



多阶段堆叠融合下的敦煌壁画修复

鄢开俊¹, 单宏全^{1*}, 田彬¹, 俞天秀², 周颜林^{1,2}

(1. 兰州交通大学 电子与信息工程学院, 兰州 730070; 2. 敦煌研究院 文物数字化研究所, 敦煌 730030)

摘要 为解决敦煌破损壁画修复难的问题, 提出一种多阶段堆叠融合下的壁画修复模型, 在第一阶段进行粗修复以及最大化地进行所需特征信息的提取时采用 Unet 结构, 为增强该阶段的修复效果, 又加入对抗网络 (LsGan); 在第二阶段进行精细化修复, 主要修复纹理以及细化遮挡区域, 引入多头连接和多尺度分支堆叠模块并对其进行改进, 对壁画进行多阶段信息提取并对其进行融合; 最后为了修复全局细节, 采用大感受野的编解码器结构并且引入轻量级通用可视化 (Mobile ViT) 模块, 同时分离了壁画图像的通道并进行分析, 引入对通道不敏感的极化注意力。实验结果表明, 该模型很好地解决了壁画纹理以及细节修复的难题, 相较于所选修复算法最优数据, 在掩码面积为 5%~20% 时, 峰值信噪比提高了 3.312, 在掩码面积为 5%~80% 时, 平均峰值信噪比提高了 1.02。

关键词 敦煌壁画修复; 多阶段堆叠; 大感受野; 多头嵌套; 多分支融合

中图分类号 TP391.41 文献标志码 A DOI 10.12178/1001-0548.2023173

Inpainting of Dunhuang Murals Under Multi-Stage Stacking Fusion

WU Kaijun¹, SHAN Hongquan^{1*}, TIAN Bin¹, YU Tianxiu², and ZHOU Yanlin^{1,2}

(1. School of Electronics and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China;

2. Institute of Cultural Digitization, Dunhuang Research Institute, Dunhuang 730030, China)

Abstract A fresco restoration model under multi-stage stacking fusion is proposed to solve the problem of difficult restoration of broken frescoes at Dunhuang. The Unet structure is used for coarse repair in the first stage and for maximizing the extraction of the required feature information, and the adversarial network (LsGan) is added to enhance the repair effect in this stage. In the second stage, the restoration is refined, mainly by restoring textures and refining occluded areas, introducing and improving the multi-head connection and multi-scale branching stacking modules, extracting and fusing multi-stage information from the murals. Finally, in order to consider the global details of the fix, the codec structure of the big feeling field is used and the mobile ViT module is introduced. The channels of the mural images are also separated and analyzed, introducing polarization attention that is insensitive to the channels. The experimental results show that the model can solve the difficult problem of fresco texture and detail restoration well. Compared with the optimal data of the selected restoration algorithm, the peak signal-to-noise ratio improved by 3.312 at a mask area of 5% to 20%, and the average peak signal-to-noise ratio is improved by 1.02 at a mask area ranging from 5% to 80%.

Key words Dunhuang mural inpainting; multi-stage stacking; large sensory field; multi-head nesting; multi-branch fusion.

由于敦煌壁画年代久远以及对古代壁画艺术知识的缺失, 传统的手工修复方法已经达不到人们想要的效果。数字化修复方法的出现, 正好弥补了这一缺失。数字化方法衍生出的修复方法大致分为三大类: 基于扩散和纹理的修复方法、基于补丁的修复方法、基于深度学习的修复方法。基于扩散和纹理

的修复方法主要用来修复破损面积较小的图像, 主要算法有偏微分方程修复 (Bertalmio-Sapiro-Caselles-Bellester, BSCB)、曲率扩散 (CDD) 及全变分 (TV) 等。如文献 [1-2] 提出的 BSCB 算法使用偏微分方程来描述图像中的缺失区域和纹理以及形状特征, 进而将缺失图像区域补全, 该方法可以有效

收稿日期: 2023-06-19; 修回日期: 2024-05-26

基金项目: 甘肃省自然科学基金 (21JR7RA300)

作者简介: 鄢开俊, 教授, 主要从事计算机视觉以及图像处理方面的研究。

*通信作者 E-mail: 1554435955@qq.com

处理复杂细小纹理特征,但计算量较大。文献 [3-4] 提出补丁优先级措施来更好地填充图像缺失区域并恢复图像的结构特征,但边缘存在不连续现象。为了解决这个问题,文献 [5] 提出了新补丁生成算法,该算法使用全局和局部上下文鉴别器来确保生成的补丁符合缺失区域。以上修复算法对于破损严重和纹理复杂的壁画来说,修复结果依旧存在肉眼可见的斑迹。

随着深度学习的出现,各种各样的方法被提出并用于图像修复^[6-10]。基于修复任务的一个主要方面是从已知图像内容中去预测未知区域的像素值,使得生成的像素值和纹理信息更贴近真实图像。文献 [6] 采用内容重建图像实现了缺失区域的预测,并且提出使用生成对抗网络(GAN)通过对图像打掩码来修复破损图像。然而该网络生成的补丁较为模糊,文献 [7] 又将现有的修复方法从粗到细的框架结构进行了优化,提出一种轻量级对抗并发编码器架构,该网络采用不同大小的感受野去提取特征进行融合,使得特征能够长距离传输并且聚焦图像的重要信息。文献 [11] 提出了一种基于相似图像配准的图像修复算法。首先结合图像的相似度计算和特征匹配为修复过程提供有效信息,然后利用图像单应性变化以及图像最佳匹配块实现图像的最终修复。文献 [12] 结合 Canny 边缘算子损失,优化了网络结构,修复缺失图像的纹理边缘信息。文献 [13] 采用特征重排进一步提高了修复内容的真实性,减少了边界伪影。随后多头自注意力 transformer^[14-15] 的出现,结合位置编码信息提取关键特征并进行融合,解决了特征长距离传输问题。文献 [16] 提取图像的通道特征或空间特征,辅助模型提升修复精度。

文献 [17] 为了解决壁画修复中结构信息与纹理信息的差异性,提出多尺度轮廓波分解的壁画修复方法,文献 [18] 提出了一种生成鉴别器的修复网络,首先对判别器网络进行优化,然后在损失函数计算阶段强调下采样部分的损失,使得修复效果更优。文献 [19] 为提高修复精度和速度,提出了一种既具有高计算效率又具有高恢复质量的有效方法。通过几个参数化线性滤波器以及几个参数化影响函数来扩展传统的非线性反应扩散模型。随后文献 [20] 提出一种基于线描图的计算机辅助壁画修复,以壁画线描稿为参照,很好地解决了壁画修复中存在的纹理模糊问题。文献 [21] 为解决该类问

题又提出全局一致性与局部一致性的修复方法。

在上述研究内容的基础上,本文设计了多阶段堆叠融合下的敦煌壁画修复模型。网络的第一阶段采用 Unet^[22] 结构尽可能地提取有效信息,使得损失信息较少,同时增加 LsGan 判别器,指导第一阶段特征信息的真实性。在第二阶段,提出多分支堆叠融合模块嵌入模型,同时受该模块对壁画信息堆叠融合的冗余,采用 Reconv^[23] (不提高计算量的同时可以提高模型精度的卷积)对多分支信息进行融合。同时在网络中采用 Mobile ViT Block 结构,通过研究壁画图像的特征,采用 PSA^[24] (序列极化自注意力),使得壁画修复效果得到显著提升。

1 模型介绍

1.1 多阶段堆叠融合修复模型

整体上,本文采用三阶段级联融合的网络修复框架,第一阶段由一个 Unet 结构和一个 LsGan 结构的判别器组成,该阶段主要作用是提取全局语义信息并进行简单的粗修复任务。第二阶段采用简单卷积结构进行局部细化,其中包含 4 个下采样层、4 个 Multi-Concat-blocks 层以及 4 个上采样层,主要作用是在第一阶段的修复基础上进一步提取图像的纹理以及深层次的语义信息进行细化修复。第三阶段采用编解码器结构进行全局细节补全,结合前两个修复阶段,更进一步修复遮挡壁画区域的整体细节部分。在该模型中,每个阶段的输出首先与原始不完整图像合并,然后与二进制掩码连接作为下一阶段的输入,整体模型结构如图 1 所示。

1.2 一阶段粗修复网络

第一阶段为粗修复网络阶段,主要作用为最大化地提取图像的特征信息。本阶段采用一种带跳跃连接的编码器-解码器结构,该网络由 8 个上下采样层组成,采用跳跃连接将图像信息经编码器传递到解码器部分,以恢复下采样期间丢失的图像信息。解码器端的感受野已经远远大于输入图像大小(256×256),大的感受野更有利于后阶段的图像修复工作。最后经过一个 LsGan 鉴别器结构以输出最贴近真实的修复效果图,该鉴别器以真实图像和生成图像为输入,输出形状为 32×32 的特征图,这个特征映射中的每个元素都被区分出真假。该阶段的训练损失函数由像素级重建和对抗损失组成,其中使用加权 L1 损失进行像素重建:

$$L_{\text{valid}}^C = \frac{1}{\text{sum}(\mathbf{I} - \mathbf{M})} |(\mathbf{I}_{\text{out}}^C - \mathbf{I}_{\text{gt}}) \odot (\mathbf{I}_1 - \mathbf{M})|$$

$$L_{\text{hole}}^C = \frac{1}{\text{sum}(\mathbf{M})} |(\mathbf{I}_{\text{out}}^C - \mathbf{I}_{\text{gt}}) \odot \mathbf{M}| \quad (1)$$

式中, L_{valid}^C 表示有效像素; \mathbf{I}_{gt} 为 ground-truth 图像; \odot 为逐元素积运算; $\text{sum}(\mathbf{M})$ 为 \mathbf{M} 中非零元素的个数; \mathbf{I}_1 表示第一阶段输出; \mathbf{I} 表示单位矩阵。

逐像素重构损失的表达式为:

$$L_r^C = L_{\text{valid}}^C + \lambda_h L_{\text{hole}}^C \quad (2)$$

式中, λ_h 为平衡因子; L_{hole}^C 为掩码部分像素值。对于 GAN 的损耗, 使用最小二乘损耗, 第一阶段修复网络和鉴别器对应的损失函^[21]数为:

$$\mathbf{I}_{\text{mer}}^C = \mathbf{I}_{\text{in}} \odot (\mathbf{I} - \mathbf{M}) + \mathbf{I}_{\text{out}}^C \odot \mathbf{M} \quad (3)$$

$$L_G^C = E_{\mathbf{I}_{\text{mer}}^C} \sim p_{\mathbf{I}_{\text{mer}}^C}(\mathbf{I}_{\text{mer}}^C) \left[(D(\mathbf{I}_{\text{mer}}^C) - 1)^2 \right] \quad (4)$$

$$L_D = \frac{1}{2} E_{\mathbf{I}_1} \sim p_{\text{data}}(\mathbf{I}_1) \left[(D(\mathbf{I}_{\text{gt}}) - 1)^2 \right] + \frac{1}{2} E_{\mathbf{I}_{\text{mer}}^C} \sim p_{\mathbf{I}_{\text{mer}}^C}(\mathbf{I}_{\text{mer}}^C) \left[(D(\mathbf{I}_{\text{mer}}^C))^2 \right] \quad (5)$$

式中, $\mathbf{I}_{\text{mer}}^C$ 是 ground-truth (真值图) 图像和掩码推理融合后的图像; $\mathbf{I}_{\text{out}}^C$ 为第一阶段的输出图像; D 代表鉴别器, 用来判断第一阶段输出图像的真假; L_D 为真实图与修复后的融合图的概率分布; $\frac{1}{2}$ 为融合系数。因此, 整个第一阶段的损失函为: $L_C = L_{\text{valid}}^C + \lambda_h L_{\text{hole}}^C + \lambda_g L_G^C$, 其中 $\lambda_h=6$, $\lambda_g=0.1$ 。

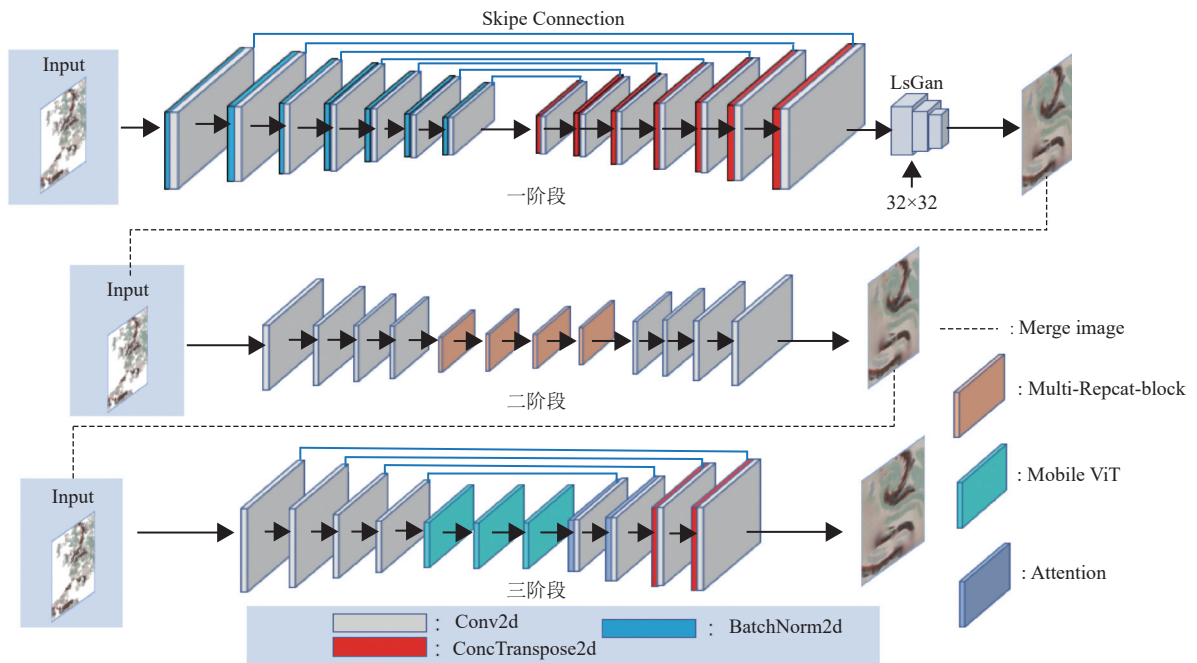


图1 整体框架图

1.3 二阶段粗修复网络

第二阶段为细化修复网络, 为此设计了一个深浅相结合的网络模型。局部细化网络包括4次下采样工作、4个 Multi-Concat-Block (主要结构如图2所示)、4次上采样工作, 同时采用滑动窗口的方式, 对第一阶段粗修复结果进行局部细节处理, 堆叠了4层 Multi-Concat-Block^[23], 并对其进行了改进。为了更加平衡模型的训练时间以及参数量, 将原有的 Conv 卷积替换成 Repconv, 该卷积在推理阶段通过堆叠多个 Repconv 卷积块以此来达到精细化提取特征的目的, 并且在推理时将分支的参数重

参数化到主分支上以减少计算量和内存消耗。如图2中 Multi-Repeat-blocks 块, 其中 Repconv 采用 Conv-3×3 与 Conv-1×1 的卷积大小, 使得模型可以提取不同尺度的特征信息, 最后通过 Fuse-RepVGG 进行融合, 整体模型依旧采用 Unet 结构。

第二阶段的训练目标为加权重构损失 L_r^L , 与式(2)相同, 根据文献[13], 总变差损失 (Total Variation, TV) 被用作平滑损失激励, 其表达式为:

$$L_{\text{tv}}^L = | \mathbf{I}_{\text{mer}}^L(i, j+1) - \mathbf{I}_{\text{mer}}^L(i, j) |_1 + | \mathbf{I}_{\text{mer}}^L(i+1, j) - \mathbf{I}_{\text{mer}}^L(i, j) |_1 \quad (6)$$

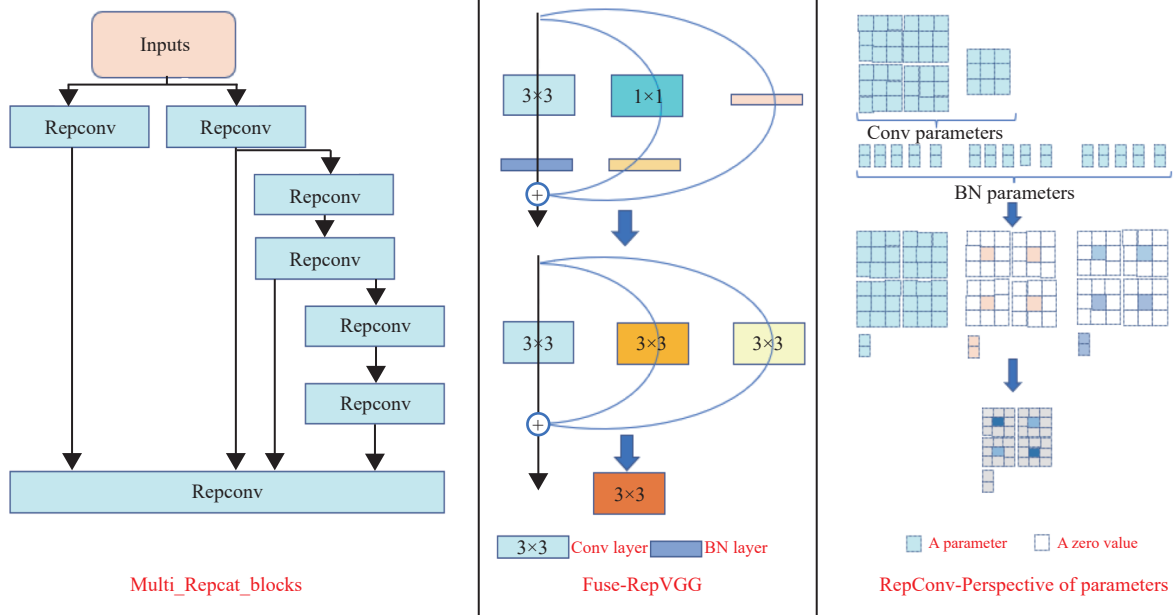


图 2 Multi-Concat-block 模型图

L_{mer}^L 的计算方式与式 (3) 相似, 此外, 还使用感知损失和风格损失来更好地恢复结构和纹理信息。与上述像素空间中进行的像素级重构损失和 TV 损失不同, 感知损失和风格损失是在特征空间中计算的, 感知损失表达式为:

$$L_{\text{per}}^L = \sum_i |f_i(I_{\text{out}}^L) - f_i(I_{\text{gt}})|_1 + |f_i(I_{\text{mer}}^L) - f_i(I_{\text{gt}})|_1 \quad (7)$$

式中, f_i 为训练 VGG16^[25] 中的第 i 层的特征映射。

风格损失为:

$$L_{\text{style}}^L = \sum_i |g_i(I_{\text{out}}^L) - g_i(I_{\text{gt}})|_1 + |g_i(I_{\text{mer}}^L) - g_i(I_{\text{gt}})|_1 \quad (8)$$

式中, $g_i = f_i(\cdot) f(\cdot)^T$ 为 Gram 矩阵, 综上所述, 第二阶段总的损失函数为:

$$L_L = L_{\text{valid}}^C + \lambda_h L_{\text{hole}}^L + \lambda_{\text{tv}} L_{\text{tv}}^L + \lambda_{\text{per}} L_{\text{per}}^L + \lambda_{\text{style}} L_{\text{style}}^L \quad (9)$$

在图像修复中, 文献 [22, 26] 是结合加权重建损失、感知损失、风格损失、TV 损失对图像修复网络进行训练的约束者。文献 [26] 通过对验证图像进行超参数搜索来选择相应的权值, 实验发现该参数可以复用, 因此采用同样的权重设置, $\lambda_h = 6$, $\lambda_{\text{tv}} = 0.1$, $\lambda_{\text{per}} = 0.05$, $\lambda_{\text{style}} = 120$ 。

1.4 三阶段全局优化网络

第三阶段为全局细节优化网络, 该阶段为全局精细化修复。采用编解码器结构, 该部分包括 4 层下采样层、3 层 Mobile ViT 结构、4 层上采样层, 采用跳跃连接连接上下采样层, 分辨率分别为

32×32 和 64×64 。Mobile ViT 采用了 CNN 和 Transformer 的混合架构, 通过将 CNN 的优点融入 Transformer 的结构中以解决 Transformer 训练困难、迁移困难的缺点, 加快网络的推理和收敛速度, 使网络运行更加高效稳定, 并且参数量更少。

经过第二阶段细化修复之后, 在周边局部区域窗口滑动的引导下, 适当消除一些视觉伪影。然而, 当从相对较大的周围区域获取信息时, 为使长距离特征传输不损失信息, 引入 Mobile ViT。

为了更好地提高模型的修复效果, 以及由于壁画的纹理细、较密集, 经过研究对比壁画的通道特征, 发现敦煌壁画图像对通道信息并不敏感, 各个通道信息高度相似, 如图 3 所示, 因此引入极化注意力 (Sequential Polarized Self Attention)。该注意力可以对高分辨率图像的输入以及输出的长距离依赖进行特征提取, 以提高细粒度的像素语义信息。同时, 该注意力使用了极化滤波, 在通道和空间注意力计算中, 保持了高内部分辨率, 将对应的维度使用正交的方式完全折叠输入张量, 保证了低参数量的同时也保证了高通道分辨率和高空间分辨率, 并且增强了合成符合细粒度回归输出分布的非线性, 使得拟合的输出更具有细腻度、更贴近真实壁画, 不仅降低了参数量, 还提高了模型的训练速度。

Mobile ViT 结构如图 4 所示, 首先特征图通过一个 $\text{conv-}n \times n$ 的卷积对图像进行局部特征提取, 然后用 $\text{conv-}1 \times 1$ 的卷积调整特征图的通道数使其

由 C 转换为 d , 接着通过“unfold-transformer-fold” (先将输入特征展开经过 transformer 再恢复) 机制的 transformer 模块进行全局特征提取。同时使用 conv-1×1 的卷积将特征图的通道数调整回 C ,

最后借助残差分支将特征图与原始的输入图沿着通道方向拼接, 并通过 conv- $n \times n$ 的卷积进行特征融合, 得到最终特征输出。其中 MV2 块主要负责上下采样处理, 其中箭头向下表示下采样处理。

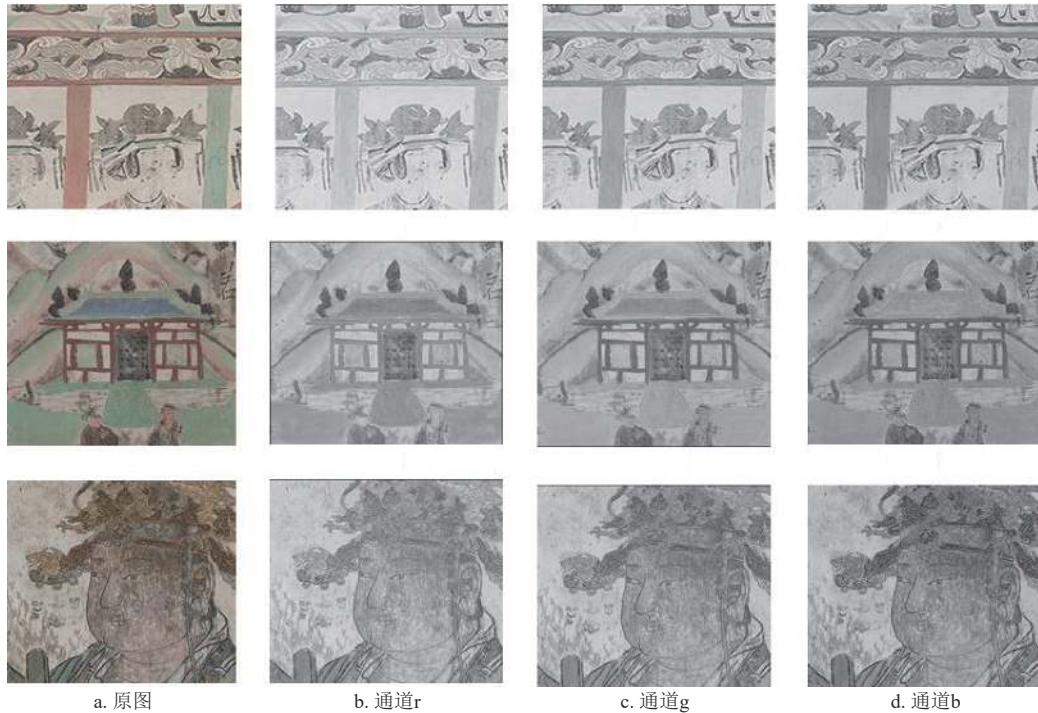


图 3 通道分离图

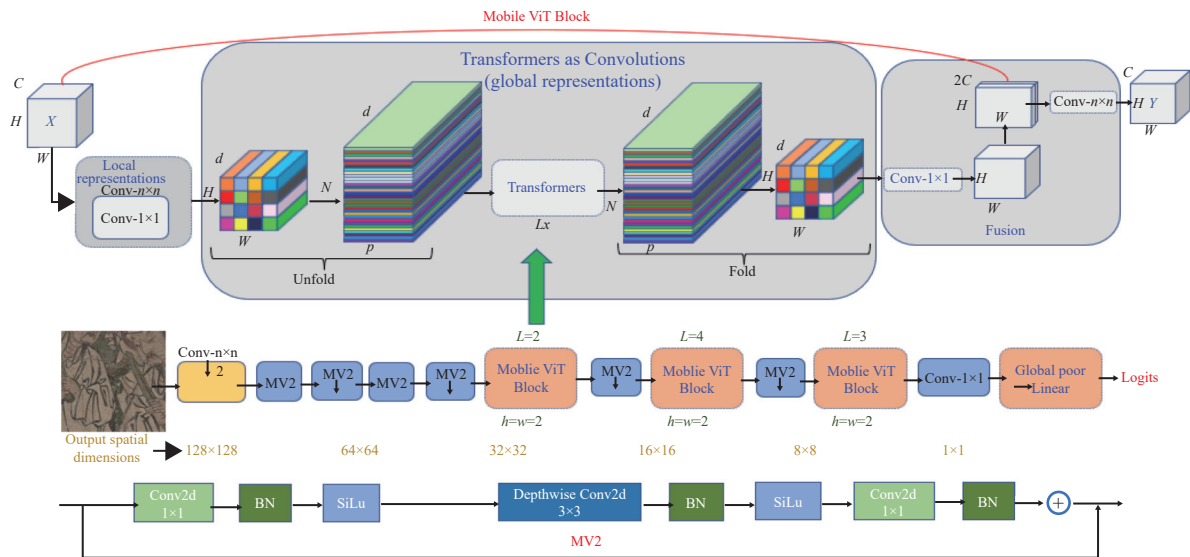


图 4 Mobile ViT 结构

在 transformer 结构中, 对于输入特征图直接展开为一维向量, 再输入 Transformer 中, 进行 self-attention, 计算时, 图中每个像素分别与其他像素进行计算。图 4 中 global representation 主要为

提取全局特征, 而在该模块中, 我们舍弃了自注意力计算的复杂性, 将特征图划分为多个 patch, 并且只对相同位置像素计算 self-attention, 这也是该模型降低训练时间的关键因素。

该阶段的训练损失函数为:

$$L_G = L_{\text{valid}}^C + \lambda_h L_{\text{hole}}^G + \lambda_{\text{tv}} L_{\text{tv}}^G + \lambda_{\text{per}} L_{\text{per}}^G + \lambda_{\text{style}} L_{\text{style}}^G \quad (10)$$

因此,总网络的损失函数为 3 个网络修复阶段的损失函数之和,即:

$$L_{\text{loss}} = L_C + L_L + L_G + L_D \quad (11)$$

注意力 PSA 的加入使得模型在细节和纹理修复方面更接近真实值。对于该注意力的具体细节可以参考文献 [24], 本文不再叙述。

2 实验

2.1 实验说明

本文所有数据集均为未经修补的原始数据集,因此数据集彩色损失严重,纹理较为模糊,其中壁画数据集包含十余种(仕女图、榜题、菩萨、莲台景观、建筑、服饰)高清版图像。大小从 512×512 到 1 024×2 048 不等,对其进行随机裁剪以及旋转等操作(如图 5 所示),随机选取了 8 000 张图像进行训练,1 000 张图像进行测试。

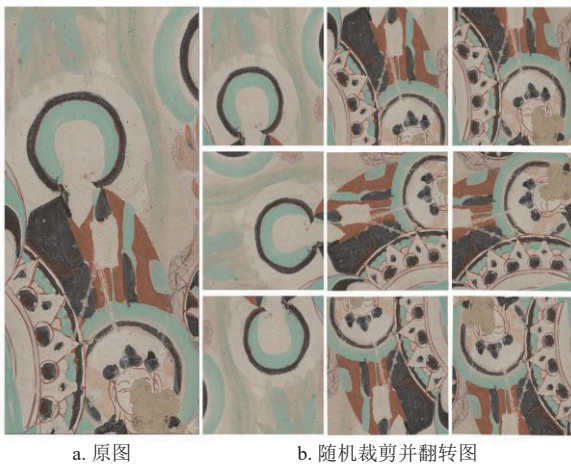


图 5 数据扩展图

为了训练网络,绘制了随机掩码(mask)数据集 10 000 余张,掩码覆盖面积为 10%~80%,并且进行了随机扩张、旋转、裁剪再次提高掩码数据集的随机性。

为验证本文提出的多阶段分支融合下的敦煌壁画修复模型的优越性,设计如下对比实验,将本文算法与以下 4 种算法进行定性以及定量分析。结果表明,在各个 mask 面积的平均指标下,本文算法修复效果最好,其中 batchsize=4,训练周期为 200 轮,在前 100 轮,学习率 lr=0.000 02,在之后

的 100 轮,学习率 lr=0.000 02,在之后的 100 轮,学习率逐渐衰减,将壁画图像统一裁剪为 256×256 像素大小。

1) Shift_Net^[13]: 一种基于 Unet 的网络结构,通过设计 shift-net 模块来使得模型的学习效果更优。

2) LGNet^[22]: 一种局部和全局细化的三阶段网络修复模型,通过引入跳跃连接以及注意力机制来增大感受野进行全局精细化修复。

3) RFR^[12]: 一种递归推理特征和知识一致性注意力(KCA)模块构建的网络,通过反复推理卷积特征图的掩码边界区域进行孔洞修复的模型。

4) CSA^[27]: 设计了新的 CSA 注意力层,该方法不仅保留上下文结构信息,还可以通过预测孔洞特征之间的语义相关性来进行更好的修复。

5) Rethinking Image Inpainting^[24]: 一种采用编解码器结构,其中解码器分为纹理部分和特征部分,将纹理和特征放在特征层进行修复,最后将两者进行融合,并且提出了双边激活函数(BPA)。

2.2 实验结果分析

1) 定量分析: 采用图像修复任务中常用的几个指标来分析验证本文模型的修复效果: L2 误差、峰值信噪比(PSNR)、结构相似性指数(SSIM)、图像相似度指标(FID),其中前两个评价指标基于像素值,后两种评价指标基于视觉感知,并且对表 1 数据的各项指标求平均数,以更直观地展示。

其中红色标记为最优指标, Avg 为各个指标在各个 mask 面积上的平均值,由此,可更直观地得出结论。从表 1 分析可知,本文的多阶段分支融合修复模型在 mask 面积为 5%~40% 上的 L2 指标最优,在 40%~80% 是次优。在 PSNR 指标上,除遮罩面积 40%~60% 外, SSIM 指标最优,在 FID 指标上,本文指标不是最优。可能是因为本文模型只注重于像素级的引导修复,尽可能地根据学习内容去补全缺失区域,而忽略了纹理级的引导策略,而 Rethinking^[24] 设计了纹理引导修复模块。

2) 定性分析: 图 6 为各算法对比结果图。本文方法与 Shift、Net、LGNet、RFR、CSA、Rethinking 在 mask 面积为 10%~80% 的修复效果。可知,各算法之间依旧存在不足与独特之处,或多或少地存在细节性的不足。

表 1 定量分析

指标	mask/%	Shift_Net	LGNet	RFR	CSA	Rethinking	本文算法
L2↓	5~20	0.019 2	0.020 7	0.016 5	0.012 5	0.014 7	0.010 1
	20~40	0.020 1	0.023 1	0.013 9	0.014 5	0.015 6	0.012 5
	40~60	0.021 4	0.029 9	0.015 3	0.016 8	0.018 8	0.016 9
	60~80	0.023 5	0.035 7	0.018 7	0.019 7	0.025 2	0.018 9
	Avg	0.021 1	0.027 4	0.016 1	0.015 9	0.018 6	0.014 6
PSNR↑	5~20	32.265 4	37.588 6	36.102 8	40.201 4	41.914 7	45.226 4
	20~40	29.365 4	35.770 1	31.248 9	36.402 6	40.012 0	41.025 9
	40~60	27.230 8	35.325 8	28.604 7	34.088 0	38.231 4	37.258 7
	60~80	26.765 4	32.379 4	23.770 9	32.955 7	35.914 3	36.825 3
	Avg	28.906 8	35.266 0	29.931 8	35.911 9	39.018 1	40.084 0
SSIM↑	5~20	0.975 4	0.983 2	0.979 8	0.992 4	0.992 7	0.993 8
	20~40	0.953 6	0.967 1	0.965 4	0.987 6	0.987 0	0.989 4
	40~60	0.924 0	0.960 8	0.932 1	0.979 6	0.950 2	0.980 1
	60~80	0.886 9	0.954 8	0.898 7	0.917 3	0.925 4	0.979 9
	Avg	0.935 0	0.966 5	0.944 0	0.969 2	0.963 8	0.985 8
FID↓	5~20	6.314 7	9.652 3	4.305 4	3.520 3	5.786 5	4.235 0
	20~40	4.207 1	19.987 0	12.954 0	8.532 1	6.920 1	9.839 4
	40~60	17.735 4	28.567 4	15.248 5	15.214 6	9.214 6	14.000 5
	60~80	19.258 3	41.265 8	28.325 0	19.365 4	16.365 4	17.256 9
	Avg	11.878 9	24.868 1	15.208 2	11.658 1	9.571 7	11.333 0

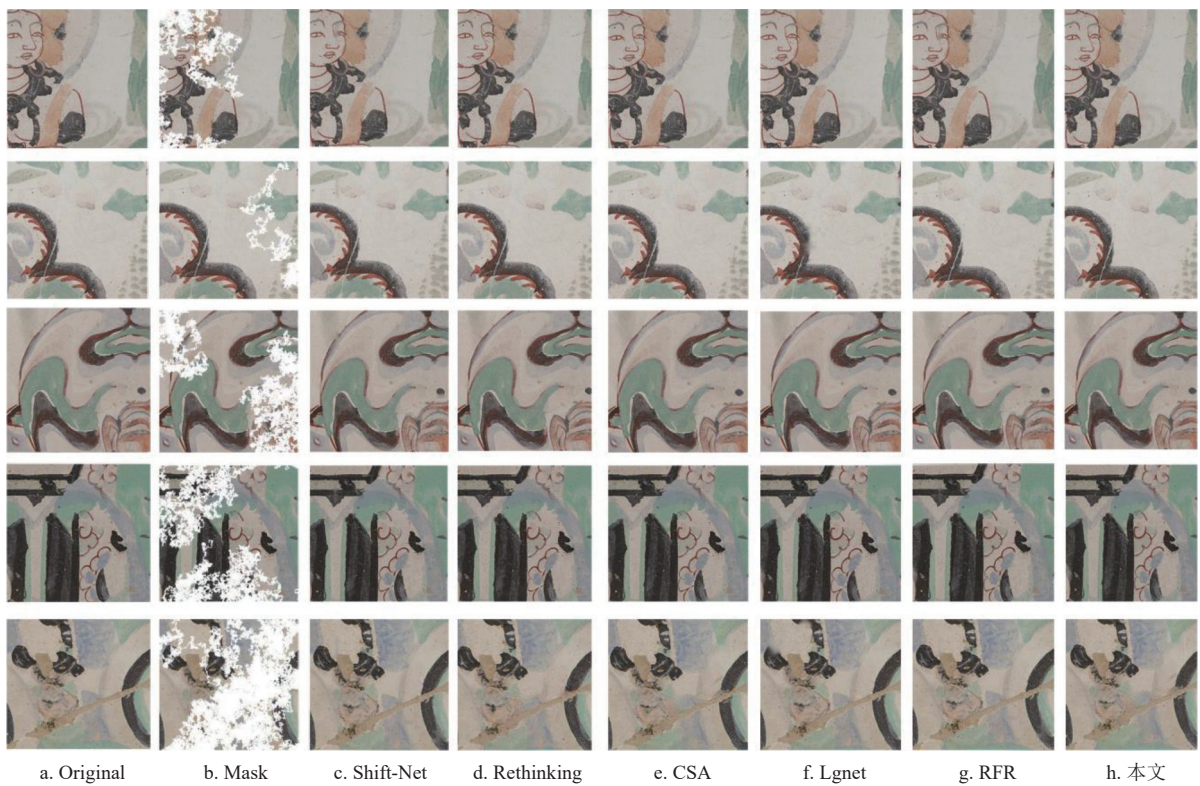


图 6 各个算法对比图

2.3 消融实验

为了验证本文在基础模型之上所作的改进对提

高图像修复结果的有效性, 进行了消融实验, 如表 2 所示。

表 2 消融实验

指标	(a)	(b)	(c)	(d)	(e)	(f)
L2↓	0.030 6	0.029 70	0.044 97	0.019 5	0.015 1	0.014 6
PSNR↑	36.712 0	38.059 0	31.632 2	39.624 8	39.725 1	40.034 0
SSIM↑	0.935 6	0.949 4	0.914 3	0.980 0	0.982 1	0.985 8
FID↓	26.985 1	23.256 4	29.248 6	15.746 5	12.918 6	11.333 0

其中 (a)~(f) 分别对应以下改进点 Multi-Branch Stacking Federation、Multi-Branch Stacking Federation+ Mobile ViT、Multi-Branch Stacking Federation+ Mobile ViT+ Multiblock(Conv)、Multi-Branch Stacking Federation+ Mobile ViT +Multiblock(Reconv)、Multi-Branch Stacking Federation+ Mobile ViT + Multiblock(Reconv)+SelfAttention、Multi-Branch Stacking Federation+ Mobile ViT+Multiblock (Reconv) +Sequential Polarized SelfAtten。通过表 2 可以清楚地看到本文模型在全局粗修复、细节修复以及细节补全上均有提高, 并且与原始 Multiblock(Conv) 进行了对比, 其中在引入 Reconv 卷积后, 各指标提升较为明显, 验证了自注意力机制 (Self Attention) 对敦煌壁

画数据的冗余性 (结合图 3)。由上表可知, 自注意力在增加参数量和训练时间的情况下, 对模型的修复效果反而不明显, 因此采用 Sequential Polarized - Self Atten 替换自注意力, 证明了模型改进的有效性。由表 2 可知, 第三阶段的全局细化网络在前两个阶段后更进一步修复了破损壁画的细节部分, 其中 FID 降低了 5.949 7, PSNR 提高了 4.399 9, SSIM 提高了 0.012 6, L2 下降了 0.009 2。

3) 复杂度分析: 这一部分评估了多阶段分支融合修复模型的计算复杂度, 并选择参数量、推理时间 (仅在 GPU 上, 其中 GPU 型号为 NVIDIA RTX3080 Ti-12G) 作为统计数据, 推理时间 Time (epoch/s) 是网络训练一轮所需要的时间, 相应的结果如表 3 所示。

表 3 复杂度分析

指标	(a)	(b)	(c)	(d)	(e)	(f)
训练时间Time/epoch·s ⁻¹	296	318	375	261	323	231
参数量Parameter/MB	115.002	90.617	120.119	94.022	95.015	99.629

模型总训练轮数为 200 轮, 由表 3 可更加直观地看出本文模型在提升修复效果的同时, 还降低了训练时间以及参数量。

3 结束语

本文提出了多阶段分支融合下的敦煌壁画修复模型, 该算法通过多阶段的特征提取学习以及融合多尺度信息, 采用大感受野很好地提取了壁画图像的纹理以及细节信息, 并对其遮罩进行预测学习, 有效解决了敦煌研究所的数字化保护存储难题。然而本文算法虽然有很好的修复效果, 但主要是建立在数据集的支持上, 然而敦煌莫高窟大多数壁画存在严重破损以及色彩脱落的情况, 且无参考数据集。今后的主要工作将偏向于研究无监督学习或半监督学习以及壁画色彩恢复, 来更好地修复敦煌莫高窟壁画存在的问题, 应用深度学习以及语义分析等方法实现更加复杂抽象的壁画修复, 以获得更好的修复结果。

参考文献

- [1] SIMS K. Artificial evolution for computer graphics[C]// Proceedings of the 18th Annual Conference on Computer graphics and Interactive Techniques. [S.l.]:[s.n.], 1991: 319-328.
- [2] 吴向中. 基于偏微分方程的数字图像修复算法研究[D]. 成都: 电子科技大学, 2016.
WU X Z. Digital image restoration algorithms based on partial differential equations[D]. Chengdu: University of Electronic Science and Technology, 2016.
- [3] CRIMINISI A, PÉREZ P, TOYAMA K. Region filling and object removal by exemplar-based image inpainting[J]. *IEEE Transactions on Image Processing*, 2004, 13(9): 1200-1212.
- [4] XU Z, SUN J. Image inpainting by patch propagation using patch sparsity[J]. *IEEE Transactions on Image Processing*, 2010, 19(5): 1153-1165.
- [5] IIZUKA S, SIMO-SERRA E, ISHIKAWA H. Globally and locally consistent image completion[J]. *ACM Transactions on Graphics (ToG)*, 2017, 36(4): 1-14.
- [6] PATHAK D, KRAHENBUHL P, DONAHUE J, et al. Context encoders: Feature learning by inpainting[C]// Proceedings of the IEEE Conference on Computer Vision

- and Pattern Recognition. [S.l.]: IEEE, 2016: 2536-2544.
- [7] PHUTKE S S, MURALA S. Diverse receptive field based adversarial concurrent encoder network for image inpainting[J]. *IEEE Signal Processing Letters*, 2021, 28: 1873-1877.
- [8] PHUTKE S S, MURALA S. Image inpainting via spatial projections[J]. *Pattern Recognition*, 2023, 133: 109040.
- [9] 柯旺. 基于深度生成模型的老照片修复及其实现[D]. 成都: 电子科技大学, 2021.
- KE W. Restoration and implementation of old photos based on deep generative models[D]. Chengdu: University of Electronic Science and Technology of China, 2021.
- [10] LI W, LIN Z, ZHOU K, et al. Mat: Mask-aware transformer for large hole image inpainting[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2022: 10758-10768.
- [11] 何凯, 刘坤, 沈成南, 等. 基于相似图像配准的图像修复算法[J]. *电子科技大学学报*, 2021, 50(2): 207-213.
- HE K, LIU K, SHEN C N, et al. Image inpainting approach using similar image registration[J]. *Journal of University of Electronic Science and Technology of China*, 2021, 50(2): 207-213.
- [12] LI J, WANG N, ZHANG L, et al. Recurrent feature reasoning for image inpainting[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2020: 7760-7768.
- [13] YAN Z Y, LI X M, LI M, et al. Shift-net: Image inpainting via deep feature rearrangement[C]//Proceedings of the European Conference on Computer Vision (ECCV). [S.l.]: [Springer], 2018: 3-19.
- [14] ZHENG C, CHAM T J, CAI J. Tfill: Image completion via a transformer-based architecture[EB/OL]. [2023-02-08]. <https://arxiv.org/pdf/1706.03762>.
- [15] YU Y, ZHAN F, WU R, et al. Diverse image inpainting with bidirectional and autoregressive transformers[C]//Proceedings of the 29th ACM International Conference on Multimedia. [S.l.]: ACM, 2021: 69-78.
- [16] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV). [S.l.]: [Springer], 2018: 3-19.
- [17] 陈永, 赵梦雪, 陶美凤. 多尺度轮廓波分解的群稀疏壁画修复算法[J]. *西安电子科技大学学报*, 2022, 49(6): 120-128.
- CHEN Y, ZHAO M X, TAO M F. Multi scale contour wave decomposition based group sparse mural restoration algorithm[J]. *Journal of Xi'an University of Electronic Science and Technology*, 2022, 49(6): 120-128.
- [18] LI J, WANG H, DENG Z, et al. Restoration of non-structural damaged murals in Shenzhen Bao'an based on a generator-discriminator network[J]. *Heritage Science*, 2021, 9: 1-14.
- [19] CHEN Y, YU W, POCK T. On learning optimized reaction diffusion processes for effective image restoration[EB/OL]. [2023-3-23]. <https://arxiv.org/pdf/1503.05768>.
- [20] 周宇. 基于线描图智能化生成的计算机辅助壁画修复[D]. 杭州: 浙江大学, 2003.
- ZHOU Y. Computer-aided mural restoration based on intelligent generation of line drawings[D]. Hangzhou: Zhejiang University, 2003.
- [21] 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.
- [22] QUAN W, ZHANG R, ZHANG Y, et al. Image inpainting with local and global refinement[J]. *IEEE Transactions on Image Processing*, 2022, 31: 2405-2420.
- [23] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2023: 7464-7475.
- [24] LIU H, JIANG B, SONG Y, et al. Rethinking image inpainting via a mutual encoder-decoder with feature equalizations[C]//16th European Conference on Computer Vision-ECCV. Glasgow: Springer, 2020: 725-741.
- [25] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. [2023-02-15]. <https://arxiv.org/pdf/1409.1556>.
- [26] LIU G, REDA F A, SHIH K J, et al. Image inpainting for irregular holes using partial convolutions[C]//Proceedings of the European Conference on Computer Vision (ECCV). [S.l.]: [Springer], 2018: 85-100.
- [27] LIU H, JIANG B, XIAO Y, et al. Coherent semantic attention for image inpainting[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. [S.l.]: IEEE, 2019: 4170-4179.