修复方法结果，可以看出该方法对于大面积破损修复出现了区域模糊现象，如第一幅面部出现了模糊现象，其原因在于稀疏表示方法对于大面积破损的壁画图像难以进行有效字典表示．图8(e)为文献[13]渐进式多级特征修复算法结果，可以看出该方法存在明显块效应问题，如第二幅图中佛像发髻部分出现了块效应，其原因是该方法通过堆叠多层普通卷积构成特征提取网络与解码网络，但是普通卷积对于全局特征提取能力较弱，易导致修复结果出现块效应现象．图8(f)为文献[14]超图卷积双阶段修复结果，可以看出该方法修复后存在较多残留，如第三幅和第四幅均出现了较多残留，导致佛像发髻与裙带部分语义丢失，其原因为该方法在修复过程中未考虑全局语义特征一致性，修复结果易出现边界效应与内容模糊等现象．图8(g)为文献[15]纹理结构双生成修复结果，可以看出该方法修复后存在细节丢失的问题，如第一幅菩萨图存在明显的鼻子等面部语义信息丢失的问题，主要是因为文献[15]采用普通卷积进行特征提取，存在特征提取不充分的问题，此外该方法在修复过程中会对纹理特征进行压缩，容易导致壁画修复结果出现内容模糊和细节丢失问题．图8(h)为文献[20]边缘连接修复方法的修复结果，可以看出该方法较其他对比方法较好，如第三幅佛像发髻部分拟合较好，但对于线条结构复杂壁画存在修复后结构扭曲的问题，这是因为该方法通过边缘预测生成来引导后续修复，当边缘无法有效预测时，易造成修复结果出现累计误差．图8(i)为文献[22]循环特征推理修复的结果，修复结果中明显存在区域模糊的问题，如第一幅和第五幅出现了明显的块效应现象，而第六幅出现了结构断裂的问题，其原因是该方法在特征提取时通过级联标准卷积层进行编解码，缺乏对像素信息的有效预测和控制，易引入无效像素参与修复过程，导致出现修复模糊的问题．图8(j)为文献[40]结构生成与纹理合成算法修复结果，可以看出修复效果较文献[13]与文献[14]有所提升，但修复结果仍存在纹理模糊的问题，其原因为该方法在生成阶段缺乏对局部信息的获取，导致修复结果出现局部模糊与块效应现象．图8(k)为本文算法修复结果，可以看出本文方法较对比方法修复效果更优，如第一、二幅佛像面部眼睛、鼻子等语义特征修复效果较好，第四、五幅壁画修复结果未出现结构断裂与语义丢失现象，对于第六幅结构与纹理复杂的壁画，结构修复连贯，纹理清晰．本文方法修复后细节更加完整且视觉效果更加连贯，这是因为采用本文提出的Transformer全局语义特征修复模块，利用双向自回归机制与掩码语言模型可以提高对全局语义特征的修复能力，以及通过大核注意力机制与快速傅里叶卷积减少了局部细节信息的丢失，提升修复壁画局部和整体特征的一致性，修复后整体效果更优．



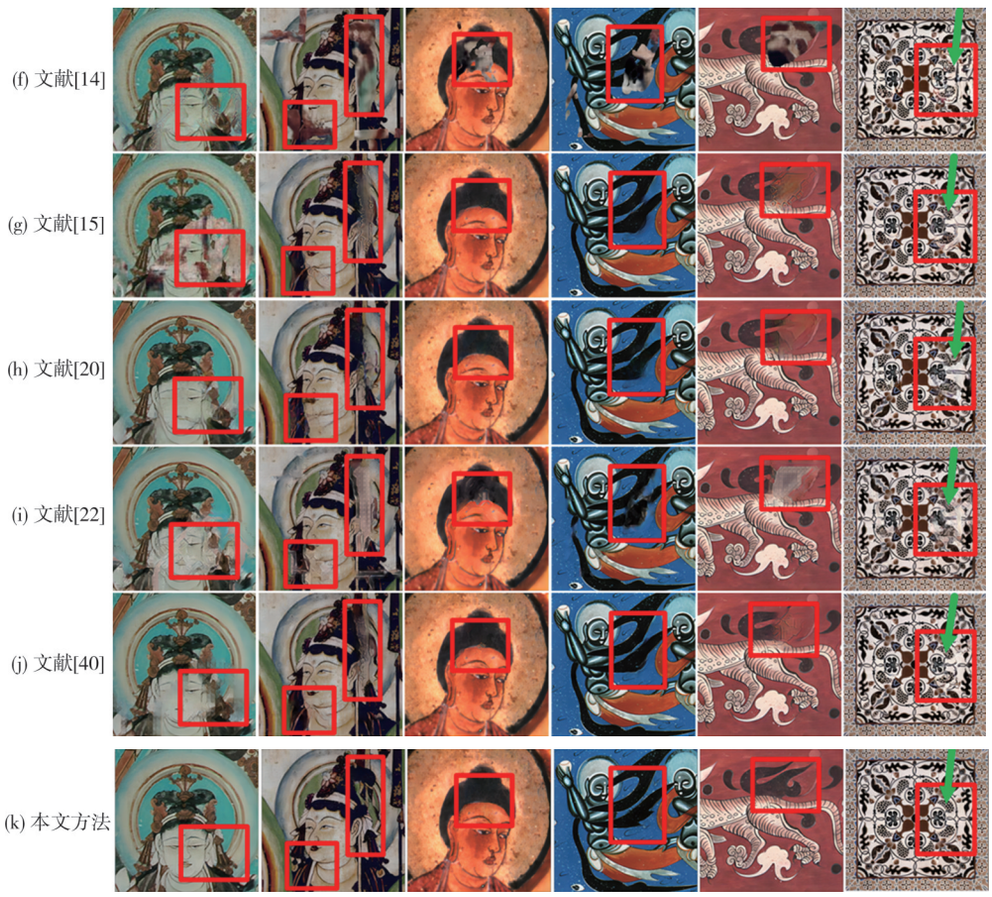


图 8 人为破损壁画修复实验对比

Fig.8 Comparison of results for artificially damaged murals

为了验证壁画修复结果的质量,下面采用峰值信噪比(PSNR)和结构相似性(SSIM)对图 8 的修复结果进行客观定量评价,比较结果见表 1 与表 2. 其中 PSNR 和 SSIM 值越大,表示修复后失真越少,修复效果越好. 从表 1 与表 2 可以看出,所提方法均优于其他对比算法,表明本文方法修复质量更好.

表 1 人为破损壁画修复实验 PSNR 客观对比

Tab. 1 PSNR objective comparison on inpainting artificially damaged murals

图像	文献[10]	文献[11]	文献[13]	文献[14]	文献[15]	文献[20]	文献[22]	文献[40]	本文方法
Image1	15.457 4	20.772 4	21.248 6	18.512 0	21.560 9	21.618 5	20.215 2	20.629 6	22.898 7
Image2	23.115 1	19.683 2	22.989 6	17.496 6	20.661 1	20.200 8	18.468 5	23.336 0	31.591 5
Image3	28.131 2	28.117 3	27.584 6	18.632 6	26.108 9	25.958 6	26.594 8	27.872 2	34.728 0
Image4	27.149 0	28.182 0	26.255 0	20.210 4	27.475 0	26.902 2	24.217 9	27.683 5	28.946 3
Image5	24.768 7	26.476 2	27.601 6	21.993 2	25.336 4	25.321 2	22.988 1	28.222 6	35.794 3
Image6	20.348 7	20.319 1	21.076 5	18.767 7	20.398 7	18.378 1	18.019 2	20.576 2	21.307 7

表 2 人为破损壁画修复实验 SSIM 客观对比

Tab. 2 SSIM objective comparison on inpainting artificially damaged murals

图像	文献[10]	文献[11]	文献[13]	文献[14]	文献[15]	文献[20]	文献[22]	文献[40]	本文方法
Image1	0.769 1	0.818 0	0.819 7	0.721 4	0.802 3	0.805 0	0.778 3	0.808 6	0.824 6
Image2	0.812 7	0.861 3	0.840 9	0.691 1	0.763 2	0.755 6	0.757 5	0.850 9	0.866 2
Image3	0.971 4	0.971 5	0.968 5	0.891 2	0.944 2	0.939 7	0.950 6	0.966 0	0.985 1
Image4	0.960 1	0.967 7	0.949 6	0.889 1	0.963 1	0.956 0	0.922 7	0.965 4	0.967 6
Image5	0.947 3	0.967 5	0.955 4	0.930 9	0.952 7	0.954 0	0.943 2	0.953 2	0.989 7
Image6	0.918 3	0.920 8	0.927 5	0.911 7	0.921 6	0.891 7	0.864 3	0.928 6	0.930 8

在完成对壁画修复结果PSNR和SSIM客观定量评价后，为了说明本文对破损壁画结构修复的有效性，进一步采用梯度相似性（gradient magnitude similarity deviation，GMSD）评价指标进行定量比较，GMSD公式如下：

$$\mathrm{GMSD}(x,y)=\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\mathrm{GMS}\left(x_i,y_i\right)-\overline{\mathrm{GMS}}\right)^2} \tag{25}$$

式中：$\mathrm{GMS}\left(x_i,y_i\right)=\dfrac{2G_xG_y+C}{G_x^2+G_y^2+C}$；$G_x$和$G_y$分别是图像$x$和$y$的梯度幅值；$\overline{\mathrm{GMS}}$是GMS的均值；$C$是一个常数，用于避免分母为零.

GMSD反映了图像修复前后的结构信息偏差，其值越小，表明修复后结构和语义变化越小，越接近于真实值，其修复性能越优.对于图8破损区域不同方法GMSD比较结果如表3所示.从表3可以得出，本文所提方法GMSD值最小，说明本文方法对于壁画破损区域的结构信息修复性能更优.

表3　人为破损壁画修复实验GMSD客观对比

Tab. 3　GMSD objective comparison on inpainting artificially damaged murals

图像	文献[10]	文献[11]	文献[13]	文献[14]	文献[15]	文献[20]	文献[22]	文献[40]	本文方法
Image1	0.336 48	0.328 93	0.322 54	0.205 05	0.233 15	0.228 41	0.260 30	0.324 24	0.191 36
Image2	0.287 25	0.236 88	0.247 90	0.294 03	0.278 60	0.278 91	0.290 27	0.240 79	0.203 11
Image3	0.192 70	0.192 41	0.191 76	0.210 56	0.192 40	0.181 82	0.224 14	0.183 31	0.149 25
Image4	0.240 11	0.247 96	0.238 52	0.288 26	0.240 65	0.245 54	0.295 10	0.240 80	0.226 62
Image5	0.272 16	0.251 73	0.227 26	0.243 24	0.232 56	0.231 97	0.238 89	0.225 12	0.161 30
Image6	0.185 53	0.191 56	0.177 49	0.206 49	0.182 66	0.197 13	0.221 81	0.175 30	0.166 97

3.3　真实破损壁画修复实验

为了进一步验证有效性，选取了四组真实破损敦煌壁画进行修复实验，修复结果如图9所示.对于第一幅“观无量寿经变之观音菩萨榆林二五窟”弟子图，文献[10]方法修复区域出现块效应现象，文献[20]的算法出现像素填充错误的现象，文献[22]方法的修复结果中出现模糊的问题.对于第二幅“安西榆林第二五窟”弟子图，文献[11]算法的修复出现纹理模糊现象，文献[14]算法的修复结果较差，破损区域未能修复完成，出现白色掩码残留现象，文献[40]算法的修复结果相较于其他文献较好.对于第三幅壁画，从公牛腿部裂痕的修复结果可以看出，其余对比方法均未能完成有效修复，且存在不同程度的掩码残留，而本文方法较好地完成了公牛腿部的裂痕修复.对于第四幅“安西榆林第一五窟”壁画，文献[10]、文献[11]、文献[13]、文献[14]和文献[20]算法的修复结果中裙带部分存在修复不彻底的问题，仍存在部分灰色的裂缝.而本文方法通过解码像素序列得到全局语义修复特征图，通过全局语义特征修复模块，提高对全局语义特征的修复能力，修复更加彻底且细节更清晰.



(a)真实破损　(b)文献[10]　(c)文献[11]　(d)文献[13]　(e)文献[14]



(f)文献[15]　(g)文献[20]　(h)文献[22]　(i)文献[40]　(j)本文方法

(Ⅰ)第一幅图



(a)真实破损 (b)文献[10] (c)文献[11] (d)文献[13] (e)文献[14]



(f)文献[15] (g)文献[20] (h)文献[22] (i)文献[40] (j)本文方法

(Ⅱ)第二幅图

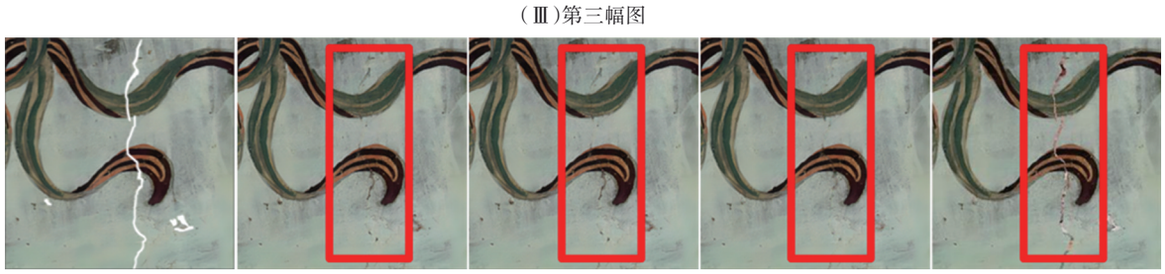


(a)真实破损 (b)文献[10] (c)文献[11] (d)文献[13] (e)文献[14]

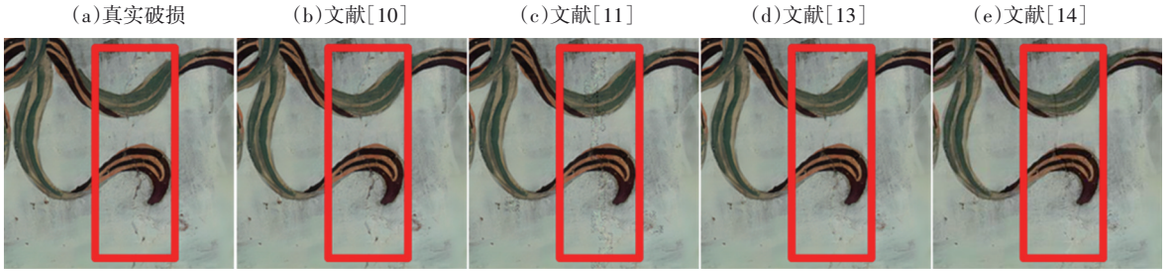


(f)文献[15] (g)文献[20] (h)文献[22] (i)文献[40] (j)本文方法

(Ⅲ)第三幅图



(a)真实破损 (b)文献[10] (c)文献[11] (d)文献[13] (e)文献[14]



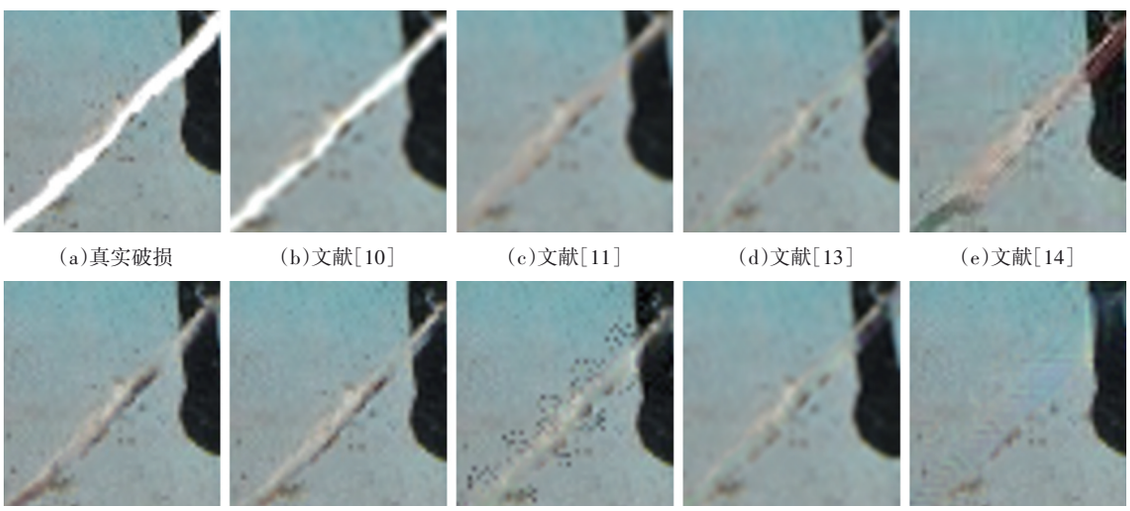
(f)文献[15] (g)文献[20] (h)文献[22] (i)文献[40] (j)本文方法

(Ⅳ)第四幅图

图 9 真实破损壁画修复实验对比图

Fig.9 Comparison of repair experiments for real broken murals

为了更加直观地比较图9实验修复差异,选择第三幅图像进行局部放大比较.将图9的第三幅进行局部放大后,如图10所示,可以看出本文方法虽然仍有少量痕迹,但整体完成了修复,而对比方法可以看出均未能有效完成修复,修复后仍存在大量残留.



(a)真实破损　(b)文献[10]　(c)文献[11]　(d)文献[13]　(e)文献[14]

(f)文献[15]　(g)文献[20]　(h)文献[22]　(i)文献[40]　(j)本文方法

图 10　修复结果局部放大对比图

Fig.10　Enlarged comparison of mural inpainting results

一般采用信息熵与图像模糊度对无参照真实破损壁画进行客观评价,对图9修复结果进行量化比较,如表4所示.其中信息熵代表图像信息丰富程度,其值越大,表明修复结果越清晰.图像模糊度用来衡量图像中细节丢失情况,其值越小,表明越清晰.从表4得出,在9种方法中,本文算法的图像信息熵值最大,并且图像的模糊度值最低,从而验证了所提方法对于真实壁画修复的有效性.

表4　真实破损壁画修复结果评价对比

Tab.4　Evaluation and comparison of restoration results of real damaged mural paintings

方法	Image 1		Image 2		Image 3		Image 4	
	信息熵	模糊度	信息熵	模糊度	信息熵	模糊度	信息熵	模糊度
文献[10]	6.996 3	13.812 4	6.896 2	11.540 2	6.946 1	9.635 4	6.810 1	9.288 7
文献[11]	6.966 3	13.796 8	6.926 2	11.568 9	6.936 8	9.687 5	6.699 0	9.284 4
文献[13]	7.028 9	13.712 5	6.908 2	11.540 4	6.987 4	9.385 7	6.708 2	9.288 5
文献[14]	6.905 8	13.804 1	6.941 0	11.573 2	6.989 9	9.563 3	6.759 2	9.515 5
文献[15]	6.884 6	13.757 4	6.936 2	11.583 4	6.967 2	9.573 9	6.742 2	9.322 2
文献[20]	7.021 8	13.721 5	6.939 6	11.584 3	6.968 1	9.580 4	6.742 9	9.319 7
文献[22]	6.971 1	13.744 7	6.946 3	11.655 5	6.976 5	9.721 4	6.763 3	9.614 8
文献[40]	7.060 9	13.574 8	6.953 4	11.514 5	6.963 2	9.687 4	6.720 2	9.521 4
本文方法	7.077 2	13.706 7	7.038 2	11.505 2	6.994 4	9.357 3	6.822 6	9.278 3

4　结 论

本文提出了一种双向自回归Transformer与快速傅里叶卷积增强的壁画修复方法.首先,设计基于Transformer结构的全局语义特征修复模块,提高对壁画全局语义特征的修复能力.其次,设计了由门控卷积和扩展卷积残差块组成的全局语义增强模块,采用大核注意力机制与快速傅里叶卷积提高了壁画细节的修复能力.最后通过敦煌壁画修复实验,结果表明,所提方法修复性能更优,在主客观评价方面均优于比较算法.

参考文献

[1]　潘云鹤,鲁东明. 古代敦煌壁画的数字化保护与修复[J]. 系

统仿真学报,2003,15(3):310-314.
PAN Y H,LU D M. Digital protection and restoration of Dunhuang mural[J]. Journal of System Simulation,2003,15(3):310-314.(in Chinese)

[2] 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.

[3] SCHAEFER K,WEICKERT J. Diffusion - shock inpainting[M]// Scale Space and Variational Methods in Computer Vision. Cham: Springer International Publishing,2023:588-600.

[4] 陈永,艾亚鹏,郭红光. 改进曲率驱动模型的敦煌壁画修复算法[J]. 计算机辅助设计与图形学学报,2020,32(5):787-796.
CHEN Y,AI Y P,GUO H G. Inpainting algorithm for Dunhuang mural based on improved curvature-driven diffusion model[J]. Journal of Computer-Aided Design & Computer Graphics,2020,32(5):787-796.(in Chinese)

[5] 李丽,高若婉,梅树立,等. 基于Shannon-Cosine小波精细积分法的壁画降噪修复方法[J]. 浙江大学学报(理学版),2019,46(3):279-287.
LI L,GAO R W,MEI S L,et al. Mural image de-noising based on Shannon-Cosine wavelet precise integration method[J]. Journal of Zhejiang University (Science Edition),2019,46(3):279-287.(in Chinese)

[6] BHELE S,SHRIRAMWAR S,AGARKAR P. An efficient texture-structure conserving patch matching algorithm for inpainting mural images[J]. Multimedia Tools and Applications,2023,82(30):46741-46762.

[7] 焦莉娟,王文剑,李秉婧,等. 改进的块匹配五台山壁画修复算法[J]. 计算机辅助设计与图形学学报,2019,31(1):118-125.
JIAO L J,WANG W J,LI B J,et al. Wutai mountain mural inpainting based on improved block matching algorithm[J]. Journal of Computer-Aided Design & Computer Graphics,2019,31(1):118-125.(in Chinese)

[8] LECOUAT B,PONCE J,MAIRAL J. Fully trainable and interpretable non-local sparse models for image restoration[M]// Computer Vision - ECCV 2020. Cham: Springer International Publishing,2020:238-254.

[9] GUO K H,LABATE D,RODRIGUEZ AYLLON J P. Image inpainting using sparse multiscale representations: image recovery performance guarantees[J]. Applied and Computational Harmonic Analysis,2020,49(2):343-380.

[10] 王欢,李利,李庆,等. 一种结合全局一致性与局部连续性的壁画修复方法[J]. 湖南大学学报(自然科学版),2022,49(6):135-145.
WANG H,LI L,LI Q,et al. A global uniform and local continuity repair method for murals inpainting[J]. Journal of Hunan University (Natural Sciences),2022,49(6):135-145.(in Chinese)

[11] 陈永,杜婉君,赵梦雪. 改进多重字典联合自适应学习的稀疏壁画修复[J]. 湖南大学学报(自然科学版),2023,50(12):1-9.
CHEN Y,DU W J,ZHAO M X. Improved sparse mural restoration algorithm using joint adaptive learning of multiple dictionaries[J]. Journal of Hunan University (Natural Sciences),2023,50(12):1-9.(in Chinese)

[12] YANG J,RUHAIYEM N I R,ZHOU C C. A 3M-hybrid model for the restoration of unique giant murals: a case study on the murals of Yongle Palace[EB/OL].[2024-04-20]. https://arxiv.org/abs/2309.06194v1.

[13] 赵磊,林思寰,林志洁,等. 中国古画渐进式多级特征修复算法[J]. 计算机辅助设计与图形学学报,2023,35(7):1040-1051.
ZHAO L,LIN S H,LIN Z J,et al. Progressive multilevel feature inpainting algorithm for Chinese ancient paintings[J]. Journal of Computer-Aided Design & Computer Graphics,2023,35(7):1040-1051.(in Chinese)

[14] WADHWA G,DHALL A,MURALA S,et al. Hyperrealistic image inpainting with hypergraphs[C]//2021 IEEE Winter Conference on Applications of Computer Vision (WACV). Waikoloa,HI,USA. IEEE,2021:3911-3920.

[15] GUO X F,YANG H Y,HUANG D. Image inpainting via conditional texture and structure dual generation[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal,QC,Canada. IEEE,2021:14114-14123.

[16] WANG N,LI J Y,ZHANG L F,et al. MUSICAL: multi-scale image contextual attention learning for inpainting[C]//Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. Macao,China. 2019:3748-3754.

[17] YANG J,QI Z Q,SHI Y. Learning to incorporate structure knowledge for image inpainting[J]. Proceedings of the AAAI Conference on Artificial Intelligence,2020,34(7):12605-12612.

[18] YU Y S,WANG H,LUO T J,et al. MaGIC: multi-modality guided image completion[EB/OL]. (2023-05-19)[2023-11-04]. https://doi.org/10.48550/arXiv.2305.11818.

[19] LIU W J,SHI Y Q,LI J C,et al. Multi-stage progressive reasoning for Dunhuang murals inpainting[C]//2023 IEEE 4th International Conference on Pattern Recognition and Machine Learning (PRML). Urumqi,China. IEEE,2023:211-217.

[20] NAZERI K,NG E,JOSEPH T,et al. EdgeConnect: generative image inpainting with adversarial edge learning[EB/OL].[2024-04-20]. https://arxiv. org/abs/1901.00212v3.

[21] LI L X,ZOU Q,ZHANG F,et al. Line drawing guided progressive inpainting of mural damage[EB/OL].[2024-04-20]. https://arxiv.org/abs/2211.06649v2.

[22] LI J Y,WANG N,ZHANG L F,et al. Recurrent feature reasoning for image inpainting[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle,WA,USA. IEEE,2020:7757-7765.

[23] 赵磊,吉柏言,邢卫,等. 基于多路编码器和双重注意力的古画修复算法[J]. 计算机研究与发展,2023,60(12):2814-2831.
ZHAO L,JI B Y,XING W,et al. Ancient painting inpainting algorithm based on multi-channel encoder and dual attention[J]. Journal of Computer Research and Development,2023,60(12):2814-2831.(in Chinese)

[24] ZENG Y, LIN Z, LU H C, et al. CR-fill: generative image inpainting with auxiliary contextual reconstruction [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada. IEEE, 2021: 14144-14153.

[25] LI W B, 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). New Orleans, LA, USA. IEEE, 2022: 10748-10758.

[26] DENG Y, HUI S Q, ZHOU S P, et al. T-former: an efficient transformer for image inpainting [C]//Proceedings of the 30th ACM International Conference on Multimedia. Lisboa, Portugal. ACM, 2022: 6559-6568.

[27] ZENG Y H, FU J L, CHAO H Y, et al. Aggregated contextual transformations for high-resolution image inpainting [J]. IEEE Transactions on Visualization and Computer Graphics, 2023, 29(7): 3266-3280.

[28] 王真言, 蒋胜丞, 宋奇鸿, 等. 基于Transformer的文物图像修复方法[J]. 计算机研究与发展, 2024, 61(3): 748-761.
WANG Z Y, JIANG S C, SONG Q H, et al. Transformer-based image restoration method for cultural relics [J]. Journal of Computer Research and Development, 2024, 61(3): 748-761. (in Chinese)

[29] LIU Q K, TAN Z T, CHEN D D, et al. Reduce information loss in transformers for pluralistic image inpainting [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA. IEEE, 2022: 11337-11347.

[30] PENG Z L, GUO Z H, HUANG W, et al. Conformer: local features coupling global representations for recognition and detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(8): 9454-9468.

[31] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NYCurran Associates Inc, 2017: 6000-6010.

[32] LI Z C, ZHANG S T, ZHAO H, et al. BatGPT: a bidirectional autoregessive talker from generative pre-trained transformer [EB/OL]. [2024-04-20]. https://arxiv. org/abs/2307.00360v2.

[33] DEVLIN J, CHANG M W, LEE K T, et al. BERT: pretraining of deep bidirectional transformers for language understanding [C]// Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics. Minnesota, MN: Association for Computational Linguistics, 2019: 4171-4186.

[34] ZHOU J H, WEI C, WANG H Y, et al. iBOT: image BERT pre-training with online tokenizer [EB/OL]. [2024-04-20]. https://arxiv.org/abs/2111.07832v3

[35] QIU X P, SUN T X, XU Y G, et al. Pre-trained models for natural language processing: a survey [J]. Science China Technological Sciences, 2020, 63(10): 1872-1897.

[36] YU J H, LIN Z, YANG J M, et al. Free-form image inpainting with gated convolution [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South). IEEE, 2019: 4470-4479.

[37] LI Y H, ZHANG X F, CHEN D M. CSRNet: dilated convolutional neural networks for understanding the highly congested scenes [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. IEEE, 2018: 1091-1100.

[38] GUO M H, LU C Z, LIU Z N, et al. Visual attention network [J]. Computational Visual Media, 2023, 9(4): 733-752.

[39] SUVOROV R, LOGACHEVA E, MASHIKHIN A, et al. Resolution-robust large mask inpainting with Fourier convolutions [C]//2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa, HI, USA. IEEE, 2022: 3172-3182.

[40] JAIN J, ZHOU Y Q, YU N, et al. Keys to better image inpainting: structure and texture go hand in hand [C]//2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa, HI, USA. IEEE, 2023: 208-217.

[41] ILYA L, FRANK H .Decoupled weight decay regularization [EB/OL]. (2019-01-04) [2023-11-23]. https://doi. org/10.48550/arXiv. 1711.05101.