



基于孪生级联空间滤波的中国传统画修复

薛文喆¹, 董兴宇¹, 胡琦瑶¹, 曹瑞¹, 彭先霖²

(1. 西北大学 信息科学与技术学院, 陕西 西安 710127; 2. 西北大学 艺术学院, 陕西 西安 710127)

摘要 中国传统画作为宝贵的文化遗产, 历经时间沉淀以及各种自然因素的影响, 常出现开裂、破损和褪色等问题。尽管一些深度学习框架在自然图像修复领域取得了显著进展, 但其大多过度依赖卷积权重共享和平移不变性, 在处理布局复杂、结构抽象的绘画图像时, 难以捕捉其独特的空间特性。针对此问题, 提出一种孪生级联空间滤波(twin cascade spatial filtering, TCSF)预测方法用于中国传统画的修复。TCSF采用层级解码策略, 从多尺度解析绘画图像的层次特征, 并级联空间滤波预测方法得到修复核, 从而由粗到细地复原缺失区域的像素。为了在特征信息匮乏的区域精确地复原缺失的结构和笔触信息, 进一步引入空间编码机制。通过对滤波特征图空间编码得到坐标矩阵, 并在滤波预测过程中注入坐标信息编码, 用于缺失像素点恢复时提供空间信息参照, 进而提升修复结果的精确度与视觉效果。实验中, 选取了具有代表性的中国传统画图像进行训练, 并增加壁画数据集和 Places 数据集测试模型的泛化性能。与现有工作使用掩码不同, 该研究在实验中提取部分真实绘画图像的破损掩码, 以更逼真地模拟破损情况。定性和定量实验结果表明, 该方法在中国传统画恢复任务中取得了较好的修复结果, 为数字艺术修复和文化遗产保护提供了有益的启示。

关键词 图像修复; 空间滤波预测; 中国传统画修复; 文物图像破损掩码

中图分类号: TP391.7 **DOI:** 10.16152/j.cnki.xdxbr.2025-01-013

Restoration of traditional Chinese painting based on twin cascade spatial filtering

XUE Wenzhe¹, DONG Xingyu¹, HU Qiyao¹, CAO Rui¹, PENG Xianlin²

(1. School of Information Science and Technology, Northwest University, Xi'an 710127, China;

2. School of Art, Northwest University, Xi'an 710127, China)

Abstract Traditional Chinese paintings are invaluable cultural legacies, but they often suffer from issues such as cracking, damage, and fading due to the effects of time and various natural factors. While some deep learning frameworks have made significant progress in natural image restoration, they tend to rely heavily on convolutional weight sharing and translational invariance. This reliance may limit their ability to fully capture

收稿日期: 2024-06-08

基金项目: 国家自然科学基金(62471390, 62306237); 陕西省重点研发计划(2024GX-YBXM-149); 西北大学研究生创新项目(CX2024204, CX2024206)。

第一作者: 薛文喆, 男, 从事计算机视觉领域中的图像修复、图像推开研究, xwz@stumail.nwu.edu.cn。

the unique spatial characteristics of paintings with intricate layouts and abstract structural information. To address this issue, this paper proposes a Twin Cascade Spatial Filtering (TCSF) prediction method for the restoration of traditional Chinese paintings. The TCSF adopts a hierarchical decoding strategy that analyzes the hierarchical features of painting images across multiple scales. It cascades a spatial filtering prediction approach to obtain restoration kernels, restoring missing region pixels from coarse to fine detail. Furthermore, in order to precisely restore the missing structural and brushstroke information in areas where feature information is sparse, this paper introduces a spatial encoding mechanism. By spatially encoding the filter feature maps into coordinate matrices and infusing coordinate information encoding into the filtering prediction process, this paper provides spatial reference information for the recovery of missing pixels, thereby enhancing the accuracy and visual quality of the restoration outcomes. In the experiments, the model was trained using representative images of traditional Chinese paintings, and the mural datasets and Places datasets were added to test the model's generalization ability. In contrast to existing work that utilizes synthetic masks, this paper extracted actual damage masks from real painting images in order to more realistically simulate damage scenarios. The qualitative and quantitative experimental results demonstrate that the proposed method achieves favorable restoration results in traditional Chinese painting recovery tasks and provides useful insights for digital art restoration and cultural heritage protection.

Keywords image restoration; spatial filter prediction; restoration of traditional Chinese paintings; broken images of cultural relics mask

中国传统画是珍贵的世界文化遗产,承载着丰富的历史、文化和艺术价值,但由于人为和环境等因素的影响,致使画面出现裂纹、脱落等问题(见图 1)。为解决此问题,传统方式常通过人工

修复,但由于修复人员习惯的差异,不仅难以还原出原始画作细腻的笔触和艺术风格,而且可能带来二次损害。因此,采用数字化修复方式还原破损画作对文化遗产保护来说具有重要意义。



图 1 一些受损的中国传统绘画作品

Fig. 1 Some damaged traditional Chinese painting works

在计算机视觉领域,图像修复一直是研究的热点,旨在根据已有像素信息来预测缺失信息。随着深度学习方法在视觉领域中的不断普及,现有方法已可以将一张破损的真实世界图像,通过

生成或差值预测方式还原其高质量样貌。其中,卷积神经网络(convolutional neural networks, CNNs)的应用在自然世界图像的补全任务中获得了一些成果^[14],这主要受益于卷积操作在提取不

用区域特征的平移等变性。但是,该方法面对包含复杂的纹理特征信息的中国传统画时,模型无法感知当前特征和全局结构的关联性,而在绘画图像的修复任务中,学习绘画图像的结构布局信息对于模型理解绘画的深层含义至关重要。尽管现有的方法已经在真实世界的图像上取得了出色的性能,但中国传统画通常具备独特的艺术特征,并包含复杂的物象排列方式和布局格式,如何在破坏原始画作艺术特点的基础上修复其缺失的结构和高频细节是一个艰巨的挑战。

目前,对于中国传统画、壁画等艺术类图像的修复任务已经取得了一定进展。在壁画修复任务中,Li 等人提出一种渐进式 GAN 结构的修复网络,通过提取文物图像的彩绘素描和梯度图,并融合两者进行信息互补从而改善壁画图像的修复效果^[5]。曹建芳等人使用经典的迁移模型 CycleGAN^[6]进行风格调制,并结合协调注意力机制缩小修复结果与原始壁画的色彩和风格差异^[7]。此外,胡升等人还使用快速傅里叶卷积(FTR)^[8]残差块和门控卷积(Gated-Conv)^[9],以增强图像关键特征的表达并且过滤噪声,从而恢复壁画中缺失的结构^[10]。相较于壁画修复的研究,目前对于中国传统画的修复工作还相对较少,赵磊等人采用多路编码器和双重注意力机制从多个尺度学习古画的语义特征,并从风格和内容两方面保证全局一致性,有效提升了古画修复的质量,拓展了这一领域的工作^[11]。尽管上述方法通过生成或风格转换的方式一定程度上改善了艺术类图像的修复效果,但通过生成方式来填补缺失区域的方式与纯粹的修复任务目标存在明显差异。不合理的推理过程可能得到模糊和扭曲的结果,不仅会使深层语义信息丢失,还会破坏作品原有的意境和韵味。

为了解决上述问题,本研究提出了一种新的方法用以改善中国传统画修复的效果。首先,本研究构建了一个高分辨率传统画数据集和真实的传统画掩码图像集用以满足模型的训练需求。其次,提出了一个适用于中国传统画的修复网络——孪生级联空间滤波(twin cascade spatial filtering, TCSF),通过并行两个分支,在多尺度下,采用进交互的方式恢复传统画的复杂结构和语义级信息。在此基础上使用空间滤波核模块(spatial kernel module, SKM)将层次特征坐标化,在滤波预测过程中融入坐标空间坐标编码,提升模型的

空间感知力。最终的定性和定量实验结果表明,在不同类型和范围的遮盖下,TCSF 均取得了出色的修复效果,并在各项评估指标上实现了提升。

1 相关工作

1.1 图像修复

传统的图像修复方法通常将全局搜索的相似内容直接复制到缺失区域^[12-15],或通过周围未损失信息扩散到缺失区域进行修复^[16-17]。这些方法通常能完成简单场景和小范围缺失的补全任务,但对于包含复杂场景的图像,其恢复结果不尽人意。

近年来,深度学习方法在图像修复领域取得了巨大成功。Pathak 等人基于生成对抗方式学习图像的深层语义信息,完成大范围的填补^[18]。此后,许多学者使用更复杂的网络或学习方式来生成高保真的图像,包括风格迁移^[19]、Transformer^[20-22]、上下文注意力^[23-26]、傅里叶卷积^[26-27]和扩张卷积^[27-29]等。另外,还有学者使用一些额外线索如物体边缘^[30-31]、结构图^[20]和语义分割图^[32]引导进行图像的修复。

虽然上述方法在真实世界图像缺失内容的填补中取得了一定效果,但它们侧重于在缺失区域上限制内容合理地生成,并非忠实地修复原始内容。在传统画修复任务中更注重保留原作品的艺术价值,而非仅仅恢复图像的完整性。为此,本研究旨在利用有效输入像素信息而重建缺失区域,在不破坏原始画作笔触和风格的基础上,实现对传统画图像中的缺失部分进行合理准确的修复。

1.2 滤波修复

图像修复的经典范式是搜寻相似信息进行补充,这种填充方式通常从已知图像的上下文或外部数据中粘贴相似的信息。然而,该方法无法实现对传统画中复杂场景进行语义级修复:由于对低级特征的过度依赖,从而导致模型无法合成在已知图像上下文中不存在相似补丁的内容。为此,一些深度生成工作通过编码和解码的方式来生成相似内容^[33-35]。虽然这些工作在一些具有挑战的真实世界图像上改善了传统修复方法的不足,但是这些方法在处理语义信息复杂的传统画图像时还存在不足。在传统绘画的修复过程中,需要尽可能明确表达图像中残留的先验信息,而不仅是简单地生成任务。

近来,一些视觉任务尝试使用滤波方式还原图像信息,如去噪^[36-37]、去模糊^[38]、去雨^[39]。Guo 等人首次将深度滤波预测用于自然图像的修复任务中,有效改善了图像的局部伪影,提高了图像的修复质量^[40]。Li 等人细化了滤波核等级,使用一种双流网络连接结构和纹理特征以恢复图像的语义信息和高频细节^[41]。

预测滤波的方式是用邻域像素的数值信息对缺失或损坏的像素进行估计,预测缺失像素的值从而重建图像,计算公式可以表示为

$$\hat{I}_q = K_w \times \sum_{p \in I_{p-q}}^{I_{p+q}} \times I_p \quad (1)$$

式中: $\hat{I}_q$ 表示缺失的区域的像素, $K_w \in R^{k \times k}$,表示大小为 k 的滤波核, w 表示可学习的滤波核参数, I_p 表示缺失区域的临近像素。

预测滤波方法使用深度卷积网络学习合适的参数而实现动态的调整滤波核的目的,这个过程可以表示为

$$K = F_{\theta}(\bar{I}) \quad (2)$$

这里 K 为可变滤波核, $\bar{I}$ 为破损图像, F_{θ} 是滤波预测网络, θ 为可学习参数。

总的来说,预测滤波的目标是构建一个合适的深度预测网络,对图像保留区域的非线性特征

进行学习,从而利用预测得到的滤波核构建缺失特征。为充分学习传统画复杂的布局结构和技法特点,本研究通过级联多次滤波并为滤波核增加显示的空间编码,以提升模型对绘画图像中复杂空间的感知力。

2 TSCF 网络修复方法

2.1 网络总览

本研究提出的孪生级联空间滤波网络 (TSCF) 整体结构图如图 2 所示。首先,从残缺画作中提取真实破损的掩码数据,与完好的绘画图像拼接得到 TSCF 的输入数据。其中 TSCF 通过两个阶段完成了修复过程。

1) 上层利用可见区域信息,预测用于缺失区域修复的滤波核参数;另外,预测过程中接受下层 CNN 编码器反馈的特征,以补偿丢失的大范围特征;最后,将得到滤波核与预测特征的空间编码融合,以增强模型对缺失像素在整体结构中的相对位置的学习能力,从而在核预测过程中感知图像的整体布局,实现空间滤波预测的目的。

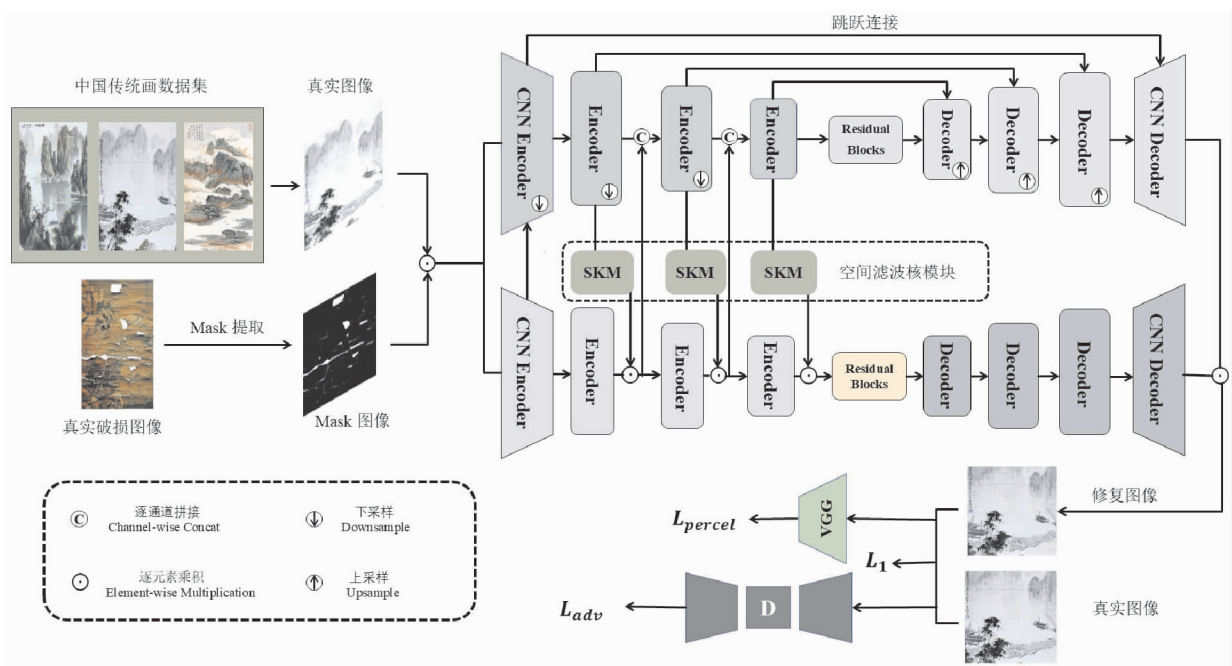


图 2 TSCF 整体框架图

Fig. 2 The overall framework of TSCF

2) 下层使用 U-net 结构的网络,逐层编码图像不同尺度下的层次特征,并多次利用从上层预测得到的不同尺度的空间滤波核进行阶段修复;

完成当前阶段的修复后,将得到的修复结果传回上层用于新一轮的滤波核预测,反复多次,最终实现缺失区域的恢复。

下面详细介绍 TCSF 中包含的组件和损失函数。

2.2 基于空间滤波预测修复

在传统绘画修复中,还原大范围缺失区域中的高分辨细节是一项重大挑战。由于画家们通常使用精细的笔触来渲染作品的意境和氛围,故画面中包含丰富的艺术特征和抽象的表现手法,而在恢复这些高频细节时需要依赖丰富的低级特征,为此,TCSF 通过在下层分支设置 4 个 CNN 编码器(包括 1 个 3×3 卷积层,一个归一化层和一个 ReLU 激活层),逐层提取不同层次的图像特征,并反馈到核预测分支,使网络逐步学习复杂和抽象的特征表示。具体而言,对于给定的一幅被遮挡的传统画图像 $\tilde{I}$,可表示为:

$$\tilde{I} = I \odot I_{\text{mask}} \quad (3)$$

其中: $\tilde{I}, I, I_{\text{mask}} \in R^{H \times W \times C}$, 分别为原始图像和 Mask 图像,将得到的缺失图像同时输入到网络的两个分支中。其中,下分支接受缺失图像 $\tilde{I}$,并通过下 CNN 编码器将图像特征从 RGB 空间映射到特征空间,并得到输出特征 F_i ,将 F_i 先通过 Avgpool 层进行特征图尺寸降采样处理,以保留图像的重要特征,之后将 F_i 送入空间滤波核的预测分支用于滤波预测操作:

$$E_{\text{Down}}^i = \text{Avgpool}(\varphi_i(\tilde{I}_{H \times W \times C})), (i = 1, 2, 3) \quad (4)$$

式中: $E_{\text{Down}}^i \in R^{H/2^i \times W/2^i \times C/2^i}$ 为下层输出的编码特征, $\varphi_i(\cdot)$ 表示下层第 i 次编码操作, $\text{Avgpool}(\cdot)$ 表示全局平均池化操作。

与此同时,上分支先将从下层得到的特征 E_{Down}^i 和自身编码器的输出特征 E_{Up}^i 逐通道拼接(channel-wise concat)后,再通过一个上层的编码器得到输出的滤波特征 E_{Filter}^i ,并将滤波特征送入 SKM,以生成该阶段的空间滤波核,此过程可以表示为

$$E_{\text{Filter}}^i = \xi_i(E_{\text{Up}}^i \oplus E_{\text{Down}}^i) \quad (5)$$

$$\hat{k}_i = \delta_{\text{SKM}}(E_{\text{Filter}}^i), (i = 1, 2, 3) \quad (6)$$

式中: $\hat{k}_i$ 表示第 i 阶段的空间滤波核, $\xi(\cdot)$ 表示空间滤波预测分支的编码操作, $\delta_{\text{SKM}}(\cdot)$ 表示空间编码操作。最后,下分支使用得到的空间滤波核进行逐元素相乘(element-wise multiplication),即滤波操作后完成本阶段修复,这个过程可以表示为

$$E_{\text{Down}}^i = \varphi^j(\cdots \varphi^2(\varphi^1(E_{\text{Down}}^0) \downarrow \odot k_1) \downarrow \odot k_2 \cdots) \downarrow \odot k_j, \quad (j = 1, 2, 3, \cdots) \quad (7)$$

$$E_e = \varphi^j(\cdots \varphi^2(\varphi^1(E_{\text{Down}}^0) \downarrow \odot k_1) \downarrow \odot k_2 \cdots) \downarrow \odot k_j, \quad (8)$$

式中: E_e 表示下分支最后一次编码的输出, $\downarrow$ 表示通过全局平均池化层进行下采样操作, E_e 为生成分支通过多次空间滤波操作后得到的编码特征,将得到的编码特征通过中间设置的多个残差块(residual blocks)学习高层特征信息,最终通过解码器解码得到最终修复结果。

2.3 空间编码融合

中国传统画注重空间透视和景物的层次感,因此空间感知能力对模型理解传统画的复杂布局 and 结构至关重要。此前,已经有一些研究工作验证了在模型拟合过程中注入空间几何信息的有效性^[42]。

为了增强模型对传统画空间结构的感知能力,本研究对输入的预测滤波特征使用空间滤波核模块(spatial kernel module, SKM)进行空间编码,以获取缺失像素的坐标信息,具体流程如图 3 所示。通过这种方式来为预测的滤波提供空间先验信息,使网络在卷积操作时更加关注缺失区域在画面中的相对位置,从而更好地捕捉到图像中的空间结构和局部特征。具体而言,SKM 模块先接收上层输出的一个滤波特征 E ,编码后得到生成 X, Y 方向的坐标信息,可表示为

$$(X_i, Y_i) = \frac{2 \times E_{h_{\text{dim}}, w_{\text{dim}}}}{E_{(h_{\text{dim}}, w_{\text{dim}}) - 1}} - 1, (i = 1, 2, 3) \quad (9)$$

式中: $X_i \in R^{B \times H \times 1 \times 1}$, $Y_i \in R^{B \times 1 \times W \times 1}$ 分别第 i 阶段为 X 方向上和 Y 方向上的编码得到的坐标信息, $E_{h_{\text{dim}}, w_{\text{dim}}}$ 表示第 i 次预测输入的滤波特征图。

其次,将滤波特征与 X, Y 方向上的坐标信息进行拼接,并使用单层卷积融合坐标信息。最后,使用 ReLU 函数激活关键特征区域得到空间滤波核 $\hat{K}_i$,此过程可表示为

$$\hat{K}_i = \text{ReLU}(\text{Conv}(E_{\text{Filter}}^i \oplus X_i \oplus Y_i)), \quad (i = 1, 2, 3) \quad (10)$$

式中: E_{Filter}^i 表示预测滤波特征; $\text{Conv}(\cdot)$, $\text{ReLU}(\cdot)$ 分别表示核大小为 3×3 的卷积和激活操作; $\hat{K}_i$ 表示第 i 次预测滤波得到空间滤波核。

2.4 损失函数

在传统画修复任务中,损失函数的选择直接影响模型重建缺失特征的过程,为确保修复结果有更高的准确性和保真度,同时忠实地保留艺术品的原貌,本研究采用 L1 损失、感知损失^[43]、对

抗损失^[44]和风格损失^[45]来指导模型修复。

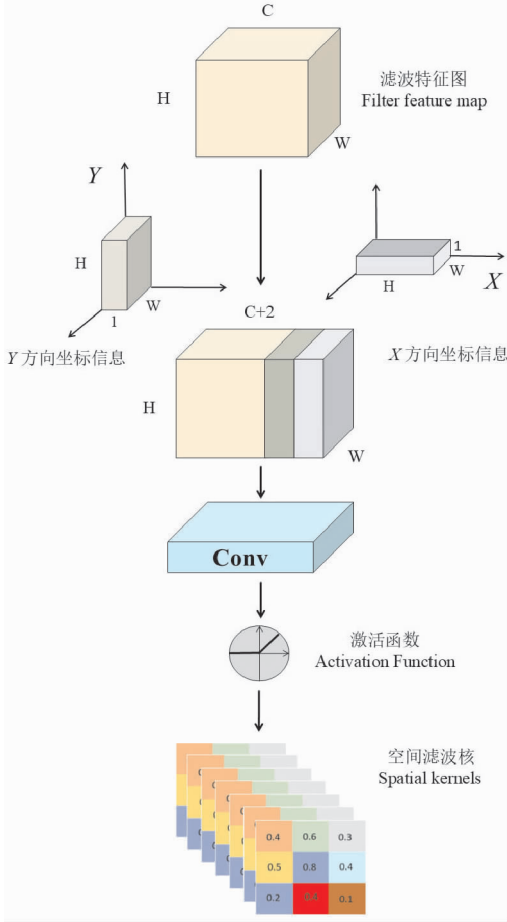


图 3 空间滤波核模块

Fig. 3 Spatial kernel module

1) L1 损失函数。L1 损失用于捕获修复图像与原始图像之间像素级别的差异,用于优化修复图像的整体结构和细节,可表示为

$$L_{l1} = \frac{1}{N} \sum_{i=1}^N |I_{gt} - I_o| \quad (11)$$

式中: n 表示样本数量; I_{gt} 表示原始图像; I_o 是修复后的图像。

2) 感知损失。感知损失有利于对齐修复图像与原始图像之间的高级语义特征,通常使用经过预训练的卷积神经网络在特征空间比较两张图像的差异,以此提高修复结果的感知质量,可表示为

$$L_{per} = \frac{\sum_{i=1}^N \|\varphi_i(I_{gt}^{(i)}) - \varphi_i(I_o^{(i)})\|_1}{N} \quad (12)$$

式中: $\varphi(\cdot)$ 为预训练 vgg-19 网络^[43]第 i 层的激活特征图, N 为样本数量。

3) 对抗损失。对抗损失用于激励生成网络生成与原始图像相似的特征分布,通过逐步优化生

成的图像,使其更符合原始图像的结构和语义,具体可表示为

$$L_{adv} = -E[\log D(I_{gt})] - E[\log(1 - D(I_o))]. \quad (13)$$

4) 风格损失。风格损失衡量修复结果与原始图像的风格差异,确保修复后的传统画图像在艺术风格上与原始图像保持一致,可表示为

$$L_{style} = E_i[\|A_i(I_o)^T A_i(I_o) - A_i(I_{gt})^T A_i(I_{gt})\|_1] \quad (14)$$

式中: $A_i(\cdot)^T A_i(\cdot)$ 为第 i 层特征图的 Gram 矩阵,整个网络的优化目标为上述损失的加权组合,可表示为

$$L_{total}(I, \tilde{I}) = \lambda_{l1} L_{l1} + \lambda_{per} L_{per} + \lambda_{adv} L_{adv} + \lambda_{style} L_{style} \quad (15)$$

式中: λ_{l1} 、 λ_{per} 、 λ_{adv} 和 λ_{style} 超参数在实验过程中分别设置为 $\lambda_{l1} = 1$, $\lambda_{style} = 250$, $\lambda_{adv} = \lambda_{per} = 0.1$ 。

3 实验设置与分析

3.1 实验准备

在实验中,本研究使用了多种数据集评估模型的性能,包括传统画数据集、壁画数据集、Places 数据集以及 Mask 数据集。为了专门测试模型在特殊数据集上的修复能力,本研究构建了传统画数据集和壁画数据集。此外,还利用公开的 Places 数据集检验本模型在真实世界图像中的泛化能力。

1) 传统画数据集。在传统画修复任务中,训练过程中所需的高质量图像数据通常非常稀缺,而高分辨率图像需要在专业的条件下获取,能满足模型训练要求的数据更是稀缺。为满足模型训练的需求,本研究构建了一个中国传统绘画数据集,其中包括来自不同朝代、画家和不同题材共 5 500 张高分辨率训练集图像以及 121 张测试集图像。

2) 壁画数据集。除使用传统画数据集以外,本研究还构建了一个壁画数据集来验证模型在其他艺术图像数据中的修复效果。其中,壁画数据集包括一千余张不同类型的壁画图片(唐卡壁画、洞穴壁画和墓葬壁画)。为测试模型的泛化性能,在实验中选择部分具有代表性的图像进行展示。

3) Places^[46]。其包含 180 万个不同地方的图像样本,覆盖了超过 400 种不同的场景类别,如

街道、海滩、卧室、森林等,从而提供了极其多样化的真实世界中的场景。

4) Mask 图像设置。区别于之前工作中^[47-49]的掩码规则,为了更加真实地模拟绘画作品的受损情况,本研究使用阈值法定位传统画图像中的破损位置,并根据真实的传统画破损样例定制了更加逼真的 Mask 数据集共 1 200 张,用以丰富现有的 Mask 数据集。数据展示如图 4 所示。

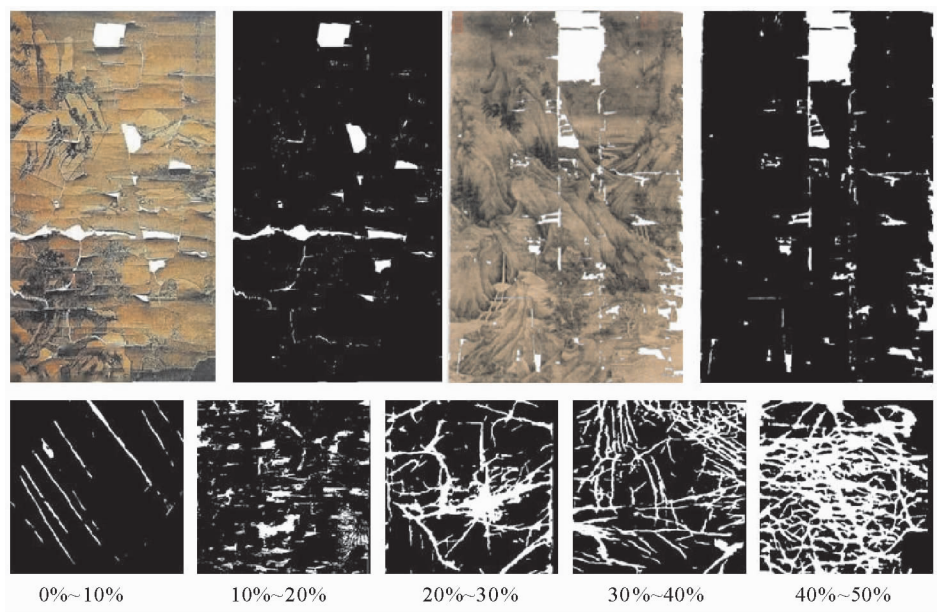


图 4 模拟真实破损的掩码图像
Fig.4 Simulation of real broken mask images

5) 评估指标。本研究遵循图像修复任务中最常见的评价设置,分别使用峰值信噪比 (PSNR)、结构相似度指数 (SSIM)、 L_1 距离和感知相似度^[50] (LPIPS) 来评估图像修复的质量。其中 L_1 距离用于评估重建像素的准确性,PSNR 和 SSIM 用于从亮度、对比度中评估图像恢复的质量,LPIPS 指标结合了人类视觉感知的特性,用来评估恢复的图像和真实图像之间的感知差异程度。

3.2 实施细节

本研究提出的 TCSF 网络基于 Pytorch^[51] 实现,所有的训练和测试在一张 RTX3090 (24 GiB) 显卡上完成,训练时间花费了 3 d,共进行了 500 000 次迭代。在训练过程中,设置模型学习率为 0.000 1,批次大小为 12,并使用 Adam^[52] 优化器训练模型,其中参数 $\beta_1 = 0.1, \beta_2 = 0.9$ 。

3.3 对比方法

以下是参与比较的各先进方法的简要概述。

1) Pconv^[49]:提出一种部分区域卷积方式,

在训练阶段和测试阶段中,为了和其他方法在性能评估上保持一致,本研究将构建的 Mask 数据集和 Liu 等人提出的自由孔洞数据集结合使用,并保持遮盖率在 0% ~ 50% 范围内,且每个区域段 (0% ~ 10%, 10% ~ 20%, 20% ~ 30%, 30% ~ 40%, 40% ~ 50%) 设置的 Mask 数据量相同。

将遮盖和保留区域隔离操作,改善了修复区域与保留区域的颜色差异。

2) HAN^[35]:一种 Transformer 结构的网络,在注意力机制中引入 Laplace 距离先验,用于在推理过程中融入图中的距离信息。

3) MISF^[41]:使用一种双分支架构,并提出语义滤波来实现破损图像的高保真修复。

4) AOT-GAN^[53]:提出一种上下文聚合模块以捕获丰富的远程信息,完成上下文推理并改善修复结果的纹理细节。

5) CoordFill^[54]:提出了一种基于连续隐式表示的图像修复方法,并通过自适应参数生成网络和多层感知器,实现高效的高分辨率图像补全。

3.4 定量实验比较

为评估 TCSF 在图像修复中的性能,本研究对构建的传统画数据集进行了不同比例的遮盖 (0% ~ 10%, 10% ~ 20%, 20% ~ 30%, 30% ~ 40%, 40% ~ 50%),并与其他先进方法进行定量比较。为了比较公平,参与测试的所有方法使用

相同的图像-Mask 对,具体实验结果见表 1。

考虑到真实的传统画破损区域多为 30% 以下,TCSF 追求在 0% ~30% 遮盖率下实现最好的效果。从表 1 中可以看出,在遮盖率为 0% ~30% 时,与最优的方法比较,本方法得到了最高的 PSRN,SSIM 和 L_1 指标的值。在遮盖 10% ~20% 下,比 CoordFill 在 PSNR 指标上提升了 1.103 9

dB,比最高模型提升了 0.235 9 dB。在 0% ~10% 遮盖下,SSIM 指标相较于最优模型提高了 0.001 3,同时其他指标上都有提升。综合分析可知,TCSF 在各个评估指标上都取得了较好的结果,由此可以证明本方法相较于其他方法,在传统画图像的修复任务中可以得到更好的定量评估结果。

表 1 TCSF 与其他方法在不同遮盖比例下的定量比较结果

Tab. 1 Results of quantitative comparison between TCSF and other methods at different masking ratios

Mertics	Masks	PConv	HAN	MISF	CoordFill	AOT-GAN	Ours
PSNR	0% ~10%	25.805 2	30.277 7	31.809 7	29.433 0	30.700 2	32.014 7
	10% ~20%	22.256 2	27.260 2	27.616 0	26.691 3	27.559 3	27.795 2
	20% ~30%	21.226 5	25.593 3	25.621 7	25.130 4	25.500 2	25.753 7
	30% ~40%	19.676 7	23.181 0	23.228 5	23.150 4	22.766 5	23.311 9
	40% ~50%	18.028 4	21.881 1	21.498 5	21.671 9	20.962 4	21.582 5
SSIM	0% ~10%	0.844 7	0.940 4	0.956 1	0.898 1	0.918 1	0.957 4
	10% ~20%	0.718 0	0.877 3	0.894 5	0.835 9	0.884 3	0.897 5
	20% ~30%	0.690 0	0.840 8	0.850 1	0.798 7	0.841 5	0.853 5
	30% ~40%	0.629 5	0.770 9	0.767 3	0.731 8	0.764 5	0.771 2
	40% ~50%	0.548 3	0.696 0	0.683 4	0.661 4	0.680 5	0.688 5
L_1	0% ~10%	0.027 0	0.007 8	0.006 2	0.020 3	0.019 3	0.006 1
	10% ~20%	0.043 6	0.016 1	0.014 6	0.027 0	0.023 2	0.014 3
	20% ~30%	0.049 2	0.020 9	0.020 3	0.031 5	0.028 4	0.019 9
	30% ~40%	0.061 8	0.031 0	0.035 8	0.040 3	0.038 7	0.030 9
	40% ~50%	0.080 4	0.041 5	0.042 8	0.049 7	0.049 3	0.042 3
LPIPS	0% ~10%	0.148 6	0.045 2	0.044 7	0.075 1	0.062 2	0.043 7
	10% ~20%	0.262 8	0.099 8	0.099 7	0.114 7	0.099 6	0.099 4
	20% ~30%	0.280 1	0.133 8	0.130 3	0.144 1	0.142 6	0.130 2
	30% ~40%	0.326 1	0.196 0	0.191 0	0.193 0	0.190 7	0.189 7
	40% ~50%	0.386 7	0.253 0	0.249 2	0.249 3	0.259 1	0.249 0

3.5 定性实验比较

除了定量评估以外,本研究将 TCSF 的修复效果与其他方法进行定性的对比,定性的可视化结果展示在图 5、图 6、图 7 和图 8 中。其中,图 5 为在 10% ~20% 破损情况下的模型修复效果的全局展示,图 6,7 中分别从细节上展示了在 10% ~30%,20% ~30% 损情况下的修复效果,图 8 中展示了一些在更大遮盖比例下的修复结果。

如图 5 所示,尽管部分卷积 PConv 接受并记忆掩码图像,能够在推理时忽略覆盖的无效像素,但这种做法可能使模型受掩码的影响,从可视化结果可以看出,在处理“划痕式”的破损时出现了明显的噪点和修复失败的区域。相比之下,Co-ordFill 通过隐式编码建立缺失像素和已有像素间的联系,有效避免了噪点产生,但在树干修复结果中出现上下文不衔接的问题。另外,从图 5(a)的

山体上可以看出另外两种方法在处理山体纹理时都复原失败,而本方法通过图像空间信息补偿有效地改善了这些问题。从图 5(b)中的树干和树叶的修复结果可以看出,其他方法得到的结果中出现了缺失和模糊的情况,而采用本方法,优化了线条的流畅度,同时实现了纹理细节的复原,在整体视觉效果上更加细腻真实。

传统画中常用各种植物元素来表现自然景观的美感,而在修复过程中,由于草木的枝条纹理细腻,故恢复难度较大。从图 6(a)中红色方形框标记的部分可以看出,本方法还原出了树枝和树叶的轮廓结构,并且纹理相比于其他方法得到的结果更加清晰,这是由于滤波预测中融入的空间编码信息能有效缓解修复结果与临近像素间语义特征不一致的问题。图 6(b)展现了一种水墨风格的景观画,从房子中的景观来看,本研究修复后得

到的画中人物轮廓更加清晰。

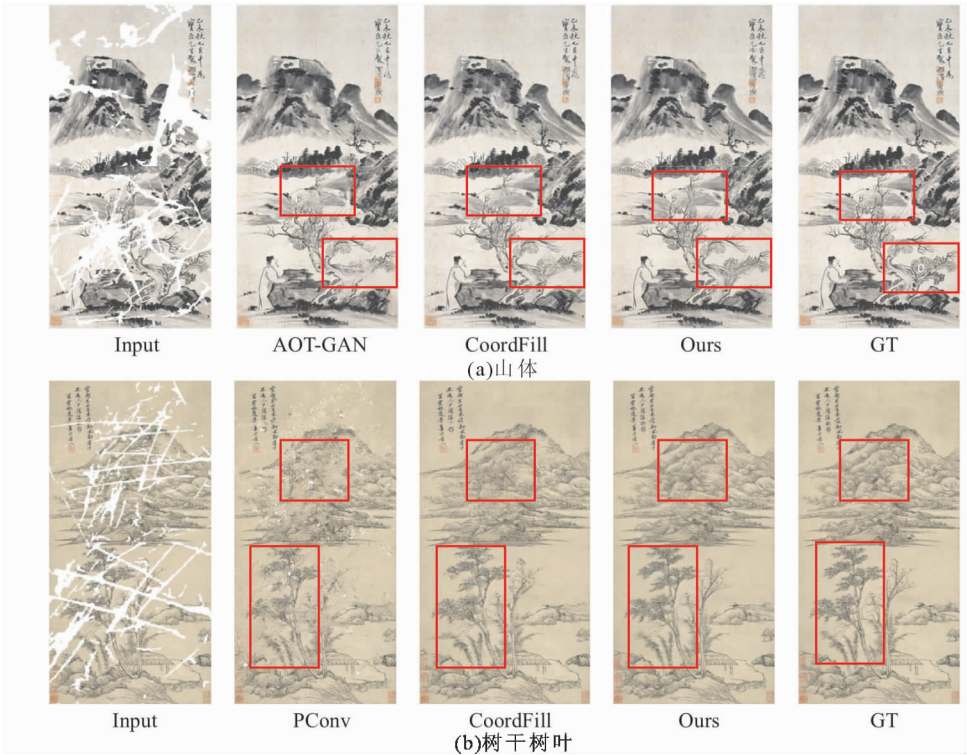


图 5 遮盖率为 10% ~20% 时修复结果全局展示
Fig. 5 Results of qualitative comparison with other methods at masking ratios of 10% ~20%

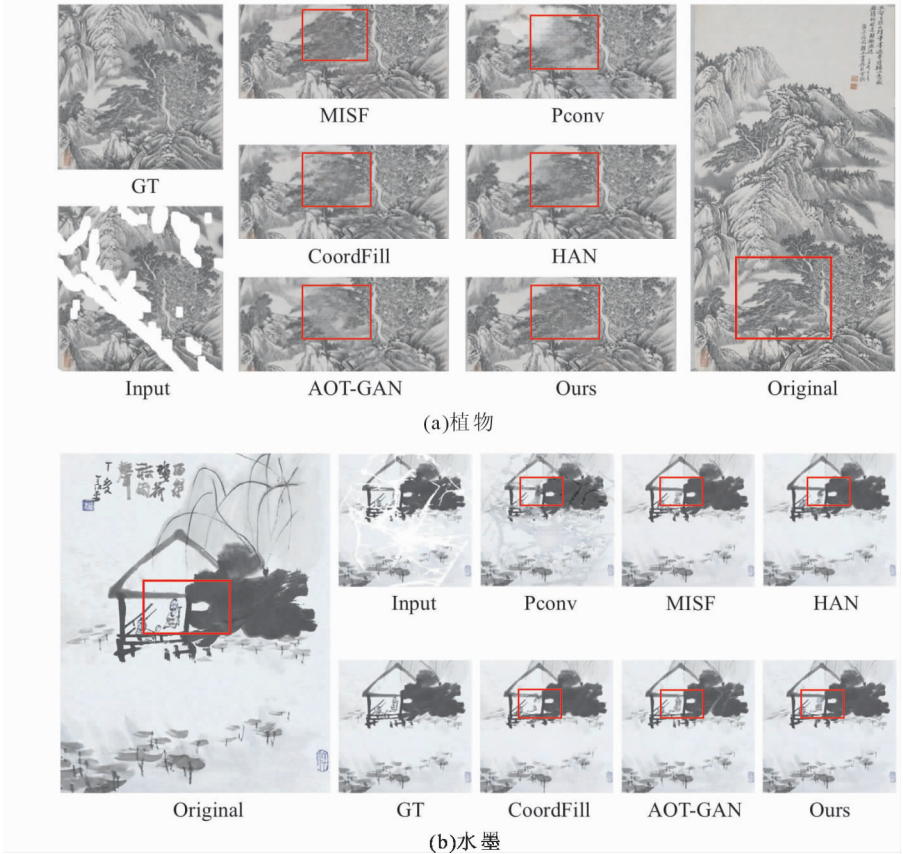


图 6 遮盖率为 10% ~20% 时与其他方法的定性比较结果
Fig. 6 Results of qualitative comparison with other methods at masking ratios of 10% ~20%

从图 7(a)中对于树干的修复结果可以看出, TCSF 能够重建出视觉合理的上下文结构,重构结果与真实图像相似度高。在待修复区域周围可参照信息较少时,本方法也能最大程度上利用已有的像素内容来预测缺失区域。图 7(b)中展示对高密度像素区域的修复结果,其中树叶和花瓣间具有复杂的纹理信息,包括叶脉和花瓣的轮廓,其他方法得到的结果在复原这些细节时过于平滑,

缺乏层次感,同时由于相互遮挡关系错误,导致画面元素堆叠方式不自然。本方法通过对图像语义级信息进行学习,提供了更佳的恢复效果。

另外,图 7(c)展示了在图中江面上的树木信息较少时,TCSF 通过空间编码信息补充推理过程中的信息缺口,为缺失区域的推理注入了关键的非局部特征提示,展示了 TCSF 在对非局部信息建模的有效性。

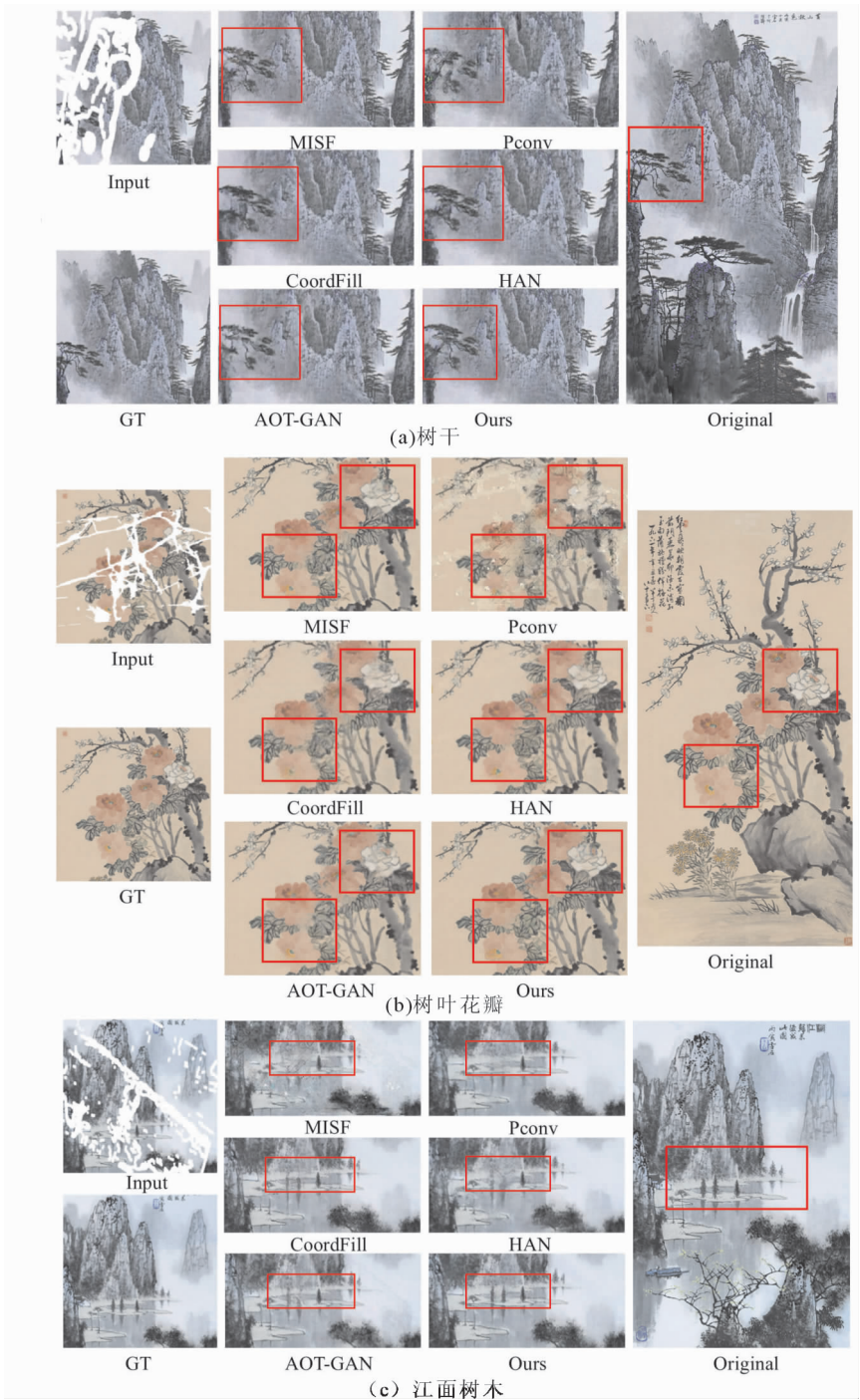


图 7 遮盖率为 20% ~ 30% 时与其他方法的定性比较结果
Fig. 7 Results of qualitative comparison with other methods at masking ratios of 20% ~ 30%

为了验证模型在大范围信息缺失情况下的修复性能,本研究设置了更大范围的遮盖来探索 TCSF 的重构潜力。图 8 中展示了在更大破损条件下的恢复效果。从图 8(a)和图 8(b)结果可以看出,即使在大范围缺失的情况下,本方法也能有

效地恢复出合理并连贯的结构,有力证明了 TCSF 相较于 PConv、AOT-GAN 等同类方法的显著优势。其他方法得到的结果轮廓模糊,原有结构信息被破坏,而 TCSF 得到的结果轮廓清晰、纹理自然,得到了良好视觉效果。

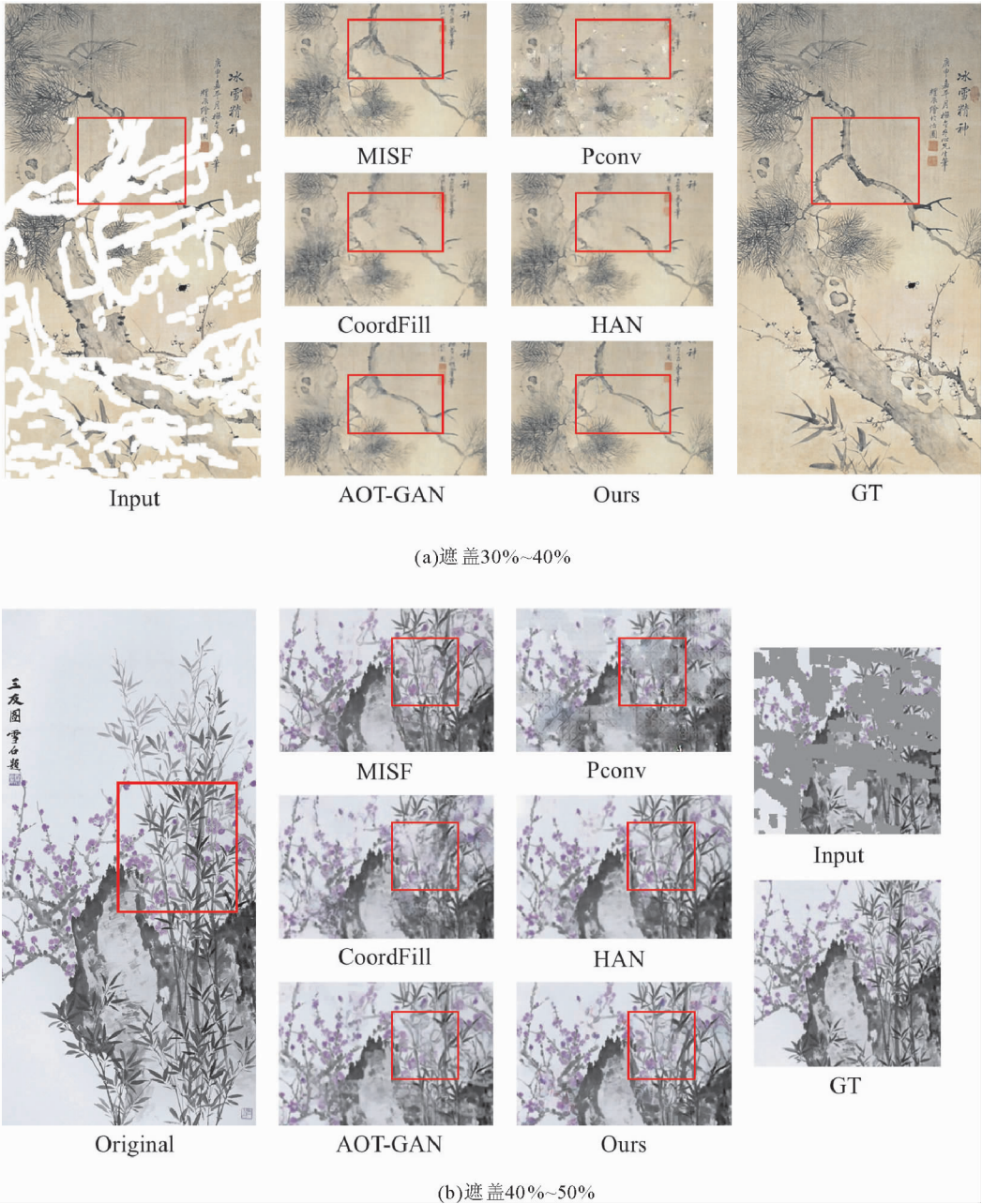


图 8 与其他模型在其他遮盖范围下的定性比较结果
Fig. 8 Qualitative comparison results with other models under other coverage ranges

3.6 消融实验分析

3.6.1 组件消融实验

为证明空间编码在滤波修复过程中的有效性,本研究通过消融实验来评估这一组件,结果展示在表 2 和图 9 中。从表 2 的结果可以看出,在不同破损范围下,注入空间编码后指标都有显著

的提升,在 0% ~ 10% 遮盖下, L_1 距离损失优化了 1.6%,在 40% ~ 50% 的遮盖下优化了 11.8%。在遮盖 30% ~ 40% 下,SSIM 指标提升 0.001 6,并且其他指标也都有提升。由此可以证明,和标准结构相比,显式的位置信息补偿有助于约束滤波修复过程,从而使修复结果更加准确。

表 2 空间滤波核模块消融实验结果
Tab.2 Ablation experiment results of spatial kernel module

Model	MASK	PSNR	SSIM	L_1	LPIPS
(w/o) SKM	0% ~ 10%	31.965 2	0.956 6	0.006 2	0.043 9
	10% ~ 20%	27.770 6	0.895 8	0.014 5	0.100 0
	20% ~ 30%	25.747 1	0.851 6	0.020 1	0.130 7
	30% ~ 40%	23.275 5	0.769 6	0.031 7	0.189 8
	40% ~ 50%	21.564 8	0.687 1	0.042 8	0.247 9
Ours	0% ~ 10%	32.014 7	0.957 4	0.006 1	0.043 7
	10% ~ 20%	27.795 2	0.897 5	0.014 3	0.099 4
	20% ~ 30%	25.753 7	0.853 5	0.019 9	0.130 2
	30% ~ 40%	23.311 9	0.771 2	0.030 9	0.189 7
	40% ~ 50%	21.582 5	0.688 5	0.042 3	0.249 0

对于组件消融学习的可视化结果展示在图 9 中。从图 9(a)的修复结果可以看出,空间滤波修复得到的山体轮廓更加清晰,并且从竹叶的修复结果可以看出,TCSF 学习到正确的图像信息,相较于标准模型得到的结果更加准确。从图 9(b)中树枝和树叶的修复结果可以看到,本方法得到

的结果更加贴合原始图像,而标准模型恢复失败。由此可以看出,空间滤波核的设计能有效地补充传统画图像的全局特征信息,并增强模型上下文的推理能力,从而在复杂图像的修复过程中得到连贯的结构和清晰的纹理。

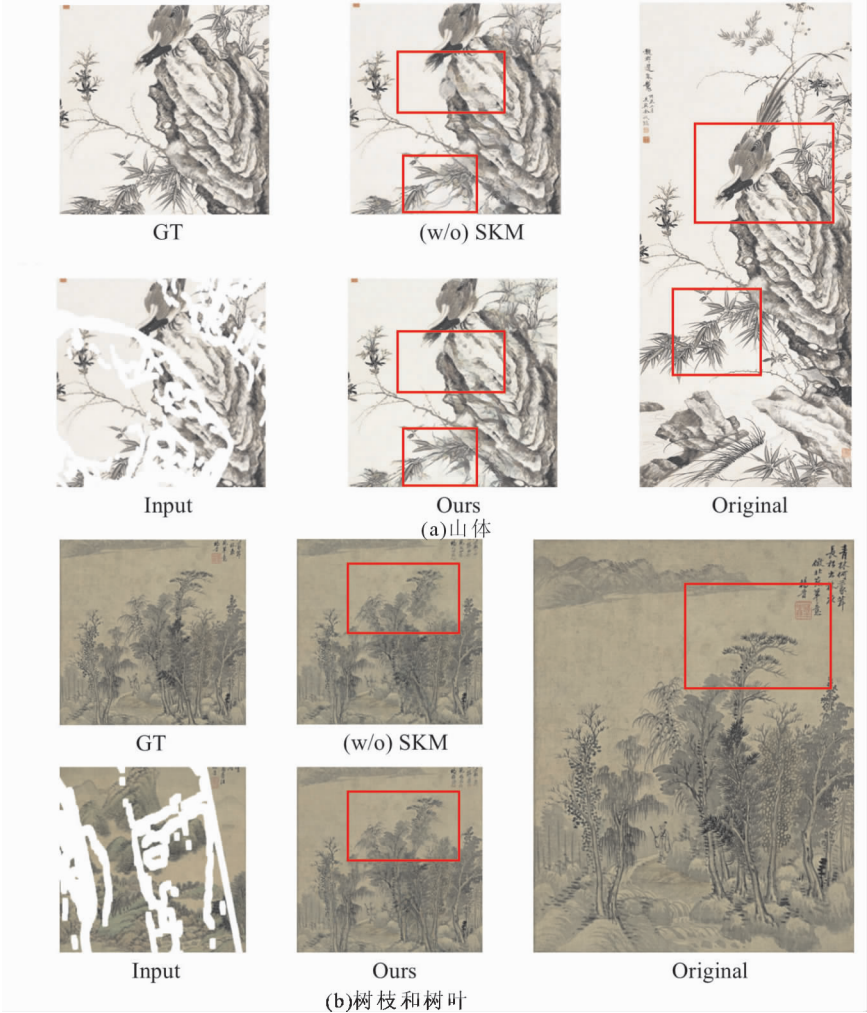


图 9 组件消融实验可视化结果
Fig.9 Visualization results of component ablation experiment

在递进的滤波核生成过程中引入坐标信息可直接为推理过程提供图像的空间位置信息,这将增强模型对图像空间结构的理解能力,从而有利于准确地恢复画作中不同元素的位置信息和形状信息。

另外,显示提供的空间信息将有助于提升模

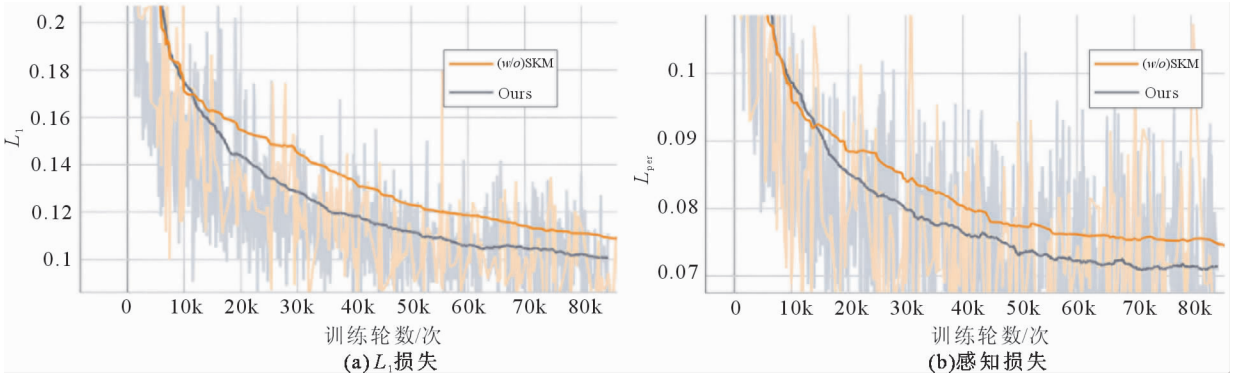


图 10 添加空间滤波预测前后 L_1 损失和感知损失的下降曲线

Fig. 10 The decline curves of L_1 loss and perceptual loss before and after adding spatial filtering prediction

3.6.2 损失函数消融实验

为了探索训练中各种损失对模型修复性能的影响,本研究设计了一组实验对使用的损失函数进行消融学习。具体来说,在实验中分别比较了去除 L_1 损失、感知损失和风格损失以及在原有损失的基础上添加了总变分损失 (total variation loss) 的几种情况,其中总变分损失是通过正则项降低相邻像素的差异,从而降低噪声并保持图像的光滑性。

有关损失函数的消融结果在表 3 和图 11 中展示。从结果可以看出, L_1 损失对于异常值有较强的鲁棒性,在数据存在异常值时可以提供更稳定的结果,同时 L_1 损失有助于生成边缘更清晰的

型在训练过程中的收敛速度和效果。在此,本研究展示了添加空间滤波预测前后 L_1 损失和感知损失的消融实验的可视化过程(见图 10)。从图 10 中可以看出,在空间编码的辅助下,模型的收敛更快。

图像,在去掉 L_1 损失时其他指标也都受到不同程度的影响。如图 11 所示,去掉 L_1 损失后图像出现了模糊,反观其他对照组的结果更加清晰。另外,风格损失用于确保修复图像保留原作的风格特征,通常与内容损失搭配使用,而去掉风格损失后得到的整体图像偏暗,与原始图像风格不符。其次,本研究测试了添加 TV 损失对模型修复性能的影响,从图 11 可以看,添加 TV 损失后能改善图像中的噪点现象,但会出现过度平滑的结果。而本方法结合各种损失,既考虑了修复结果的精确度,也兼顾了图像的感知质量、风格特征和视觉真实性,从而实现对原作的高质量恢复。

表 3 不同损失函数的消融实验结果

Tab. 3 Experimental results of ablation with different loss functions

Loss	PSNR	SSIM	L_1	LPIPS
(w/o) L_{l_1}	25.637 4	0.850 9	0.023 1	0.131 4
(w/o) L_{perc}	25.626 4	0.848 9	0.022 7	0.135 3
(w/o) L_{style}	25.710 6	0.850 5	0.022 8	0.133 7
+ L_{tv}	25.725 3	0.851 8	0.022 5	0.134 5
Ours	25.753 7	0.853 5	0.019 9	0.130 2

3.7 其他数据集集中的适用性测试

3.7.1 壁画数据集测试

壁画数据通常包含复杂的图案和细腻的细节,对修复模型提出了较高的要求,要求模型不仅

能恢复缺失的信息,还要能保持原有的艺术风格。在实验中,本研究构建了几种不同类型的壁画绘画数据集并进行了评估。修复结果如图 12 所示。从图 12 (a) 中的服饰细节信息的还原和图 12

(b)、(c)中对图像结构和色彩的恢复可以看出，好的视觉效果,进一步证明了模型的广泛适用性。本方法在其他艺术类型的数据集上依旧能得到良

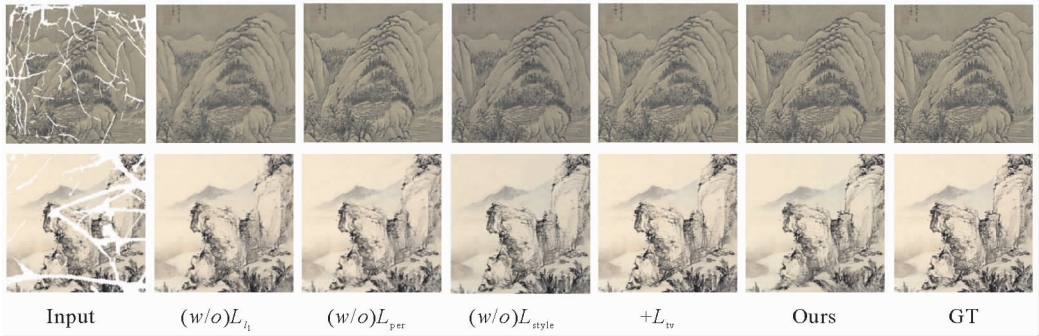


图 11 对不同损失函数消融的可视化展示

Fig. 11 Visual display of ablation of different loss functions

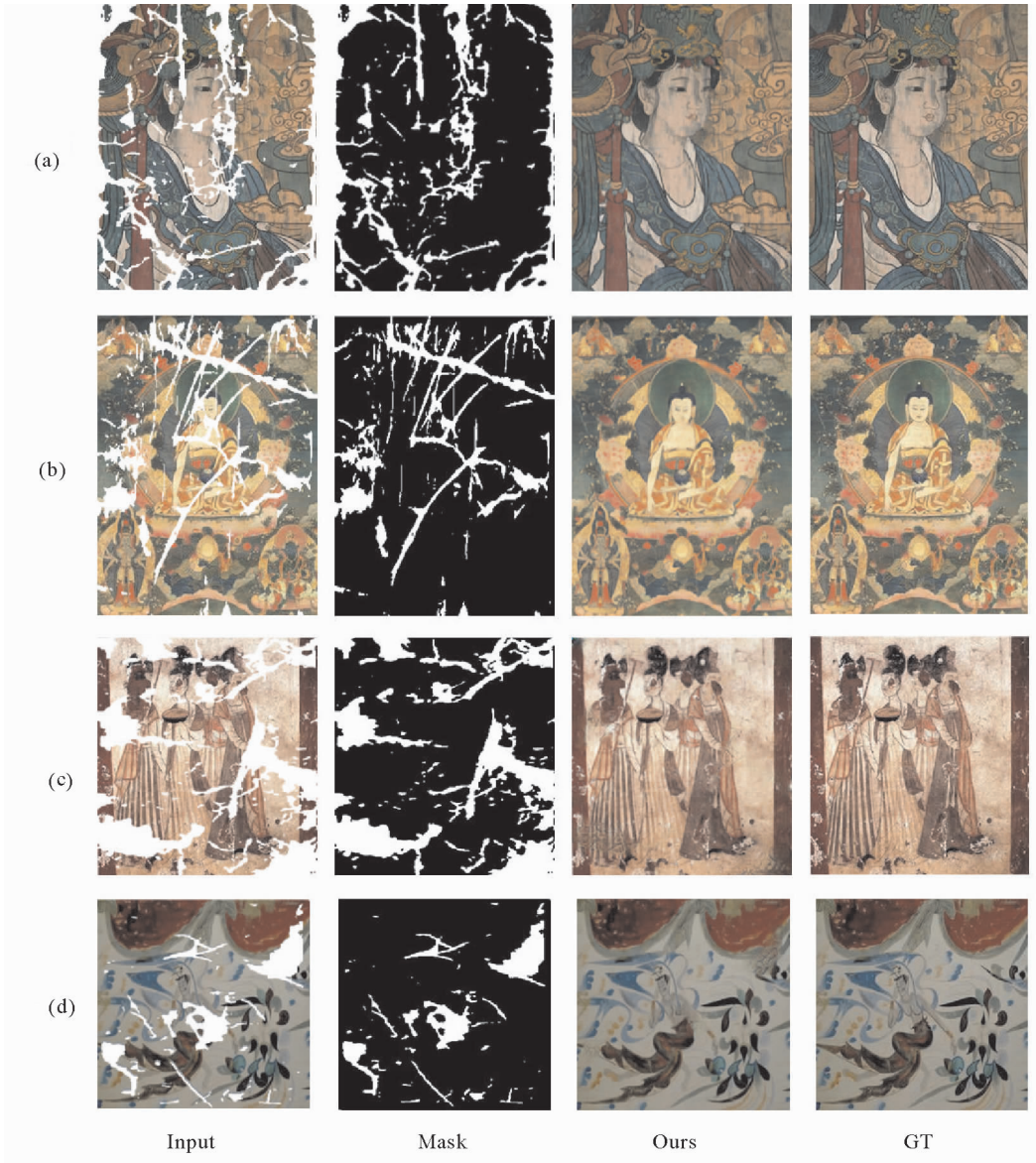


图 12 在壁画数据集集中的修复结果可视化

Fig. 12 Visualization of restoration results in mural datasets

3. 7. 2 Places 数据集测试

Places 数据集在图像修复任务中被广泛使用,其中包含丰富多样的环境,提供了丰富的视觉元素和结构变化。本研究在实验中选择部分有挑战的数据来检测本模型在面对未知环境的修复能力。结果如图 13 所示。图 13(a)、(b)中展示了

对街景的恢复,可以看出本方法恢复出了缺失的部分。在图 13(c)、(e)中展示了对复杂的建筑物外景的修复,从结果可以看出,TCSF 同样能保持原有结构的连贯性和视觉和谐。综上所述,在不同数据集上的修复结果有力地证明了本模型的泛化能力和出色的修复性能。

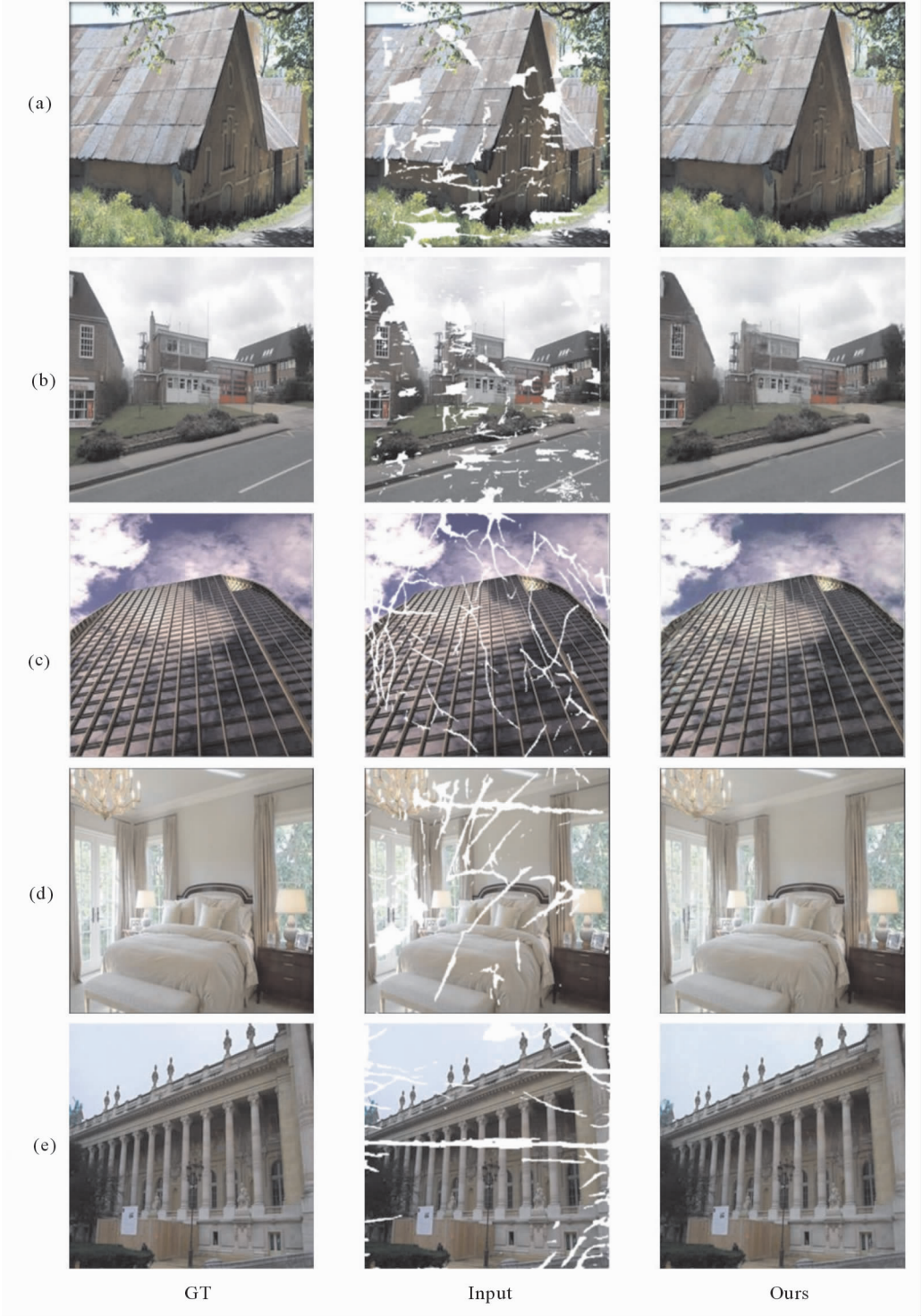


图 13 在 Places 数据集上的修复结果可视化
Fig. 13 Visualization of restore results in the Places datasets

4 结语

针对现有深度学习方法在具有复杂图像结构、表达方式抽象的中国传统画图像的修复中结果不理想的问题,提出了一种孪生级联空间滤波网络(TCSF),验证了空间预测滤波方法在特殊主题图像修复任务中的有效性。通过在滤波生成过程中使用空间编码机制捕捉传统画图像中像素间的空间关系和结构信息,实现了高保真的传统画图像细节填补和出色的视觉效果。同时,本方法修复的准确性以及指标方面与现有的方法相比均取得了显著的改进,展示了更好的中国传统图像修复方法。

后续将继续开展两个方向的研究:进一步优化滤波修复模型,提升修复质量和速度,通过调整网络结构、损失函数设计、参数调优等方式来提升对传统画笔触和风格的恢复性能;增加数据集的规模和多样性,包括不同风格、时代和主题的传统画图像,以提高模型的泛化能力和修复效果。

参考文献

- [1] 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. Seattle: IEEE, 2020: 7760-7768.
- [2] 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. Waikoloa: IEEE, 2022: 2149-2159.
- [3] WAN Z, ZHANG J, CHEN D, et al. High-fidelity pluralistic image completion with transformers[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021: 4692-4701.
- [4] LIAO L, XIAO J, WANG Z, et al. Guidance and evaluation: Semantic-aware image inpainting for mixed scenes[C]//Computer Vision-ECCV 2020: 16th European Conference. Cham: Springer, 2020: 683-700.
- [5] LI L, ZOU Q, ZHANG F, et al. Line drawing guided progressive inpainting of mural damages [EB/OL]. (2022-11-12) [2024-05-15]. <https://arxiv.org/abs/2211.06649>.
- [6] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 2223-2232.
- [7] 曹建芳,靳梦燕,李朝霞,等. 基于循环生成对抗网络的壁画色彩修复算法[J]. 山东科技大学学报(自然科学版), 2023, 42(4): 101-112.
CAO J F, JIN M Y, LI Z X, et al. Mural color restoration algorithm based on cyclic generative adversarial network[J]. Journal of Shandong University of Science and Technology (Natural Science), 2023, 42(4): 101-112.
- [8] CHI L, JIANG B, MU Y. Fast fourier convolution [J]. Advances in Neural Information Processing Systems, 2020, 33: 4479-4488.
- [9] YU J, LIN Z, YANG J, et al. Free-form image inpainting with gated convolution[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 4471-4480.
- [10] 胡升,薛涛,季虹. 多尺度信息融合的生成对抗网络壁画修复[J]. 国外电子测量技术, 2024, 43(4): 30-38.
HU S, XUE T, JI H. Mural inpainting algorithm for generative adversarial network with multi-scale information fusion [J]. Foreign Electronic Measurement Technology, 2024, 43(4): 30-38.
- [11] 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.
- [12] 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): 1-11.
- [13] CHO T S, BUTMAN M, AVIDAN S, et al. The patch transform and its applications to image editing[C]//2008 IEEE Conference on Computer Vision and Pattern Recognition. Anchorage: IEEE, 2008: 1-8.
- [14] KWATRA V, ESSA I, BOBICK A, et al. Texture optimization for example-based synthesis [J]. ACM SIGGRAPH 2005 Papers, 2005, 24(3): 795-802.
- [15] DARABI S, SHECHTMAN E, BARNES C, et al. Image melding: Combining inconsistent images using patch-based synthesis [J]. ACM Transactions on Graphics (TOG), 2012, 31(4): 1-10.
- [16] ZOME T. Learning how to inpaint from global image

- statistics[C] // Proceedings Ninth IEEE International Conference on Computer Vision. Nice; IEEE, 2003; 305-312.
- [17] 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.
- [18] 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. Las Vegas; IEEE, 2016; 2536-2544.
- [19] 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. New Orleans; IEEE, 2022; 10758-10768.
- [20] DONG Q, CAO C, FU Y. Incremental transformer structure enhanced image inpainting with masking positional encoding[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans; IEEE, 2022; 11358-11368.
- [21] KO K, KIM C S. Continuously masked transformer for image inpainting[C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris; IEEE, 2023; 13169-13178.
- [22] CHEN B W, LIU T J, LIU K H. Lightweight image inpainting by stripe window transformer with joint attention to CNN[C] // 2023 IEEE 33rd International Workshop on Machine Learning for Signal Processing (MLSP). Rome; IEEE, 2023; 1-6.
- [23] 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. Seoul; IEEE, 2019; 4170-4179.
- [24] YAN Z, LI X, LI M, et al. Shift-net: Image inpainting via deep feature rearrangement[C] // Proceedings of the European Conference on Computer Vision (EC-CV). Cham; Springer, 2018; 1-17.
- [25] WANG N, LI J, ZHANG L, et al. MUSICAL: Multi-scale image contextual attention learning for inpainting[C] // Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI). Macao; IJCAI, 2019; 3748-3754.
- [26] YU J, LIN Z, YANG J, et al. Generative image inpainting with contextual attention[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City; IEEE, 2018; 5505-5514.
- [27] ZHENG H, LIN Z, LU J, et al. Image inpainting with cascaded modulation gan and object-aware training[C] // European Conference on Computer Vision. Cham; Springer, 2022; 277-296.
- [28] 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. Waikoloa; IEEE, 2022; 2149-2159.
- [29] IIZUKA S, SIMO-SERRA E, ISHIKAWA H. Globally and locally consistent image completion[J]. ACM Transactions on Graphics (ToG), 2017, 36(4): 1-14.
- [30] GUO X, YANG H, HUANG D. Image inpainting via conditional texture and structure dual generation[C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal; IEEE, 2021; 14134-14143.
- [31] ZENG Y, LIN Z, LU H, et al. Cr-fill: Generative image inpainting with auxiliary contextual reconstruction[C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal; IEEE, 2021; 14164-14173.
- [32] SONG Y, YANG C, SHEN Y, et al. Spg-net: Segmentation prediction and guidance network for image inpainting[EB/OL]. (2018-08-06) [2024-05-15]. <https://arxiv.org/abs/1805.03356>.
- [33] ZHENG C, CHAM T J, CAI 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. New Orleans; IEEE, 2022; 11512-11522.
- [34] DENG Y, HUI S, ZHOU S, et al. T-former: An efficient transformer for image inpainting[C] // Proceedings of the 30th ACM International Conference on Multimedia. Lisboa; ACM, 2022; 6559-6568.
- [35] DENG Y, HUI S, MENG R, et al. Hourglass attention network for image inpainting[C] // European Conference on Computer Vision. Cham; Springer, 2022; 483-501.
- [36] MILDENHALL B, BARRON J T, CHEN J, et al. Burst denoising with kernel prediction networks[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City; IEEE, 2018; 2502-2510.
- [37] XIE C, TIAN X, JIANG R, et al. Dilated kernel prediction network for single-image denoising[J]. Journal

- of Electronic Imaging, 2021, 30(2): 1-15.
- [38] CARBAJAL G, VITORIA P, LEZAMA J, et al. Blind motion deblurring with pixel-wise kernel estimation via kernel prediction networks[J]. IEEE Transactions on Computational Imaging, 2023, 9: 928-943.
- [39] GUO Q, SUN J, JUEFEI-XU F, et al. Efficientderain: Learning pixel-wise dilation filtering for high-efficiency single-image deraining[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2021: 1487-1495.
- [40] GUO Q, LI X, XU F, et al. Jpgnet: Joint predictive filtering and generative network for image inpainting[C]//Proceedings of the 29th ACM International Conference on Multimedia. New York: ACM, 2021: 386-394.
- [41] LI X, GUO Q, LIN D, et al. Misf: Multi-level interactive siamese filtering for high-fidelity image inpainting[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 1869-1878.
- [42] LIU R, LEHMAN J, MOLINO P, et al. An intriguing failing of convolutional neural networks and the coord-conv solution[J]. Advances in Neural Information Processing Systems, 2018, 31: 9605-9616.
- [43] JOHNSON J, ALAHI A, FEI-FEI L. Perceptual losses for real-time style transfer and super-resolution[C]//Computer Vision-ECCV 2016: 14th European Conference. Cham: Springer, 2016: 694-711.
- [44] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[J]. Advances in Neural Information Processing Systems, 2014, 27: 2672-2680.
- [45] GATYS L A, ECKER A S, BETHGE M. Image style transfer using convolutional neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 2414-2423.
- [46] ZHOU B, LAPEDRIZA A, KHOSLA A, et al. Places: A 10 million image database for scene recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(6): 1452-1464.
- [47] SIMONYAN K. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-10) [2024-05-15]. <https://arxiv.org/abs/1409.1556>.
- [48] YI Z, TANG Q, AZIZI S, et al. Contextual residual aggregation for ultra high-resolution image inpainting[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 7508-7517.
- [49] 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). Cham: Springer, 2018: 85-100.
- [50] ZHANG R, ISOLA P, EFROS A A, et al. The unreasonable effectiveness of deep features as a perceptual metric[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 586-595.
- [51] PASZKE A, GROSS S, MASSA F, et al. Pytorch: An imperative style, high-performance deep learning library[J]. Advances in Neural Information Processing Systems, 2019, 32: 8026-8037.
- [52] LOSHCHILOV I. Decoupled weight decay regularization[EB/OL]. (2019-01-04) [2024-05-15]. <https://arxiv.org/abs/1711.05101>.
- [53] ZENG Y, FU J, CHAO H, et al. Aggregated contextual transformations for high-resolution image inpainting[J]. IEEE Transactions on Visualization and Computer Graphics, 2022, 29(7): 3266-3280.
- [54] LIU W, CUN X, PUN C M, et al. Coordfill: Efficient high-resolution image inpainting via parameterized coordinate querying[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Washington: AAAI Press, 2023, 37(2): 1746-1754.

(编辑 雷雁林)