

基于多尺度注意力机制的两阶段文物图像修复方法

刘浩威, 姚镜池, 刘波, 毕秀丽, 肖斌

引用本文

刘浩威, 姚镜池, 刘波, 毕秀丽, 肖斌. 基于多尺度注意力机制的两阶段文物图像修复方法[J]. 计算机科学, 2023, 50(6A): 220600129-8.

LIU Haowei, YAO Jingchi, LIU Bo, BI Xiuli, XIAO Bin. [Two-stage Method for Restoration of Heritage Images Based on Muti-scale Attention Mechanism](#) [J]. Computer Science, 2023, 50(6A): 220600129-8.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于多特征融合的GRU-LSTM大学生就业动态预测](#)

College Students Employment Dynamic Prediction of Multi-feature Fusion Based on GRU-LSTM
计算机科学, 2023, 50(6A): 220500056-6. <https://doi.org/10.11896/jsjcx.220500056>

[基于深度学习的超高频标签识别系统](#)

Tag Identification for UHF RFID Systems Based on Deep Learning

计算机科学, 2023, 50(6A): 220200151-6. <https://doi.org/10.11896/jsjcx.220200151>

[CT影像阶段化目标检测方法研究](#)

Study on Phased Target Detection in CT Image

计算机科学, 2023, 50(6A): 220200063-10. <https://doi.org/10.11896/jsjcx.220200063>

[基于深度学习的摩托车车道实时检测](#)

Real-time Detection of Motorcycle Lanes Based on Deep Learning

计算机科学, 2023, 50(6A): 220200066-5. <https://doi.org/10.11896/jsjcx.220200066>

[基于改进YOLOv5的电动车头盔佩戴检测算法](#)

Electric Bike Helment Wearing Detection Alogrithm Based on Improved YOLOv5

计算机科学, 2023, 50(6A): 220500005-6. <https://doi.org/10.11896/jsjcx.220500005>

基于多尺度注意力机制的两阶段文物图像修复方法

刘浩威 姚镜池 刘波 毕秀丽 肖斌

图像认知重庆市重点实验室 重庆 400065

(liu_haow@163.com)

摘要 文物常因保存或物理修复手段不当而受到损坏,使用虚拟技术对其进行修复很重要,而现有传统图像修复技术和基于深度学习的修复方法主要针对结构纹理简单、破损区域较小的图像或是破损区域规则的自然图像,无法直接应用于文物图像。针对文物图像结构纹理复杂、破损区域不规则及现存文物图像数据集较小等问题,以山水画图像修复为例提出了一种基于多尺度注意力机制的两阶段文物图像修复方法。首先基于全局注意力机制对文物图像的整体结构和基础色调进行粗粒度修复,然后使用局部注意力机制和残差模块对文物图像的小型结构和细节纹理进行局部细粒度修复,并在粗粒度修复的结果上使用上下文注意力机制从文物图像远距离精确借用信息,对图像的大型结构和纹理进行全局细粒度修复,最后将局部和全局的修复结果进行特征融合,实现文物图像的修复。针对文物图像特殊的破损类型,修复的文物图像伪迹较少,色彩均匀,结构纹理清晰,相比对比方法,在峰值信噪比上平均提高了 3.76 dB,在结构相似性上平均提高了 0.034。实验结果的主观和客观分析表明,与主流图像修复方法相比,在语义合理性、信息准确性和视觉自然性上都具有一定优势,在文物修复领域有较大应用价值。

关键词: 文物图像;图像修复;深度学习;两阶段模型;多尺度注意力机制

中图分类号 TP391

Two-stage Method for Restoration of Heritage Images Based on Multi-scale Attention Mechanism

LIU Haowei, YAO Jingchi, LIU Bo, BI Xiuli and XIAO Bin

Chongqing Key Laboratory of Image Cognition, Chongqing 400065, China

Abstract The use of virtual technology is important for the restoration of relics, which are often damaged by improper preservation or physical restoration methods. Existing traditional image restoration techniques and deep learning-based restoration methods are mainly suitable for images with simple structural textures, small damaged areas, or natural images with regular damage, and cannot be directly applied to heritage images. Using landscape painting image restoration as an example, a two-stage method for restoration of heritage images based on a multi-scale attention mechanism is proposed in this paper to address the problems of complex structural textures, discreet colouring and small size of existing datasets of heritage images. The method firstly performs coarse restoration of the overall structure and base tones of the image based on the global attention mechanism, then performs local fine restoration of small structures and fine textures of the image using the local attention mechanism and the residual module, as well as global fine restoration of large structures and textures using the contextual attention mechanism on the result of coarse restoration to borrow information accurately at a distance. Finally, the local and global fine restoration results are fused to achieve the restoration of heritage images. The proposed method has the advantage of improving the peak signal-to-noise ratio by 3.76 dB and the structural similarity by 0.034 compared with the comparative methods on average. Both the subjective and objective analysis of the experimental results show that the method has some advantages in semantic rationality, information accuracy and visual naturalness compared with the existing methods, and has a high potential for application in the field of heritage restoration.

Keywords Heritage image, Image restoration, Deep learning, Two stage model, Multi-scale attention mechanism

1 引言

文物是人类社会发展变迁过程的实物遗存,具有重要的历史文化价值。然而,在过去的很长一段时间里,因为技术有限,很多文物不仅无法得到有效的保存和修复,有时甚至因为工作人员操作不当,受到二次毁坏。因此,利用数字图像修复技术对其进行虚拟修复一直是国内外学者比较关注的问题^[1]。

目前主要的文物图像修复都是基于传统的图像修复技术,如 Bertalmio 等^[2]基于有界变差函数提出了偏微分方程的方法;Barnes 等^[3]、Criminisi 等^[4]、Huang 等^[5]等基于图像自相似性提出了样本块匹配方法;Starck 等^[6]、Mairal 等^[7]基于图像的冗余信息提出基于稀疏表达的方法。总体而言,这类方法在破损面积不大、结构简单、具有重复纹理的图像上可以取得很好的效果。但是,因其主要依据单张图像局部特征,不仅缺乏内容表达能力,且修复结果语义的合理性不足。

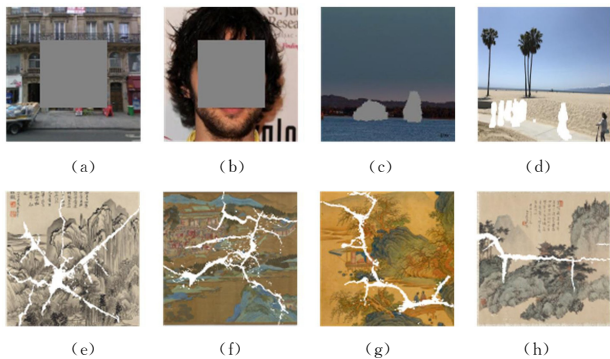
基金项目:国家级大学生创新创业训练计划支持项目(202110617006)

This work was supported by the National-level Student Innovation and Entrepreneurship Training Program(202110617006).

通信作者:刘波(boliu@cqupt.edu.cn)

深度学习通过学习同类图像的多种特征,具备较强的生成泛化能力,可在语义层面修复图像整体内容。目前文物图像修复领域暂时还没有针对性的深度学习方法,但是很多基于深度学习的自然图像修复方法也会测试其应用于文物图像修复的效果。这些方法可以分为一阶段修复方法^[8-10]、两阶段修复方法^[11-19]以及多阶段修复方法^[20-24]3类。一阶段修复方法的网络框架大多采用 U-Net^[25]结构,或基于一对编、解码器^[10,26-27],以端到端的方式实现图像修复,这类方法能较好地复原图像的基本语义信息,但部分修复结果纹理细节模糊,存在较多伪迹;两阶段修复方法在一阶段方法的基础上对其结果进行进一步处理,得益于第一阶段填充的大量信息,两阶段方法的修复结果比较稳定,结构纹理连贯,具有相对一阶段方法更好的视觉效果。受传统方法中样本块匹配思想的启发,两阶段方法引入注意力机制^[11,14-15],基于一阶段修复结果,计算破损区域内外的相似度,用已知区域特征块重建缺失区域内容,使图像修复的效果达到了一定的高度;而多阶段方法则是延续两阶段方法的思想,采用尽可能多的生成器^[21,24],对信息进行由粗到细的处理,但此类方法在显著增加计算量的同时对图像修复的效果提升并不明显。

从自然图像修复领域迁移过来的方法不能很好地实现对文物图像的修复,主要原因如下。首先,自然图像和文物图像的破损类型和整体内容存在一定差异。如图 1(a)、图 1(b)所示,自然图像的破损大多是规则的,且这些图像的纹理结构大多比较简单,或具有固定结构,而文物图像的破损多是不规则的。虽然已有方法针对不规则破损图像补全任务进行研究并取得了不错成果,但其主要应用于前景物体及水印移除等任务,如图 1(c)、图 1(d)所示,这些破损多数面积较小,且集中在局部区域,并没有涉及到较多的结构和纹理。而如图 1(e)一图 1(h)所示,需要修复的文物图像具有纹理结构复杂、破损区域不规则且大面积贯穿全图的特点,导致自然图像修复领域的方法对其修复效果欠佳,无法直接应用于文物修复。其次,现有文物图像数据集规模较小,而大多数基于深度学习的修复方法严重依赖于大型数据集以拟合图像特征分布,导致这些方法在小规模文物数据集上无法学习到足够的生成泛化能力,修复结果存在明显的模糊和破损痕迹。最后,文物修复工作对保真度要求较高,现有的自然图像修复方法优先考虑图像的主观视觉自然性,而非客观真实性,对文物修复任务并不适用。



注:图 1(a)一图 1(d)是自然图像,图 1(e)一图 1(h)是文物图像

图 1 自然图像破损与文物图像破损对比

Fig. 1 Comparison between natural image damage and heritage image damage

针对上述问题及文物图像的特点,本文提出了一种针对文物图像修复的深度学习方法。该方法使用基于多尺度注意力机制的两阶段模型对图像进行由粗到细的修复。实验结果表明,对于小面积不规则破损,本文方法的修复结果主观视觉效果好,结构纹理清晰,颜色均匀;对于大面积破损,本文方法能更好修复图像的轮廓,语义更加连贯合理,在文物修复领域具有较高的应用价值。

本文第 2 节将介绍所提出的网络框架和具体模型,讨论如何选定网络的基本架构,并对粗粒度网络和细粒度网络进行具体介绍;第 3 节设计了 4 组实验,以主观视觉评定和客观指标测试两种方式,从消融实验和方法对不同类型破损修复的有效性两个角度来验证本文方法的性能;最后总结全文。

2 基于多尺度注意力机制的两阶段文物图像修复方法

目前自然图像修复领域内的主流方法,大多都是基于生成对抗网络^[28](Generative Adversarial Network, GAN),通过生成器与判别器的博弈学习,使得生成器能够通过已知区域生成破损区域的内容,从而实现图像修复^[10,14-16,29]。但前期实验结果表明,在小规模数据集下,基于 GAN 机制的方法修复效果并不理想,如图 2 所示,本文在纹理识别数据集^[30](Describable Textures Dataset, DTD)上基于 Yu 等^[14]提出的网络架构,分别在含有对抗损失的 GAN 结构和使用多种其他损失的联合损失替代对抗损失的卷积神经网络^[31](Convolutional Neural Networks, CNN)结构下对网络进行训练和测试。实验结果表明,在小规模数据集下训练测试时,基于 GAN 机制的网络结构不仅训练不稳定,且修复结果存在过多不合理的填充和视觉伪影,如图 2(b)第二行填补内容的蓝色伪影,相比之下,基于 CNN 结构的修复结果虽然细节信息更少,存在一定的模糊现象,但没有产生明显的修补痕迹,相比基于 GAN 机制的网络结构的修复结果整体更加符合语义,注重填补内容的真实性,符合文物图像修复任务的要求。

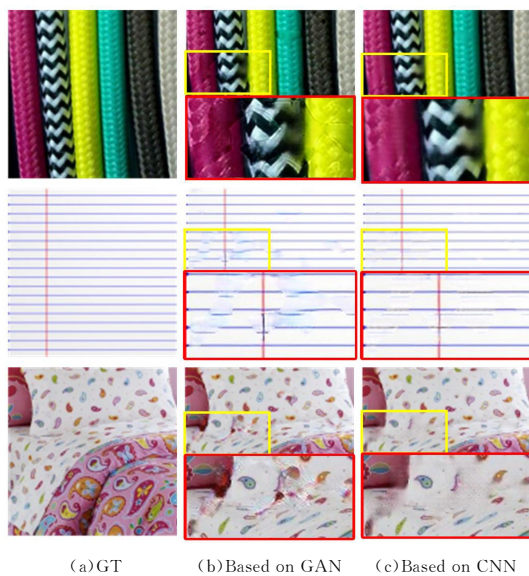


图 2 基于 GAN 或 CNN 结构修复效果的对比

Fig. 2 Comparison of restoration results based on GAN and CNN

特征提取的过程对修复至关重要,但大多方法采用的普通卷积对图像没有区分破损区域与未破损区域,导致网络将

破损区域也作为图像特征进行提取,这可能导致修复结果产生更多的伪迹。针对这个问题,Liu等^[9]提出了部分卷积,将破损图像所有像素分为有效像素和无效像素,在进行特征提取时过滤无效像素的影响,但随着网络的深入,图像的无效像素将全部转变为有效像素,这意味着掩码信息将在网络深层消失,无法有针对性地对缺失区域进行修复。因此,Yu等^[14]在其基础上提出的门控卷积加入了动态特征选择机制,用可学习的软门控替代部分卷积中的掩码更新规则,有效提高了针对不规则掩码的图像修复质量,更加适合修复文物图像的不规则破损。

除此之外,我们认为当前的深度学习方法都过度依赖于大型的数据,而少有研究工作关注如何根据实时任务对已有数据进行更深入和有针对性的开发和利用。如在文物图像修复任务上,本文认为应该回归到文物图像的创作过程,首先对其进行先整体后细节的基础修复,再对其进行语义层面的约束并逐步完成对局部结构纹理信息的加强和修正。

综上所述,本文方法以CNN为基本结构,避免GAN网络

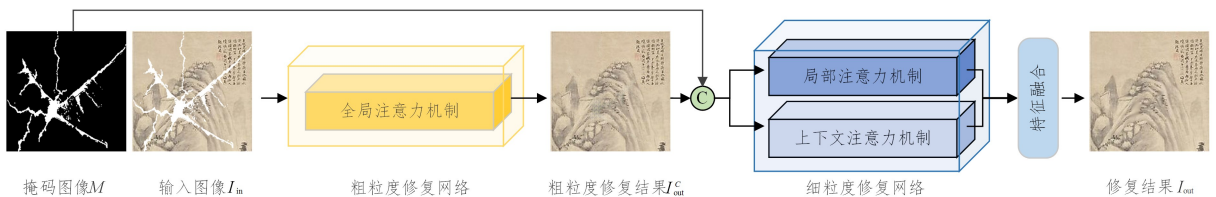


图3 基于多尺度注意力机制的两阶段文物图像修复方法流程图

Fig. 3 Flowchart of two-stage method for restoration of heritage images based on multi-scale attention mechanism

2.1 粗粒度修复网络

粗粒度修复网络的主要作用为修复破损图像的整体结构和纹理,其输入为破损的文物图像及对应的掩码,输出为经粗粒度网络初步修复的图像。粗粒度修复网络采用带有跳跃连接的编解码结构,基于全局注意力机制对文物图像的整体结构纹理及基础颜色进行修复。

如图4所示,粗粒度网络使用8层门控卷积作为编码器,逐步将输入的分辨率为 256×256 的图像下采样为 1×1 的特征图。在此过程中,编码器通过连续的下采样卷积操作提取输入图像的逐级特征映射,在提取不同尺度特征的同时,依据破损区域周围的信息对其进行初步填充。从卷积层的浅层到深层,网络将逐步提取到图像的更高级特征,而全局注意力的实质,指网络的深层的神经元学习到了图像的全部信息,且位于不同空间位置的神经元对同一特征的表达能力不同。

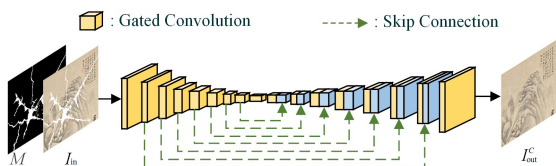


图4 粗粒度修复网络的结构

Fig. 4 Structure of coarse restoration network

解码器的网络层数与编码器保持一致,将 1×1 的特征上采样解码恢复至原始大小,重建出完整文物图像。此外,在对应尺度特征编码器与解码器之间加入跳跃连接以传递细节信息,减少下采样过程信息的丢失。

粗粒度网络的结构如表1所列,其中形如 $(64, 64, 48)$ 表示输出大小为 64×64 、具有48个通道的特征图,gconv $(4, 2, 48)$

训练过程中的不稳定性,采用门控卷积作为网络中卷积层对图像进行更有效的特征提取,以从粗到细的两阶段修复方式对文物图像进行修复。与Yu等^[14]提出的方法不同,本文方法进一步加深了粗网络的深度以提取多尺度的特征信息,并在细修复网络中使用全局-局部两个分支对补全内容进行细化。方法的具体流程如图3所示,其中掩码M是根据文物破损规律制作的掩码,方法包括粗粒度修复和细粒度修复两部分,其中细粒度修复网络由两个子分支组成,具体过程如下:

(1)利用粗粒度修复网络,修复文物图像的整体纹理结构及基础的颜色信息。

(2)使用细粒度全局修复分支在粗粒度修复结果的基础上修复大型结构纹理;使用细粒度局部修复分支修复文物图像的小型结构和纹理细节,并改善粗粒度修复结果的伪影模糊现象。

(3)将细粒度全局和局部修复结果在特征层次进行融合,再将其转换为RGB图像输出,实现对文物图像由粗到细的两阶段修复。

表示数量为48、大小为 4×4 的卷积核,进行步长为2的门控卷积操作,gdconv(192)表示数量为192、大小为 4×4 ,进行步长为1/2的上卷积操作,elu和tanh是两种激活函数。

表1 粗粒度修复网络明细

Table 1 Details of coarse restoration network

Input	Operation	Output
(256, 256, 3)	gconv(4, 2, 48), elu	(128, 128, 48)
(128, 128, 48)	gconv(4, 2, 48), elu	(64, 64, 48)
(64, 64, 48)	gconv(4, 2, 96), elu	(32, 32, 96)
(32, 32, 96)	gconv(4, 2, 192), elu	(16, 16, 192)
(16, 16, 192)	gconv(4, 2, 192), elu	(8, 8, 192)
(8, 8, 192)	gconv(4, 2, 192), elu	(4, 4, 192)
(4, 4, 192)	gconv(4, 2, 192), elu	(2, 2, 192)
(2, 2, 192)	gconv(4, 2, 192), elu	(1, 1, 192)
(1, 1, 192)	gdconv(192), elu	(2, 2, 192)
(2, 2, 384)	gdconv(192), elu	(4, 4, 192)
(4, 4, 384)	gdconv(192), elu	(8, 8, 192)
(8, 8, 384)	gdconv(192), elu	(16, 16, 192)
(32, 32, 384)	gdconv(192), elu	(64, 64, 192)
(64, 64, 384)	gdconv(64), elu	(128, 128, 48)
(128, 128, 92)	gdconv(3), tanh	(256, 256, 3)

2.2 细粒度修复网络

细粒度修复网络用于对粗粒度的输出结果进行进一步细化,减弱或消除明显的伪迹。细粒度修复网络采用两个平行修复分支,分别基于不同的注意力机制对粗粒度修复结果进行调整和细化,并在网络的最后进行特征融合,将两个分支的输出特征图通过卷积转化为RGB图像,实现对文物图像整体结构和局部细节的修复和完善。

细粒度全局修复分支利用上下文注意力机制从远距离精确借用信息来进一步构建大型结构纹理。该分支采用如图4

所示的粗粒度修复网络的结构,并分别在 $16 \times 16, 32 \times 32, 64 \times 64$ 这 3 个尺度的特征图上采样前加入对应尺度的上下文注意力层^[11] (Contextual Attention, CA), 利用粗粒度修复的结果,在不同尺度特征对缺失区域进行重建,保证修复内容在全局上的语义合理性。

细粒度局部修复分支基于局部注意力机制去除由远距离错误信息带来的伪迹,并对模糊细节进行补充。如图 5 所示,为减少下采样编码过程中的细节信息损失,局部分支采用较浅的网络结构,仅对图像进行两次下采样,以提取文物图像的浅层细节信息。此外,为更好地提取图像特征,本文在编码器和解码器间加入 4 个残差模块,每一模块使用两个级联卷积层对输入特征图进行特征提取,逐步扩大单个神经元的感受野,该分支的网络明细如表 2 所列。

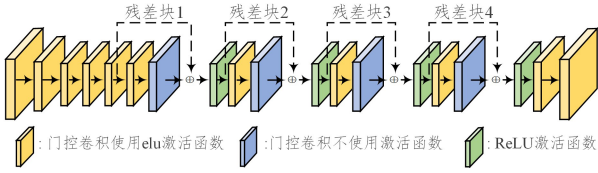


图 5 细粒度局部分支网络结构图

Fig. 5 Structure of fine restoration network

表 2 细粒度局部分支网络明细

Table 2 Details of local branch of fine restoration network

Input	Operation	Output
(256, 256, 3)	gconv(7, 1, 48), elu	(256, 256, 48)
(256, 256, 48)	gconv(3, 2, 96), elu	(128, 128, 96)
(128, 128, 96)	gconv(3, 2, 192), elu	(64, 64, 192)
(64, 64, 192)	Residual Block	(64, 64, 192)
(64, 64, 192)	Residual Block	(64, 64, 192)
(64, 64, 192)	Residual Block	(64, 64, 192)
(64, 64, 192)	Residual Block	(64, 64, 192)
(64, 64, 192)	gdconv(96)	(128, 128, 96)
(128, 128, 96)	gdconv(48)	(256, 256, 48)

2.3 损失函数

结合图像的低层像素信息和高层语义信息,本文采用的联合损失函数如式(1)所示:

$$L_{\text{total}} = \lambda_1 L_{\text{rec}} + \lambda_2 L_{\text{tv}} + \lambda_3 L_{\text{perc}} + \lambda_4 L_{\text{style}} \quad (1)$$

式(1)等式右边 4 项分别为重建损失、TV 损失、感知损失和风格损失及其加权系数。

在训练细节上,对于粗粒度网络,仅使用重建损失对修复结果进行约束,而对于细粒度网络,使用 4 个损失的加权和修复结果进行约束。

2.3.1 重建损失

对于低层像素信息,采用重建损失 L_{rec} 如式(2)所示:

$$L_{\text{rec}} = \lambda_v \cdot L_{\text{valid}} + \lambda_h \cdot L_{\text{hole}} \quad (2)$$

其中, L_{valid} 和 L_{hole} 分别为已知区域和未知区域的 L_1 损失,定义如式(3)所示:

$$\begin{cases} L_{\text{valid}} = \|(1-M) \odot (I_{\text{out}} - I_{\text{gt}})\|_1 \\ L_{\text{hole}} = \|M \odot (I_{\text{out}} - I_{\text{gt}})\|_1 \end{cases} \quad (3)$$

其中, \odot 为逐像素点乘, I_{gt} 是原图。

2.3.2 感知损失

对于高层特征,添加感知损失 L_{per} 与风格损失 L_{style} 。感知损失是原图与修复图像相似特征之间的 L_1 距离,其定义如式(4)所示:

$$L_{\text{per}} = \sum_i (\|\psi_i(I_{\text{out}}) - \psi_i(I_{\text{gt}})\|_1 + \|\psi_i(I_{\text{comp}}) - \psi_i(I_{\text{gt}})\|_1) \quad (4)$$

其中, ψ_i 是 VGG16 网络第 i 层输出的特征图, $i \in (5, 10, 17)$, I_{comp} 是原图修复融合后的图像,其定义如式(5)所示:

$$I_{\text{comp}} = I_{\text{in}} \odot (1-M) + I_{\text{out}} \odot M \quad (5)$$

2.3.3 风格损失

风格损失是原图和修复图像对应的格拉姆(Gram)矩阵间的 L_1 距离,如式(6)所示:

$$L_{\text{style}} = \sum_i \|G_i(I_{\text{out}}) - G_i(I_{\text{gt}})\|_1 + \sum_i \|G_i(I_{\text{comp}}) - G_i(I_{\text{gt}})\|_1 \quad (6)$$

其中, Gram 矩阵 $G_i(\cdot) = \psi_i(\cdot) \psi_i(\cdot)^T$ 。

2.3.4 TV 损失

为促进生成图像的空间平滑性,减小噪声的干扰,添加 TV 损失的定义如式(7)所示:

$$L_{\text{tv}} = \sum_{(i,j) \in P, (i,j+1) \in P} \|I_{\text{comp}}^{i,j+1} - I_{\text{comp}}^{i,j}\|_1 + \sum_{(i,j) \in P, (i+1,j) \in P} \|I_{\text{comp}}^{i+1,j} - I_{\text{comp}}^{i,j}\|_1 \quad (7)$$

其中, P 包含缺失部分的所有像素。

3 实验结果与分析

本节设计消融实验来验证所提方法中多尺度注意力机制的有效性,并利用本文方法和 Barnes^[3] 方法、Yan^[8] 方法、Liu^[9] 方法、Yu^[14] 方法、Zeng^[19] 法针对不同类型破损的修复结果进行定性和定量的对比分析,以展现本文提出的网络结构在文物图像小规模数据集上修复质量的优越性。对比实验中其他的修复方法的参数都设置为复现算法最佳实验结果时的参数值。

3.1 实验平台及参数

本文所有的实验硬件环境为 Intel(R) Xeon(R) Gold 5218 CPU 和单个 GeForce RTX 2080 Ti(11G)。实验平台软件环境为 TensorFlow 1.1.5, CUDA 10.0。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} ,一阶动量为 0.5,二阶动量为 0.999。

3.2 实验数据库及对比方法

在上述硬件平台以及参数设置下,分别在纹理识别数据集 DTD 与中国山水画数据集^[32] TCLPD(Traditional Chinese Landscape Painting Dataset)上,将各数据集下的图像以 4 张为一批次送入网络。训练和测试所用的图像及掩码大小均为 256×256 。训练集及测试集的划分如表 3 所列。

表 3 数据集划分

Table 3 Division of dataset

Dataset	Trainset	Testset	Total
DTD	5580	220	5800
TCLPD	1800	100	1900

本文采用以下 5 种主流图像修复方法与本文方法进行对比。

(1) Barnes 等^[3]提出的方法:一种典型的基于块匹配的方法,从图像已知区域搜索、匹配相似块,并将其复制到缺损区域。

(2) Yan 等^[8]提出的方法:一种基于 U-Net 的一阶段修复方法,引入移位连接(shift-connection),将编码层已知特征信息移入对应的编码层,引导解码空间中缺失区域内部特征的修复。

(3)Liu 等^[9]提出的方法:一种使用有效像素作为输入的方法,其掩码更新机制实现具有不规则掩码的图像的修复。

(4)Yu 等^[14]提出的方法:一种分为粗细两阶段训练的模型方法,采用两个编码器串联的形式,在部分卷积的基础上提出软门控机制并在细粒度网络中使用上下文注意力机制。

(5)Zeng 等^[19]提出的方法:一种使用可学习的损失函数替代注意力机制的两阶段模型,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借用特征进行修复的能力。

3.3 评价指标

对于破损图像修复,目前尚无领域公认的客观评价指标。本文选取最常用的峰值信噪比^[33](Peak Signal to Noise Ratio, PSNR)和结构相似性^[34](Structural Similarity Index, SSIM)作为本文方法及对比方法修复结果的客观评价指标,

数值越高说明修复质量最好,两指标的对应表达式如下:

$$PSNR = 20 \cdot \log_{10} \left(\frac{MAX_I}{\sqrt{MSE}} \right) \quad (8)$$

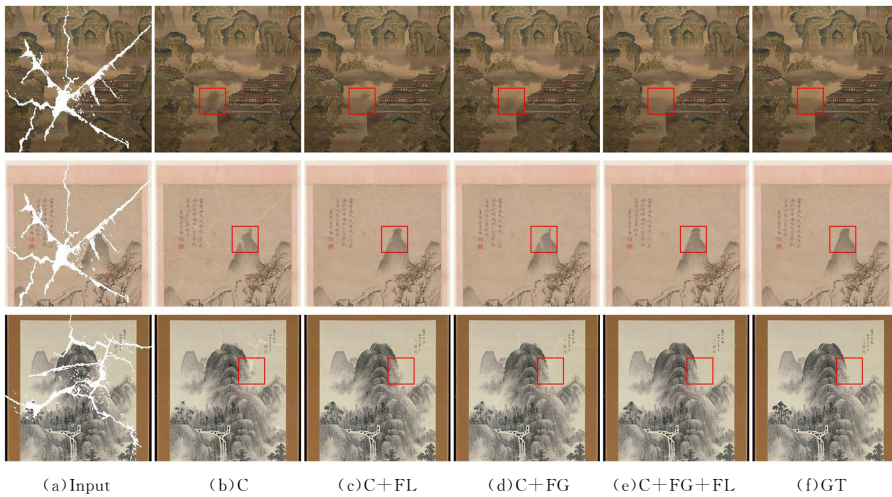
$$SSIM(y_i, \hat{y}_i) = \frac{(2\mu_{y_i}\mu_{\hat{y}_i} + C_1)(\sigma_{y_i\hat{y}_i} + C_2)}{(\mu_{y_i}^2 + \mu_{\hat{y}_i}^2 + C_1)(\sigma_{y_i}^2 + \sigma_{\hat{y}_i}^2 + C_2)} \quad (9)$$

$$MSE = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W (I_{out}^{i,j} - I_{comp}^{i,j})^2 \quad (10)$$

其中, MSE 是均方误差; MAX_I 是输入图像的最大像素值,其值为255; μ 表示图像的像素均值; σ_y 和 $\sigma_{y_i\hat{y}_i}$ 分别表示方差和协方差; C_1 和 C_2 为常数; H 和 W 分别是图像的宽和高。

3.4 消融实验

为验证多尺度注意力机制的有效性,本节设计消融实验分别验证粗粒度网络、细粒度网络局部分支和全局分支,即全局注意力机制、局部注意力机制和上下文注意力机制在修复中的作用。



注:C表示粗粒度修复网络,FL和FG分别表示细粒度修复网络的局部和全局分支

图6 多尺度注意力机制消融实验结果的对比图

Fig. 6 Comparison of results of multi-scale attention mechanism ablation experiments

如图6(b)所示,粗粒度网络对图像实现了基本的修复,而如图6(c)、图6(d)所示,网络增加细粒度网络的分支后,图像的整体轮廓改变较小,粗粒度修复结果的大部分内容被保留,说明粗粒度修复阶段在整个图像修复过程中影响较大,该阶段的错误填补内容会严重影响后续的修复过程。

此外,细粒度局部分支和全局分支的作用显然是不同的。如图6(c)所示,加入局部分支后,修复结果的颜色一致性更好,破损区域的伪迹消失,且局部结构和细节得到合理的完善,但由于其可利用的信息范围过小,无法结合远距离的信息判断填补结果的语义合理性,导致修复内容仍与原始图像有一定差距。如图6(d)所示,仅有粗粒度网络和全局分支的情况下,缺乏对局部特征信息的使用和约束,文物图像的局部缺失区域无法得到很好的修复,与已知区域存在明显色差,但

由于注意力机制从已知区域借用特征信息对缺失区域进行修复,图像整体的结构和纹理变得更加柔和清晰。

如图6(e)所示,同时具备3种注意力机制的网络,对缺失区域局部特征信息进行充分的学习和利用,并通过上下文注意力机制从图像的远距离空间获取信息,对图像进行语义层面的约束和修复,整体修复结果语义平滑,视觉连贯,仅通过肉眼难以分辨原图与修复后的图像。

如表4所列,粗粒度网络的修复结果在PSNR和SSIM上都取得了较好的数值,并在加入局部分支后分别在PSNR和SSIM上取得了次优和最优的结果,展现出了方法优越的性能,由表5最后一行可知,在细粒度网络同时保留全局分支和局部分支时PSNR提高了0.148~0.224,而在SSIM上仅与只有局部分支时相差0.002,使修复结果达到最高水平。

表4 消融实验修复结果的客观评价

Table 4 Objective evaluation value of ablation experiment

Index Mask	PSNR [↑]						SSIM [↑]					
	M _s 1	M _s 2	M _s 3	M _s 4	M _s 5	Avg	M _s 1	M _s 2	M _s 3	M _s 4	M _s 5	Avg
C	35.22	35.69	38.82	36.30	36.35	37.08	0.907	0.898	0.941	0.927	0.959	0.926
C+FL	37.59	38.64	41.75	38.85	41.45	39.66	0.942	0.943	0.969	0.954	0.971	0.956
C+FG	36.78	37.70	40.74	37.90	40.75	38.78	0.929	0.938	0.958	0.942	0.967	0.945
C+FL+FG	37.74	38.79	41.52	39.05	41.67	39.75	0.938	0.943	0.964	0.952	0.971	0.954

3.5 对比实验

本节进行实验,从主观定性评价和客观定量评估两个方面对比各方法在 TCLPD 数据集上针对 3 种不同类型破损的修复结果。

3.5.1 小面积不规则破损修复

本节将测试各方法对于小面积不规则破损的修复效果,测试采用根据文物破损规律自行制作的 5 张掩码,主要特征是破损面积较小,形状不规则,且贯穿全图。图 7 给出了各方法的修复结果。如图 7(a)所示,Barnes 等^[3]提出的方法整体修复质量较高,但在已知区域不存在与缺失区域相匹配的内容时修复效果则下降明显。而在选取的深度学习对比方法中,如图 7(c)、图 7(d)、图 7(f)所示,Yan 等^[8]提出的方法、Liu 等^[9]提出的方法、Zeng 等^[19]提出的方法的修复结果填补内容语义不连贯,存在明显的

伪迹。如图 7(e)、图 7(g)所示,Yu 等^[14]提出的方法与本文方法的修复效果较好。上述结果说明上下文注意力机制的加入可以提高修复结果的语义合理性,但由于 Yu 等^[14]提出的方法未结合局部信息对图像进行细化,部分区域仍存在模糊和扭曲现象。由图 7(g)可知,本文方法修复结果语义连贯,伪迹较少,具有较好的视觉效果,相比 Barnes 等的^[3]方法,能修复已知区域未出现过的内容,与其他深度学习方法相比,对文物图像的结构和纹理修复效果更好,并在语义合理性上有一定优势。

各方法修复结果客观评价指标如表 5 所列,对于小面积不规则的破损,本文方法在 PSNR 上表现最优,相比其他 5 种方法在 PSNR 上提高了 0.582~8.68 dB;在 SSIM 上取得了次优的结果,相比在此类破损表现最好的 Barnes 等^[3]提出的方法仅相差 0.001。

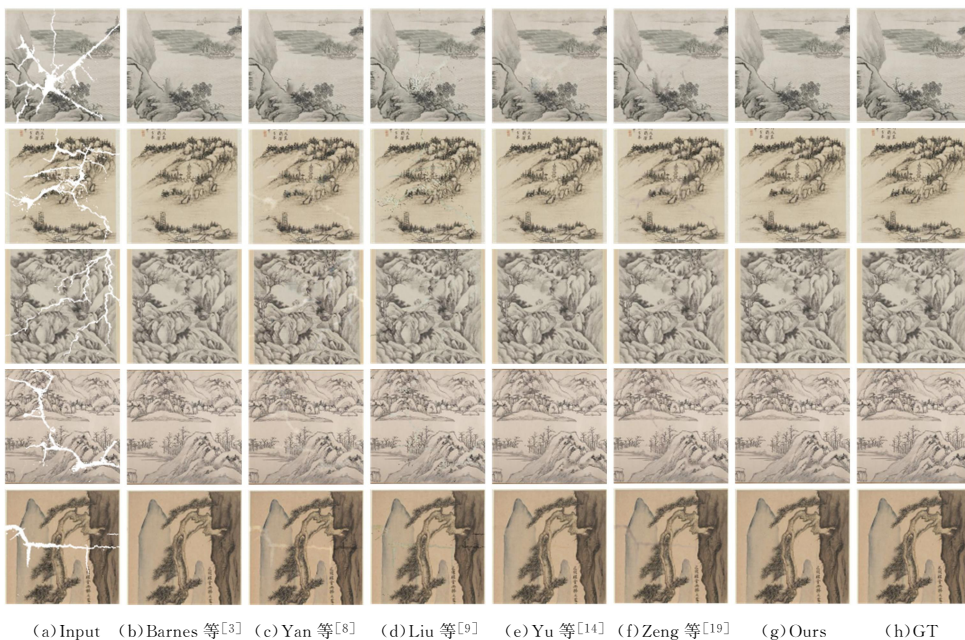


图 7 小面积不规则破损文物图像修复结果对比

Fig. 7 Comparison of image restoration results for small irregularly broken heritage images

表 5 小面积不规则破损修复客观评价价值

Table 5 Objective evaluation value of small and irregular area broken

Index	PSNR [↑]						SSIM [↑]						
	Mask	M _{s1}	M _{s2}	M _{s3}	M _{s4}	M _{s5}	Avg	M _{s1}	M _{s2}	M _{s3}	M _{s4}	M _{s5}	Avg
Barnes 等 ^[3]		36.64	37.80	40.76	37.59	40.54	38.67	0.935	0.946	0.972	0.950	0.969	0.954
Yan 等 ^[8]		34.33	33.15	31.77	34.35	35.69	33.86	0.877	0.852	0.842	0.891	0.920	0.876
Liu 等 ^[9]		30.41	28.02	32.47	32.78	34.78	31.69	0.859	0.814	0.899	0.898	0.938	0.882
Yu ^[14]		37.28	38.36	41.01	38.27	40.94	39.17	0.937	0.946	0.966	0.949	0.969	0.953
Zeng ^[19]		34.72	35.69	38.44	35.65	38.37	36.58	0.909	0.915	0.946	0.927	0.958	0.931
Ours		37.74	38.79	41.52	39.05	41.67	39.75	0.938	0.943	0.964	0.952	0.971	0.953

3.5.2 大面积不规则破损修复

为进一步验证本文方法的修复性能,本节进行实验测试所提方法在更大破损面积下的修复效果,采用 Liu 等^[9]提出的方法所制作的掩码数据集中 3 张较大面积的掩码对各方法进行测试分析。

各方法修复结果如图 8 所示,由图 8(b)、图 8(e)可知,Barnes 等^[3]提出的方法、Yu 等^[14]提出的方法的修复效果较好,但前者缺乏语义信息的指导,修复结果有明显边缘不连续现象,后者对于细节区域的处理不足,局部纹理不连贯。如图 8(c)所示,Yan 等^[8]的方法受限于一阶段

的结构,修复内容整体趋于模糊,且修补痕迹明显。而图 8(d)、图 8(f)所示的 Liu 等^[9]、Zeng 等^[19]的方法则受限于数据集大小,无法从小规模文物数据集中学习到足够的生成泛化能力,导致修补结果语义的合理性较差。相比之下,本文方法修复结果的整体结构更加完整,尤其对船体、山的轮廓等细节实现了较好的修复,整体的视觉效果优于其他方法。

如表 6 所列,对于大面积不规则的破损,本文方法在 PSNR 上取得了次优的结果,除了 Yu 等^[14]的方法,相比其他方法分别提高了 0.547~3.226 dB,并在 SSIM 上取得了最优

的效果,相比其他方法平均提高了 0.04,表明本文方法对于

具有较大破损的文物图像也具有较好的修复效果。

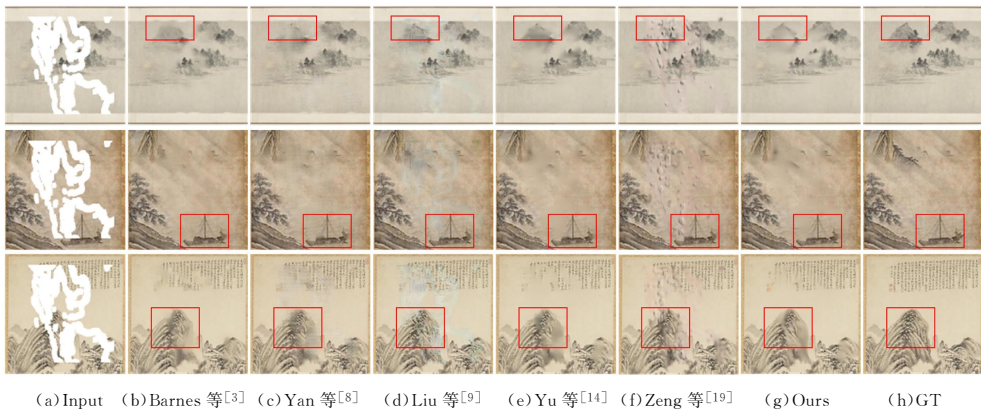


图 8 大面积不规则破损文物图像修复结果对比

Fig. 8 Comparison of image restoration results for large irregularly broken heritage image

表 6 大面积不规则破损修复客观评价值

Table 6 Objective evaluation value of large and irregular area broken

Index	PSNR [↑]				SSIM [↑]			
	M _s 1	M _s 2	M _s 3	Avg	M _s 1	M _s 2	M _s 3	Avg
Barnes 等 ^[3]	30.03	29.75	29.23	29.67	0.735	0.765	0.747	0.749
Yan 等 ^[8]	30.38	29.40	29.60	29.79	0.750	0.710	0.745	0.735
Liu 等 ^[9]	27.87	26.95	26.51	27.11	0.724	0.751	0.736	0.737
Yu 等 ^[14]	30.89	30.38	30.07	30.45	0.769	0.780	0.773	0.774
Zeng 等 ^[19]	27.89	28.09	27.64	27.87	0.726	0.759	0.744	0.743
Ours	30.88	30.29	29.84	30.34	0.781	0.799	0.781	0.787

3.5.3 矩形规则破损修复

由于目前大多数基于深度学习的图像修复方法的研究对象都包含矩形规则破损修复,因此本文也测试了所提方法在该任务上的修复效果,并选取上两组实验中表现较好的 Barnes 等^[3]的方法和 Yu 等^[14]的方法进行对比,结果如图 9 所示。由图 9(b)可知,Barnes 等^[3]的方法对缺失区域进行了详实的填充,但整体视觉效果并不好,与原图差距也较大。而如图 9(c)、图 9(d)所示,受限于数据集的大小,Yu 等^[14]的方法和本文方法修复结果都比较模糊,但得益于对局部信息和上下文信息的综合使用,本文方法对物体的轮廓的刻画能力更好,修复结果更为合理。

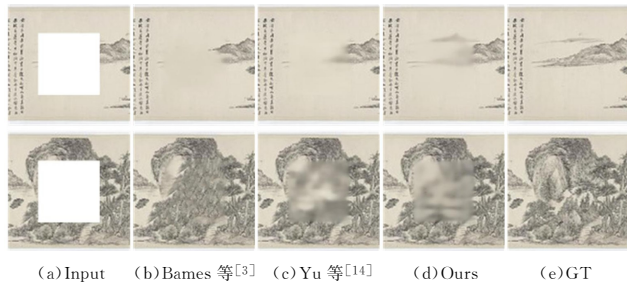


图 9 矩形破损文物图像修复结果对比

Fig. 9 Comparison of image restoration results of rectangular broken heritage images

此外,各方法的修复速度与掩码的大小密切相关,在掩码增大时,各方法的修复速度有不同程度的下降,其中基于块匹配的 Barnes 等^[3]的方法下降最为明显。在小面积掩码实验中,Barnes 等^[3]的方法修复速度快于其余对比方法,但在大面积不规则掩码与矩形掩码的对比实验中,Yu 等^[14]的方法与本文方法修复速度均优于 Barnes 等^[3]的方法。相比 Yu 等^[14]的方法,本文方法加深了粗网络的深度,并在细网络全局分支上使用了多个上下文注意力机制,因此参数量稍大于

Yu 等^[14]的方法,但能够更好地利用图像已知区域信息,从多尺度完成结构和纹理的修复。

结束语 本文针对文物图像特殊的破损类型,提出了一种基于多尺度注意力机制的两阶段文物图像修复方法,该方法依据文物图像的构造及破损规律,在图像修复的不同的阶段针对不同的特征信息进行提取和使用,逐步完成对结构、内容、纹理的生成和调整细化。

本文方法的修复效果稳定,能够实现对文物图像的高质量修复,适用于训练数据集少、图像结构纹理信息复杂、破损区域不规则等大多数现有图像修复方法不适用的情况。但本文方法受限于数据集大小,对较大面积破损的修复达不到理想状态,修补的内容存在较严重的模糊。针对以上问题,今后的研究工作将集中在对文物图像数据集的收集、分类和预处理,并尝试将数字图像超分辨重建技术与图像补全结合,以提高修复结果的分辨率和观赏价值。

参考文献

- [1] MA G C. Analysis of the application of digital image restoration technology in the conservation of cultural relics [J]. Cultural Heritage Identification and Appreciation, 2021, 214 (19): 106-108.
- [2] BERTALMIO M, SAPIRO G, CASELLES V, et al. Image inpainting [C] // Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, 2000.
- [3] BARNES C, SHECHTMAN E, FINKELSTEIN A, et al. Patch-Match: A randomized correspondence algorithm for structural image editing [J]. ACM Trans Graph, 2009, 28(3): 24.
- [4] 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.
- [5] HUANG J B, KANG S B, AHUJA N, et al. Image completion

- using planar structure guidance [J]. *ACM Transactions on Graphics(TOG)*, 2014, 33(4): 1-10.
- [6] STARCK J L, ELAD M, DONOHO D L. Image decomposition via the combination of sparse representations and a variational approach[J]. *IEEE Transactions on Image Processing*, 2005, 14(10): 1570-1582.
- [7] MAIRAL J, ELAD M, SAPIRO G. Sparse representation for color image restoration[J]. *IEEE Transactions on Image Processing*, 2007, 17(1): 53-69.
- [8] 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(ECCV)*. 2018.
- [9] 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)*. 2018.
- [10] 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*. 2016.
- [11] 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*. 2019.
- [12] 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*. 2020.
- [13] NAZERI K, NG E, JOSEPH T, et al. Edgeconnect: Structure guided image inpainting using edge prediction[C]// *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*. 2019.
- [14] 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*. 2019.
- [15] 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*. 2018.
- [16] IIZUKA S, SIMO-SERRA E, ISHIKAWA H. Globally and locally consistent image completion[J]. *ACM Transactions on Graphics(ToG)*, 2017, 36(4): 1-14.
- [17] REN Y, YU X, ZHANG R, et al. Structureflow: Image inpainting via structure-aware appearance flow[C]// *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019.
- [18] WANG Y, CHEN Y C, TAO X, et al. Vcnet: A robust approach to blind image inpainting[C]// *Proceedings of the European Conference on Computer Vision*. 2020.
- [19] 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*. 2021.
- [20] ZENG Y, LIN Z, YANG J, et al. High-resolution image inpainting with iterative confidence feedback and guided upsampling [C]// *Proceedings of the European Conference on Computer Vision*. 2020.
- [21] GUO Z, CHEN Z, YU T, et al. Progressive image inpainting with full-resolution residual network[C]// *Proceedings of the 27th ACM International Conference on Multimedia*. 2019.
- [22] LI J, HE F, ZHANG L, et al. Progressive reconstruction of visual structure for image inpainting[C]// *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019.
- [23] 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*. 2020.
- [24] ZHANG H, HU Z, LUO C, et al. Semantic image inpainting with progressive generative networks[C]// *Proceedings of the 26th ACM International Conference on Multimedia*. 2018.
- [25] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]// *Proceedings of the International Conference on Medical Image Computing and Computer-assisted Intervention*. 2015.
- [26] YANG C, LU X, LIN Z, et al. High-resolution image inpainting using multi-scale neural patch synthesis[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017.
- [27] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. *arXiv:151106434*, 2015.
- [28] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. *Communications of the ACM*, 2020, 63(11): 139-144.
- [29] LIU H, JIANG B, SONG Y, et al. Rethinking image inpainting via a mutual encoder-decoder with feature equalizations[C]// *Proceedings of the European Conference on Computer Vision*. 2020.
- [30] SHARAN L, ROSENHOLTZ R, ADELSON E H. Accuracy and speed of material categorization in real-world images[J]. *Journal of Vision*, 2014, 14(9): 12.
- [31] FUKUSHIMA K, MIYAKE S. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition [M]. *Competition and Cooperation in Neural Nets*. Springer. 1982: 267-85.
- [32] XUE A. End-to-end chinese landscape painting creation using generative adversarial networks[C]// *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2021.
- [33] HORE A, ZIOU D. Image quality metrics: PSNR vs. SSIM [C]// *Proceedings of the 2010 20th International Conference on Pattern Recognition*. 2010.
- [34] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity[J]. *IEEE Transactions on Image Processing*, 2004, 13(4): 600-612.



LIU Haowei, born in 2001, undergraduate student. His main research interests include image processing and deep learning.



LIU Bo, born in 1987, Ph.D, lecturer, is a member of China Computer Federation. His main research interests include multimedia forensics and computer vision.