计算机研究与发展
Journal of Computer Research and Development
DOI: 10.7544/issn1000-1239.202220623
61(3): 748−761, 2024

基于 Transformer 的文物图像修复方法

王真言　蒋胜丞　宋奇鸿　刘　波　毕秀丽　肖　斌
（图像认知重庆市重点实验室（重庆邮电大学）　重庆　400065）
（704717215@qq.com）

Transformer-Based Image Restoration Method for Cultural Relics

Wang Zhenyan, Jiang Shengcheng, Song Qihong, Liu Bo, Bi Xiuli, and Xiao Bin
（*Chongqing Key Laboratory of Image Cognition*(*Chongqing University of Posts and Telecommunications*), *Chongqing* 400065）

Abstract　Cultural relics are prone to partial losses of structure or texture due to improper preservation. In order to restore the original image of cultural relics, we propose a new method for restoring cultural relics by using Transformer's global structure understanding ability to restore the coherent structure of cultural relics and using convolutional neural networks' local texture understanding ability to restore the delicate texture of cultural relics. To achieve this goal, the restoration work is divided into two steps: the first step is to use Transformer to reconstruct the rough image and restore the coherent structure; the second step is to use a convolutional neural network to enlarge the rough image and restore the fine texture of the missing area. Considering that there is no high-quality, large-scale heritage database in China and abroad, a new heritage image database is also proposed. The experimental results show that the restoration results outperform the current image restoration algorithms in both subjective and objective evaluations in both breakage restoration experiments and large-area breakage restoration experiments that match realistic scenes. At the same time, the proposed method supports diversified output, which provides diverse references for restorers and improves the restoration efficiency.

Key words　cultural relic database; cultural relic image completion; Transformer; convolutional neural networks; super-resolution; virtual restoration

摘　要　文物极易因为保存不当而导致部分结构或纹理缺失,而现有的图像修复技术由于受到先验信息和卷积操作的局限而无法直接应用于文物图像修复,为更合理地恢复文物图像原貌,提出了一种新的文物图像修复方法,将文物图像修复工作分为 2 个步骤:第 1 步使用 Transformer 进行粗略的图像重建并恢复连贯的结构;第 2 步使用卷积神经网络将粗略的重建图像进行上采样并恢复缺失区域的精细纹理.考虑到目前国内外没有高质量的大型文物数据库,因此也提出了一个新的高质量大型文物图像数据库.最终实验结果表明,在符合现实场景的破损修复实验和大面积破损修复实验中,修复效果在主观和客观评估中均优于当前图像修复算法.同时,支持多元化输出,为修复人员提供多样化参考,极大地提升了文物修复效率.

关键词　文物数据库;文物图像补全;Transformer;卷积神经网络;超分辨;虚拟修复

收稿日期:2022-07-07;**修回日期**:2023-06-30

基金项目:国家自然科学基金项目 (62172067, 61976031);国家重点研发计划项目 (2019YFE0110800);重庆市杰出青年自然科学基金项目 (CSTB2022NSCQ-JQX0001);重庆市教委科技研究计划项目 (KJQN202200635)

This work was supported by the National Natural Science Foundation of China (62172067, 61976031), the National Key Research and Development Program of China (2019YFE0110800), the Natural Science Foundation of Chongqing for Distinguished Young Scholars (CSTB2022NSCQ-JQX0001), and the Science and Technology Research Program of Chongqing Municipal Education Commission(KJQN202200635).

通信作者:刘波(boliu@cqupt.edu.cn)

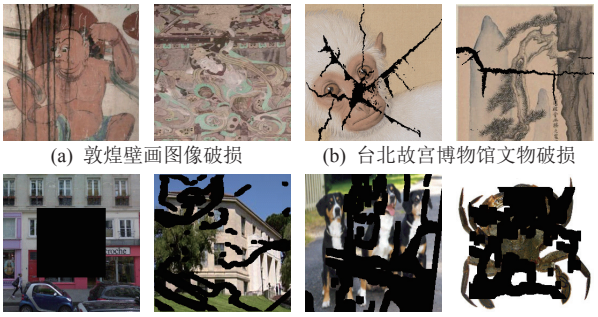
中图法分类号 TP391

文物是国家的瑰宝，其蕴含着一个民族独有的精神价值、思维方式和想象力，具有极高的历史、艺术与科学研究价值. 由于材质的特殊性，文物极易受到潮湿发霉、虫蛀鼠咬、人为破坏等威胁. 纸张的自然老化也会产生破洞、碎裂、褪色等问题. 如果不及时修复，文物的价值就会下降. 传统破损文物修复技术经过数代人继承和发展已经逐渐成熟，但现有修复技术都会对文物本身造成一定的损害. 此外，传统文物修复行业还面临供需不平衡、人才紧缺、技能门槛高等一系列问题. 因此，国内外博物馆都倾向使用数字图像修复技术对文物进行虚拟化修复，该技术已成为当前的研究热点[1-4].

目前主要的文物图像修复技术都是基于传统的图像处理算法. 例如：陈永等人[5]针对敦煌壁画裂纹修复问题提出了一种改进曲率驱动扩散的修复算法；Criminisi 等人[6]提出的基于块匹配的方法，以及 Barnes 等人[7]提出的基于随机采样的块匹配方法 PatchMatch，在文物修复领域也有很广泛的应用[8-9]. 这类方法的主要思想是通过迭代地执行待修复样本块相似度计算、最佳匹配样本块搜索、待修复样本块填充 3 个步骤来完成修复任务. 但是这类方法把图像的局部特征作为修复的主要依据，忽视了上下文的特征信息.

深度学习技术通过对大规模数据进行自监督训练，并使用学习到的高维特征映射修复缺失区域，能结合语义信息来恢复图像，具有较强的生成泛化能力. 目前，使用深度学习技术进行文物图像修复成为一种趋势. 例如，2022 年 Yu 等人[10]使用现代计算机视觉和机器学习技术对敦煌文化遗产进行修复. 但是，大多数方法都使用自然图像修复中常见的卷积神经网络(convolutional neural network，CNN). 同时，网络框架也大多采用编解码器的结构. 这类方法能够结合上下文语义信息进行修复. 但是，由于卷积操作的局部相关性和平移不变性，网络难以利用图像的全局结构. 因此，重复的纹理和伪影经常出现在被修复的区域中. 近年来，Transformer[11]在计算机视觉领域取得了极其优异的成绩. 与 CNN 相比，它克服了卷积的局部相关性，通过自注意力模块来获取图像的全局信息，其应用在文物图像修复任务中可更好地恢复出图像的完整结构. 同时，使用多头注意力机制可以实现修复结果的多元化输出. 但是，Transformer 也存在着计算量过大的问题.

值得注意的是，自然图像修复与文物图像修复存在一定的区别，如图 1 所示. 从破损结构而言，自然图像的破损大多具有固定的结构，而文物图像的破损多是不规则且连续的. 虽然有针对不规则破损图像补全的方法，并已实现了良好的补全效果，但并没有运用到图像修复领域中；从破损面积而言，自然图像破损面积更大且破损区域往往并不连续，而文物图像的破损面积都较小；从纹理复杂程度而言，文物图像的纹理比自然图像更加复杂. 同时，基于深度学习的方法大多以数据驱动，但目前国内外并没有高质量的大型文物数据集，导致网络模型无法学习到足够的文物图像特征. 因此，从自然图像修复领域直接迁移的模型并不能很好地实现对文物图像的修复.



(a) 敦煌壁画图像破损 (b) 台北故宫博物馆文物破损

(c) 自然图像破损

Fig. 1 Comparison of damaged images

图 1 破损图像对比

综上所述，针对文物图像修复任务，本文结合文物图像的特点和艺术家工作时“先结构后细节”的工作流程提出了一种基于 Transformer 的文物图像修复方法，将文物图像修复工作分为 2 个步骤：第 1 步使用 Transformer 进行多元化结构修复；第 2 步使用卷积神经网络进行上采样并恢复缺失区域的纹理，方法流程图如图 2 所示. 实验结果表明，在符合现实场景的破损文物修复实验和文物大面积破损修复实验中，本文方法的修复结果视觉效果更好，客观指标也高于代表性方法. 同时，支持多元化输出，为修复人员提供了多样化的参考，在文物修复领域具有较高的价值.

1 相关工作

本节主要介绍图像修复领域的相关工作及其在文物修复领域的应用，并讨论这些方法的优缺点. 图像修复方法可以分为传统方法和基于深度学习的方法.

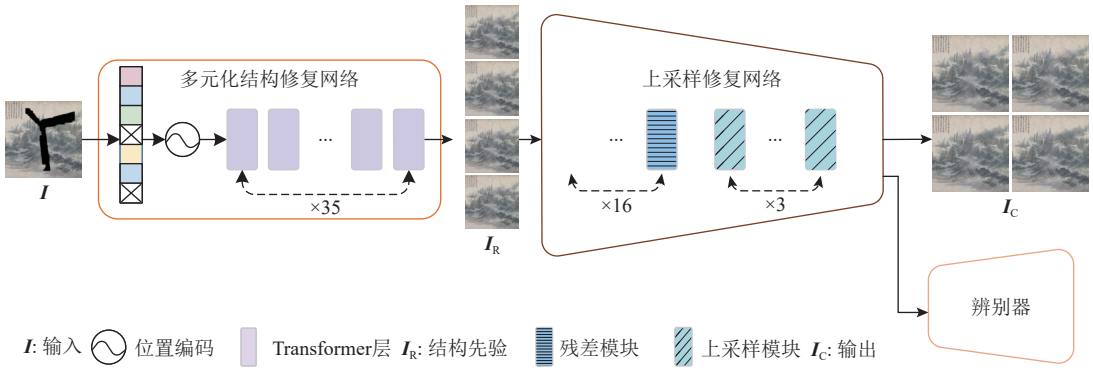


Fig. 2 Transformer-based image restoration method for cultural relics

图 2 基于 Transformer 的文物图像修复方法

1.1 传统的图像修复方法

传统的图像修复方法主要包括基于扩散的修复方法和基于样本块匹配的修复方法. 目前主要的文物图像修复都是基于这 2 类修复技术.

基于扩散的修复方法通过设计的扩散函数将相邻区域的像素块传递到缺失区域内[12-14]. 陈永等人[5]针对壁画裂纹修复问题提出了改进曲率驱动扩散的敦煌壁画修复算法, 使曲率扩散的扩散项更合理.

基于样本块匹配的修复方法是从同一幅图像的相似区域选择关联度高的样本块并填充到缺失区域[6,15-16]. 其中, 2004 年提出的 PatchMatch[6] 通过使用快速最近邻算法可以很好地完成图像修复任务. 然而实现这种方法的前提是能够在图像的已知区域找到缺失区域的相似纹理, 但并不是所有待修复图像都能满足这样的条件. 在文物图像修复领域, 2019 年 Yao[9] 在修复唐卡图像的过程中, 在 Criminisi 算法的基础上引入结构信息对匹配策略进行优化, 尽可能避免了错误匹配的问题.2019 年 Wang 等人[17] 针对敦煌壁画修复问题提出了结合多个候选区的稀疏模型来保证纹理的相似性和结构的连续性.

基于扩散的修复方法和基于样本块匹配的修复方法主要依赖单张图像的局部特征, 很难恢复出符合上下文特性的缺失区域.

1.2 基于深度学习的修复方法

深度学习技术通过对大规模数据进行自监督训练, 并使用学习到的高维特征映射修复缺失区域, 相较于传统方法其能够学习到更高维度的特征[18], 并且能够在缺失的区域生成连贯的结构. 基于 CNN 的方法占主导地位, 可以分为基于编解码结构的单阶段修复模型和提供先验信息的修复模型.

最先使用基于编解码结构的单阶段修复模型是 2016 年提出的 Context Encoder[19], 编码器映射图像缺失区域到低维特征空间, 解码器用来构造输出图像. 然而, 输出图像的恢复区域通常包含视觉伪影且模糊. 2018 年 Liu 等人[20] 为解决普通卷积特征提取不足的问题创造了“部分卷积”, 把传统卷积层替换为部分卷积层, 将编码器层的深层和浅层特征作为 CNN 的输入, 使得修复后的图像纹理更加一致. 2021 年, Zeng 等人[21] 提出一种使用可学习的损失函数替代注意力机制的 2 阶段模型. 基于上下文重建损失, 无注意力机制的生成器也能学习到从已知区域匹配特征进行修复的能力. 在文物图像修复领域, 2021 年 Zhang 等人[22] 针对古代石刻碑文保护的问题提出了一种基于多尺度特征融合的石刻图像去噪与修复方法.

提供先验信息的修复方法[23-24] 在单阶段方法的基础上加入了先验信息, 因此效果更好. 2018 年 Contextual Attention[25] 采用的策略为: 第 1 阶段使用简单的膨胀卷积网络粗略地恢复缺失内容; 第 2 阶段的细化网络使用上下文注意力机制来优化结果. 2018 年 Shift-Net[26] 受样本块匹配思想的启发, 在 U-Net 模型的基础上增加了 Shift 连接层, 可以计算每个缺失区域样本块与已知区域的相似度系数, 并在编码器特征上引入了引导损失, 提升了修复的精度. 2020 年 Zeng 等人[27] 利用深度卷积神经网络对破损图像进行粗略修复, 然后利用最近邻像素匹配进行可控制的再次修复, 使得修复的图像更具真实感. 2021 年 Qin 等人[28] 提出了基于多尺度注意力网络的修复模型, 通过引入多尺度注意力组来提高修复后图像的真实性. 在文物图像修复领域, 2019 年曹建芳等人[29] 针对古代壁画起甲、脱落等问题提出一种基于增强一致性生成对抗网络的图像修复算法, 提高了壁画修补区域与全局的一致性.

将文献 [23−29] 所述的方法直接应用到文物图像修复领域虽然能够修复大面积破损的图像, 但由于

文物的破损往往存在细节丢失、特征不足等问题[5,30]，难以恢复连贯结构[28-29]. 同时卷积的局部连接和平移不变性导致其对图像的全局结构很难把握，也容易产生重复的纹理. 随着深度学习理论的发展，Transformer 在计算机视觉领域内的广泛应用，一些初步的工作[31-33]也证明了它在自然图像合成方面有非常强的能力. 如 2021 年 ICT(image completion transformer)[33]采用 Transformer 的输出作为结构先验以实现高保真度的图像补全. 此外，Transformer 中的多头注意力机制可以实现结果的多样化输出. 但是，由于自注意力机制每次都要计算所有块之间的注意力，计算复杂度为输入长度的平方，因此处理高分辨率图像较为困难[34].

2 文物图像修复方法

本文受到提供先验信息的深度学习方法的启发，为避免卷积操作难以恢复全局结构的缺点，先采用 Transformer 模型来恢复整体的结构信息，再通过一个具有上采样功能的修复网络进行纹理修复，使得修复后的文物图像整体结构连贯，同时避免出现伪影、模糊等现象，并且使修复结果更加逼近原始图像. 文物图像修复的目的是将有缺失像素的输入图像 $\boldsymbol{I}_{\mathrm{D}}=\boldsymbol{I}\odot(1-M)$ 通过预测转化成完整图像 $\boldsymbol{I}_{\mathrm{C}}$. 因此，本文方法将文物图像修复任务分成多元化结构修复和上采样纹理修复 2 个阶段，网络结构分别如图 3、图 4 所示. 第 1 阶段中，Transformer 模型将输入图像 $\boldsymbol{I}_{\mathrm{D}}$ 变换为具有连贯结构的中间修复结果 $\boldsymbol{I}_{\mathrm{R}}$，此过程可表示为 $p(\boldsymbol{I}_{\mathrm{R}}|\boldsymbol{I}_{\mathrm{D}})$. 第 2 阶段中，$\boldsymbol{I}_{\mathrm{R}}$ 通过 CNN 模型学习到 $\boldsymbol{I}_{\mathrm{C}}$ 的特征映射，将 $\boldsymbol{I}_{\mathrm{R}}$ 进行上采样的同时修复精细纹理，并在特征层次进行融合得到输出图像 $\boldsymbol{I}_{\mathrm{C}}$，实现对文物图像的修复，此过程可表示为 $p(\boldsymbol{I}_{\mathrm{C}}|\boldsymbol{I}_{\mathrm{R}},\boldsymbol{I}_{\mathrm{D}})$. 综上，整个修复过程可表述为

$$p(\boldsymbol{I}_{\mathrm{C}}|\boldsymbol{I}_{\mathrm{D}})=p(\boldsymbol{I}_{\mathrm{R}}|\boldsymbol{I}_{\mathrm{D}})\times p(\boldsymbol{I}_{\mathrm{C}}|\boldsymbol{I}_{\mathrm{R}},\boldsymbol{I}_{\mathrm{D}}). \tag{1}$$

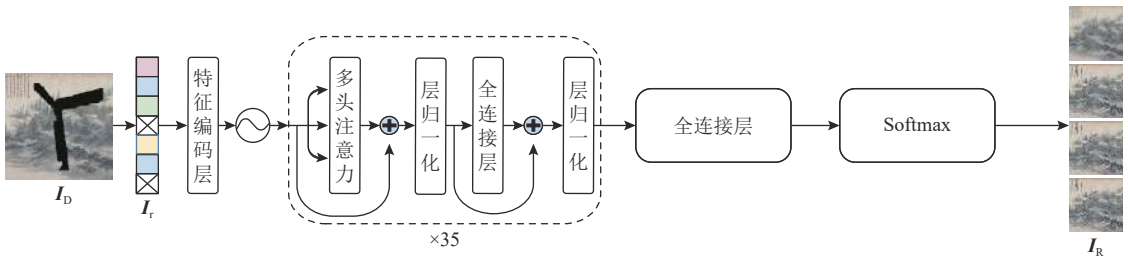


Fig. 3 Diversified structure repair network structure diagram

图 3 多元化结构修复网络结构图

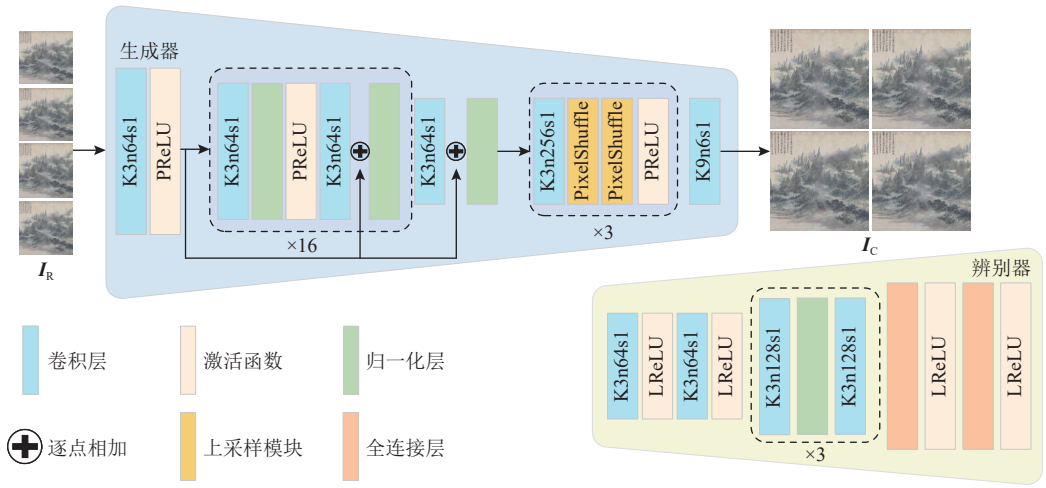


Fig. 4 Upsamping texture repair network structure diagram

图 4 上采样纹理修复网络结构图

2.1 多元化结构修复网络

Transformer 摒弃了 CNN 的局部相关性和平移不变性，通过多个自注意力模块来获取图像的全局信息[11]，其应用在文物图像修复任务中可更好地恢复

出图像的完整结构.

本文使用的 Transformer 模块结构如图 3 所示，与 GPT-2 模型[35]的做法相同，仅使用 Transformer 的解码器，其计算过程可表述为

$$\begin{aligned} F^{l-1} &= LN\left(MSA\left(T^{l-1}\right)\right)+T^{l-1},\\ T^{l-1} &= LN\left(MLP\left(T^{l-1}\right)\right)+F^{l-1}, \end{aligned} \tag{2}$$

其中 LN, MSA, MLP 分别代表层归一化、多头自注意力模块、全连接层. 其中，MSA 可表示为

$$MSA(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = Concat(head_1,\cdots,head_n)MLP^{O}, \tag{3}$$

$$head_i = softmax\left(\frac{\boldsymbol{QK}}{\sqrt{d_{\boldsymbol{K}}}}\right)\boldsymbol{V}.$$

其中 n 代表注意力头的个数；$\boldsymbol{Q}$, $\boldsymbol{K}$, $\boldsymbol{V}$ 代表 3 个可学习的映射矩阵；$d_{\boldsymbol{K}}$ 为 $\boldsymbol{K}$ 的特征维度；MLP^{O} 是一个连接不同自注意力模块的全连接层.

由于 GPT 模型是通过单向注意力预测缺失区域的元素，仅能关注到前序序列的信息. 为了使每个样本块都能关注到所有位置的信息，本文采用与 BERT[36]类似的遮蔽语言模型(mask language model, MLM)来优化 Transformer 模型，确保生成像素可以捕获所有可用的上下文信息，从而使生成内容与已知区域相一致. 具体来说，让 $\boldsymbol{\Pi}=(\pi_1,\pi_2,\ldots,\pi_m)$ 表示离散化输入中标记为掩膜的索引，其中 m 是被屏蔽的标记数量. MLM 的目标是使所有观察区域条件下 $X_{\boldsymbol{\Pi}}$ 的负对数似然最小，用公式表示为

$$L_{\mathrm{MLM}} = \underset{X}{E}\left[\frac{1}{m}\sum_{k=1}^{m}-\log p\left(x_{\boldsymbol{\Pi}_k}|X_{-\boldsymbol{\Pi}},\theta\right)\right], \tag{4}$$

其中 θ 代表 Transformer 模型所学习到的参数. MLM 和双向注意力机制相结合，确保了网络可以利用所有位置的信息来预测缺失区域的内容.

在输入到 Transformer 之前，图像首先被编码成为一个离散化的序列，表示为 $\boldsymbol{I}_{\mathrm{r}}=(i_1,i_2,\cdots,i_{len})$，其中 len 代表 $\boldsymbol{I}_{\mathrm{r}}$ 的长度. 为了对图像的空间信息进行编码，将一个可学习位置的特征选择器添加到每个位置的标记特征中，然后再将这个离散化序列通过预学习编码映射到一个高维的特征向量中，最后再组成 Transformer 模型的输入. 考虑到 Transformer 的计算复杂度与输入长度的平方成正比，且多头注意力机制[11]的计算复杂度与输入长度的 4 次方成正比，如果采用正常的 RGB 像素表达矩阵，那么计算量将过于巨大. 为了降低计算成本，采用与 ICT 相似的做法，使用原始文物图像对应的低分辨率版本来表示其连贯结构的先验，即使用大小为 32×32 的块来表示原始图像的结构信息和粗略的纹理. 为了进一步降低维度，本文还使用 K-Means 聚类算法在文物图像数据集上构造大小为 512×3 的 RGB 像素词表. 可以通过搜索这个词表来重建每个连贯结构的先验信息.

2.2 上采样纹理修复网络

由于 Transformer 中多头注意力机制的存在使 $\boldsymbol{I}_{\mathrm{R}}$ 可以很容易地输出多种合理的结果，针对每一种结果，在重建低维结构先验信息之后，本文方法还学习了一个确定的映射将低维图像 $\boldsymbol{I}_{\mathrm{R}}$ 放大为输入图像的大小，并且没有改变空洞区域和非掩码区域的边缘结构信息. 为了尽可能恢复图像的纹理细节和高频信息，本文使用了传统 CNN 卷积及残差结构，并使用联合感知损失函数对细节修复的结果进行约束，在上采样的过程中尽可能恢复高频细节.

在上采样纹理修复网络中，双层残差结构在保证输入先验结构信息不被篡改的同时，使修复结果更加符合上下文语义信息. 然后利用 PixelShuffle 像素重组技术对图像进行上采样，在不损失清晰度与真实细节信息的前提下对图像进行超分辨.

为达到这个目的，本文方法的第 2 阶段训练了前馈式 CNN 的生成网络 G_θ，网络结构如图 4 所示. 其中生成器网络的训练过程可以表示为

$$\hat{\theta}_G = \arg\min_{\theta_G}\frac{1}{N}\sum_{n=1}^{N} l^{\mathrm{SP}}\left(G_{\theta_G}\left(\boldsymbol{I}_n^{\mathrm{R}}\right),\boldsymbol{I}_n^{\mathrm{GT}}\right), \tag{5}$$

其中联合损失函数 l^{SP} 是若干损失函数的加权组合以模拟恢复图像的不同特征. $\theta_G=\{W_{1:L};b_{1:L}\}$ 代表第 N 层网络通过联合损失函数 l^{SP} 优化后的权重和偏置. 为恢复更加逼真的纹理细节，在网络训练阶段将生成器参数 $\hat{\theta}_G$ 和辨别器参数 $\hat{\theta}_D$ 进行交替优化，优化表达式为：

$$\begin{aligned} &\max_{D} E_{\boldsymbol{I}^{\mathrm{H}}\sim p(\boldsymbol{I}^{\mathrm{H}})}\left[\log D\left(\boldsymbol{I}^{\mathrm{H}}\right)\right]+E_{\boldsymbol{I}^{\mathrm{L}}\sim p(\boldsymbol{I}^{\mathrm{L}})}\left[\log\left(1-D\left(\boldsymbol{I}^{\mathrm{L}}\right)\right)\right],\\ &\min_{G} E_{\boldsymbol{I}^{\mathrm{L}}\sim p(\boldsymbol{I}^{\mathrm{L}})}\left[\log\left(1-D\left(\boldsymbol{I}^{\mathrm{L}}\right)\right)\right], \end{aligned} \tag{6}$$

其中 $\boldsymbol{I}^{\mathrm{H}}$ 为高分辨率图像，$\boldsymbol{I}^{\mathrm{L}}$ 为低分辨率图像.

2.3 联合感知损失函数

为了更好地完成基于结构先验的修复任务，本文在 Ledig 等人[37]的基础上设计改进了更适用于文物图像修复任务的联合感知损失函数，其表达式为

$$l^{\mathrm{SP}} = l_{\mathrm{context}} + 10^{-3}l_{\mathrm{adv}} + 10^{-3}l_{\mathrm{TV}}. \tag{7}$$

联合感知损失函数由 3 部分组成：内容感知损失 l_{context}、生成对抗损失 l_{adv}、全变分损失 l_{TV}. 无论是在像素层面还是感知层面都能很好地提升修复图像的质量.

2.3.1 内容损失函数

内容损失 l_{context} 主要分为 2 部分，即像素层面的均方差损失 $l^{\mathrm{SP}}_{\mathrm{MSE}}$ 和特征层面上提出 VGG 损失 $l^{\mathrm{SP}}_{\mathrm{VGG}/i,j}$.

像素层面的均方差损失 $l_{\mathrm{MSE}}^{\mathrm{SP}}$ 的表达式为：

$$l_{\mathrm{MSE}}^{\mathrm{SP}}=\frac{1}{R^2WH}\sum_{x=1}^{RW}\sum_{y=1}^{RH}(\boldsymbol{I}_{x,y}^{\mathrm{H}}-(G_{\theta_G}(\boldsymbol{I}^{\mathrm{L}})_{x,y}))^2, \tag{8}$$

其中 R, W, H 分别表示图像缩放系数、图像宽度、图像高度. 这是图像修复领域运用最广泛的损失函数. 然而实验结果表明虽然修复结果具有较高的峰值信噪比(peak signal noise rating, PSNR)指标，但使用其进行优化往往会导致修复结果缺乏高频细节，丧失图像原本的真实性. 因此前人在特征层面提出 VGG 损失 $l_{\mathrm{VGG}/i,j}^{\mathrm{SP}}$ 为：

$$l_{\mathrm{VGG}/i,j}^{\mathrm{SP}}=\frac{1}{W_{i,j}H_{i,j}}\sum_{x=1}^{W_{i,j}}\sum_{y=1}^{H_{i,j}}(\phi_{i,j}(\boldsymbol{I}_{\mathrm{R}})_{x,y}-\phi_{i,j}(G_{\theta_G}(I_{\mathrm{C}})_{x,y})^2, \tag{9}$$

其中 VGG/i, j 中的 i, j 代表本文所选用的卷积层，$\phi_{i,j}$ 为第 i 个池化层前的第 j 个卷积层，$W_{i,j}$ 和 $H_{i,j}$ 表示 VGG 网络中特征图的宽度与高度. 这种方法得到的 PSNR 较高，纹理也相对真实.

2.3.2 对抗损失函数

本文方法在训练阶段还额外添加了一个对抗损失来产生更加真实的纹理：

$$l_{\mathrm{adv}}=E\left[\log\left(1-D_{\omega}\left(\boldsymbol{I}_{\mathrm{pred}}\right)\right)\right]+E\left[\log D_{\omega}(\boldsymbol{I})\right], \tag{10}$$

其中 D 是参数为 ω 的辨别器. 交替共同训练生成器网络 F 和辨别器 D 来求解优化问题：

$$\min_{F}\max_{D}l_{\mathrm{up}}(\delta,\omega)=\alpha_1 l_{\ell_1}+\alpha_2 l_{\mathrm{adv}}, \tag{11}$$

其中 l_{ℓ_1} 为 L_1 损失函数，α_1=1.0, α_2=0.1.

2.3.3 全变分损失函数

使用 2.3.2 节的损失函数能够产生较好的修复结果，但在文物图像破损区域仍然存在部分失真与不平滑的区域. 因此，本文引入了全变分损失函数 l_{TV}[38]，使修复图像和掩膜的边缘更加平滑. 其表达式为：

$$l_{\mathrm{TV}}(x)=\sum_{i,j}^{u_{i,j}}\left(\left|x_{i+1,j}-x_{i,j}\right|^2+\left|x_{i,j+1}-x_{i,j}\right|^2\right)^{\frac{\beta}{2}}, \tag{12}$$

其中 u 为支持域，β=2.0.

3 实验结果与分析

为了讨论和验证本文提出的修复方法，并将其与已有的修复方法进行对比和分析，本节将从 5 个角度进行实验. 首先，对比各方法在现实场景下的破损修复效果；然后对比各方法对大面积破损的修复效果、验证本文方法在不同数据集上的修复效果、验证本文方法在自然图像上的修复效果；最后讨论多元化文物图像修复的优点.

3.1 实验数据库

目前文物图像修复领域并没有高质量的大型公开数据集，而大多数基于深度学习的方法均以数据为驱动，如果训练集过小则会导致网络性能不佳、修复效果不理想等问题. 为解决文物图像数据库过小的问题，本文对中国台北故宫博物馆官网提供的 5 000 张分辨率为 3 000×4 000 的文物图像进行处理，最终得到的 88 000 张分辨率为 256×256 的文物图像作为本次实验的基础数据集. 具体处理过程为：首先将 5 000 张分辨率为 3 000×4 000 的图像进行人工筛选，剔除颜色过于单一、无关内容太多的图像；然后通过随机剪裁分成分辨率为 256×256 的图像；再进行一次人工筛选，最终构成包含山水画、人物画像、壁画、花鸟画 4 类共计 80 000 张的大型文物图像数据集. 此外，为了降低数据集原始样本的影响，本文在训练样本中随机抽取 8 000 张图像进行数据增广，详细的数据集划分与增广操作如表 1 所示.

Table 1 Statistics for Using Datasets

表 1 使用数据集的统计信息

数据划分	数据增广		训练数据	测试数据
	高斯噪声 K=3 σ_X=2 σ_Y=2	高斯模糊 μ=0 σ=0.01		
山水画	1 000	1 000	41 900	100
人物像	1 000	1 000	11 900	100
壁画	1 000	1 000	11 900	100
花鸟画	1 000	1 000	21 900	100

3.2 评价指标

在图像修复领域，目前最常用的评估指标有峰值信噪比和结构相似性(structural similarity index measure, SSIM). 近年来，为更深层次地评价修复图像的质量，常使用学习感知图像块相似性[39](learned perceptual image patch similarity, LPIPS)来作为评估指标.

3.3 实验细节

本文实验的硬件环境为 Intel® Xeon® Platinum 8255C 和 2 块 Tesla V100.实验平台的软件环境为 Pytorch 1.8.0, CUDA 11.1.1.训练过程中采用 Adam 优化器对参数进行优化，学习率为 1×10^{-4}，1 阶动量为 0.5，2 阶动量为 0.1.

3.4 实验对比与分析

本节在符合现实场景破损和大面积破损 2 种情

况下, 与基于块匹配的传统方法代表 PM[6]、结合先验信息的深度学习方法代表 Shift-Net[26]、双阶段修复模型 CRF[21] 和 EC[23]、使用不同类型卷积的单阶段深度学习方法代表 PC[20] 、基于 Transformer 和 CNN 的双阶段修复方法代表 ICT[33] 和 EC[33] 进行对比, 以验证本文方法的有效性.

3.4.1 符合现实场景破损的修复实验

本节将对比各种方法在符合现实场景的皴裂山水画的修复效果. 结合现实中文物破损面积小、破损不规则、破损区域连续等特点, 本文专门设计了大小在 10%~15% 且破损区域连续、大小依次递增的 5 张掩膜, 表示为 M_s-1~5.本节实验均采用这 5 张特定的掩膜. 同时, 由于山水画更加强调结构的连贯性与纹理的细腻性, 修复难度更高, 因此实验在本节与 3.4.2 节均在山水画中进行对比.

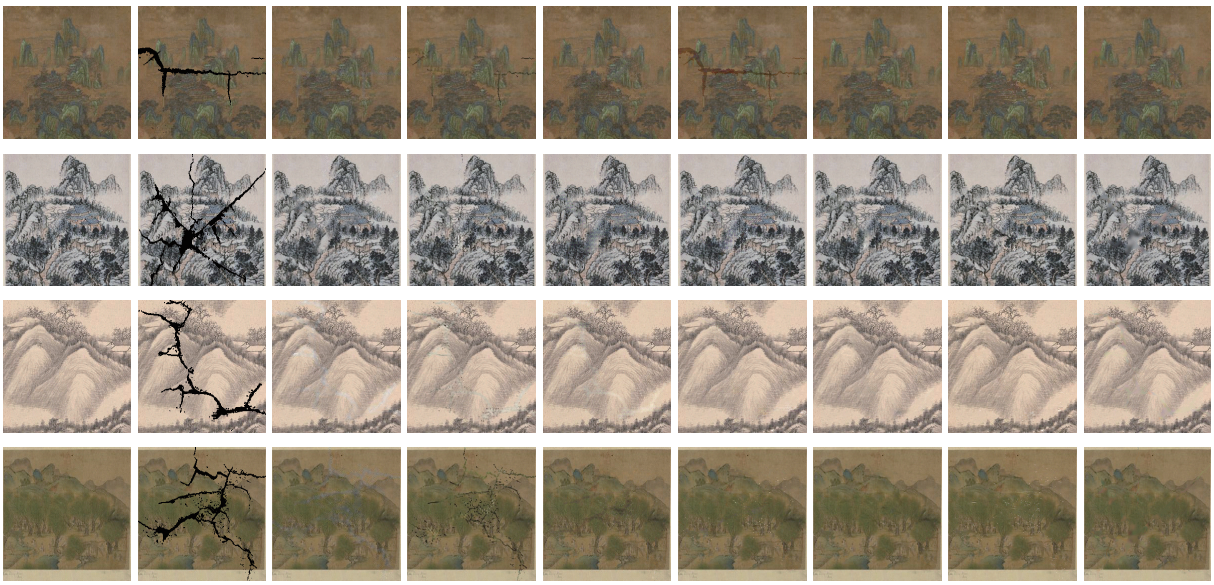
图 5 展示了各方法的修复结果. 由图 5(c)可见, PM 整体表现优异, 但修复区域缺乏上下文的语义信息. 在已知区域无法提供足够多的先验信息时尤为明显. 由图 5(d)(e)所示, PC 和 Shift-Net 的修复结果在连贯结构的恢复上并不合理, 存在重复的纹理和伪影. 由图 5(f)(g)所示, EC, CRF 的修复结果较为优秀, 但在某些特定掩膜下会出现严重的伪影. 相对这些方法, 本文的 2 阶段修复模型的图像处理效果均比较理想, 阶段性的修复过程更加适用于文物图像修复任务.图 5(h)的 ICT 与本文方法的修复质量相对较高, 由此可见, 基于 Transformer 的修复方法可以给模型提供更好的先验信息. 但 ICT 未将上采样和修复过程结合, 导致其在上采样过程中丢失的关键像素无法被修复, 部分区域仍存在颜色失调的现象. 而本文方法先使用 Transformer 进行结构先验, 再结合图像超分辨的思想, 把上采样和修复的过程结合起来, 增强了网络上采样的能力, 使网络在上采样的过程中能保留更多的关键信息. 同时在联合损失函数的约束下, 文物图像修复的质量得到极大提升, 修复结果语义连贯、伪影和重复的纹理较少, 取得了较好的指标和视觉效果.

各种方法的修复指标如表 2 所示, 本文方法在 PSNR, SSIM 上表现均为最优. 相较于 PM, PC, Shift-Net, EC, CRF, ICT 这 6 种方法, SSIM 指标分别提升了 13.2 个百分点、11.7 个百分点、11.9 个百分点、1.3 个百分点、2.7 个百分点、0.8 个百分点; PSNR 指标分别提升了 14.4 个百分点、14.5 个百分点、22.3 个百分点、9.7 个百分点、6.0 个百分点、11.5 个百分点. 随着掩膜逐渐增大, PSNR, SSIM 这 2 项指标均有所下降, 但本文方法相较于其他方法下降趋势更加平稳.

3.4.2 针对大面积破损的修复实验

3.4.1 节实验中使用的是特定的掩膜. 为进一步验证本文方法对大面积破损文物修复的有效性, 本节使用 3 张 20%~25% 的随机掩膜进行测试, 表示为 M_L-1~3.

各种方法的修复结果如表 3 所示. 可以看出, PM, PC, Shift-Net 这三者的指标相较于 3.4.1 节实验中的



(a) 原图 (b) 掩膜 (c) PM (d) PC (e) Shift-Net (f) EC (g) CRF (h) ICT (i) 本文方法

Fig. 5 Comparison of repair results of different algorithms for realistic damaged scenes

图 5 不同算法对现实破损场景的修复结果对比

Table 2 Results of Our Method and Other Methods for Repairing Real-Life Damaged Scenes

表 2 本文方法与其他方法针对现实破损场景修复的结果

指标	PSNR/dB						SSIM					
	M_S-1	M_S-2	M_S-3	M_S-4	M_S-5	均值	M_S-1	M_S-2	M_S-3	M_S-4	M_S-5	均值
PM[6]	32.26	32.27	34.59	33.59	36.71	33.89	0.88	0.82	0.88	0.88	0.88	0.870
Shift-Net[26]	34.33	33.15	31.77	34.35	35.69	33.86	0.89	0.85	0.84	0.84	0.89	0.882
PC[20]	30.41	28.01	32.47	32.79	34.78	31.69	0.90	0.81	0.90	0.90	0.90	0.880
EC[23]	36.86	33.29	35.85	34.88	35.69	35.31	**0.98**	0.96	0.97	0.97	**0.98**	0.972
CRF[21]	34.72	35.69	38.44	35.65	38.37	36.58	0.98	0.95	0.95	0.94	0.97	0.958
ICT[33]	33.55	32.58	36.57	34.72	36.45	34.77	0.98	0.97	0.98	0.98	0.97	0.976
本文方法	**36.99**	**36.07**	**40.25**	**37.97**	**40.86**	**38.76**	**0.98**	**0.98**	**0.99**	**0.99**	**0.98**	**0.984**

注：黑体数值为最优值.

值出现了比较明显的下降，而基于 Transformer 的 2 种修复方法在面对大小不同的掩膜时，指标下降的趋势相对平稳，由此可以验证基于 Transformer 的模型的泛化性、鲁棒性更好. 本文方法相较于 PM, PC, Shift-Net, EC, CRF, ICT 这 6 种方法，SSIM 指标分别提升了 27.8 个百分点、30.9 个百分点、30.9 个百分点、3.2 个百分点、2.0 个百分点、1.0 个百分点；PSNR 指标分别提升了 21.1 个百分点、20.7 个百分点、32.6 个百分点、17.8 个百分点、15 个百分点、11.6 个百分点. 由表 4 左侧可以看出，在破损区域较为集中时，得益于 Transformer 模型提供的先验信息和上采样修复模型的约束，本文方法可以很好地恢复文物图像的底色和山的轮廓，局部纹理也清晰可见，伪影较少. 由表 4 右侧中可以看出，在破损区域较为随机时，本文

Table 3 Results of Our Method and Other Methods for Repairing Large Areas of Damage

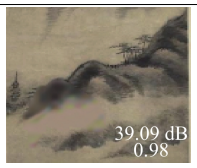
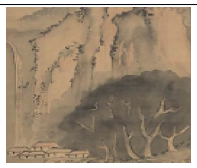
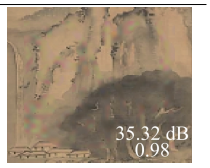
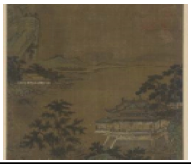
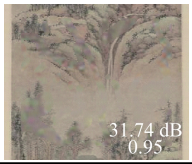
表 3 本文方法与其他方法针对大面积破损修复的结果

指标	PSNR/dB				SSIM			
	M_L-1	M_L-2	M_L-3	均值	M_L-1	M_L-2	M_L-3	均值
PM[6]	30.03	29.75	29.23	29.67	0.74	0.77	0.75	0.753
Shift-Net[26]	30.38	29.39	29.60	29.79	0.75	0.71	0.75	0.736
PC[20]	27.87	26.95	26.51	27.11	0.72	0.75	0.74	0.736
EC[23]	30.80	28.09	32.64	30.51	0.95	0.89	0.96	0.933
CRF[21]	30.45	29.91	33.28	31.21	0.95	0.93	0.95	0.944
ICT[33]	32.23	31.54	32.95	32.24	0.96	0.94	0.96	0.953
本文方法	**36.91**	**33.80**	**37.17**	**35.96**	**0.97**	**0.95**	**0.97**	**0.963**

注：黑体数值为最优值.

Table 4 Large Area Damage Repair Renderings

表 4 大面积破损修复效果

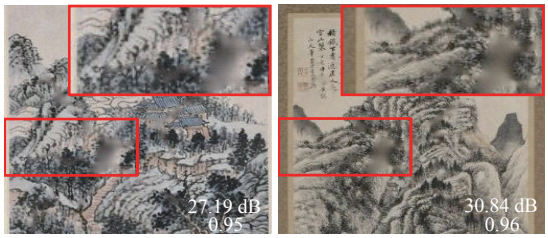
破损区域较集中			破损区域较随机		
原图	掩膜	修复结果	原图	掩膜	修复结果
					
					

注：“修复结果”中位于图像右下角的数值分别为当前图像的 PSNR（dB 值）和 SSIM 评估指标值.

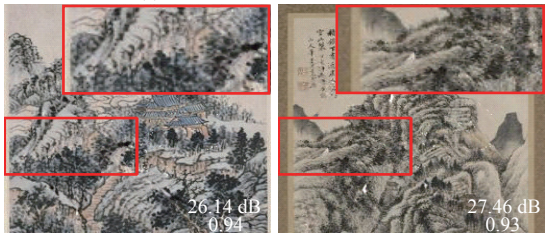
方法恢复的结构很好地把握了全局的信息, 局部纹理也非常逼真.

表 3 和表 4 表明本文方法对大面积破损文物的结构恢复能力突出, 纹理、颜色的恢复也较为合理, 具有很好的修复效果.

在实验过程中发现, 少部分指标较高的图片却存在修复区域模糊的现象, 而符合人类视觉感知的修复结果其指标反而更低, 如图 6 所示.



(a) 高指标低视觉质量修复结果



(b) 低指标高视觉效果修复结果

注: 位于图像右下角的数值分别为当前图像的 PSNR (dB 值) 和 SSIM 评估指标值.

Fig. 6 Comparison chart of abnormal indicators

图 6 异常指标对比图

本文认为利用单一损失函数约束的回归模型在 PSNR, SSIM 两个指标上能得到提升, 但其修复结果并不是人类视觉感知上最好的结果. 并且损失函数的部分约束计算和 PSNR, SSIM 这 2 个指标计算类似, 因此会影响网络的训练环节, 导致部分文物图像修复结果出现轻微模糊、但其图像指标颇高的现象. 为进一步验证本文方法的有效性, 本文引入更深层次衡量修复图像质量的指标 LPIPS[39]. 初步实验结果证明, LPIPS 指标更符合人类的视觉感知, 其值越低表示 2 张图像越相似, 结果如图 7 所示. 本文方法相较于基于深度学习的 Shift-Net, PC, EC, CRF, ICT 这 5 种方法, LPIPS 指标分别下降了 41.1 个百分点、70.7 个百分点、27.1 个百分点、1.8 个百分点、17.2 个百分点.

综上所述, 本文方法在 PSNR, SSIM, LPIPS 这 3 种指标上表现均优于其他具有代表性的方法.

3.4.3 数据库中其他类型图片的修复实验

本节将验证本文方法在人物画像、壁画、花鸟画中的修复效果. 人物画像、壁画、花鸟画绘画技法都以线为主勾勒表现绘画对象, 对整体结构的连贯性要求比较小, 颜色的层次更少, 纹理相较于山水画更简单. 因此本节实验中分别使用数据集中的人物画像、壁画、花鸟画进行训练, 实验细节与 3.4.1 节、3.4.2 节完全一致. 表 5 中, 对于花卉修复结果, 无论掩膜区域相对集中还是随机, 本文方法都能恢复出连贯的结构和合理的颜色; 人物画像修复结果主要

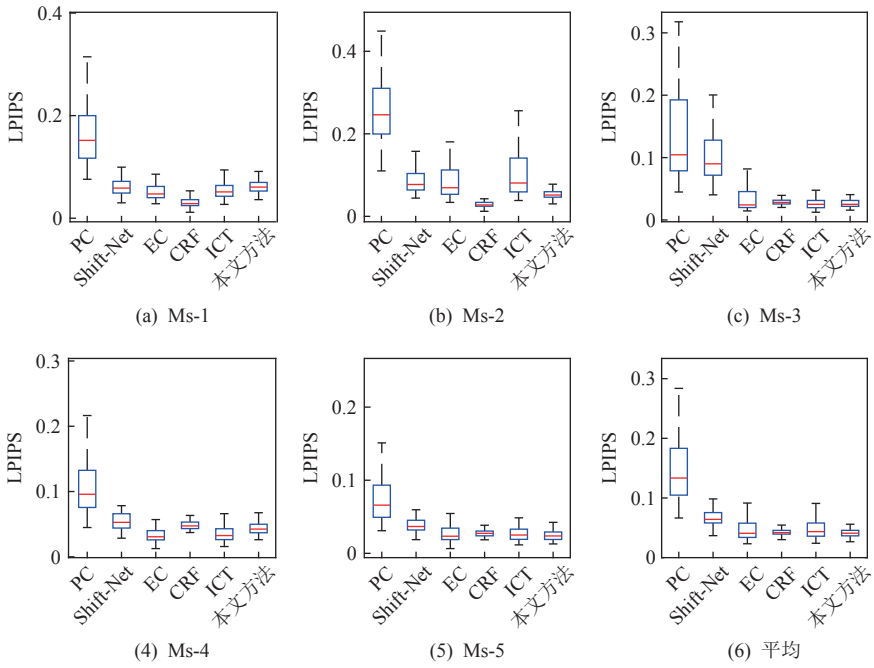


Fig. 7 Comparison of LPIPS values of different masks

图 7 不同掩膜的 LPIPS 值对比

Table 5 Restoration Renderings of Other Types of Cultural Relics

表 5 其他类型文物修复效果图



注：“修复结果”中位于图像右下角的数值分别为当前图像的 PSNR（dB 值）和 SSIM 评估指标值.

结构相对集中，在主体结构缺失的情况下，本文方法恢复的结果语义相对连贯，轮廓、细节都能得到很好的恢复；壁画的结构相对简单，颜色相对单一，本文方法恢复的结果伪迹较少，具有良好的视觉效果；但由于根据类别划分的数据集样本较少，以及 Transformer 多元化输出的特点，本文方法在修复大面积破损图像时会出现轻微失真的情况. 对于此现象带来的优缺点，本文将在 3.6 节中对多元化修复作进一步阐述.

以上结果表明在对不同类别文物图像的修复中，本文方法在主观和客观指标上都具有较好的修复效果.

3.5 常规图像修复

由于目前大多数基于深度学习的图像修复方法针对的都是常规图像，因此本文也验证了所提方法对自然图像的修复效果. 由表 6 可见，针对缺失面积较大的图像，本文方法恢复的图像结构连贯，与原图相对一致，轮廓清晰，局部纹理逼真. 针对缺失面积较小的图像，本文方法恢复的图像结构连贯且符合上下文语义信息.

3.6 多元化修复

文物修复往往需要将修复人员的主观认知和原始参照物结合起来. 但在实际的文物修复任务中，存在无对照样本的情况，导致修复结果无法验证合理性. 针对此问题，本文方法通过多头注意力机制实现多样化输出. 多元化的修复结果如表 7、表 8 所示. 值得注意的是，本文方法虽然在进行大面积修复时偶尔会出现轻微失真、模糊的情况，但可以为专家提供多种修复参考，为后序修复提供决策依据，在降低文物修复的主观性、随机性的同时加强修复的准确性，极大地提升了文物修复效率.

3.7 阶段性子网络分析实验

本节通过讨论 2 阶段网络的修复结果，分析论证各阶段子网在文物修复任务中的不同作用. 在实际修复任务中，文物图像常常会因老化而产生破洞、裂痕等问题，使图像整体结构受到破坏，如图 8 所示. 第 1 阶段利用 Transformer 对破损文物图像进行整体结构先验信息修复的结果如图 8(b)所示；第 2 阶段利用多重残差卷积网络对第 1 阶段输出图像进行纹

Table 6 Conventional Image Restoration Renderings

表 6 常规图像修复效果图



注：“修复结果”中位于图像右下角的数值分别为当前图像的 PSNR（dB 值）和 SSIM 评估指标值.

Table 7 Diversified Restoration Renderings

表 7 多元化修复效果图

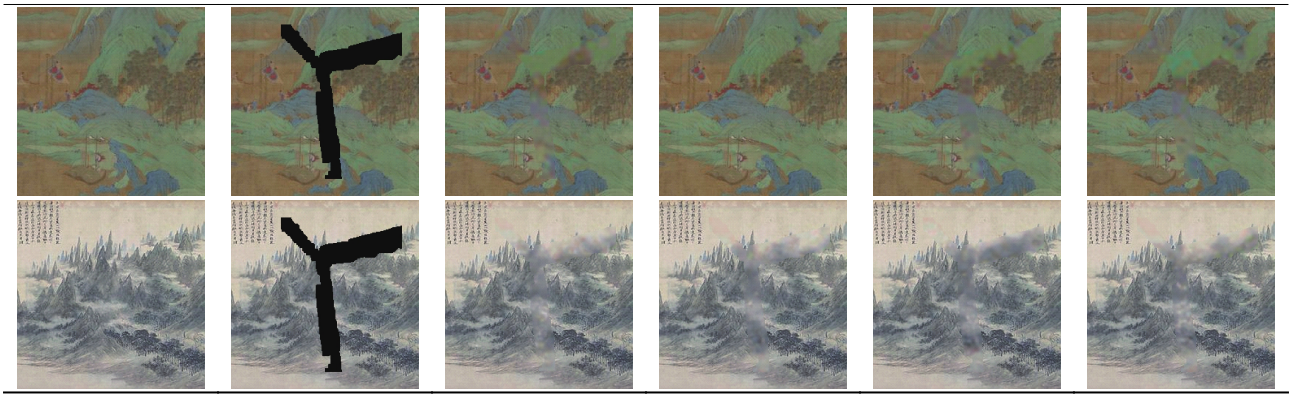
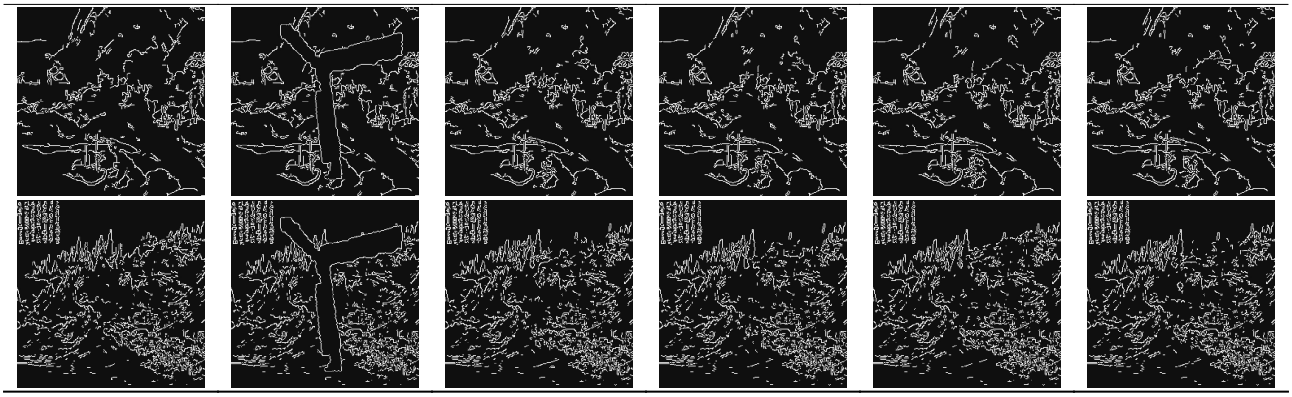
原图	掩膜	修复结果 1	修复结果 2	修复结果 3	修复结果 4
					

Table 8 Diversified Restoration Edge Detection Renderings

表 8 多元化修复边缘检测效果图

原图	掩膜	修复结果 1	修复结果 2	修复结果 3	修复结果 4
					

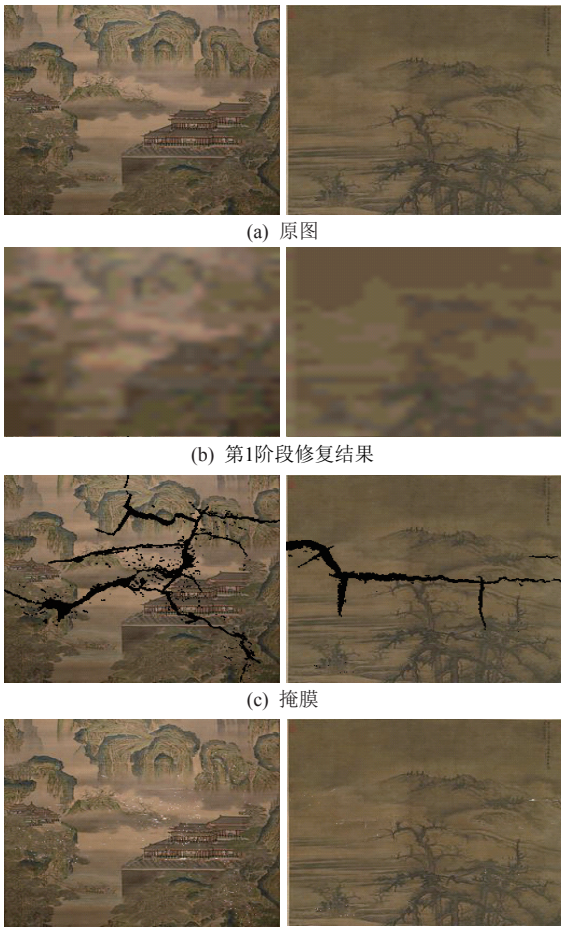
理修复的结果如图 8(d)所示.

本文方法从文物图像的特性出发,在修复过程中充分考虑了图像的整体结构与局部纹理. 如果在文物图像修复任务中仅仅使用第 1 阶段的子网络,虽然可以生成连贯且符合语义的整体结构,但为减少计算量的降维操作往往会导致修复结果较为模糊、缺乏相应的细节语义信息,如图 8(b)所示;如果在文物图像修复任务中仅仅使用第 2 阶段的子网络,虽然可以生成有效的局部纹理,但修复的图像会缺乏正确的上下文语义信息,如图 8(d)所示.

综合上述对阶段性子网络的分析表明,本文方法只有在 2 阶段网络共同作用下才能实现最优修复效果.

4 总 结

在文物图像补全领域,长期存在着既要实现足够的多样性又要求修复效果逼真的困境. 本文针对文物图像修复任务提出了一种基于 Transformer 的修复方法,该方法将 Transformer 和 CNN 的优点结合;利用 Transformer 的全局结构理解能力和多元化输出以及 CNN 较强的局部感知能力,本文方法实现了对文物图像的高质量修复. 同时,本文提出了一个新的高质量文物数据库,解决了国内外缺乏相关数据库的问题. 大量实验表明,本文方法在主客观效果上均优于现有方法,并实现了多元化输出,提升了文物修复效率. 本文方法仍有需要深入研究与改进的方面,如图 8 的修复结果在原始缺损区域附近产生模糊等问题. 此外,本文设计的 2 阶段网络能够生成多元化的结果,最大程度地为文物修复提供指导与参考,但在修复大面积破损图像时也会产生与原始图像不一致的结果. 最后,由于评估指标的缺陷,实验中会出现高指标、低主观感知质量和低指标、高主观感知质量的修复结果,对网络训练的非线性回归产生影响. 后续工作考虑在第 2 阶段子网络设计方面对图像进行多尺度特征融合与去噪[25];探索更能反映文物图像修复质量的评价指标,加强对网络的约束并



(a) 原图

(b) 第1阶段修复结果

(c) 掩膜

(d) 第2阶段修复结果

Fig. 8 Effect of the various stages of cultural relics restoration

图 8 各阶段文物修复效果

提升文物图像修复质量. 也可借助门控卷积[40]网络的设计思想对第 2 阶段进行改进, 使其更能适应缺损面积较大的文物图像修复任务. 还可以进一步提高文物图像数据集的数量与质量, 将各类文物图像进行更加细致地划分, 以适应更加复杂的修复场景; 更新专项训练策略, 提高修复结果的分辨率和观赏价值.

作者贡献声明：王真言负责部分实验开发任务、数据整理分析并提供论文修改意见；蒋胜丞完成部分实验开发任务、数据集的创建并撰写论文；宋齐鸿提供论文修改意见；刘波、毕秀丽和肖斌提供实验开发思路，给予工作支持和指导意见.

参 考 文 献

[1] Wang Hua, Li Qing, Jia Songtao. A global and local feature weighted method for ancient murals inpainting[J]. International Journal of Machine Learning and Cybernetics, 2020, 11(6): 1197−1216

[2] Chen Hao, Zhao Lu, Wang Zhenyu, et al. DualAST: Dual style-learning networks for artistic style transfer[C]//Proc of the 34th IEEE/CVF Int Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE , 2021: 872−881

[3] Haliassos A, Barmpoutis P, Stathaki T, et al. Classification and detection of symbols in ancient papyri[C]//Proc of the 5th Visual Computing for Cultural Heritage. Berlin: Springer, 2020 : 121−140

[4] Opgenhaffen L. The impact of digital technology on archaeological recording strategies and ensuing open research archives[J]. Digital Applications in Archaeology and Cultural Heritage, 2022, 27: 1−15

[5] Chen Yong, Ai Yapeng, Guo Hongguang. Improved curvature driven model for Dunhuang Fresco restoration algorithm[J]. Journal of Computer Aided Design & Computer Graphics, 2020, 32(5): 787−796(in Chinese)
(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)

[6] Criminisi A, Perez P, Toyama K. Region filling and object removal by exemplar-based image inpainting[J]. IEEE Transactions on Image Processing, 2004, 13(9): 1200−1212

[7] Barnes C, Shechtman E, Finkelstein A, et al. PatchMatch: A randomized correspondence algorithm for structural image editing[J]. ACM Transactions on Graphics, 2009, 28(3): 24−34

[8] Lu Xiaobao, Wang Weilan. Improvement of damaged Thangka image repair algorithm based on sample block[J]. Journal of Computer Applications, 2010, 30(4): 943−946(in Chinese)
(卢小宝, 王维兰. 基于样本块的破损唐卡图像修复算法的改进[J]. 计算机应用, 2010, 30(4): 943−946)

[9] Yao Feng. Damaged region filling by improved Criminisi image inpainting algorithm for Thangka[J]. Cluster Computing, 2019, 22(6): 13683−13691

[10] Yu Tong, Lin Chuan, Zhang Shuo, et al. Artificial intelligence for dunhuang cultural heritage protection: The project and the dataset[J]. International Journal of Computer Vision, 2022, 130(11): 2646−2673

[11] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Proc of the 31st Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT, 2017: 6000–6010

[12] Bertalmio M, Sapiro G, Caselles V, et al. Image inpainting[C]//Proc of the 27th Annual Conf on Computer Graphics and Interactive Techniques. New York: ACM, 2000: 417−424

[13] Efros A A, Freeman W T. Image quilting for texture synthesis and transfer[C]//Proc of the 28th Annual Conf on Computer Graphics and Interactive Techniques. New York: ACM, 2001: 341−346

[14] Ballester C, Bertalmio M, Caselles V, et al. Filling-in by joint interpolation of vector fields and gray levels[J]. IEEE Transactions on Image Processing, 2001, 10(8): 1200−1211

[15] Darabi S, Shechtman E, Barnes C, et al. Image melding: Combining inconsistent images using patch-based synthesis[J]. ACM Transactions on Graphics, 2012, 31(4): 1−10

[16] Hays J, Efros A A. Scene completion using millions of photographs[J]. ACM Transactions on Graphics, 2007, 26(3): 4−33

[17] Wang Hua, Li Qing, Zou Qian. Inpainting of Dunhuang murals by

sparsely modeling the texture similarity and structure continuity[J]. Journal on Computing and Cultural Heritage, 2019, 12(3): 1−21

[18] Xu Shoujin, Liu Dong, Xiong Zhihong. E2I: Generative inpainting from edge to image[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(4): 1308−1322

[19] Pathak D, Krähenbühl P, Donahue J, et al. Context Encoders: Feature learning by inpainting[C] //Proc of the 29th IEEE/CVF Int Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 2536−2544

[20] Liu Guilin, Reda F A, Shih K J, et al. Image inpainting for irregular holes using partial convolutions[C] //Proc of the 15th IEEE/CVF European Conf on Computer Vision. Piscataway, NJ: IEEE, 2018: 89−105

[21] Zeng Yuhui, Lin Zhouchen, Lu Hongbin, et al. CR-Fill: Generative image inpainting with auxiliary contextual reconstruction[C] //Proc of the 18th IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE 2021: 14144−14153

[22] Zhang Hui, Qi Yuchao, Xue Xiaoyan, et al. Ancient stone inscription image denoising and inpainting methods based on deep neural networks[J]. Discrete Dynamics in Nature and Society, 2021, 2021: 1−12

[23] Nazeri K, Ng E, Joseph T, et al. EdgeConnect: Structure guided image inpainting using edge prediction[C] //Proc of the 17th IEEE/CVF Int Conf on Computer Vision Workshop. Piscataway, NJ: IEEE, 2019: 3265−3274

[24] Yeh R A, Chen C, Lim T Y, et al. Semantic image inpainting with deep generative models[C] //Proc of the 30th IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 6882−6890

[25] Yu Jiahui, Lin Zhe, Yang Jimei, et al. Generative image inpainting with contextual attention[C] //Proc of the 31st IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 5505−5514

[26] Yan Zhaoyi, Li Xiaodong, Li Mu, et al. Shift-Net: Image inpainting via deep feature rearrangement[C]//Proc of the 16th IEEE/CVF European Conf on Computer Vision. Piscataway, NJ: IEEE, 2018: 3−19

[27] Zeng Yijun, Gong Yi, Zeng Xianglong. Controllable digital restoration of ancient paintings using convolutional neural network and nearest neighbor[J]. Pattern Recognition Letters, 2020, 133: 158−164

[28] Qin Jie, Bai Honggang, Zhao Yizhou. Multi-scale attention network for image inpainting[J]. Computer Vision and Image Understanding, 2021, 204: 1−12

[29] Cao Jianfang, Li Yanfei, Cui Hongyan, et al. Restoration of ancient temple murals by adaptive sample block local search algorithm[J]. Journal of Computer Aided Design & Computer Graphics, 2019, 31(11): 2030−2037(in Chinese)
(曹建芳, 李艳飞, 崔红艳. 自适应样本块局部搜索算法对古代寺观壁画的修复 [J]. 计算机辅助设计与图形学学报, 2019, 31(11): 2030−2037)

[30] Li Qingquan, Wang Huan, Zou Qin. A mural restoration algorithm based on sparse representation model[J]. Geomatics and Information Science of Wuhan University, 2018, 43(12): 1847−1853(in Chinese)
(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)

[31] Chen M, Radford A, Child R, et al. Generative pretraining from pixels[C] //Proc of the 37th Int Conf on Machine Learning. New York: ACM, 2020: 1691−1703

[32] Parmar N, Vaswani A, Uszkoreit J, et al. Image Transformer[C] //Proc of the 35th Int Conf on Machine Learning. New York: ACM, 2018: 4055−4064

[33] Wan Zijian, Zhang Jing, Chen Dongdong, et al. High-fidelity pluralistic image completion with transformers[C] //Proc of the 18th IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE 2021: 4672−4681

[34] Yang Yuting, Jiao Licheng, Liu Xu, et al. Transformers meet visual learning understanding: A Comprehensive Review[J]. arXiv preprint, arXiv: 2203. 12944, 2022

[35] Radford A, Wu J, Child R, et al. Language models are unsupervised multitask learners[C/OL]. 2019[2021-04-17].https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

[36] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[C] //Proc of the 17th Conf of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: ACL, 2019: 4171−4186

[37] Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network[C] //Proc of the 30th IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 105−114

[38] Rudin L I, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms[J]. Physica D: Nonlinear Phenomena, 1992, 60(1): 259−268

[39] Zhang R, Isola P, Efros A A, et al. The unreasonable effectiveness of deep features as a perceptual metric[C] //Proc of the 31st IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 586−595

[40] Yu Jiahui, Lin Zhe, Yang Jimei, et al. Free-form image inpainting with gated convolution[C] //Proc of the 17th IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2019: 4471−4480



Wang Zhenyan, born in 1999. Bachelor. His main research interest includes image processing.
王真言, 1999 年生. 学士. 主要研究方向为图像处理.



Jiang Shengcheng, born in 2000. Bachelor. His main research interest includes image processing.
蒋胜丞, 2000 年生. 学士. 主要研究方向为图像处理.



Song Qihong, born in 2001. Bachelor. His main research interest includes image processing.

宋奇鸿, 2001 年生. 学士. 主要研究方向为图像处理.



Liu Bo, born in 1987. PhD, lecturer. Member of CCF. His main research interests include image processing and multimedia forensics.

刘　波, 1987 年生. 博士, 讲师. CCF 会员. 主要研究方向为图像处理、多媒体取证.



Bi Xiuli, born in 1982. PhD, associate professor, PhD supervisor. Member of CCF. Her main research interests include image processing and multimedia information security.

毕秀丽, 1982 年生. 博士, 副教授, 博士生导师. CCF 会员. 主要研究方向为图像处理、多媒体信息安全.



Xiao Bin, born in 1982. PhD, professor, PhD supervisor. Member of CCF. His main research interests include image processing and pattern recognition.

肖　斌, 1982 年生, 博士. 教授, 博士生导师. CCF 会员. 主要研究方向为图像处理、模式识别.