武汉大学学报(信息科学版)
Geomatics and Information Science of Wuhan University
ISSN 1671-8860,CN 42-1676/TN

《武汉大学学报(信息科学版)》网络首发论文

题目：文本模态辅助引导的壁画修复算法
作者：陈永，杜婉君，张世龙
DOI：10.13203/j.whugis20240251
收稿日期：2025-01-28
网络首发日期：2025-03-13
引用格式：陈永，杜婉君，张世龙．文本模态辅助引导的壁画修复算法[J/OL]．武汉大学学报(信息科学版). https://doi.org/10.13203/j.whugis20240251

DOI:10.13203/j.whugis20240251

引用格式：

陈永，杜婉君，张世龙. 文本模态辅助引导的壁画修复算法[J].武汉大学学报（信息科学版）, 2025, DOI:10.13203/J.whugis20240251（CHEN Yong, DU Wanjun, ZHANG Shilong. Text Modality Assisted Guided Mural Inpainting Algorithm[J]. Geomatics and Information Science of Wuhan University, 2025, DOI:10.13203/J.whugis20240251）

文本模态辅助引导的壁画修复算法

陈永[1,2]　杜婉君[1]　张世龙[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，仅考虑壁画图像先验信息，缺乏文本信息引导壁画修复，导致修复结果出现语义不一致，细节缺失等问题，提出了一种基于文本模态辅助引导的壁画修复算法。首先，提出一个文本引导的壁画修复网络，利用文本信息作为壁画修复的辅助控制引导，为壁画图像提供上下文引导修复信息。其次，构建文本语义过滤模块，利用掩膜图像和互补图像过滤得到破损区域的文本特征，并设计了语义增强模块，对过滤后的文本特征进行增强，提高文本语义与图像语义的一致性。然后，设计上采样纹理细节修复网络，实现对壁画纹理细节的修复。最后，采用谱归一化判别器博弈对抗，完成修复。通过对真实敦煌壁画的修复实验表明，所提方法能够有效完成破损壁画的修复，在主客观评价方面均优于比较算法。

关键词：壁画修复；文本引导；跨模态修复；语义过滤；语义增强

Text Modality Assisted Guided Mural Inpainting Algorithm

CHEN Yong[1,2]　*DU Wanjun*[1]　*ZHANG Shilong*[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, Lanzhou 730070，China

Abstract: Objectives: In order to solve the problems of existing deep learning algorithms in the inpainting of murals, which only consider the prior information of mural images to guide the inpainting of damaged areas, and lack of text information to guide the inpainting of murals, resulting in semantic inconsistency and lack of details in the inpainting results, a text modality-assisted guided mural inpainting algorithm is proposed. **Methods:** First, a text modality-assisted guided mural inpainting network is proposed, which uses the text information as the control guidance for mural inpainting, and provides context-guided repair information for mural images. Secondly, a text filtering module is constructed to obtain the text features of the damaged area by filtering the mask image and the complementary image, and a cross-modal semantic enhancement module is designed to enhance the filtered text features to improve the consistency between text semantics and image semantics. Then, the upsampled texture detail inpainting network is designed to achieve bidirectional fusion of shallow and deep features to obtain mural images with fine-grained features. Finally, the normalized discriminator is used to match the recovered murals to the game, and the recovered murals are obtained. **Results:** The experimental results of real mural restoration show that this

收稿日期：2025-01-28

项目资助：教育部人文社会科学研究青年基金资助项目(19YJC760012)；兰州交通大学基础研究拔尖人才项目(2022JC36)；兰州交通大学重点研发项目(ZDYF2304)。

第一作者：陈永，博士，教授，研究方向为图像处理、计算机视觉。edukeylab@126.com

通信作者：陈永，博士，教授。edukeylab@126.com

method can effectively repair damaged murals, and is superior to comparative algorithms in subjective and objective evaluation. **Conclusions:** The proposed method can effectively repair damaged murals, achieving better visual perception and coordination.

Key words: mural inpainting; text guide; cross-modal inpainting; semantic filtering; semantic enhancement

壁画作为独特古老的绘画艺术形式，是不可再生的文化遗产，是千余年历史文化的重要载体，在美术史上占有重要地位。古代壁画有着辉煌的历史，包括远古岩画、墓室壁画、石窟壁画和庙堂壁画等，其风格独特，绘画技巧精湛，具有重要的艺术和考古学价值[1]。而敦煌壁画是目前世界上现存规模最大的古壁画资源，其通过巧妙构图、绘画技法等独特的艺术语言将古老传说、历史文化等生动地描绘于窟壁上，饱含着深厚的文化底蕴，是人类文明传承的宝贵财富。

然而，由于年代久远、气候变化、人为破坏等客观原因，古代壁画出现了不同程度的病害，如裂缝、霉变、脱落等亟待解决的问题。对于古壁画遗产的保护，为了减缓壁画老化和病害的侵蚀，目前主要通过安装传感器实时采集和数据分析，以及文化遗产数字化智能保护处理技术，及时预警潜在风险，为壁画预防性保护提供预警监测。数字化修复可以较好的克服人工修复风险大的不足，采用数字化修复技术实现对文化遗产的保护，已成为当前的研究热点问题[2]。

图像数字化修复是在缺失区域生成逼真视觉内容的任务，同时保持语义正确性和连贯性[3]。目前，图像修复方法分为传统方法[4-5]和深度学习方法[6]。传统修复方法一般采用传播机制或者复制粘贴的方式寻找最佳匹配块对缺失区域进行填补[7]，但该类方法未考虑壁画图像的高级语义信息，对破损区域较大的图像修复效果不佳。

基于深度学习的图像修复方法通过学习图像特征信息，实现对破损像素的修复和生成。目前，基于深度学习修复方法主要分为引导性修复和非引导性修复方法，非引导性修复一般是利用破损图像的先验特征直接进行修复。如 Zeng 等[8]采用金字塔上下文编码器对破损图像进行修复，学习图像深层特征中的已知内容，然后迁移至浅层特征，逐层完成图像修复。Xiao 等[9]提出一种改进残差集的图像修复方法，利用残差学习从层次特征中学习特征信息，但由于在修复过程中对高级语义信息考虑不足，导致修复结果存在语义断裂的问题。Li 等[10]提出了一种视觉结构重建网络，通过在编解码器中添加视觉结构重建层来重建破损信息，但该方法忽略了纹理信息的重要性，修复后存在纹理模糊的问题。Suvorov 等[11]利用快速傅里叶卷积扩大图像范围的感受野，捕获图像复杂的周期结构，但该方法没有考虑图像的边界语义约束，该模型会产生结构淡化模糊的问题。Zheng 等[12]提出了一种先生成粗略修复图像，然后细化第一阶段修复图的算法，但该方法由于将完好区域和破损区域分开建模，忽略了图像的全局语义约束，导致修复结果存在全局语义不一致的问题。以上修复方法仅利用图像像素信息对破损区域进行修复，缺乏高级语义信息对修复过程进行指导，导致修复结果存在语义不一致等问题。

引导修复目的是利用已知先验信息引导破损图像完成修复，如利用图像的结构信息、统计信息、语义信息等特征完成对破损图像进行引导修复。Nazeri 等[13]提出了利用边缘轮廓引导图像修复的模型 EdgeConnect，通过先修复破损的边缘图像，再利用边缘引导图像内容补全，但该算法的内容补全依赖拟合的边缘轮廓，易出现语义修复错误的问题。Zhao 等[14]提出使用图像块引导图像修复的方法，但该方法对上下文语义考虑不足，易存在修复结果过渡不自然、结构不协调的问题。

Guo 等[15]提出了一种纹理和结构联合约束的修复方法，但忽略了图像高级语义信息引导作用，导致修复结果存在全局语义不一致的问题。Cao 等[16]引入了一种多尺度草图张量修复网络，通过学习草图张量空间来恢复损坏图像中的边缘、线和连接，但修复过程中仅考虑图像结构特性，忽略了图像的语义特性，导致修复结果存在语义不一致的问题。Zhang 等[17]提出了一种文本引导的双重注意力 TDANet 修复模型，该模型利用破损区域的显示语义信息指导图像修复，但是该模型在修复过程中忽略了图像细节信息的修复，导致修复结果存在细节缺失、纹理模糊的问题。

综上所述，目前的图像修复方法在修复破损壁画时，仅根据壁画图像先验信息对破损区域进行推断和补全，缺乏引导性文本信息指导图像完成修复，导致修复结果易出现语义不一致、细节缺失等问题。因此，本文提出了一种基于文本引导的跨模态壁画修复算法。本文主要工作包括：

(1)针对壁画修复时，缺少引导性文本信息指导修复，导致出现语义不一致的问题，提出一个基于文本引导的跨模态修复网络，利用文本信息引导完成对破损壁画内容信息的修复。

(2)为了提高文本引导的精确性，构建了文本语义过滤模块，利用掩膜图像和互补图像辅助过滤得到破损区域的文本特征，并设计语义增强模块，增强文本语义与图像语义的一致性。

(3)针对文本引导壁画修复过程中对壁画细节修复不足的问题，设计了上采样纹理细节修复模块，通过壁画深层特征与浅层特征的双向融合，完成对壁画图像的细粒度修复。

1 本文方法

1.1 整体网络框架

壁画图像具有丰富的语义信息，但现有的壁画图像修复方法中，忽略了壁画图像的文本信息属性，仅从像素层面对破损壁画进行修复，导致修复结果存在语义不一致，细节缺失的问题。基于此，本文提出了一种基于文本模态辅助引导的壁画图像修复算法，利用文本信息和壁画图像信息，实现对破损壁画的跨模态修复，整体修复网络模型如图 1 所示。模型工作时，首先，将掩膜壁画和互补壁画分别输入到图像编码器中提取得到图像特征，将文本描述输入到双向 LSTM 网络中，得到单词特征和句子特征。其次，将图像特征和单词特征输入到文本语义过滤模块，通过注意力过滤机制提取得到壁画破损区域的单词特征信息。然后，将过滤得到的单词特征与图像特征输入到语义增强模块，对文本语义特征进行增强，提高文本语义与图像语义的一致性。最后，利用上采样纹理细节修复网络进行局部细节修复，得到修复后的壁画图像。

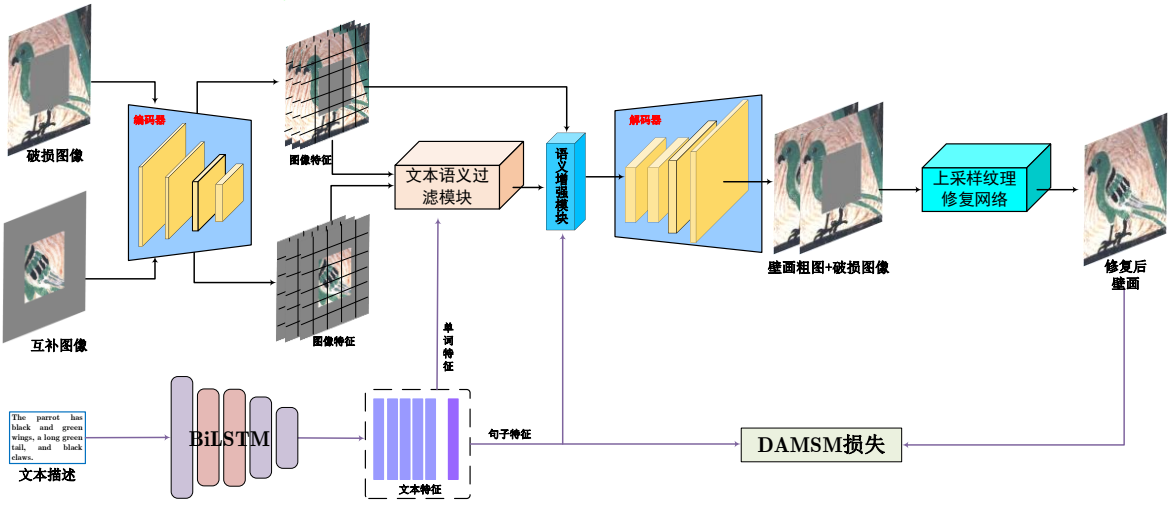


图 1 整体框架图
Fig. 1 Model Framework of Proposed Method

1.2 特征提取

由于文本和壁画图像是异构的，处于不同的特征空间，需要分别提取文本和壁画图像的特征信息，然后将这些特征投射到共享的语义空间，从而实现文本语义与壁画图像语义的一致性。使用双向长短期记忆网络(Bidirectional Long Short-Term Memory, BiLSTM)将文本序列编码成向量表示，同时使用图像编码器将壁画图像编码成向量表示。然后，将这两种向量在相同的语义空间进行一致性比较，从而实现文本语义与壁画图像语义的对齐。具体流程为，首先提取文本特征，如图 1 所示，采用双向长短期记忆网络 BiLSTM 编码器进行文本特征提取。其次，将壁画破损图像和互补图像分别输入到图像编码器中，得到破损图像的图像特征和互补图像的特征。

1.2.1 文本特征提取

为了更好的学习文本上下文信息，采用 BiLSTM 编码器进行文本特征提取。对于输入文本语句 $T(w_1,w_2,\cdots,w_L)$，通过 BiLSTM 计算单词的特征 W 和句子特征 s，过程为：

$$W,s=\mathrm{F}_{BiLSTM}(T) \tag{1}$$

其中，$s\in R^D$ 是句子向量，$W\in R^{D\times L}$ 为单词特征矩阵，D 为维数，L 是单词的数量，式 $\mathrm{F}_{BiLSTM}(\cdot)$ 为双向长短时记忆网络。

1.2.2 图像特征提取

本文使用经过预训练的 Inception-v3 模型作为图像编码网络，利用图像编码器将壁画图像特征映射到与文本特征公共的语义空间。首先，利用编码网络学习得到破损壁画图像特征矩阵 $f_m\in R^{D\times N}$ 和互补图像特征矩阵 $f_c\in R^{D\times N}$，其中壁画图像特征矩阵 f_m 由 N 个子区域列向量构成。然后，通过感知器层，将图像特征矩阵 f_m 映射到文本特征的公共语义空间，该过程如下式：

$$F_m=Q(f_m) \tag{2}$$

其中，$F_m\in R^{D\times N}$ 是破损壁画图像特征，$Q(\cdot)$ 代表感知器层运算。同理，也得到互补图像映射后的互补图像的特征 $F_c=Q(f_c)$。

1.3 文本语义过滤模块

在得到文本特征和图像特征之后，利用文本特征引导壁画图像进行修复。文本引导壁画修复的目的是使修复后的壁画破损区域与完好区域相一致，因而为了进一步提高文本引导壁画修复的精确性，设计了文本语义过滤模块，利用注意力过滤机制，过滤得到破损壁画缺失部分的文本语义信息，从而实现对于破损区域的精准引导。文本语义过滤模块分别计算互补壁画特征和破损壁画特征与文本特征的注意力图，得到文本特征与破损壁画之间为负对的单词向量，和文本特征与互补壁画之间为正对的单词向量，将这两个向量进行比对，即可过滤得到缺失部分对应的文本语义。文本语义过滤模块如图 2 所示。

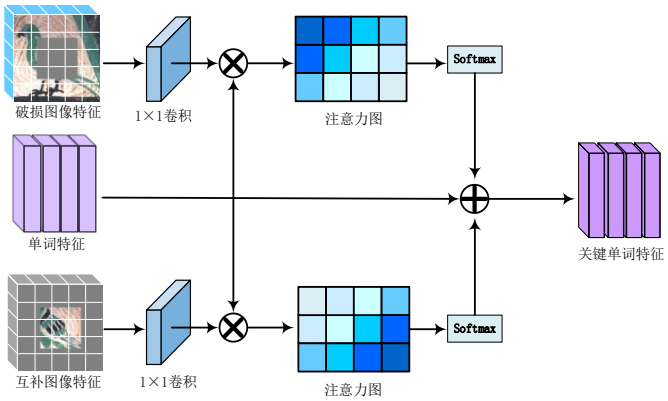


图 2 文本语义过滤模块
Fig. 2 Text Semantic Filtering Module

首先，将破损图像特征 F_m、单词特征 W 和互补图像的特征 F_c 输入到文本语义过滤模块，然后分别计算图像特征和单词特征之间的注意力权重，计算公式如下：

$$A^c_{i,j} = M_i^{'} Q(F_c)^T W_j \tag{3}$$

$$A^m_{i,j} = -Q(F_m)^T W_j + M_i \tag{4}$$

其中，$A^c_{i,j}$ 表示互补特征与单词特征之间的注意力图，$A^m_{i,j}$ 表示破损壁画特征与单词特征之间的注意力图，$M^{'}$ 表示输入的二值掩膜图像，掩膜处像素为 0，其他地方像素为 1，$Q(\cdot) = Conv(\cdot)$，$Conv(\cdot)$ 是1×1的卷积滤波器。

然后将注意力图分别输入到 softmax 函数中，求得文本嵌入的权重，如式(5)和式(6)所示：

$$\beta^c_{i,j} = \frac{\exp(A^c_{i,j})}{\sum_{i=1}^{C} \exp(A^c_{i,j})} \tag{5}$$

$$\beta^m_{i,j} = \frac{\exp(A^m_{i,j})}{\sum_{i=1}^{C} \exp(A^m_{i,j})} \tag{6}$$

其中，C 是特征图的面积，$\beta^*_{i,j}$ 表示注意力权重，根据权重，过滤得到壁画破损区域单词的权重，如式(7)所示：

$$\beta_{i,j} = Filter(\beta^c_{i,j}, \beta^m_{i,j}) \tag{7}$$

其中，$\beta_{i,j}$ 为单词权重，Filter(·) 为过滤器函数，最后，计算加权后的单词表示：

$$W_{ei} = \sum_{j=1}^{L} \beta_{i,j} W_j \tag{8}$$

其中，$W_e = (W_{e1}, W_{e2}, \cdots W_{eL})$ 是加权单词特征。

1.4 跨模态融合解码修复

壁画跨模态解码修复部分以图像和文本特征作为先验，将文本特征和图像特征进行融合后输出修复壁画图像。为了提高文本语义与图像语义的一致性，进行更好的融合，设计语义增强模块，将过滤得到的单词特征和句子特征与图像特征进行增强，然后使用解码网络进行修复。整体架构如图 3 所示，主要由语义增强融合网络，条件增强模块和解码网络构成。

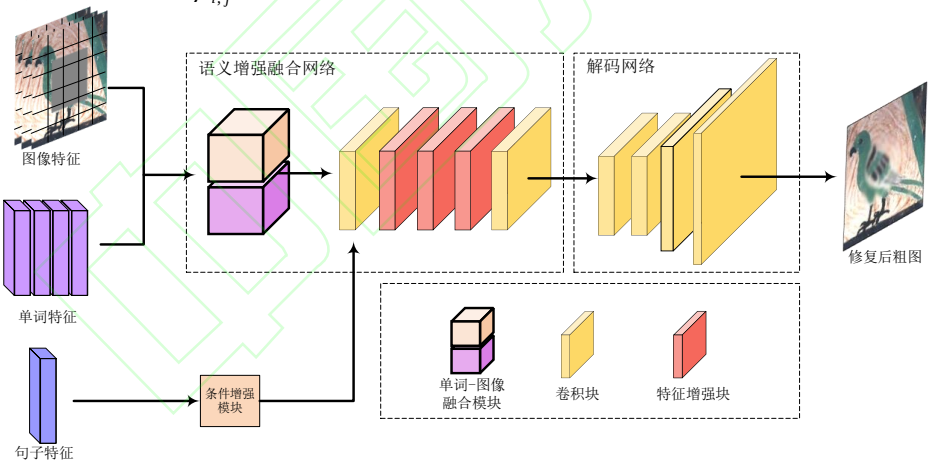


图 3 跨模态融合解码修复网络

Fig. 3 Cross-Modal Fusion Decoding Repair Network

1.4.1 语义增强融合网络

在图 3 跨模态融合解码修复网络中，为了提高文本语义与图像语义的一致性，将文本特征信息有效融合到图像特征中，本文设计语义增强融合网络，如图 4 所示。首先，计算破损壁画图像特征 F_m 和过滤后的加权单词特征 W_e 的通道注意力矩阵 a，即 $a = (W_e)^T F_m$。然后，将通道注意力矩阵 a 传入到 softmax 函数中得到每一个单词与图像子区域的关系权重：

$$m_{j,i} = \frac{e^{a_{j,i}}}{\sum_{k=1}^{L} e^{a_{j,k}}} \tag{9}$$

其中，$m_{j,i}$ 是第 i 个单词与第 j 个图像子区域的关系权重，$e^{a_{j,i}}$ 是第 i 个单词与第 j 个图像子区域的通道注意力指数值。然后在式(9)计算结果的基础上，以加权求和的方式求出单词关系权重 $m_{j,i}$ 与每一个图像子区域特征 F_{mi} 的一致性表示：

$$p_j = \sum_{i=1}^{N} m_{j,i} F_{mi} \tag{10}$$

其中，p_j 表示第 i 个单词特征与图像特征 F_m 一致性特征。最后，进行拼接，得到文本语义与图像语义对齐后的中间融合特征 $p = (p_0, p_1, \cdots, p_{N-1}) \in R^{D \times N}$ 。

由于句子特征嵌入以非线性的方式转换为生成条件的潜在变量，但潜在特征表示是高维的，其语义空间的不连续性，会造成修复生成的效果不理想[18]。为了解决该问题，增强文本语义分布的稠密性，设计条件语义增强模块对文本句子进行重采样来提高模型的性能。条件语义增强后的句子特征 s_{ca} 如下式所示：

$$s_{ca} = \mathrm{F}_{ca}(s) \tag{11}$$

其中，$s_{ca} \in R^{D'}$，D' 是条件增强后句子向量的维数。

最后，将中间融合特征 p 和句子特征 s_{ca} 输入到特征增强模块，进行卷积和哈达玛积运算，关联文本信息和图像区域。其过程可表示为：

$$h = s_{ca} \odot w(p) + b(p) \tag{12}$$

其中，h 为增强融合后的特征，$w(p)$ 为权重，$b(p)$ 为偏差，$\odot$ 为哈达玛积。

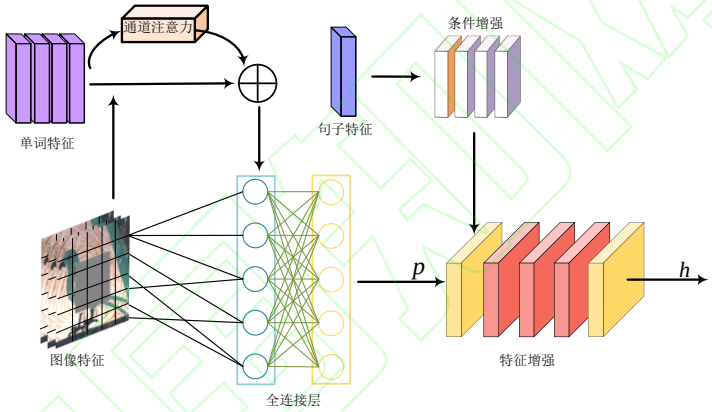


图 4 语义增强融合模块

Fig. 4 Semantic Enhancement Fusion Module

1.4.2 解码修复

将多模态特征 h 作为先验条件输入到解码网络中并投影至潜在空间。假设该潜在空间是高斯分布，并利用解码网络来预测潜在空间的一组参数，融合过程可以表示为：

$$\mu, \sigma = Decoder(h) \tag{13}$$

其中，μ 和 σ 分别为预测高斯分布的均值和方差，$Decoder(\cdot)$ 表示解码网络。

然后，通过从分布中对潜在变量进行采样，并与 h 通过式(13)进行连接，得到特征表示 r：

$$r = h + Gaussian(\mu, \sigma) \tag{14}$$

最后，通过解码器对特征解码上采样获得粗略的壁画修复结果 I_{gen1}。

为了直观验证文本引导壁画修复功能和融合模块的有效性，下面进行图像文本注意力图[19]可视化实验，显示文本描述中感兴趣的对象。选择两幅壁画，分别为壁画中菩萨的形象和动物鹦鹉的形象，对其可视化结果进行展示，如图 5 所示。在图 5 中，热力图中越偏蓝绿色说明该区域越符合文本描述。

bodhisattva　black　mouth　eyes　ears

图 5 注意力可视化图

Fig. 5 Attention visualization diagram

1.5 上采样纹理修复网络

在完成跨模态语义增强后，为了提高对壁画局部纹理细节的修复性能，进一步设计了上采样纹理修复网络，实现对壁画局部纹理细节的精细化修复，如图 6 所示。上采样纹理修复网络主要由编码器和特征双向融合解码器构成，其输入为文本引导阶段修复后的壁画粗图和掩码图像，最后得到修复后的壁画图像。

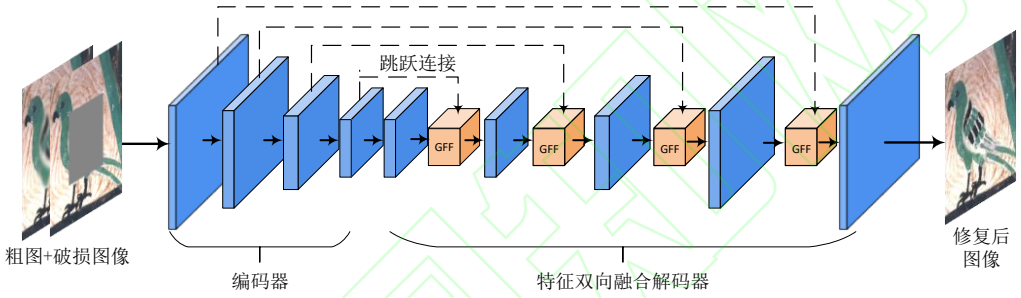


图 6 上采样纹理修复网络

Fig. 6 Texture Repair Network

图 6 中设计了特征双向门控融合模块 GFF，双向门控融合模块将编码器提取得到的壁画的浅层特征和解码器提取得到的深层特征进行融合，这样在保留壁画图像原始特征的同时，进一步修复壁画的纹理细节信息，如图 7 所示。

修复过程为，编码器特征 I_{en} 与解码器特征 Y_{de} 输入到特征双向融合模块中，利用门控机制分别学习其门控权重，如下式所示：

$$g_{en} = \delta(Conv_{1\times1}(I_{en})) \tag{15}$$

$$g_{de} = \delta(Conv_{1\times1}(Y_{de})) \tag{16}$$

其中，$\delta(\cdot)$ 为门控卷积，$Conv_{1\times1}(\cdot)$ 表示 1×1 卷积。

然后，将门控权重与编解码器特征进行交互融合，让 I_{en} 与 Y_{de} 互相学习各自的特征信息，得到更好的修复特征，其过程为：

$$I_{en}^{'} = I_{en} \oplus (I_{de} \odot g_{en}) \tag{17}$$

$$Y_{de}^{'} = Y_{de} \oplus (Y_{en} \odot g_{de}) \tag{18}$$

最后，将融合后的特征 $I_{en}^{'}$ 和 $Y_{de}^{'}$ 按照维度进行逐通道拼接，得到修复后的融合输出。

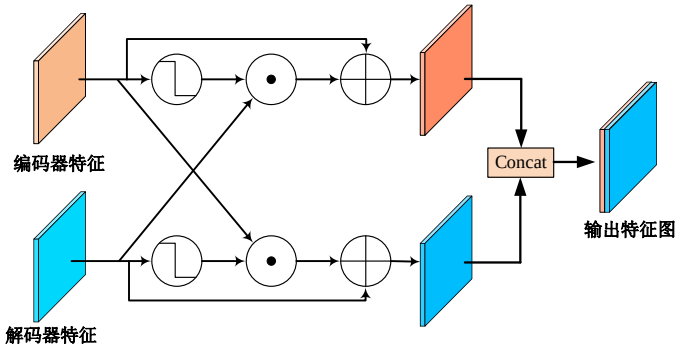


图 7 特征双向融合模块

Fig. 7 Feature Bidirectional Fusion Module

1.6 损失函数

为了使修复后的壁画图像更加真实自然，本文采用文本引导注意力损失、重建损失、文本图像匹配损失、判别器损失和感知损失构成的联合损失函数[19-20]对本文网络进行训练。

1.6.1 文本引导注意力损失

在文本引导壁画修复中，为了提高修复后壁画与文本描述之间的相关性，训练过程中引入文本引导的注意力损失[19]。文本引导注意力损失通过计算修复后壁画特征与单词特征之间的语义差异，引导网络在修复过程中更加注重文本描述的内容。这有助于确保修复后的壁画在语义上与文本描述保持一致，提高修复结果的准确性。其表达式为：

$$L_{TGA} = \| A_{i,j} I_{gen} - A_{i,j} I_{gt} \|_1 \tag{19}$$

其中，$A_{i,j}$ 为获得单词上下文特征注意力图，I_{gen} 为修复后的壁画图像和 I_{gt} 为真实壁画图像。

1.6.2 重建损失

所提方法在文本引导壁画修复阶段对壁画内容进行修复，上采样纹理修复阶段进行纹理细节修复，为了约束壁画在各个阶段的修复质量，采用 L1 重建损失进行约束。重建损失用于计算修复后壁画与真实壁画之间的像素差异，以确保网络在训练过程中不断减少修复后壁画与真实壁画之间的差异，从而提高修复图像的质量。其表达式为：

$$L_{rec} = \| I_{gen1} - I_{gt} \|_1 + \| I_{gen} - I_{gt} \|_1 \tag{20}$$

其中，I_{gen1} 文本引导阶段的修复结果，I_{gen} 为修复后的壁画图像，I_{gt} 为真实壁画图像。

1.6.3 文本图像匹配损失

为了进一步细化修复后壁画和文本的关系，在文本图像相似度匹配中，采用深度注意双重模态相似性模型(Deep Attention Multimodal Similarity Model, DAMSM)损失[20]进行约束。DAMSM 损失是通过图像编码器从修复后的壁画中提取特征，并将这些特征与文本特征进行比较，定义 DAMSM 损失为：

$$L_{DAMSM} = L^w + L^s \tag{21}$$

其中，w 和 s 分别代表单词和句子，L^w 为修复后壁画与对应的单词描述匹配的负对数后验概率，L^s 为修复后壁画与句子描述匹配的负对数后验概率。

1.6.4 判别器损失

为了确保修复后壁画的视觉真实性和语义合理性，在修复路径中使用谱归一化马尔可夫判别器 (Spectral Normalized Markovian Discriminator, SN-PatchGAN)进行判别约束。在壁画修复中，引入谱归一化有助于防止判别器过强或修复模型崩溃的情况，使模型训练过程更加稳定，并且 SN-PatchGAN 在约束过程中不仅关注壁画的局部细节，还可以捕获壁画图像的全局信息。这种全局与局部信息的平衡有助于模型在修复过程中同时考虑整体结构和局部细节，从而生成更加自然和真实的修复壁画。其表示为：

$$L_{adv} = E_{I_{gt} \sim P_{data}} [\sigma(1 - D_l(I_{gt})] + E_{I_{gen} \sim p_z} [\sigma(1 + D_l(I_{gen})] \tag{22}$$

式中：$D_l(\cdot)$ 为谱归一化网络，$\sigma(\cdot)$ 为 ReLU 激活函数，p_z 为修复壁画的分布。

1.6.5 感知损失

感知损失有助于更好地捕捉修复后壁画图像的高级语义特征，所以采用感知损失从修复后的壁画与真实壁画中提取语义特征，保证图像语义的正确性。其公式表示为：

$$L_p = \frac{1}{C_i H_i W_i} \| \phi_i(I_{gen}) - \phi_i(I_{gt}) \|_2^2 \tag{23}$$

其中，$\phi_i(\cdot)$ 表示固定特征映射，C_i, H_i, W_i 分别表示第 i 个特征图的通道数、高度和宽度。

最后，总损失可以表示为：

$$L = \lambda_{TGA} L_{TGA} + \lambda_{rec} L_{rec} + \lambda_{DAMSM} L_{DAMSM} + \lambda_{adv} L_{adv} + \lambda_p L_p \tag{24}$$

其中，λ_{TGA}、λ_{rec}、λ_{DAMSM}、λ_{adv} 和 λ_p 分别为文本引导注意力损失、重构损失、文本匹配损失、对抗损失和感知损失对应的权重。

2 实验结果与分析

2.1 数据集和实验设置

本文以敦煌壁画数据集为主要来源，构成了

自制敦煌壁画数据集，共 12000 张，其中 70%用于训练，30%用于测试。文中所使用的敦煌壁画文本数据集主要来源于对《敦煌壁画解读》书籍的文本描述，构成了与壁画图像对应的文本数据集。此外，为了减少壁画文本描述切词的歧义性，本文采用 BiLSTM 更易处理的英文词嵌入方法，直接将描述单词转换为向量的方式进行处理。

对于壁画修复任务来说，采用大小不同，形态各异的掩膜更能模拟壁画的真实破损情况，更具备现实意义。为了验证所提模型对于不同程度破损壁画的修复效果，使得修复实验更加贴近真实场景，因而对壁画采取中心掩膜破损、随机不规则掩膜破损以及真实破损壁画进行修复实验。随机不规则掩膜使用不规则掩膜图像数据集[21]，并使用学习率为 10^{-4} 的 Adam 优化器进行训练。同时分别与文献[13]基于边缘轮廓的修复方法，文献[15]基于纹理结构引导的修复方法和文献[17]文本引导的修复方法进行实验对比分析。实验环境为 Intel(R) Core i7-10700K CPU，32.0 GB RAM，NVIDIA GeForce RTX 2060 SUPER，使用 pycharm 搭建的 python3.7 及 Pytorch1.13 的深度学习框架。

2.2 壁画中心掩膜修复实验

为了模拟壁画大面积破损，首先进行中心掩膜破损修复实验，验证所提算法的修复性能，结果如图 8 所示。图 8(a)为文本描述，图 8(b)为壁画原图，图 8(c)为壁画掩膜图像，图 8(d)是文献[13]基于边缘轮廓引导的修复结果，由于该方法未充分考虑壁画图像的语义信息，且修复结果过度依赖一阶段的边缘轮廓信息，在中心掩膜修复过程中，未能拟合边缘轮廓，导致存在修复未完成的问题。如图 8 中前三幅佛像均存在修复未完成的现象。图 8(e)是文献[15]纹理结构联合引导的壁画修复算法，由于该方法在壁画修复的过程中对全局语义考虑不足，导致修复结果出现了语义断裂、内容模糊的问题，如第一幅壁画存在语义断裂，第二幅壁画存在语义内容模糊的问题。图 8(f)是文献[17]文本引导图像修复算法，相较文献[13]和文献[15]在语义修复方面取得了较好的效果，但是该方法在修复过程中忽略了图像细节信息的修复，导致修复结果存在细节缺失、纹理模糊的问题，如第二幅和第五幅壁画存在纹理模糊的问题，第四幅壁画存在细节缺失的问题。图 8(g)为本文结果，可以看出较对比算法，所提方法取得了更好的视觉修复效果，是因为本文首先用文本引导壁画图像修复，其次再对壁画图像的纹理细节进行修复，较好的解决了修复结果语义不一致，纹理细节缺失的问题。

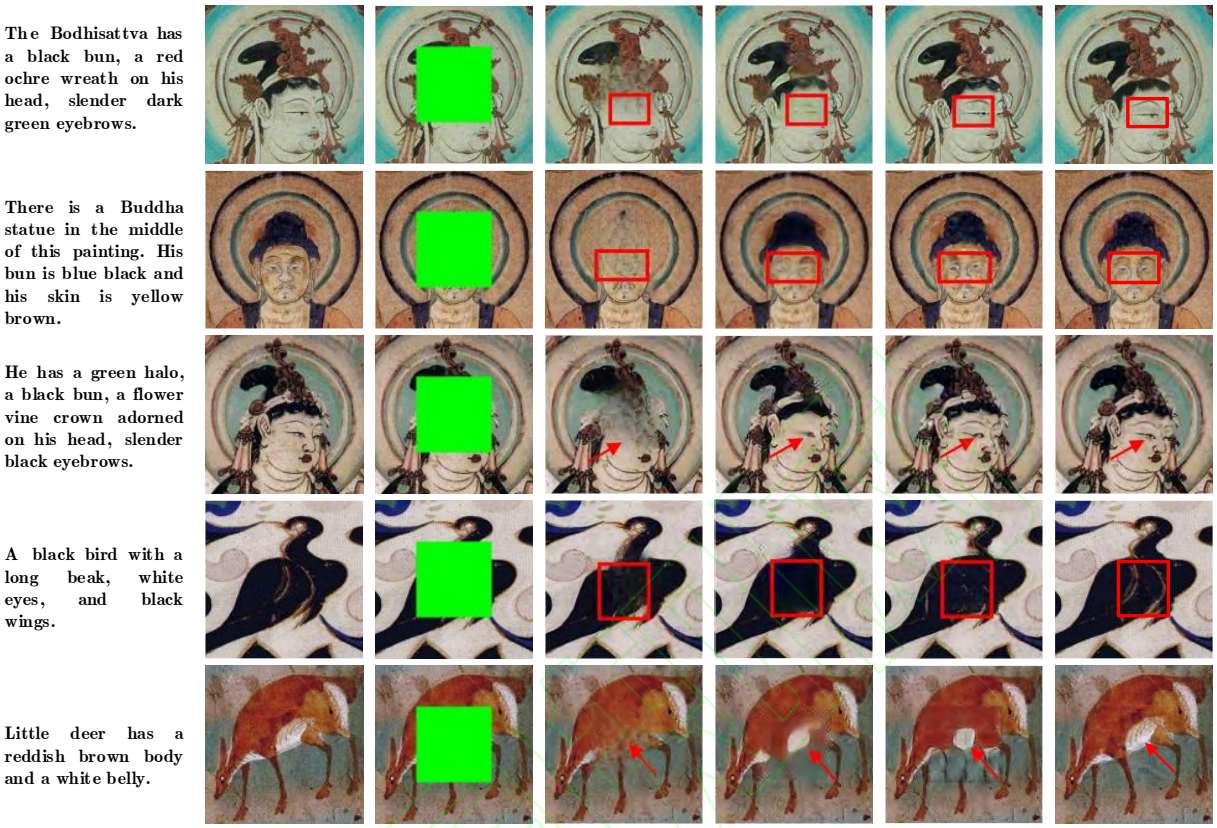


图 8 壁画中心破损修复实验

Fig. 8 Comparison of Repair Experiments Results for Center Damaged Murals

为了对图 8 进行定量评价，采用峰值信噪比 PSNR 和结构相似性 SSIM 对比分析。这两个指标的值越高，表明修复效果越好。由表 1 可以看出，本文算法相较对比算法取得了较好的客观定量评价指标。可以表明，本文算法能够较好地完成中心破损壁画的修复，且取得可较好的视觉和客观效果。

表 1 不同算法对中心破损壁画修复结果 PSNR 和 SSIM 对比

Table 1 Comparison of PSNR and SSIM Inpainting Results of Center Damaged Murals using Different Algorithms

壁画	文献[13]		文献[15]		文献[17]		本文算法	
图像	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
1	22.1388	0.8634	26.2267	0.8961	28.1157	0.9069	**30.2035**	**0.9362**
2	19.7033	0.8725	25.9058	0.9096	26.0267	0.9073	**30.4586**	**0.9598**
3	20.8262	0.8395	25.3996	0.8531	24.7835	0.8799	**29.9257**	**0.9352**
4	24.2441	0.8814	27.0957	0.8889	26.7259	0.8897	**28.9625**	**0.9305**
5	24.2835	0.8960	27.8531	0.9102	26.2760	0.9033	**29.9867**	**0.9402**

2.3 壁画随机破损掩膜修复实验

其次，为了模拟对于不同程度破损壁画的修复性能，下面对壁画图像进行随机破损掩膜修复实验，结果如图 9 所示。图 9(a)为文本描述，图 9(b)为真实图像，图 9(c)为掩膜图像，图 9(d)为文献[13]边缘轮廓引导修复方法，图 9(e)为文献[15]

结构纹理联合修复方法，图 9(f)为文献[17]文本引导修复方法，图 9(g)为本文方法。由修复结果可以看出，图 9(d)边缘轮廓引导修复结果中，第一幅鹦鹉图和第三幅双鸽图存在结构断裂的问题，第二幅存在语义错误的问题。图 9(e)结构纹理联合修复结果中第一幅、第四幅和第五幅壁画存在棋盘格和块效应。图 9(f)文本引导的修复方法引入了外部文本语义信息，但由于在文本特征和图像特征融合过程中，忽略了文本冗余特征的干扰性，且未考虑壁画图像的细节修复，导致修复结果存在像素模糊，细节缺失的问题，如第一幅图像存在像素模糊的问题，第三幅存在细节缺失的问题。图 9(g)为本文算法的修复结果，本文算法首先利用文本过滤模块过滤掉文本干扰信息，其次在引导修复后进一步对细节信息进行修复，所以相较其他对比算法取得了较好的视觉修复效果。

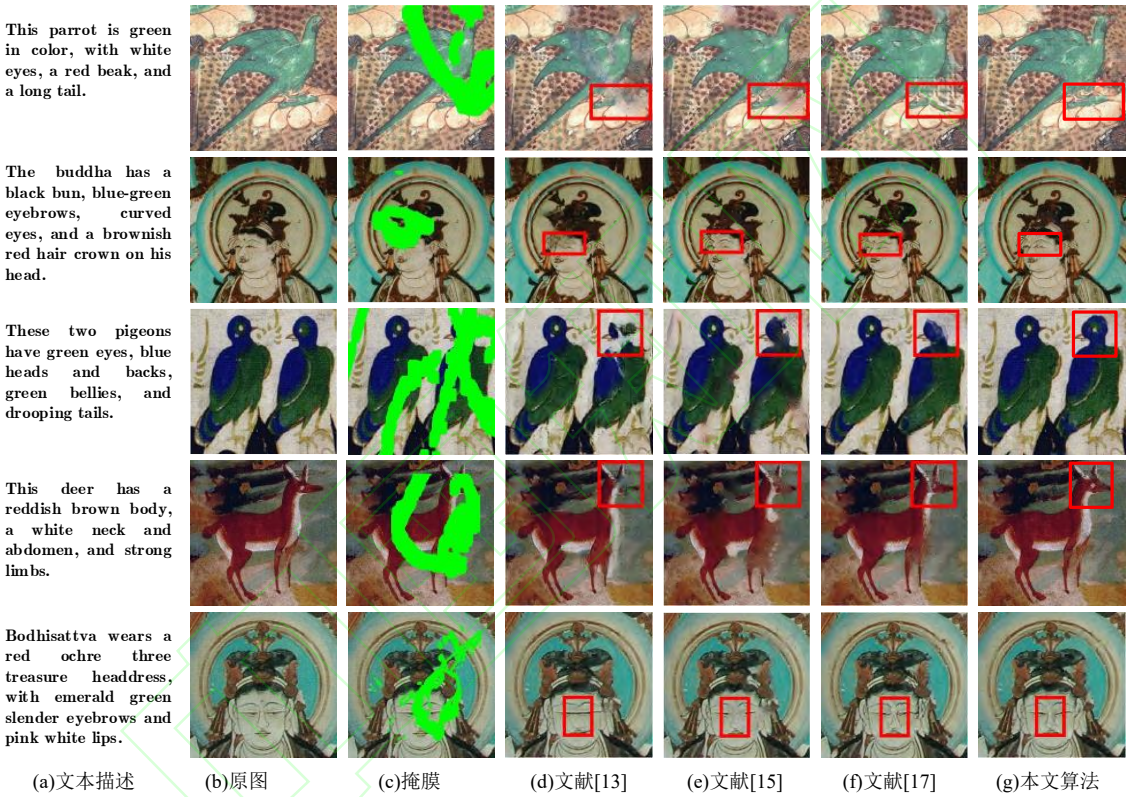


图 9 壁画随机破损修复实验结果

Fig. 9 Comparison of Repair Experiments Results for Random Damaged Murals

对图 9 定量评价，如表 2，可以发现，本文算法较对比算法均更优，验证了本文的有效性。

表 2 不同算法对随机破损壁画修复结果 PSNR 和 SSIM 对比

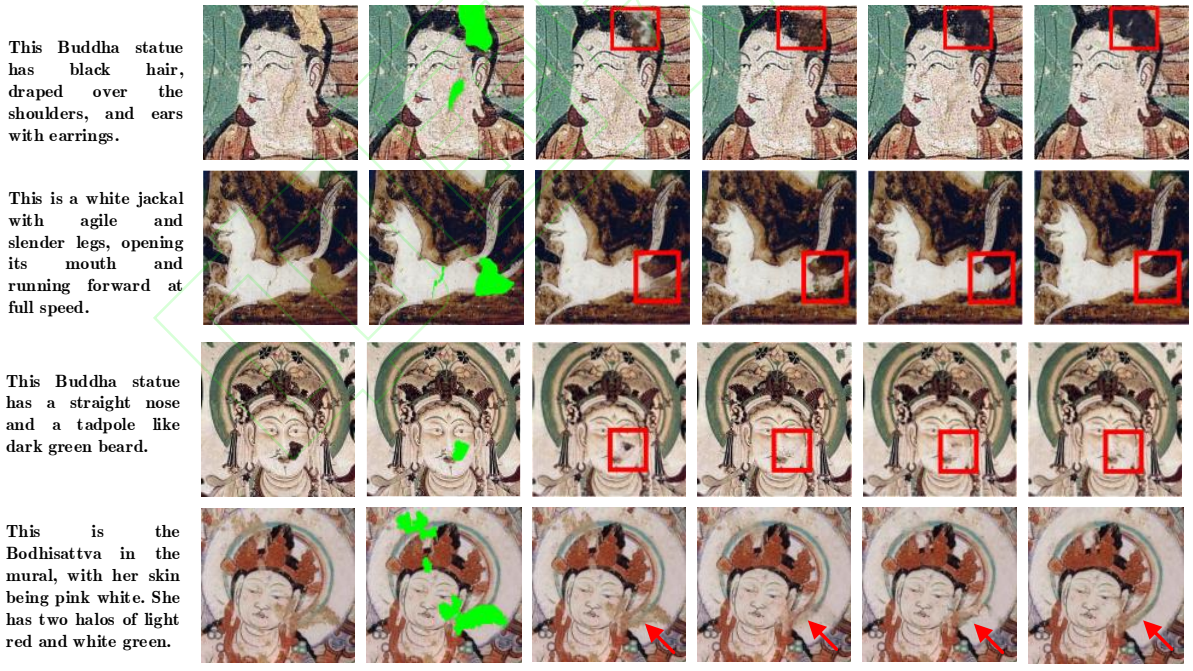
Table 2 Comparison of PSNR and SSIM Inpainting Results of Random Damaged Murals using Different Algorithms

壁画	文献[13]		文献[15]		文献[17]		本文算法	
图像	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
1	25.6819	0.8865	25.5076	0.9070	25.6659	0.8842	**26.6138**	**0.9185**
2	26.9130	0.8561	30.7534	0.9661	29.2837	0.9622	**33.2862**	**0.9776**
3	19.7242	0.8135	23.3606	0.8656	23.7827	0.8757	**27.2031**	**0.9310**
4	25.7038	0.8821	25.5672	0.9029	28.6264	0.9381	**30.8953**	**0.9597**
5	29.8625	0.9545	30.9278	0.9478	31.1542	0.9502	**33.4721**	**0.9802**

对比表 1 和表 2 可以发现，本文方法对于壁画中心破损的修复量化结果较随机破损修复结果有更好的性能提升。其主要原因为：虽然中心掩膜破损区域较大，但其破损位置相对集中且位置固定，采用所提方法中的文本语义过滤模块，更易定位于破损区域，利用注意力过滤机制，可以实现对于固定破损区域的精准引导。而对于随机破损修复的实验，其破损位置随机生成，且破损形状、大小和位置均具有不确定性，这种不确定性增加了修复的难度。此外，在文本辅助引导修复过程中，对于输入的文本描述基本以壁画图像中心主体为描述，因而对于处于壁画图像中心的中心破损壁画修复，文本辅助引导能够对中心主体破损起到更好的修复作用。而对于位置和形状不定的随机掩膜破损的壁画，文本对于随机破损壁画的文本描述较中心描述更有限，因而本文方法对于随机破损修复性能较中心破损结果略有逊色。

2.4 壁画真实破损修复实验

最后进行真实破损壁画修复实验，修复结果如图 10 所示，图 10(a)为壁画对应的文本描述；图 10(b)为真实破损的壁画原图，图 10(c)为真实破损区域掩膜图，图 10(e)-(g)为文献[13]、文献[15]和文献[17]的修复结果。由修复结果可以看出，图 10(d)边缘轮廓引导修复方法在第一幅、三幅和第四幅壁画图像存在修复未完成的问题。图 10(e)为结构纹理联合约束的修复结果，如第二幅图像存在拟合未完成的问题，第四幅壁画图像存在结构断裂的问题。图 10(f)为文本引导图像修复方法的结果，第二幅图像破损区域存在错误像素溢出的问题，第三幅壁画图像存在细节缺失的问题。图 10(g)为本文算法的修复结果，由于本文算法在壁画修复过程中既关注了壁画的语义信息又考虑到了壁画图像的细节信息，所以本文算法取得了更好的修复效果。



(a)文本描述 (b)原图 (c)掩膜 (d)文献[13] (e)文献[15] (f)文献[17] (g)本文算法

图 10 真实破损壁画修复实验结果

Fig. 10 Comparison of Repair Experiments for Real Damaged Murals

对于真实破损壁画修复结果的评价，由于缺少相应的标准参考图像，故采用无参考主观评价指标(Mean Opinion Score, MOS)对修复后的壁画进行评价[22]。 MOS 值越大，表明视觉修复效果越好。不同修复方法 MOS 评价结果如表 3 所示，可以看出本文方法评价优于对比方法，说明其修复后视觉连贯性更强。综合该评价，表明所提方法 MOS 指标均优于对比方法，验证了对于真实

壁画修复的有效性。

表 3 真实破损壁画修复结果 MOS 评价对比

Table 3 Comparison of MOS Evaluation of the Inpainting Results of Real Damaged Murals

壁画图像	文献[13]	文献[15]	文献[17]	本文算法
1	1.9333	3.2000	3.7333	4.4666
2	2.9233	2.7333	2.7666	3.9333
3	2.2666	2.3333	3.6000	3.7333
4	2.3333	2.4666	2.5333	3.8666

3 结 语

针对现有算法在壁画修复过程中仅考虑壁画先验信息对破损区域的修复，缺乏文本信息引导性修复，导致修复结果存在语义不一致、细节缺失的问题，提出了一种基于文本模态辅助引导的壁画修复算法，引入文本信息，指导壁画缺失区域的修复。构建了文本语义过滤模块，得到壁画缺失区域的文本语义，利用该语义指导壁画修复，解决修复结果语义不一致的问题。其次，设计语义增强模块，提高文本语义与图像语义的一致性。然后，针对文本引导修复过程中可能会丢失图像细节的问题，设计上采样纹理细节修复网络，通过深浅层特征的双向融合，对壁画细粒度特征进行修复，有效保持修复结果的细节完整性。最后，采用谱归一化判别器对修复后的壁画图像进行博弈对抗，得到修复后的壁画图像。通过对真实破损壁画修复实验，表明所提方法取得了更好地视觉感和协调性，在主客观评价方面均优于比较算法。

参考文献

[1] LI Qingquan, WANG Huan, ZOU Qin. A Murals Inpainting Algorithm Based on Sparse Representation Model[J]. *Geomatics and Information Science of Wuhan University*, 2018, 43(12): 1847-1853.(李清泉, 王欢, 邹勤. 一种基于稀疏表示模型的壁画修复算法[J]. 武汉大学学报 (信息科学版), 2018, 43(12): 1847-1853.)

[2] Chen Yong, Ai Yapeng, Guo Hongguang. Improved Curvature-driven Model of Dunhuang Mural Restoration Algorithm[J]. *Journal of Computer Aided Design and Graphics*,2020,32(5):787-796.(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.)

[3] Zhang X B, Zhai D H, Li T R, et al. Image Inpainting based on Deep Learning: A review[J]. *Information Fusion*, 2023, 90: 74-94.

[4] Li R, Zheng B. A Spatially Adaptive Hybrid Total Variation Model for Image Restoration Under Gaussian Plus Impulse noise[J]. *Applied Mathematics and Computation*, 2022, 419: 126862 -126883.

[5] Jing Huiying, Liu Xinyi, Zhang Yongjun, et al. Low-Rank Matrix Aided Automatic Texture Inpainting of Building Facades from UAV Images [J]. Geomatics and Information Science of Wuhan University,2023,DOI:10.13203/J.whugis20220399 (景慧莹,刘欣怡,张永军,等.低秩矩阵辅助的无人机影像建筑物立面纹理自动修复[J/OL].武汉大学学报(信息科学版),2023,DOI:10.13203/J.whugis20220399.)

[6] LI Xiaolin, LI Gang, ZHANG Enqi, et al. Determinant Point Process Sampling Method for Text-to-Image Generation[J]. Geomatics and Information Science of Wuhan University, 2024, 49(2): 246-255.(李晓霖, 李刚, 张恩琪, 等. 行列式点过程采样的文本生成图像方法[J]. 武汉大学学报 (信息科学版), 2024, 49(2): 246-255.)

[7] Han R Y, Liu X, Liao S H, et al. Adaptive Image Inpainting Algorithm Based on Sample Block by Kriging Pretreatment and Facet Model[J]. *Journal of Electronic Imaging* 30 (2021): 043021 - 043021.

[8] Zeng Y H, Fu J L, Chao H Y, et al. Learning Pyramid-context Encoder Network for High-quality Image Inpainting[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 1486-1494.

[9] Xiao Q G, Li G Y, Chen Q C. Image Inpainting Network for Filling Large Missing Regions using Residual Gather[J]. *Expert Systems with Applications*, 2021, 183: 115381.

[10] Li J Y, He F X, Zhang L F, et al. Progressive Reconstruction of Visual Structure for Image Inpainting[J], *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), 2019:5961-5970.

[11] Suvorov R, Logacheva E, Mashikhin A, et al. Resolution-robust Large Mask Inpainting with Fourier Convolutions[C]//Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2022: 2149-2159.

[12] Zheng C X, Song G X, Cham T J, et al. Bridging Global Context Interactions for High-fidelity Image Completion[C].Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (CVPR), New Orleans, LA, USA, 2022:

11502-11512.

[13] Nazeri K, Ng E, Joseph T, et al. Edgeconnect: Generative Image Inpainting with Adversarial Edge Learning[J]. *Computer Vision and Pattern Recognition*,2019: 3265-3274.

[14] Zhao Y N, Price B, Cohen S, et al. Guided Image Inpainting: Replacing an Image Region by Pulling Content from Another Image[C]. 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2019：1514-1523.

[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, 2021:14114-14123.

[16] Cao C J, Fu Y W. Learning a Sketch Tensor Space for Image Inpainting of Man-made Scenes[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 14509-14518.

[17] Zhang L S, Chen Q C, Hu B T, et al. Text-guided Neural Image Inpainting[C]. Proceedings of the 28th ACM International Conference on Multimedia. 2020: 1302-1310.

[18] Yu Kai, Bin Yi, Zheng Ziqiang, et al. Text-to-image Generation with Conditional Semantic Augmentation[J]. *Journal of Software*. 2024,35(5):2150-2164. (余凯,宾燚,郑自强,等.基于条件语义增强的文本到图像生成[J].软件学报,2024,35(5):2150-2164.)

[19] Lin Q, Yan B, Li J C, et al. MMFL: Multimodal Fusion Learning for Text-Guided Image Inpainting[C]. Proceedings of the 28th ACM international conference on multimedia. 2020: 1094–1102.

[20] Tao X, Zhang P C, Huang Q Y, et al. Attngan: Fine-grained Text to Image Generation with Attentional Generative Adversarial Networks[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 1316-1324.

[21] Liu G L, Read F A, Shih K J, et al. Image Inpainting for Irregular Holes using Partial Convolutions[M]. Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018:89-105.

[22] Ye Yuqi, Design and Realization of No-reference Thangka Image Quality Evaluation System[D]. Lanzhou: Northwest Minzu University,2020.(叶雨琪. 无参考唐卡图像质量评价系统的设计与实现[D]. 兰州: 西北民族大学,2020.)

网络首发：

标题：文本模态辅助引导的壁画修复算法

作者：陈永，杜婉君，张世龙

收稿日期：2025-01-28

DOI:10.13203/j.whugis20240251

引用格式：

陈永，杜婉君，张世龙. 文本模态辅助引导的壁画修复算法[J].武汉大学学报（信息科学版）, 2025, DOI:10.13203/J.whugis20240251（CHEN Yong, DU Wanjun, ZHANG Shilong. Text Modality Assisted Guided Mural Inpainting Algorithm[J]. Geomatics and Information Science of Wuhan University, 2025, DOI:10.13203/J.whugis20240251）

武汉大学学报(信息科学版), 2024, 49(3): 376-388.
http://ch.whu.edu.cn/article/doi/10.13203/j.whugis20220613

面向城市增强现实信息标注的建筑物场景结构提取方法

徐旺, 游雄, 张威巍, 陈冰, 胡宗敏
武汉大学学报(信息科学版), 2023, 48(6): 926-935.
http://ch.whu.edu.cn/article/doi/10.13203/j.whugis20200373