

## 基于多路编码器和双重注意力的古画修复算法

赵磊<sup>1</sup> 吉柏言<sup>1</sup> 邢卫<sup>1</sup> 林怀忠<sup>1</sup> 林志洁<sup>2</sup>

<sup>1</sup>(浙江大学计算机科学与技术学院 杭州 310027)

<sup>2</sup>(浙江科技学院信息与电子工程学院 杭州 310023)

(cszh@zju.edu.cn)

## Ancient Painting Inpainting Algorithm Based on Multi-Channel Encoder and Dual Attention

Zhao Lei<sup>1</sup>, Ji Boyan<sup>1</sup>, Xing Wei<sup>1</sup>, Lin Huaizhong<sup>1</sup>, and Lin Zhijie<sup>2</sup>

<sup>1</sup>(School of Computer Science and Technology, Zhejiang University, Hangzhou 310027)

<sup>2</sup>(School of Information and Electronic Engineering, Zhejiang University of Science and Technology, Hangzhou 310023)

**Abstract** Painting is an important form of culture and art. For thousands of years, a large number of paintings have been produced in ancient China. And paintings contain rich cultural, artistic, scientific, and historical values. But due to natural disasters (earthquake) and various reasons such as natural weathering and more and more human economic activities, some paintings are more or less damaged or missing in large pieces, which seriously affects the appreciation, cultural creativity, cultural communication and other activities based on these paintings. Compared with natural images, ancient painting images usually have high self-similarity, obvious style characteristics, rich cultural connotation and delicate texture. Although impressive progress has been made in the inpainting technology of natural images, these methods cannot be directly applied to the inpainting of ancient Chinese paintings. Combined with the characteristics of ancient Chinese paintings, we design the algorithm and model structure and proposes a Chinese ancient painting inpainting algorithm based on a multi-channel encoder and dual attention module. The goal is to automatically repair the damaged ancient paintings. In order to better repair the ancient paintings from multiple scales, we use a multi-channel encoder to learn the semantic features of ancient paintings at different scales and repair ancient paintings through the learned macro, meso, and micro semantic features, which solves the difficult problem of rich and delicate texture repair of ancient paintings. In order to better learn the global semantic features of ancient paintings and conduct the harmony and consistency of the repaired ancient paintings, we use the dual attention module to learn the global semantic features of ancient paintings from two aspects: style and content. In order to verify the advanced nature of the algorithm, an ancient painting data set is produced. Experiments on this dataset prove that the algorithm proposed in this paper has better repair quality than the SOTA algorithms.

收稿日期: 2022-07-24; 修回日期: 2023-01-11

基金项目: 浙江省尖兵计划项目(2022C01222); 国家重点研发计划项目(2020YFC1522704); 国家自然科学基金项目(62172365); 浙江省自然科学基金项目(LY21F020005, LY19F020049); 国家社科基金重大项目(19ZDA197); 浙江省文物保护科技项目(2019011); 石窟寺文物数字化保护国家文物局重点科研基地项目; 浙江大学教育部脑与脑机融合前沿科学中心项目(2021008)

This work was supported by the Zhejiang Elite Program(2022C01222), the National Key Research and Development Program of China(2020YFC1522704), the National Natural Science Foundation of China(62172365), the Natural Science Foundation of Zhejiang Province(LY21F020005, LY19F020049), the Key Program of the National Social Science Foundation of China(19ZDA197), the Zhejiang Cultural Relics Protection Science and Technology Project(2019011), the Key Scientific Research Base for Digital Conservation of Cave Temples of State Administration for Cultural Heritage and the Project of MOE Frontier Science Center for Brain Science & Brain-Machine Integration(Zhejiang University)(2021008).

通信作者: 林怀忠(linhz@zju.edu.cn)

**Key words** painting image inpainting; multi-channel encoder; dual attention mechanism; style expression; multi-scale reatoration

**摘 要** 绘画是重要的文化艺术形式,数千年以来,我国古代产生了大量的绘画作品,包含有丰富的文化、艺术、科学与历史价值,但是由于自然灾害(地震)与自然风化以及人类越来越多的经济活动等种种原因导致部分绘画作品存在或多或少的残损或者大块缺失,严重影响了基于这些绘画作品的鉴赏、文化创意、文化传播等活动.与自然图像相比,古画图像的自相似性通常较高,有着明显的风格特点、丰富和细腻的纹理.尽管目前在自然图像上的修复技术已经取得了令人印象深刻的进展,但是这些算法还不能直接用于中国古画的修复.结合中国古画的特点对算法和模型结构进行设计,提出了基于多路编码器和双重注意力机制的中国古画修复算法,目标是对内容受损的古画进行自动化修复.为了能够较好地多个尺度来修复古画,采用了多路编码器来学习古画不同尺度的语义特征,通过学习到的宏观、中观、微观的语义特征来对古画进行修复,解决了古画丰富和细腻的纹理修复困难问题.为了更好地学习古画的全局语义特征,使得修复后的古画整体更加和谐一致,采用了双重注意力模块分别从风格和内容2个方面来学习古画的全局语义特征.为了验证提出的算法的先进性,制作了一个古画数据集,在该数据集上的实验证明,提出的算法相对于目前最先进的算法而言具有较好的修复质量.

**关键词** 绘画图像修复;多路编码器;双重注意力机制;风格表达;多尺度修复

中图法分类号 TP391

在人类文明的发展历程中,文化艺术起着非常重要的作用.而绘画是重要的文化艺术形式,它是中华民族文化的重要组成部分,根植于民族文化土壤之中,通过优美的艺术形式,来反映广泛的现实生活内容,从而体现各民族的文化面貌和审美趣味,是人类观察世界和表达世界的独特而重要的方式.数千年以来,我国产生了大量的绘画作品.对这些绘画作品的研究,是人们了解人类的历史、文化、艺术以及科技的发展史,从而进一步推动人类文明的发展的重要手段.这些绘画作品上下跨度几千年,包含有丰富的文化、艺术、科学与历史价值,但是由于自然灾害(地震)与自然风化以及人类越来越多的经济活动等种种原因导致部分绘画作品存在或多或少的残损或者大块缺失,严重影响了基于这些绘画作品的鉴赏、文化创意、文化传播等活动.因此如何利用最新的科学技术手段对这些已经残损或者大块缺失的艺术图像进行数字化的修复具有非常重要的意义.

随着计算机采集技术和硬件的发展以及国家对文化遗产保护的重视,在国家相关部门的支撑下,相关的文保部门开展了大量的绘画作品(艺术图像)高精度数字化采集工程,典型的工程如历代绘画大系、宋画全集、元画全集等,积累了海量的艺术图像数据.这些数字化数据为艺术图像的智能分析和处理提供了数据基础.同时,计算机人工智能技术特别是深度学习技术目前有了重大的突破,如基于深度学习的图像分类准确率已经达到96%、人脸识别准确

率已经达到97%以上等.计算机人工智能技术的发展为艺术图像的数字化修复提供了技术可行性.另外,目前我国的人均GDP已经超过了1万美元,根据国际经验,人均GDP达到1万美元时文化娱乐消费需求将出现爆发式增长.因此艺术图像的修复工作对当前中国文化产业而言具有非常重要的意义.

图像修复定义为:给定一个包含有噪音或者缺失像素的图像,基于图像中已有的可见像素,利用各种算法推理缺失像素,进而达到对图像的缺失内容的修复.学术界针对图像修复已经做了大量的工作.通常情况下丢失的像素构成图像中的狭长区域或者小洞,这种情况往往会表现为带有划痕和污渍的照片.基于偏微分方程(PDE)的算法可以有效地解决这类修复问题.但是,在图像对象扣除后的图像修复以及严重受损图像的修复等情况下,待修复图像中包含有较大的空洞,针对这类图像的修复来说,基于偏微分方程的算法往往会产生过平滑以及模糊的人工痕迹.基于例子的算法能够较好地解决这个问题.图像缺失的像素是按照随机形式丢失的,这种图像在实际应用中经常出现.重建随机缺失像素的算法和基于案例的算法是有区别的,因为待修复图像中没有例子可以利用,为了解决这个问题学术界提出了很多算法:全变差在已观察的图像保真度的约束下通过最小化图像整体变分(total variation, TV)能量(定义为梯度粒度的积分)来求解问题;矩阵修复算法当原始待修复图像的秩很低时修复效果较好;基于包

的(patch based)算法提供比前述的数值算法更好的修复结果,但通常需要更多的运行时间和占用更多内存.基于深度学习<sup>[1-13]</sup>的图像修复技术由于能够通过深度卷积神经网络来学习和提取图像的特征以及学习图像的先验知识而成为目前主流的图像修复算法,在自然图像的修复方面<sup>[7-40]</sup>也取得了一定的成效.针对空洞区域是光滑而且没有纹理的修复效果较好,但是针对复杂纹理图像的大块空洞,其修复后的图像存在纹理不一致、人工痕迹严重、修复结果多样性差等问题,同时上述文献算法主要针对自然图像,图像修复过程的图像先验模型主要依据人类对于图像内容完整性的视觉感知规律.

与自然图像相比,中国古画具有3个明显的特点而本文也针对这些特点设计了一些应对方案:

1)古画图像的自相似性通常较高,即画面中往往存在多处相似的局部结构或图案,比如山水画中通常有许多外观相似的岩石、树木;街景画中有许多外观相似的房屋、人物等.基于该特点,本文算法利用双重注意力模块中的空间注意力模块使图像上的每个位置建立起全局的视野,修复缺失位置的时候借鉴语义上与之相似的位置的更多信息,以已有的画意作为补缺的标准,使修复结果尽可能接近原作.

2)中国古画作为一种艺术图像,有着明显的风格特点,修复后的区域不应干扰整幅图像的风格.因此本文算法利用双重注意力模块中的风格注意力模块增强模型对图像的风格表示,同时在损失函数中引入风格损失,约束修复结果以及同时维持原图的风格.

3)自然图像相对平滑,而古画中的笔触使得图像有着丰富和细腻的纹理.模型中的多路编码器具有多个视野大小不同的分支,拥有大视野的分支网络关注宏观的结构,而具有小视野的分支网络关注细微的纹理,使得本文算法不仅能够修补受损的内容,还能相对细致地复原纹理,令接笔处与整幅画浑然一体,不露痕迹.总的来说,本文算法力求修复后的古画在内容的完整性和风格上的延续性.

古画遭受的损害通常是宣纸或绢帛表面的破损褶皱、水渍油污、霉变虫洞等,这些病害使得古画不再完整,画面上部分的内容缺失或受损影响了古画的外观和信息表达.在传统的手工修复算法中,会对受损古画进行全色.全色包括补色和接笔等步骤,补色指的是在画面缺损处填充与画面相同的底色,让整幅画的色调趋于协调,做到“纸必补处莫分”;接笔指的是在因残损造成的断笔处,按照画意填补上缺

失的内容,使画意重新衔接完整.但是手工修复的风险较大,而在计算机上利用图像修复算法对古画进行数字化修复能够避免对文物本体造成再次损伤.由于画家在创作过程中带有主观性,绘画作品较之于自然图像有较大的区别.首先,与自然图像相比,艺术图像颜色具有不同程度的退化.其次,与一般的自然图像相比,绘画作品是人工作品,图像内容经由艺术家刻意设计,具有鲜明的特点,如艺术风格基本组成是线描和色彩.自然图像主要表达真实存在的内容,无论是背景还是目标主题,都不脱离现实.绘画作品则更多地反映画家的构想,其内容可能与现实差别巨大.即使是同一主体对象,在绘画作品中也可能以不同的形式、不同的艺术风格表现出来,更遑论艺术风格本身就是主观抽象的概念.再次,艺术图像颜色组成简单(相对于自然图像的丰富色彩),而且具有明显的笔触(绘画艺术区别于摄影艺术的重要标志).如何学习艺术图像中隐含的风格、色彩等规律,并应用这些规律对图像进行修复是关键的科学难题.总的来说,相对于自然图像而言,艺术图像由于其色彩退化、噪音类型复杂等问题导致其数字化修复异常困难,但是同时由于艺术图像具有鲜明的艺术风格、内容色彩简单有规律等特点,为艺术图像的修复提供了良好的先验性约束.鉴于艺术图像的特殊性,传统的图像修复算法并不适合直接应用于艺术图像的修复.

综上所述,针对艺术图像修复仍然存在2个问题:

1)艺术作品一般都包含了作者的绘画风格和流派、意境等信息,如何对艺术图像中隐含的这些信息建模学习,并依据这些信息作为约束(如风格)来修复艺术图像是一个难题;

2)相对于自然图像,艺术图像数量较少,另外待修复的艺术图像并没有和其对应的真实(ground truth)图像用于训练深度学习模型,如何在小样本条件下构建艺术图像的无监督修复模型是一个难题.

针对这2个问题,结合中国古画的特点对算法和模型结构进行设计,采用了双重注意力模块分别从风格和内容2个方面来学习古画的全局语义特征.为了验证本文提出的算法的先进性,制作了一个古画数据集,在该数据集上的实验证明,提出的算法相对于目前最先进的算法而言具有较好的修复能力.

## 1 相关工作

图像修复技术作为数字图像处理领域中热门的

研究方向在国内外都受到了很大的关注, 图像修复<sup>[1-6]</sup>概念的提出是受到了以往手工修补艺术图像的启发, 至今已经有将近 20 年的发展历程. 目前现有的解决图像缺失算法主要有 4 类.

### 1.1 基于偏微分方程 (PDE) 的图像修复算法

基于 PDE 的图像修复算法最早是由图像修复技术概念的提出人 Bertalmio 等人<sup>[7]</sup>在 2000 年的 SIGGRAPH 大会上提出来的, 该模型又称 BSCB (Bertalmio sapiro caselles ballester) 模型<sup>[7]</sup>. Telea<sup>[8]</sup>在 BSCB 的基础上提出了快速行进算法, 通过确定待修补区域领域的等照度线方向, 层层递进对图像进行完整修复. Chan 等人<sup>[9]</sup>提出了利用欧拉-拉格朗日方程和各向异性扩散的方式来保持等照度线方向的算法, 这种算法被称为整体变分. 为了满足图像视觉连通性的要求, Chan 等人<sup>[10]</sup>提出基于曲率驱动扩散 (curvature-driven diffusion, CDD) 的修复模型. 相关算法还有: Tsai 等人<sup>[11]</sup>提出的 Mumford-Shah 模型、Shen 等人<sup>[12]</sup>提出的 Euler's Elastica 模型、Esedoglu 等人<sup>[13]</sup>提出的 Mumford-Shah-Euler 模型. 以上基于 PDE 的图像处理算法在缺失区域较小的图像中有较好的表现, 但对于图像破损区域较大的图像通常会出现一些如边界模糊、纹理延伸错误等问题.

### 1.2 基于纹理合成的图像修复算法

基于纹理合成的图像修复算法, 其起源于 Efros 等人<sup>[14-15]</sup>提出用于纹理合成的非参数算法, 通过查询样本图像找到所有相似的邻域来估计到目前为止合成的所有邻域的像素的条件分布. Bertalmio 等人<sup>[16-17]</sup>将图像分解成 2 个部分, 结构子图利用基于扩散的算法进行修复, 纹理子图则用纹理综合的算法填充. Drori 等人<sup>[18]</sup>提出了基于碎片的图像修复算法, 但这个算法处理速度慢. Criminisi 等人<sup>[19]</sup>提出了采样填充图像的修复算法, 极大地提升了图像修复的最终效果. Barnes 等人<sup>[20]</sup>提出的包匹配 (patch match) 基于快速近似最近邻的思想大大提高了算法的运行效率<sup>[21]</sup>.

### 1.3 基于稀疏表示的图像修复算法

Stark 等人<sup>[22]</sup>利用形态成分分析 (morphological component analysis) 将图像结构和纹理等部分合并得到最终的修复图像. Elad 等人<sup>[23]</sup>在 Stark 等人<sup>[22]</sup>的基础上引入了惩罚因子来约束图像结构信息, 从而较好地保持原图的结构信息, 兼顾了纹理信息. Mairal 等人<sup>[24]</sup>将学习字典应用于彩色图像的复原, 并且提出了在线字典学习<sup>[25]</sup>的算法. Shen 等人<sup>[26]</sup>则将信号标记字典应用于图像修复, 无需字典的训练来达到提高计算效率的目的. Li<sup>[27]</sup>提出的综合稀疏表示

(hybrid sparse representations) 模型是利用贝叶斯平均法来组合图像全局和局部稀疏表示的算法. Gabriel<sup>[28]</sup>将稀疏表示模型和纹理合成相结合提出一种新的修复算法. Zhang 等人<sup>[29-30]</sup>提出了基于空间变化域上联合稀疏统计修复模型来达到更好的修复效果. Zhang 等人<sup>[31]</sup>提出将纹理合成中的基于样本的修复算法和稀疏表示相结合的方法, 并将稀疏表示运用于搜索图像块局部特征中, 该方法既可以克服纹理扩散问题又可以解决稀疏表示中修复大破损区域的边界模糊问题.

### 1.4 基于深度学习的图像修复算法

传统的非深度学习的图像修复的算法主要是基于物理和数学的算法, 但是这些算法的修复往往依靠单张图像的信息, 没有办法学习某一类图像的规律, 进而导致其修复效果存在明显的人工痕迹. 近几年, 深度神经网络被引入用于图像纹理合成图像修复中, Pathak 等人<sup>[4]</sup>训练编码-解码 CNN (上下文编码器) 和对抗生成网络<sup>[32]</sup>的组合网络直接预测缺失的图像区域. 该算法的修复结构往往会因为丢失图像纹理信息导致修复结果缺乏精细的纹理细节从而在孔的边缘产生较为明显的伪影, 无法处理高分辨率图像. 为了解决高分辨率图像问题, Yang 等人<sup>[33]</sup>在上下文编码器的基础上提出了基于图像内容和纹理约束的联合优化的多尺度神经补片合成的算法, 该算法不仅保留了上下文结构, 而且通过匹配和调整补丁与最相似的中间层特征的相关性来产生高频细节, 从而能获得较高分辨率图像. Zeng 等人<sup>[34]</sup>提出了多尺度的特征金字塔的图像修复模型. Ren 等人<sup>[35]</sup>利用图像的结构信息来对图像进行修复. Xiong 等人<sup>[36]</sup>首先将待修复的图像进行前背景分离, 然后再分别进行修复, 取得了较好的修复结果. Zheng 等人<sup>[37]</sup>采用 2 条并行路径来提高图像修复的多样性, 其中一条路径学习缺失部分的先验分布, 另一条路径根据先验分布生成缺失部分. 然而, 这 2 条并行路径只能提供有限的多样性, 因为它仅仅通过学习图像缺失部分的先验分布来生成不同的图像修复结果. Zhao 等人<sup>[38]</sup>提出了无监督跨空间转换生成对抗网络来探索可能的修复结果空间, 该对抗网络表现出了较好的图像修复的多样性. Liao 等人<sup>[39]</sup>引入了语义和纹理之间的一致性先验来增加修复的纹理约束, 首先对先验进行建模, 然后以粗到细的方式对图像修复和语义分割进行相应的交错优化, 通过探索非局部语义连贯性, 跨尺度细化完成图像纹理, 有效缓解了纹理的混淆. Liu 等人<sup>[40]</sup>提出了一种用于两者联合恢复的互



编码器-解码器,使用编码器深层和浅层的卷积神经网络特征分别表示输入图像的结构和纹理,将深层特征发送到结构分支,将浅层特征发送到纹理分支,在每个分支中,都会在卷积神经网络的多个尺度特征中填充漏洞.Nazeri 等人<sup>[41]</sup>提出了一种2阶段对抗模型 EdgeConnect,该模型由边缘生成器和图像完成网络组成,边缘生成器使图像的缺失区域(规则和不规则)的边缘产生幻觉,图像完成网络使用幻觉边缘作为先验来填充缺失区域.Yu 等人<sup>[42]</sup>提出了一种生成式图像修复系统,用自由形式的掩模和引导来完成,该系统基于从数百万图像中学习的门控卷积,无需额外的标记工作.

目前,尽管深度学习算法在图像修复方面做了大量的工作,相对于传统的修复算法在修复效果上也取得了一定的突破,但是现存算法都是针对自然图像修复的,针对艺术图像修复的算法仍然存在前面所述的问题.

## 2 基于多路编码器和双重注意力机制的修复算法

### 2.1 网络模型

当前图像修复领域最先进的算法都采用生成对抗网络作为基础架构,生成对抗网络中生成器和判别器相互博弈的训练策略促使生成器去模拟真实数

据的分布,使得修复结果与真实图像更接近.本文模型也将生成对抗网络作为基础架构,模型包含生成器 $G$ 网络和判别器 $D$ 网络,在训练阶段,生成器和判别器共同参与训练;而在测试阶段,仅使用生成器对图像进行修复,不需要判别器的参与.如图1所示,进一步对生成器网络进行划分,可以分解为3个部分:位于网络前段的多路编码器、位于网络中间位置的空间-风格注意力模块和位于网络后段的解码器.生成器的输入是带有不规则形状孔洞的输入图像 $\mathbf{x}$ 以及标识缺失像素和完整像素位置的掩膜 $\mathbf{M}$ .由于不规则孔洞的数量、位置、形状、面积都具有很高的随机性,仅仅输入缺失图像 $\mathbf{x}$ ,生成器很容易将古画上的一些孔洞和原本的图案混淆起来,所以此处将 $\mathbf{M}$ 作为辅助的条件信息,输入到生成器之中.其中 $\mathbf{x}$ 尺寸为 $256 \times 256$ , $\mathbf{M}$ 是一张二值单通道掩膜,宽和高与 $\mathbf{x}$ 相同, $\mathbf{M}$ 上每个像素的值为0或1,分别对应图像上的完整像素和缺失像素. $\mathbf{M}$ 在通道维度被拼接在 $\mathbf{x}$ 的后面,形成4通道的 $[\mathbf{x}, \mathbf{M}]$ 后作为生成器的输入.算法最终输出的修复图像为 $\hat{\mathbf{y}} = G([\mathbf{x}, \mathbf{M}]) \odot \mathbf{M} + \mathbf{x} \odot (1 - \mathbf{M})$ ,其中符号 $\odot$ 表示求哈达玛积(Hadamard product),即求2个矩阵对应元素相乘(element wise)的结果.因此生成器的任务是对输入图像上的孔洞中的像素进行修复,孔洞以外的完整像素则直接输出,这样生成器不需要考虑这些孔洞以外的区域的生成问题,简化了生成器的任务.判别器的输入是修复图像 $\hat{\mathbf{y}}$ 和对应

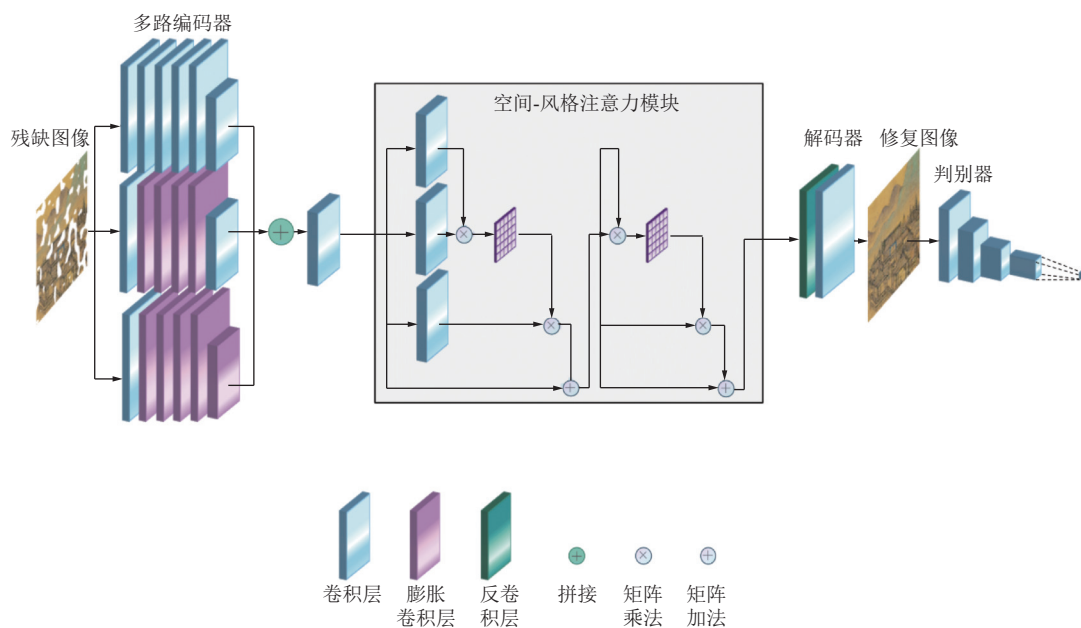


Fig. 1 Network structure diagram of Chinese ancient painting restoration algorithm based on multi-channel encoder and dual attention module

图1 基于多路编码器和双重注意力模块的中国古画修复算法的网络结构图

的真实完整图像(ground truth)  $y$ ,  $D$ 对两者分别评分, 尝试将它们区分开来. 下面对每个部分进行详细介绍.

### 2.1.1 多路编码器

卷积神经网络通过将输入图像逐层地抽象化以提取图像的特征, 在这个过程中, 卷积核尺寸(kernel size)、膨胀系数(dilation)、卷积操作的步长(stride)这些参数都会影响网络的感受野(receptive field)的大小, 即网络中某一层输出的特征图上的单个像素点对应于输入图像上映射区域的大小. 感受野大小不同, 观察到的特征也是不同的, 小的感受野关注图像局部的特征, 能够观察到更多的细节, 大的感受野则能够观察到整体的趋势. 因此网络的感受野大小决定该网络能够捕获到图像中多大尺度的信息, 可以被理解为网络的“视野”.

在古画的修复任务中, 想要获得合理且修复痕迹轻的修复结果, 需要在不同大小的视野下观察图像. 以图2(a)为例, 被虚线包围的白色区域代表图像上的孔洞.  $p1$  是孔洞中的 1 个待修复的像素点,  $\Omega1$ 指代的方形区域表示感受野, 图像修复模型要对孔洞造成的内容缺失进行修补, 需要从较大的视野里看清缺失区域所属部分的结构, 此时需要较大尺寸的感受野. 为了使修复后的结果看起来自然, 修复时还要保证修复区域的纹理被逼真地复原, 避免在修复区域留下明显的痕迹, 此时应当使用较小的视野, 关注周围邻近区域的纹理和细节, 如图2(b)所示. 本文算法中的多路编码器能够实现多尺度特征抽取和多尺度特征融合, 如图1中所展示的多路编码器结构, 它具有 3 条并行的卷积神经网络分支. 本文算法中通过使用膨胀卷积<sup>[43]</sup>来迅

速地增大部分分支的感受野, 使得 3 条分支最终的感受野大小各不相同. 膨胀卷积与同尺寸的普通卷积相比, 并没有改变实际参与卷积点乘的像素个数, 而是通过对像素点按照一定数目的间隔进行采样, 扩大了输入窗口的尺寸. 膨胀卷积扩大感受野的方式, 其中 3 张子图分别表示依次堆叠起来的 3 个卷积层, 卷积核尺寸均为  $3 \times 3$ , 步长均为 1, 而膨胀系数依次为 1, 2, 4. 第 1 层卷积层膨胀系数为 1, 表示不膨胀, 其感受野为  $3 \times 3$ ; 第 2 层卷积层膨胀系数为 2, 尽管实际只有网格标注的 9 个像素参与了卷积运算, 但是由于该层的每个像素都是前一层卷积的输出, 因此该层每个像素点的感受野都为  $3 \times 3$ , 这 9 个像素的感受野平铺后就能覆盖图上  $7 \times 7$  的区域, 使得该层达到  $7 \times 7$  的感受野; 第 3 层卷积层的膨胀系数为 4, 同理, 由于参与卷积的 9 个像素点都具有  $7 \times 7$  的感受野, 平铺后能够覆盖图上  $15 \times 15$  的区域. 普通卷积的感受野随层数增长的速度太慢, 为了获得大感受野需要一定的网络深度, 导致训练难度提高, 训练时长增加. 膨胀卷积和同尺寸的普通卷积相比, 并没有增加参数的数量和计算量, 而膨胀卷积的感受野随层数的增加呈指数级的增长, 增长速度明显快于普通卷积. 多路编码器具有 3 条并行的卷积神经网络分支, 其输入数据的形式为  $[x, M]$ , 3 个分支网络都接受相同的输入, 且除了膨胀系数不同以外, 3 个分支网络具有完全相同的结构和配置. 3 个分支均由 6 个卷积层构成, 卷积核尺寸都为  $3 \times 3$ . 由于在编码阶段多次下采样的做法容易丢失图像细节信息, 导致解码时重构图像的纹理模糊, 本文算法尽可能减少下采样的次数. 每个分支网络中, 前 5 层卷积的步长均

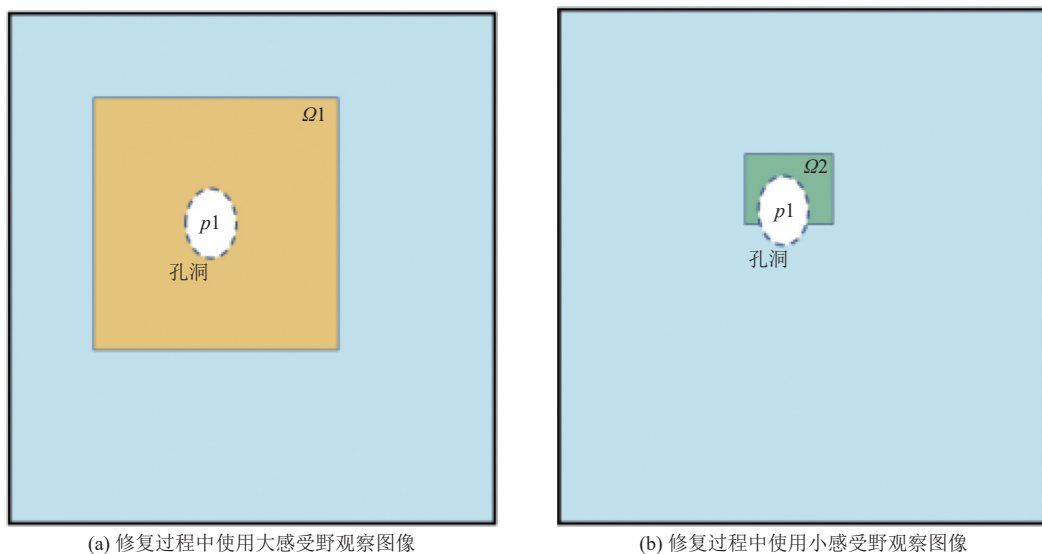


Fig. 2 Observing image using different receptive fields in the inpainting process

图2 修复过程中使用不同感受野观察图像

为1, 仅第6层卷积的步长为2, 即只在最后1层进行下采样, 目的是将多路编码器输出的特征图尺寸缩小到合适的大小, 以减少后续计算步骤所需的存储空间. 所有卷积层采用 Elus<sup>[44]</sup> 作为激活函数, 其正值部分与 ReLU 函数相同, 导数均为1, 减轻了梯度弥散的问题. 但 ReLU 的输出值域不包含负数, 所以其激活值的均值为正值, 会对下一层造成内部协变量偏移(internal covariate shift)的影响, 而 Elus 具有负值部分, 使得激活值的均值更接近0, 产生类似批归一化(batch normalization)的效果. 此外, Elus 与 ReLU 相比, 负值部分是指数函数, 而不是线性函数, 当 Elus 的输入取较小值时具有饱和的特性, 对噪声更鲁棒.

3个分支网络各层的膨胀系数设置如表1所示. 第1个分支的6个卷积层的膨胀系数均为1, 即全不膨胀, 最后1层的感受野大小为13×13; 第2个分支的6个卷积层的膨胀系数分别为1, 2, 4, 8, 16, 1, 即在中间4层进行膨胀, 最后1层的感受野大小达到65×65; 第3个分支的6个卷积层的膨胀系数分别为1, 2, 4, 8, 16, 32, 即在后5层进行膨胀, 最后1层的感受野大小达到127×127. 不同的膨胀系数配置, 使得不同分支具备大小不同的感受野, 从而关注不同尺度的特征. 而在这3个分支网络的末端, 3个输出结果

在通道维度上被拼接起来, 然后再通过1次卷积, 实现对不同尺度特征的融合, 融合后的特征输入到空间-风格注意力模块执行后续的处理.

### 2.1.2 空间-风格注意力模块

空间注意力模块为图像中的每个位置构建全局视野, 有效捕获空间上不同位置之间的长距离依赖关系, 令每个位置从语义上更相似的位置处获取更多信息, 在对缺失位置进行修复时更有依据, 增强修复结果的可靠性和准确性; 风格注意力模块能够增强模型对图像风格的表示, 其与损失函数中的风格损失配合, 能够约束修复结果和保持原图的风格. 如图3所示, 2个注意力模块以串行的方式相连接. 空间注意力模块的输入特征图A是由多路编码器输出的特征图, 形状为 $C \times H \times W$ . 现有的实验<sup>[45]</sup>表明, 在计算注意力矩阵时适当降低特征图的通道数并不会导致明显的性能下降, 因此在进行矩阵乘法之前, 先令特征图A分别经过2个独立的卷积层的处理, 将深度压缩为原来的1/8, 依次得到形状为 $\frac{C}{8} \times H \times W$ 的特征图P和Q, 从而提升空间效率. 接着将特征图P和Q的形状均调整为 $\frac{C}{8} \times N$ , 其中 $N = H \times W$ , N表示特征图上的像素总数. 然后对P的转置和Q进行矩阵乘法,

Table 1 Network Structure of Multi-Channel Encoder

表1 多路编码器网络结构

分支编号	层编号	层类型	卷积核尺寸	膨胀系数	步长	输出尺寸	输出通道数	感受野尺寸
1	1	卷积层	3×3	1	1	256×256	64	3×3
	2	卷积层	3×3	1	1	256×256	64	5×5
	3	卷积层	3×3	1	1	256×256	64	7×7
	4	卷积层	3×3	1	1	256×256	64	9×9
	5	卷积层	3×3	1	1	256×256	64	11×11
	6	卷积层	3×3	1	2	128×128	128	13×13
2	1	卷积层	3×3	1	1	256×256	64	3×3
	2	膨胀卷积层	3×3	2	1	256×256	64	7×7
	3	膨胀卷积层	3×3	4	1	256×256	64	15×15
	4	膨胀卷积层	3×3	8	1	256×256	64	31×31
	5	膨胀卷积层	3×3	16	1	256×256	64	63×63
	6	卷积层	3×3	1	2	128×128	128	65×65
3	1	卷积层	3×3	1	1	256×256	64	3×3
	2	膨胀卷积层	3×3	2	1	256×256	64	7×7
	3	膨胀卷积层	3×3	4	1	256×256	64	15×15
	4	膨胀卷积层	3×3	8	1	256×256	64	31×31
	5	膨胀卷积层	3×3	16	1	256×256	64	63×63
	6	膨胀卷积层	3×3	32	2	128×128	128	127×127
	7	卷积层	3×3	1	1	128×128	128	无

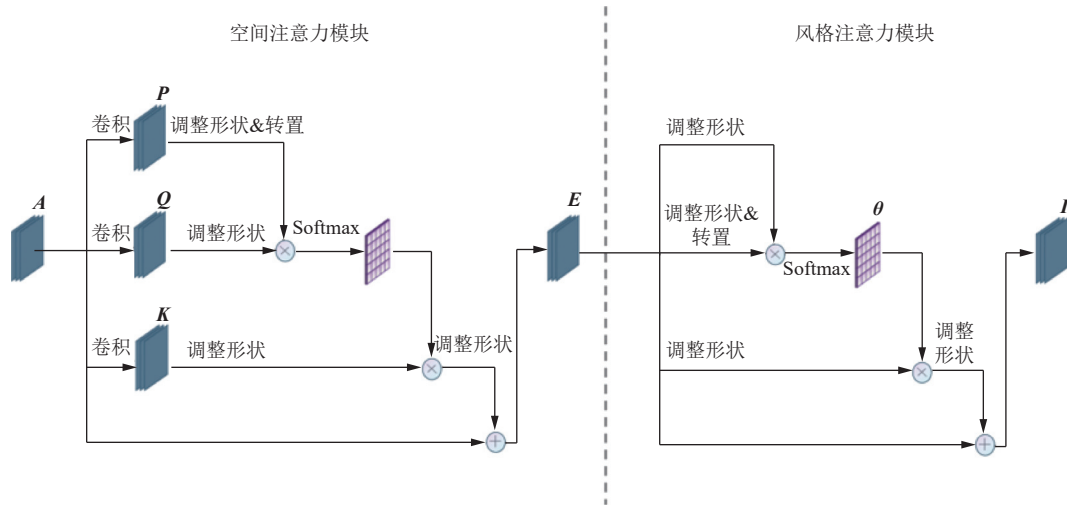


Fig. 3 Structure diagram of space-style attention module

图3 空间-风格注意力模块的结构图

得到的结果通过 Softmax 操作归一化, 得到空间注意力矩阵  $\beta \in \mathbb{R}^{N \times N}$ , 如式(1)和式(2)所示:

$$s_{ij} = P_i \cdot Q_j, \quad (1)$$

$$\beta_{ij} = \frac{\exp(s_{ij})}{\sum_{i=1}^N \exp(s_{ij})}. \quad (2)$$

由于之前经过了多路编码器的处理, 特征图上的每个点都包含了一定视野范围内的语义信息, 此时空间注意力矩阵能够反映空间上不同位置之间的语义相似性,  $\beta_{ij}$ 表示第  $i$  个位置对第  $j$  个位置的影响力权重, 进行 Softmax 操作的目的是将矩阵相乘得到的原始分值整理为所有元素之和为 1 的概率分布, 通过 Softmax 机制突出重要元素的权重. 与此同时, 将输入的特征图  $A$  再经过 1 个卷积层的处理得到特征图  $K$ , 值得注意的是这次不对  $K$  的深度进行压缩. 接着将  $K$  的形状调整为  $C \times N$  后, 对  $K$  和空间注意力矩阵执行矩阵乘法, 对特征图上的每个位置进行加权求和, 得到  $o = (o_1, o_2, \dots, o_j, \dots, o_N) \in \mathbb{R}^{C \times N}$ , 其中

$$o_j = \sum_{i=1}^N \beta_{ji} K_i. \quad (3)$$

最后, 将  $o$  的形状还原为  $C \times H \times W$ , 并以残差的形式和原始的特征图  $A$  相加, 如式(4)所示:

$$E = \alpha o + A, \quad (4)$$

其中  $\alpha$  是初始值为 0 的可学习的比例系数, 因此在训练的初始阶段, 空间注意力模块执行非常简单的任务, 即直接返回输入的特征图  $A$ , 之后随着网络的学习, 逐渐地将经过注意力加权的特征图加到原始的特征图上, 从而强调了需要施加注意力的部分.

综上所述, 空间注意力模块的主要操作是: 通过

特征图之间的矩阵乘法, 计算出空间上任意 2 个位置之间的语义相似性作为权重系数, 对原始特征图进行加权求和, 实现了基于位置之间的语义相似性来分配从每个位置获取的信息量.

卷积神经网络中特征图的每个通道可以视作对某个特定类别特征的响应, 浅层的特征主要反映图像局部的特点, 比如局部的纹理; 深层的特征则更能反映整体的特点, 如形状、轮廓等. 在风格迁移领域, 将图画中形状、色彩等特征的组合方式理解成图像的“风格”. 风格注意力模块显式地计算特征图的不同通道之间的相关性, 并让每个通道从与其关联性更强的通道中获取更多信息, 以增强网络对图像风格的表示.

风格注意力模块的输入特征图是空间注意力模块的输出  $E$ , 与计算空间注意力矩阵时不同, 输入特征图不再经过卷积层的预处理. 将输入特征图  $E$  的形状调整为  $C \times N$  得到  $F$ , 接着直接将  $F$  与它的转置相乘, 得到 Gram 矩阵  $G \in \mathbb{R}^{C \times C}$ ,  $G$  的对角线元素体现每个特征自身的强度, 非对角线元素体现不同特征间的相关度, 由于  $G$  能反映特征本身的信息及特征之间的关系, 风格迁移领域将  $G$  作为图像的风格表示. 计算得到  $G$  后利用 Softmax 操作进行归一化, 得到风格注意力矩阵  $\theta \in \mathbb{R}^{C \times C}$ , 如式(5)和式(6)所示:

$$G_{ij} = F_i \cdot F_j, \quad (5)$$

$$\theta_{ji} = \frac{\exp(G_{ij})}{\sum_{i=1}^C \exp(G_{ij})}, \quad (6)$$

其中  $\theta_{ji}$  表示第  $i$  个通道特征对第  $j$  个通道特征的影响力权重. 之后将  $\theta_{ji}$  与  $F$  相乘, 得到  $p = (p_1, p_2, \dots,$



$p_j, \dots, p_C) \in \mathbb{R}^{C \times N}$ , 其中

$$p_j = \sum_{i=1}^C \theta_{ji} F_i. \quad (7)$$

最后将  $p$  的形状还原为  $C \times H \times W$ , 乘上比例系数  $\gamma$  后和输入的特征  $E$  相加, 如式(8)所示.

$$I = \gamma p + E. \quad (8)$$

经过风格注意力模块处理后的特征图被传入解码器模块.

### 2.1.3 解码器

解码器模块由 1 个反卷积层和 1 个卷积层构成, 反卷积层的目的是将特征图尺寸上采样到原始分辨率, 卷积层的目的是将特征图的通道还原为 RGB, 网络结构如表 2 所示.

Table 2 Decoder Network Structure

表 2 解码器网络结构

层编号	层类型	卷积核尺寸	步长	输入尺寸	输出尺寸	输入通道数	输出通道数
1	反卷积层	3×3	2	128×128	256×256	128	64
2	卷积层	3×3	1	256×256	256×256	64	3

生成器的最终输出为  $G([x, M])$ . 缺失图像上原本完整的那些像素在修复结果中应该保持不变, 所以对于输入图像上孔洞之外的区域, 修复结果直接使用输入图像中的像素; 对于孔洞内的区域, 修复结果使用生成器输出图像中的像素, 算法最终产生的修复结果为  $\hat{y} = G([x, M]) \odot M + x \odot (1 - M)$ .

### 2.1.4 判别器

判别器的构成如表 3 所示, 接收的输入图像尺寸为 256×256. 网络的前 4 层为卷积层, 步长均为 2, 采

用 ReLU 作为激活函数, 逐步降低图像的分辨率, 同时提升通道数. 最后 1 层为全连接层, 输出判别的分值. 判别器的训练目标是使真实图像  $y$  的输出分值尽可能大、修复图像  $\hat{y}$  的分值尽可能小, 以区分 2 种分布. 值得注意的是, 由于本文算法采用带梯度惩罚 (gradient penalty)<sup>[46]</sup> 的 Wasserstein 损失<sup>[47]</sup> 作为对抗损失函数, 判别器的输出结果不再是简单的二分类, 而是对 Wasserstein 距离的拟合, 因此最后 1 层不使用 Sigmoid 激活函数.

Table 3 Discriminator Network Structure

表 3 判别器网络结构

层编号	层类型	卷积核尺寸	步长	输入尺寸	输出尺寸	输入通道数	输出通道数
1	卷积层	3×3	2	256×256	128×128	3	64
2	卷积层	3×3	2	128×128	64×64	64	128
3	卷积层	3×3	2	64×53	32×32	128	256
4	卷积层	3×3	2	32×32	16×16	256	512
5	全连接层			16×16	1	512	1

## 2.2 损失函数

本文算法中, 损失函数由对抗损失函数、重构损失函数和风格损失函数 3 部分构成, 下面进行介绍.

### 2.2.1 对抗损失函数

原始 GAN 的损失函数设计存在一定缺陷, 当判别器  $D$  训练到最优时, 生成器  $G$  的优化目标即为最小化真实图像分布与修复图像分布之间的 JS 散度. 然而 JS 散度在度量分布之间的距离时有一定缺陷, 当 2 个分布几乎不重叠时, 无论 2 个分布的中心距离是多少, JS 散度的值恒为常数  $\ln 2$ , 由于常数的梯度是 0, 使得采用梯度下降算法进行优化的  $G$  无法继续学习. 因此在原始 GAN 的损失函数的指导下,  $D$  训练得越好,  $G$  就越可能面临梯度弥散的问题.

本文算法采用带梯度惩罚的 Wasserstein 损失作为对抗损失函数, 如式(9)所示:

$$L_{\text{adv}} = E_{\hat{y}} [D(\hat{y})] - E_y [D(y)] + \lambda_{\text{gp}} E_{\hat{y} \sim p_{\hat{y}}} [(\|\nabla_{\hat{y}} D(\hat{y}) \odot M\|_2 - 1)^2] \quad (9)$$

其中  $E_{\hat{y}} [D(\hat{y})] - E_y [D(y)]$  表示分布之间的 Wasserstein 距离, 在 2 个分布不存在重叠的情况下, 它依然是一种有效的距离度量方式, 为  $G$  提供梯度信息.  $\lambda_{\text{gp}} E_{\hat{y} \sim p_{\hat{y}}} [(\|\nabla_{\hat{y}} D(\hat{y}) \odot M\|_2 - 1)^2]$  表示梯度惩罚项, 使  $D$  的梯度收敛在固定值 1 附近, 以满足 Lipschitz 连续条件. 式中  $\tilde{y}$  是  $\hat{y}$  和  $y$  的随机插值, 即

$$\begin{aligned} \varepsilon &\sim \text{Uniform}[0, 1], \\ \tilde{y} &= (1 - \varepsilon)y + \varepsilon\hat{y}, \end{aligned} \quad (10)$$

$M$ 表示残缺图像的掩膜。

### 2.2.2 重构损失函数

对抗损失引导生成器产生合理的修复结果以欺骗判别器, 重构损失则引导生成器产生与真实图像更接近的修复结果. 与 L2 距离相比, L1 距离更能减轻修复图像的模糊, 因此本文算法采用 L1 距离计算重构损失:

$$L_{\text{rec}} = E [\|\mathbf{y} - \hat{\mathbf{y}}\|_1]. \quad (11)$$

### 2.2.3 风格损失函数

古画修复任务除了需要修复受损的内容, 还应该保持画作原有的风格, 本文算法利用风格注意力模块增强模型对图像风格的表示, 并采用风格损失函数对修复结果进行风格约束. 如 2.1.2 节介绍风格注意力模块时提到的, 使用特征的 Gram 矩阵  $G$  来表示图像的风格. 具体来说, 将修复结果图像  $\hat{\mathbf{y}}$  和对应的真实图像  $\mathbf{y}$  分别输入在 ImageNet 数据集上预训练过的 VGG19 网络<sup>[48]</sup> 中, 然后分别提取 2 张在 VGG19 网络中 conv1\_1, conv2\_1, conv3\_1, conv4\_1, conv5\_1 这 5 个卷积层输出的特征图, 并对每层特征计算 Gram 矩阵获得该层的风格表示, 最终风格损失的计算为:

$$L_{\text{style}} = \sum_l \left[ \frac{1}{4C_l^2 N_l^2} \sum_{i,j} (Y_{ij}^l - \hat{Y}_{ij}^l)^2 \right], \quad (12)$$

其中  $\hat{Y}_{ij}^l$  和  $Y_{ij}^l$  分别表示  $\hat{\mathbf{y}}$  和  $\mathbf{y}$  第  $l$  层的 Gram 矩阵中的元素, 而  $C_l$  和  $N_l$  分别表示第  $l$  层特征的通道数和宽高乘积. 求出每一层 Gram 矩阵之间的均方误差后, 逐层累加起来得到总体的风格损失.

最后, 完整的损失函数  $L_{\text{total}}$  由对抗损失  $L_{\text{adv}}$ 、重构损失  $L_{\text{rec}}$  和风格损失  $L_{\text{style}}$  的加权和构成:

$$L_{\text{total}} = \lambda_{\text{adv}} L_{\text{adv}} + \lambda_{\text{rec}} L_{\text{rec}} + \lambda_{\text{style}} L_{\text{style}}. \quad (13)$$

## 3 实验设计

### 3.1 实验数据

目前业内关于中国古画的研究工作很少, 也没有公认的标准数据集, 所以本文自行建立了一个中国古画数据集并基于此开展实验. 本文收集了 197 幅超高分辨率的古画图像, 这些作品出自 11 位画家之手, 作品类别包括山水画、街景画、花鸟画、人物画等. 古画数据的特殊性在于每幅作品的分辨率非常高, 一些图像的长度或宽度可以达到几万像素. 现有的修复模型和主流显卡的显存均无法处理如此高的分辨率图像, 因此这里采用划分切片的数据处理算

法, 将超大尺寸的古画图像裁剪为合适大小的切片图像. 通过 PhotoShop 工具, 人工将每幅原始图像裁剪成多个互不重叠的切片, 切片的尺寸均为  $256 \times 256$ , 有些尺寸特别大的图像会包含数量巨大的切片. 通过人工的手段裁剪切片的原因是中国古画与其他国家画作的艺术风格不同, 作品中经常有大量的留白. 古画中的留白是为了衬托画面中的实体, 是一种“以虚衬实”的艺术表现手法<sup>[49]</sup>. 然而本文算法关注古画的内容修复, 画面中大面积的留白导致使用程序自动裁剪时, 得到的数据集包含大量的空白背景切片, 这些大量的空白数据会影响算法的学习效果, 所以本文人工地选择原始图像上包含内容的区域并将其裁剪为切片, 如图 4 所示. 将每幅作品处理为切片图像后, 需要进行训练集和测试集的划分. 本文实验中训练集有 1 个而测试集有 2 个, 包括测试集 I 和测试集 II. 具体划分方式为: 对表 4 中列出的 9 位画家的共计 178 幅作品中的每幅作品的切片大致按照 4:1 的比例划分到训练集和测试集 I, 最终训练集中包含 10031 张切片, 测试集 I 中包含 2516 张切片. 对表 5 中列出的 2 位画家的共计 19 幅作品中的每幅作品的所有切片划分到测试集 II, 最终测试集 II 包含 1634 张切片. 测试集 I 中的切片对于模型而言是未观测过的新鲜样本, 但是这些切片与训练集中的切片是来自相同分布(同幅画作)的采样, 所以模型对于测试集 I 中的切片具有较强的泛化能力. 测试集 I 的划分方式是合理的, 因为在实际情况中, 古画的尺寸通常很大, 画面上的受损区域在整幅画卷中所占的比例是较小的, 所以在实际应用时, 可以先对待修复的古画划分切片, 用其中内容完整的切片对模型进行微调(fine-tuning)训练, 待模型学习到古画的分布后, 再用模型修复那些受损的切片.

此外, 本文算法的模型是在来自不同画家、不同类别的作品切片上训练的, 所以也能够学习到古画领域中一些通用的信息, 模型对于未见过的作品的切片也具有一定的泛化能力. 本节通过测试集 II 来验证模型对于未观测过的作品切片的泛化能力.

### 3.2 实验流程

实验分为训练和测试 2 个阶段. 在训练阶段, 先生成带有随机不规则形状孔洞的二值单通道掩膜  $M$ ,  $M$  的尺寸为  $256 \times 256$ , 与训练集中图像的尺寸相同.  $M$  中每个位置的值为 0 或 1, 0 代表完整像素, 1 代表缺失像素. 接着通过掩膜  $M$  和训练集中的真实图像  $\mathbf{y}$  制作缺失图像  $\mathbf{x} = \mathbf{y} \odot (1 - M)$ , 并将  $\mathbf{x}$  和  $M$  在通道维度拼接起来, 形成  $[\mathbf{x}, M]$  作为生成器  $G$  的输入. 然后将生

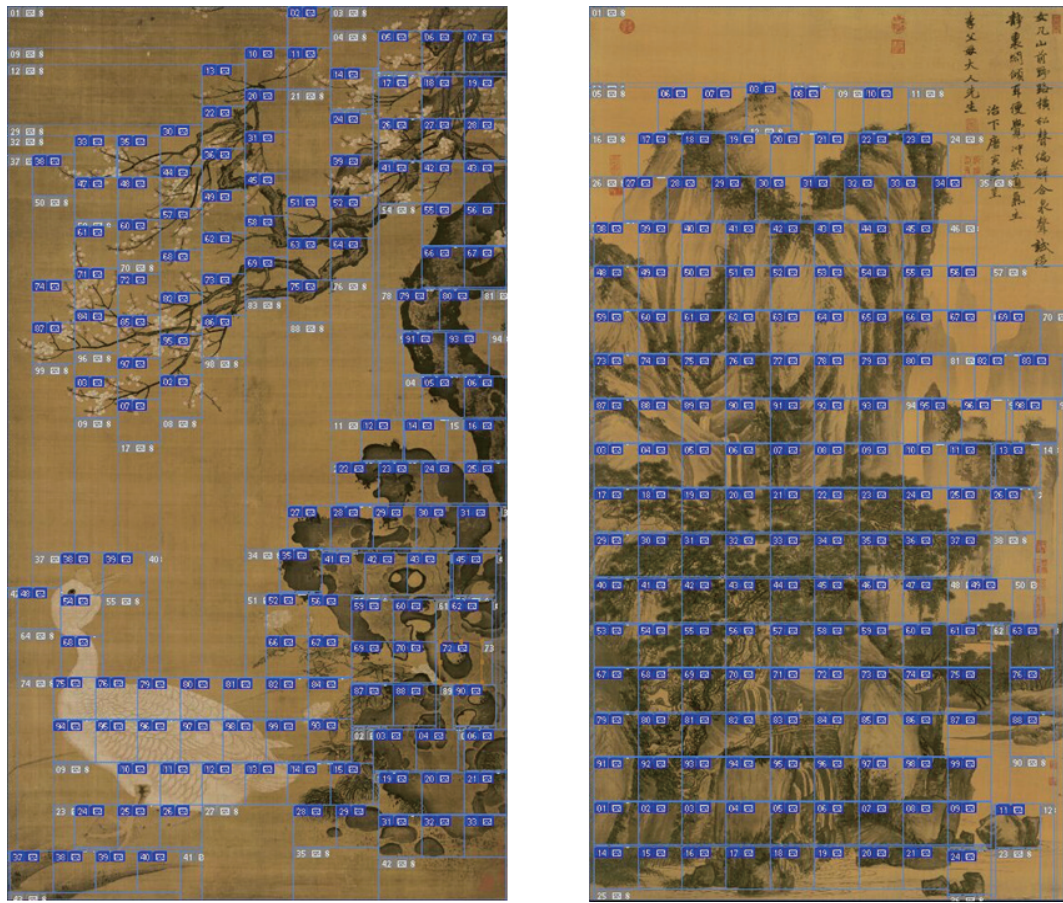


Fig. 4 Schematic diagram of ancient painting slices

图4 古画切片示意图

Table 4 Data Distribution in Training Set and Test Set I

表4 训练集与测试集 I 中的数据分布

编号	画家	作品数	训练集切片数	测试集 I 切片数
1	陈淳	7	684	171
2	陈洪绶	17	904	226
3	仇英	22	1 449	362
4	董其昌	17	892	223
5	蓝瑛	19	1 024	265
6	吕纪	9	964	241
7	沈周	30	1 166	291
8	唐寅	32	1 372	343
9	文徵明	25	1 576	394
总计		178	10 031	2 516

Table 5 Data Distribution in Test Set II

表5 测试集 II 中的数据分布

序号	画家	作品数	测试集 II 切片数
1	戴进	11	645
2	陆治	8	989
总计		19	1 634

成器的输出 $G([x, M])$ 与输入图像中的完整像素结合起来,得到修复图像 $\hat{y} = G([x, M]) \odot M + x \odot (1 - M)$ .判别器 $D$ 依次接收修复图像 $\hat{y}$ 和真实图像 $y$ 并分别给出评分.生成器和判别器交替进行训练,判别器的目标是区分两者,而生成器的目标是制造出能够迷惑判别器的修复结果.实验中使用Adam优化器对生成器和判别器的参数进行优化,学习率为0.000 1,2个Momentum参数分别为0.5和0.9. Batch Size设置为1,损失函数中的系数 $\lambda_{\text{gp}}, \lambda_{\text{adv}}, \lambda_{\text{rec}}, \lambda_{\text{style}}$ 的值分别设置为10, 0.002, 1, 10.在测试阶段,同样先生成形状随机的掩膜 $M$ ,接着通过掩膜和测试集中的真实图像 $y$ 制作缺失图像 $x = y \odot (1 - M)$ ,然后将 $[x, M]$ 作为生成器的输入,经过生成器的处理后得到修复图像 $\hat{y}$ .测试阶段不需要判别器的参与,也不计算损失函数.

## 4 实验结果与分析

在本节中,按照第3节所述流程进行实验,然后展示模型在测试集 I 和测试集 II 上的实验结果.为了更加全面地对本文算法的效果进行评估,本节还选



择了近年来图像修复领域较杰出的 7 种算法,与本文算法的实验结果进行定性和定量的对比,并通过消融实验验证了空间-风格注意力模块的有效性.

#### 4.1 实验结果

本文实验使用了 2 种不规则形状的孔洞来模拟古画上的损害,测试阶段在 NVIDIA GeForce RTX 2 080 GPU 上平均修复每张测试图像需要 105 ms. 图 5 展示了本文算法在测试集 I 上的实验结果,其中每一行展示 2 组结果,每组包含 3 张图像,从左至右依次为:原始图像、残缺图像和修复图像.从内容的修复效果上看,本文算法对山水画图像的修复效果最好,因为山水画中有很多相似的局部结构和内容,在修复受损区域时有比较丰富的信息可以借鉴,修复结果和真

实图像比较接近.对于花鸟画图像,本文算法也能较好地修复图像的内容,但人物画图像的修复效果相对逊色,未能较好地修复人物的五官,主要原因可能在于人物画面内相似的内容相对较少,缺乏可以借鉴的信息.从纹理的修复效果上看,修复区域的纹理和周围很接近,削弱了修复区域与完整区域之间的视觉差异,没有留下明显的修复痕迹.从风格保持的效果上看,修复图像整体上维持了画作原有的风格.

为了进一步评估模型对于未见过的作品的泛化能力,图 6 中展示了模型在测试集 II 上的实验结果.从实验结果可以看出,模型对于测试集 II 中的图像仍具有较好的泛化能力,基本能够修复缺失的内容.

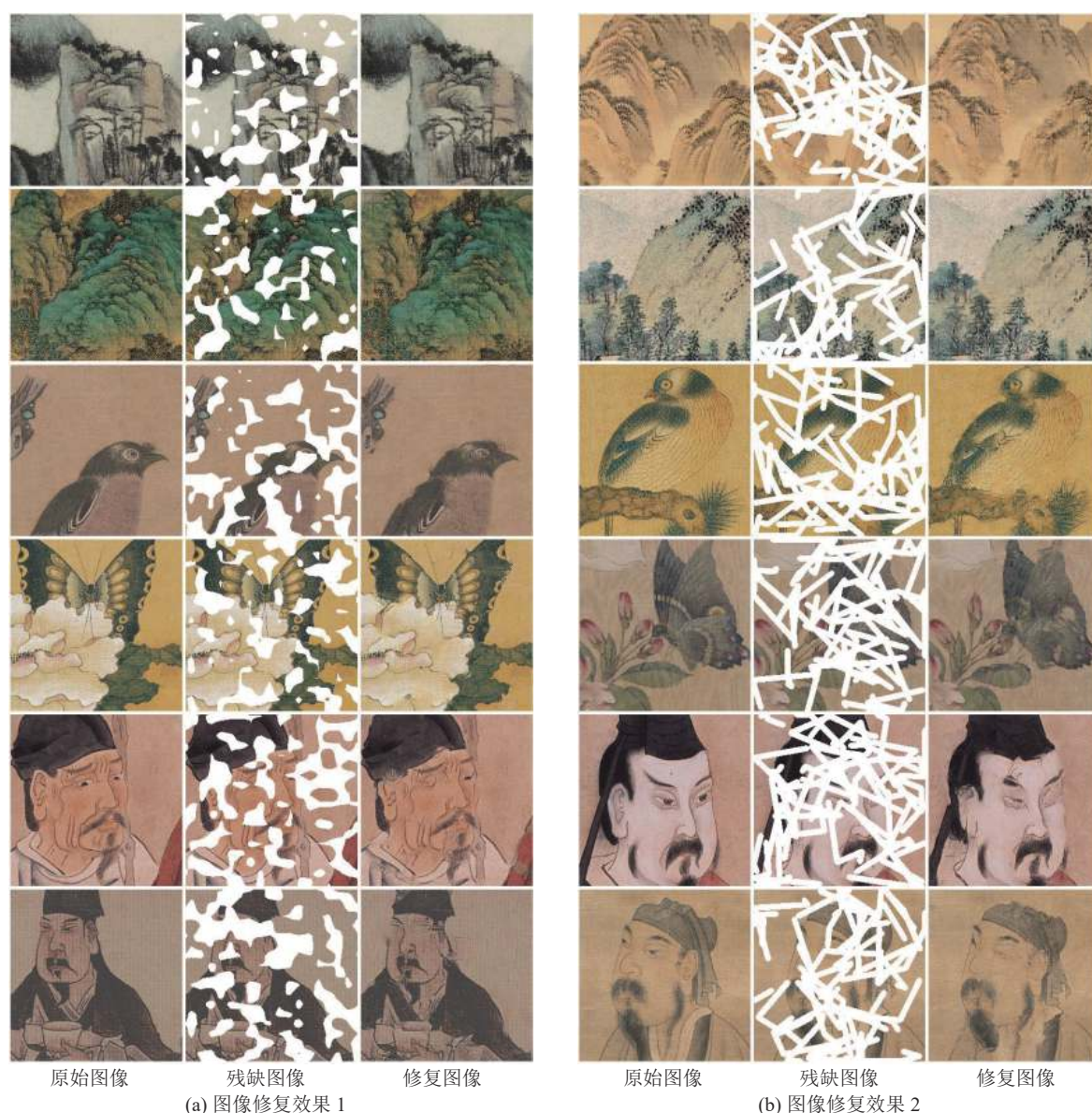


Fig. 5 Experimental results of our algorithm on test set I

图 5 本文算法在测试集 I 上的实验结果



同在测试集 I 上的情况一样,算法在山水画图像上的修复效果最好,其次是花鸟画,对人物画中人物面部的修复效果最差.

但是也存在一些问题,比如少数结果中一些断裂的结构没有被连接,如图 6(a)中的第 2 行花卉的枝干;少数结果中修复区域的色彩纹理与周围不一致,如图 6(a)中的第 3 行花瓣和叶片上留下的修复痕迹.总的来说,模型在测试集 II 上的修复稳定性略

低于测试集 I,但足够体现模型具有较强的泛化能力.此外,本文实验也尝试了对大面积内容缺失的古画图像进行修复,图 7 中展示了本文算法对缺失面积为  $128 \times 128$  的图像的修复结果.可以看到,针对大面积缺失古画的修复,本文算法的修复效果不太理想,主要原因是在本文算法中,当受损区域较小时,模型能够根据周围的信息合理地推测受损区域原本的语义,然后在图像的完整区域寻找和它在语义上相似

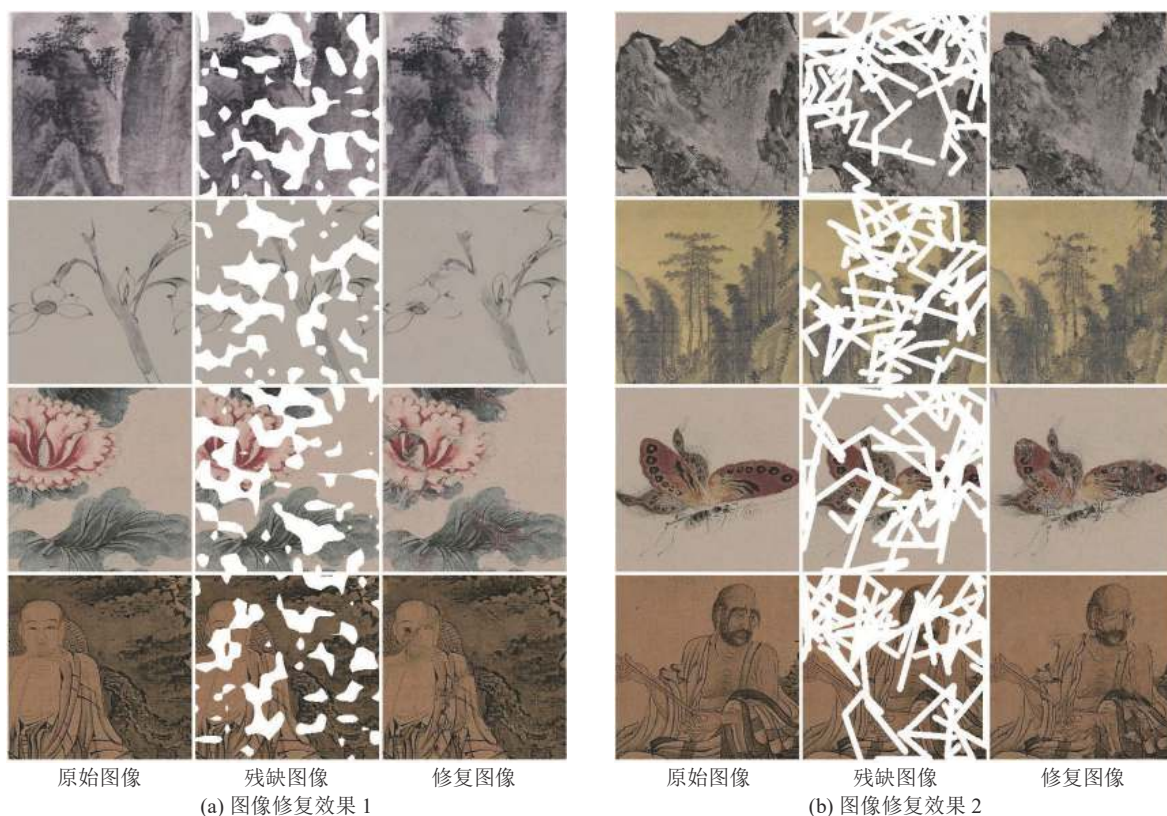


Fig. 6 Experimental results of our algorithm on test set II

图 6 本文算法在测试集 II 上的实验结果



Fig. 7 Restoration results of our algorithm for ancient painting images with large area of missing content

图 7 本文算法对大面积内容缺失的古画图像的修复结果



的位置,利用相似位置的信息进行修复;但是当图像受损面积很大时,孔洞中越靠近内部的像素,周围的信息越稀疏,模型无法准确推测这些像素的语义,导致对缺失像素的修复缺乏可靠的依据.

## 4.2 对比实验

为了更加全面地评估本算法的修复效果,本文实验选取了近年来图像修复领域较杰出的7种算法:GL(global&local)<sup>[50]</sup>,CA(context attention)<sup>[51]</sup>,PC(partial convolution)<sup>[52]</sup>,CPST(coherence priors of semantics and textures)<sup>[39]</sup>,MED(mutual encoder-decoder)<sup>[40]</sup>,EC(edge connect)<sup>[41]</sup>,GC(gated convolution)<sup>[42]</sup>在相同条件下进行实验,并将得到的结果进行对比.图8中展示了不同算法在测试集I上的修复结果,其中每行展示1组对比实验结果,从左至右依次为:原始图像、残缺图像、本文算法的修复结果、GL算法的修复结

果、CA算法的修复结果、PC算法的修复结果.从视觉效果上看,经GL算法和PC算法修复后的局部区域的色彩和纹理与周围不够一致,而经CA算法修复后的区域比较模糊,GL,CA,PC这3种算法均在视觉上造成修复区域与原图像之间的差异,干扰了原画的风格,影响了画面的连续性与美观性.相比之下经本文算法修复的局部区域纹理与周围更相符且更清晰,修复痕迹不明显;从整张图像上看,风格的一致性更高,修复质量更高.

除了从视觉效果上对修复结果进行定性分析,本文实验还选取了2种图像修复领域最常用的评价指标进行定量分析,分别是峰值信噪比(peak signal to noise ratio, *PSNR*)和结构相似性(structure similarity, *SSIM*).*PSNR*是基于原始图像和修复图像对应位置像素间均方误差的图像质量衡量指标,*PSNR*的计算

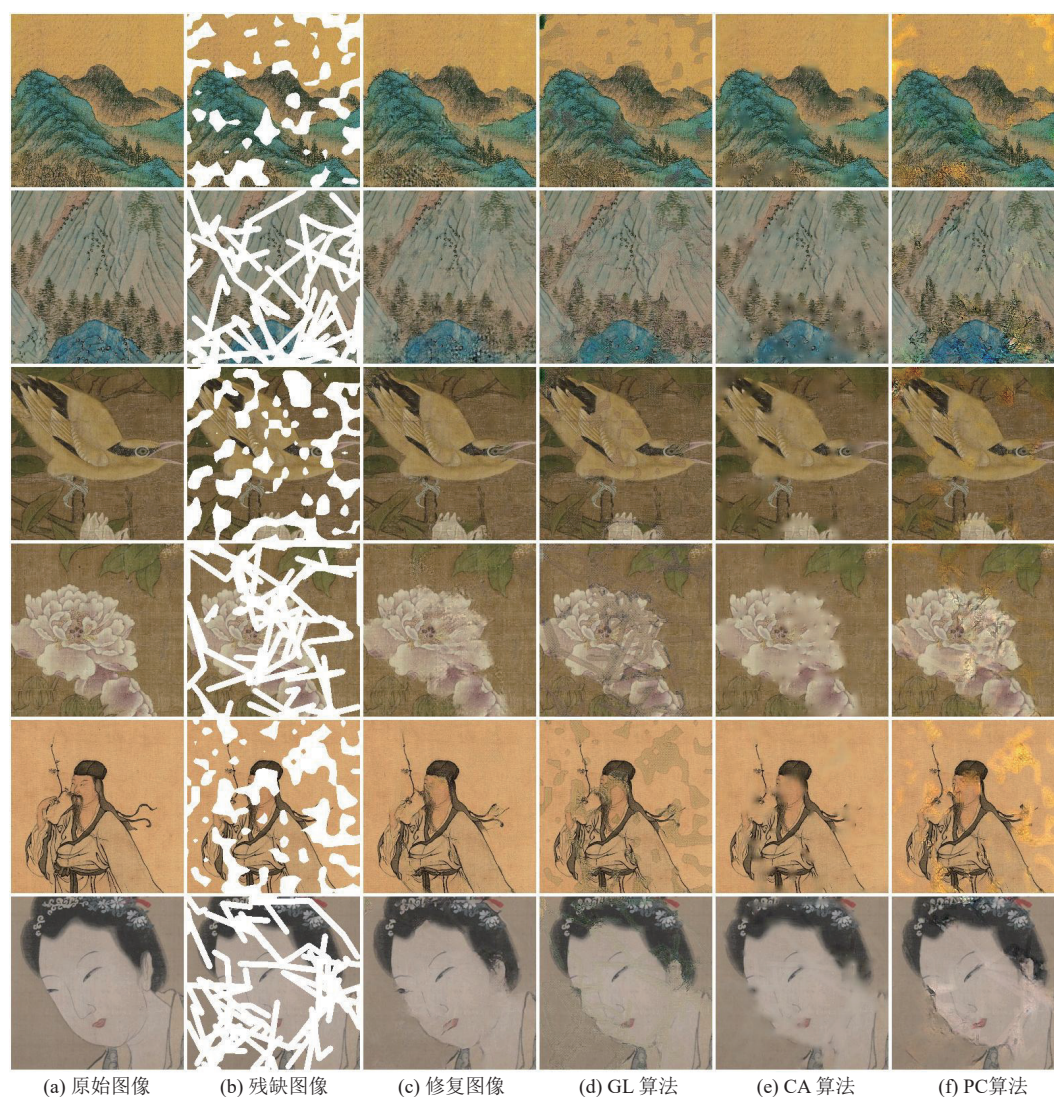


Fig. 8 Comparison of experimental results of different algorithms on test set I

图8 不同算法在测试集I上的实验结果对比

公式为:

$$PSNR(y, \hat{y}) = 10 \times \ln \left( \frac{\max_y}{MSE(y, \hat{y})} \right),$$

$$MSE(y, \hat{y}) = \frac{1}{HW} \sum_i \sum_j (y_{ij} - \hat{y}_{ij})^2, \quad (14)$$

其中, 式(14)中的  $\max$  表示像素可能的最大值, 一般情况下取 255. 因此  $PSNR$  值越大表明修复图像与原始图像的相似度越高.

$SSIM$  是一种从亮度、对比度、结构 3 个角度对修复图像和原始图像的相似程度进行评估的指标, 计算方式为:

$$SSIM(y, \hat{y}) = [l(y, \hat{y})]^\alpha [c(y, \hat{y})]^\beta [s(y, \hat{y})]^\gamma, \quad (15)$$

$$l(y, \hat{y}) = \frac{2\mu_y\mu_{\hat{y}} + c_1}{\mu_y^2 + \mu_{\hat{y}}^2 + c_1}, c(y, \hat{y}) = \frac{2\sigma_{y\hat{y}} + c_2}{\sigma_y^2 + \sigma_{\hat{y}}^2 + c_2}, s(y, \hat{y}) = \frac{\sigma_{y\hat{y}} + c_3}{\sigma_y\sigma_{\hat{y}} + c_3},$$

其中  $l(y, \hat{y})$  是亮度比较,  $c(y, \hat{y})$  是对比度比较,  $s(y, \hat{y})$  是结构比较,  $\mu_y$  和  $\mu_{\hat{y}}$  分别表示  $y$  和  $\hat{y}$  的均值,  $\sigma_y$  和  $\sigma_{\hat{y}}$  分别表示  $y$  和  $\hat{y}$  的标准差,  $\sigma_{y\hat{y}}$  表示  $y$  和  $\hat{y}$  的协方差.  $c_1, c_2, c_3$  是为了避免分母为 0 而加入的常数, 在计算中通常设定  $\alpha = \beta = \gamma = 1$ .  $SSIM$  的值域为  $[-1, 1]$ , 值越大说明

2 幅图像间的差异越小, 即修复效果越好.

本文实验中计算了 4 种算法在测试集 I 上的 2 种评价指标的数值, 整理成表 6. 表 7 整理了测试集 II 在 8 种算法上  $PSNR$  和  $SSIM$  两种评价指标的对比. 其中  $PSNR$  和  $SSIM$  的值越大表示修复质量越高, 表 6~8 中每行加粗的数值表示在该指标上取得的最优值. 从表 6 中可以看出, 本文算法在这 2 个指标上均领先于其他算法. 本文实验同样评估了 8 种算法在测试集 II 上的修复效果, 如图 9 所示. 4 种算法在测试集 II 上的修复情况与先前的对比结果大致相同. 同样, 本文算法在这 2 个指标上均领先. 因此, 在综合考虑了定性和定量分析的结果后, 本文算法在古画图像的修复结果上要优于其他 3 种图像修复算法.

Table 6 Comparison of Evaluation Indicators of Different Algorithms on Test Set I

表 6 不同算法在测试集 I 上的评价指标对比

量化指标	本文算法	GL 算法	CA 算法	PC 算法
$PSNR$	<b>26.762</b>	24.427	26.008	23.589
$SSIM$	<b>0.803</b>	0.733	0.780	0.743

注: 加粗数字表示最优值.

Table 7 Comparison of Evaluation Indicators of Different Algorithms on Test Set II

表 7 不同算法在测试集 II 上的评价指标对比

量化指标	本文算法	GL 算法	CA 算法	PC 算法	GC 算法	MED 算法	EC 算法	CPST 算法
$PSNR$	<b>25.707</b>	22.644	24.443	21.307	25.243	24.993	25.603	25.443
$SSIM$	<b>0.747</b>	0.669	0.724	0.663	0.736	0.728	0.732	0.714

注: 加粗数字表示最优值.

### 4.3 消融实验

为了验证空间-风格注意力模块对修复的提升效果, 本节中设计了相应的消融实验, 分别测试了 4 种不同注意力模块的构成方式: 不使用任何注意力模块、仅使用空间注意力模块、仅使用风格注意力模块、使用空间-风格注意力模块, 并将这 4 种情况下的模型在测试集 I 上的 2 种评价指标记录在表 8 中. 从实验结果上看, 仅使用空间注意力模块或风格注

Table 8 Comparison of Evaluation Indicators of Different Attention Module Composition Methods on Test Set I

表 8 不同的注意力模块组成方式在测试集 I 上的评价指标对比

量化指标	不使用注意力模块	仅使用空间注意力模块	仅使用风格注意力模块	使用空间-风格注意力模块
$PSNR$	24.469	25.867	25.602	<b>26.762</b>
$SSIM$	0.724	0.775	0.740	<b>0.803</b>

注: 加粗数字表示最优值.

意力模块均对修复质量有提升作用, 而联合使用这 2 个模块能够取得更好的效果. 此外, 选择了 1 组具有代表性的实验结果来更加直观地说明空间-风格注意力对古画修复的提升效果, 如图 10 所示.

为了方便观察, 图 10 中用方形框标出了存在明显区别的局部. 可以看到, 不使用注意力模块的修复结果中缺失的内容并没有得到补全, 如鸟的额头和上方的树枝, 而且补全区域的纹理和周围也不一致, 修复质量较差; 仅使用空间注意力模块时, 缺失

Table 9 Comparison of Evaluation Indicators of Different Network Architectures on Test Set I

表 9 不同的网络架构在测试集 I 上的评价指标对比

量化指标	仅使用第 1 个分支	仅使用第 2 个分支	仅使用第 3 个分支	使用全部分支
$PSNR$	25.269	25.367	25.502	<b>26.762</b>
$SSIM$	0.674	0.695	0.720	<b>0.803</b>

注: 加粗数字表示最优值.



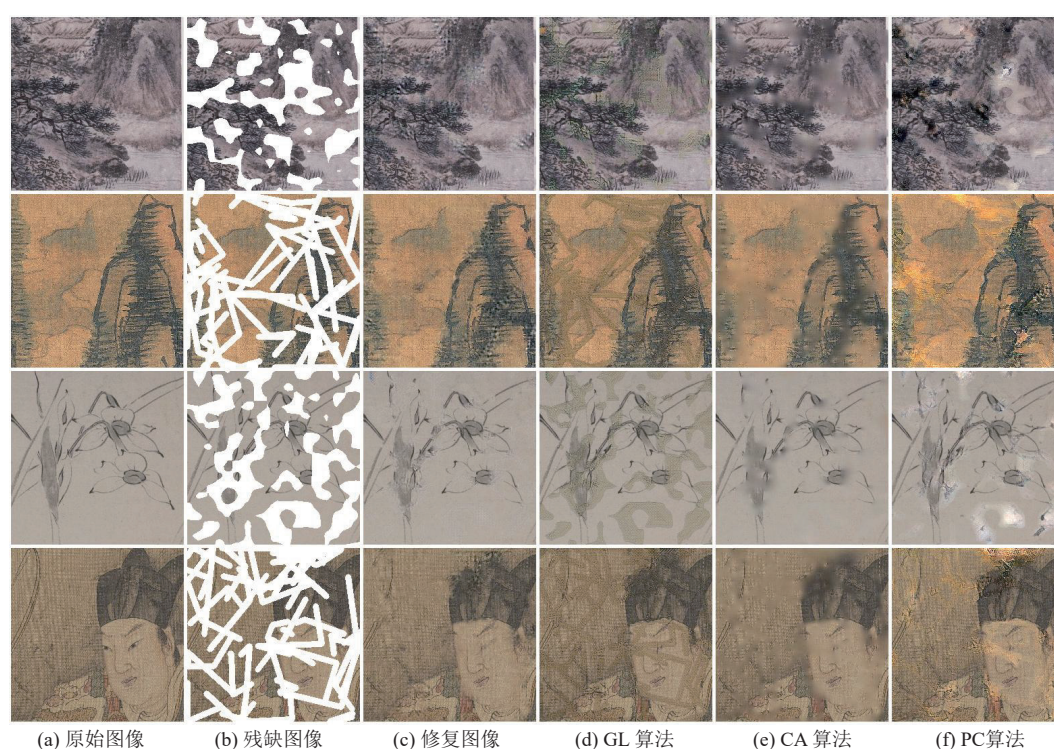


Fig. 9 Comparison of experimental results of different algorithms on test set II

图9 不同算法在测试集 II 上的实验结果对比

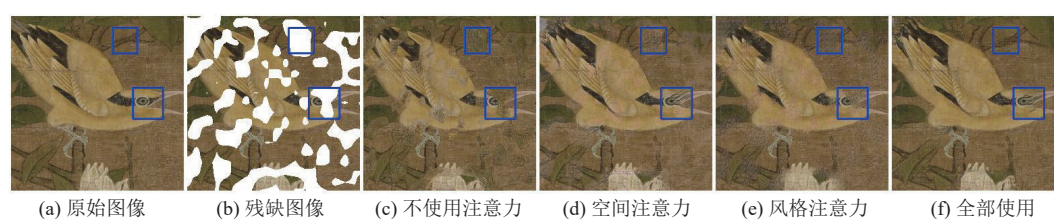


Fig. 10 Comparison of repair effects of different attention module composition methods on test set I

图10 不同的注意力模块组成方式在测试集 I 上的修复效果对比

的内容得到了补全,但是树枝附近的纹理仍与周围不一致,使修复区域显得比较突兀;仅使用风格注意力模块时,修复区域的纹理与周围比较接近,修复痕迹较轻,但是缺失的内容没有得到补全,如断开的树枝没有被连接;最后,使用空间-风格注意力模块的修复结果中,内容和纹理都得到了较好地修复,修复质量相对较高。

## 5 总结

针对古画图像的自相似性高、具有明显的风格、丰富和细腻的纹理等特点,提出了基于多路编码器和双重注意力机制的修复算法。该算法在网络架构上能够学习不同尺度的语义信息,同时利用注意力机制来学习全局结构信息。不同尺度的局部信息加上全部语义信息的学习,使得网络模型不仅仅支持

细腻的细节修复,同时也能够在宏观内容结构上也能够保持和谐一致。另外为了更好地修复色彩和纹理等风格信息,也利用了风格注意力机制来学习待修复古画图像的风格特点,有利于修复后的结果图像在风格上和待修复古画图像保持一致。未来希望利用古画图像自身隐含的各种信息来提高修复的质量,也就是利用自监督学习来学习古画领域的特征表达,通过有效的语义特征表达来进一步提高修复的质量。

**作者贡献声明:** 赵磊提出论文思路、整体架构设计;吉柏言负责整个模型实现与代码编写;邢卫负责论文整体思路的讨论提升;林怀忠负责论文的修改和提升,并对论文架构做了整体的调整与优化;林志洁负责引言部分的架构设计。



## 参 考 文 献

- [1] Nguyen A, Dosovitskiy A, Yosinski J, et al. Synthesizing the preferred inputs for neurons in neural networks via deep generator networks [C] // Proc of the 30th Advances in Neural Information Processing Systems (NIPS). La Jolla, CA: NIPS, 2017: 11–19
- [2] Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position[J]. *Biological Cybernetics*, 1980, 36(1): 189–196
- [3] LeCun Y, Boser B, Denker J, et al. Backpropagation applied to handwritten zip code recognition[J]. *Neural Computation*, 1989, 1(1): 146–156
- [4] Pathak D, Krahenbuhl P, Donahue J, et al. Context encoders: Feature learning by inpainting [C] //Proc of the 31st IEEE Conf on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2016: 1033–1038
- [5] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks [C] //Proc of the 29th IEEE Conf on Computer Vision and Pattern Recognition (CVPR). Los Alamitos, CA: IEEE Computer Society, 2014: 2672–2680
- [6] Efros A, Leung T. Texture synthesis by non-parametric sampling [C] //Proc of the 7th IEEE Int Conf on Computer Vision (ICCV). Piscataway, NJ: IEEE, 1999: 1033–1038
- [7] Bertalmio M, Sapiro G, Caselles V et al. Image inpainting [C] // Proc of the 8th ACM Int Conf on Multimedia (ACM MM). New York: ACM, 2000: 417–424
- [8] Telea A. An image inpainting technique based on the fast marching method[J]. *Journals of Graphics Tools*, 2004, 9(1): 25–36
- [9] Chan T, Shen Jiahong. Non texture inpainting by curvature-driven diffusion[J]. *Journal of Visual Communication and Image Representation*, 2001, 12(4): 436–449
- [10] Chan T, Kang S, Shen Jiahong. Euler's Elastica and curvature-based inpainting[J]. *SIAM Journal on Applied Mathematics*, 2003, 63(2): 564–592
- [11] Tsai A, Yezzi A, Willsky A. Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification[J]. *IEEE Transactions Image Processing*, 2001, 10(8): 1169–1186
- [12] Shen Jiahong, Kang S, Chan T. Euler's Elastica and curvature based inpainting[J]. *SIAM Journal on Applied Mathematics*, 2002, 63(2): 564–592
- [13] Esedoglu S, Shen Jiahong. Digital inpainting based on the Mumford-Shah-Euler image model[J]. *European Journal of Applied Mathematics*, 2002, 13(4): 353–370
- [14] Efros A, Leung T. Texture synthesis by nonparametric sampling [C] // Proc of the 7th IEEE Int Conf on Computer Vision (ICCV). Piscataway, NJ: IEEE, 1999: 1033–1038
- [15] Efros A, Freeman T. Image quilting for texture synthesis and transfer [C] //Proc of the 8th Int Conf of ACM SIGGRAPH. New York: ACM, 2001: 341–346
- [16] Bertalmio M, Vese L, Sapiro G. Simultaneous structure and texture image inpainting[J]. *IEEE Transactions on Image Processing*, 2003, 12(8): 882–889
- [17] Bertalmio M, Vese L, Sapiro G et al. Image filling in a decomposition space [C] //Proc of the 10th Int Conf on Image Proceeding. Los Alamitos, CA: IEEE Computer Society, 2003: 852–855
- [18] Drori I, Cohen D, Yeshurun H. Fragment-based image completion[J]. *ACM Transaction on Graphics*, 2003, 22(3): 303–312
- [19] 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
- [20] Barnes C, Shechtman E, Goldman D. PatchMatch: A randomized correspondence algorithm for structural image editing[J]. *ACM Transaction on Graphics*, 2009, 28(3): 2401–24011
- [21] Zeng Y, Yi G, Jin Z. Feature learning and patch matching for diverse image inpainting[J]. *Pattern Recognition*, 2021, 119(2021): 108036
- [22] Stark J, Elad M, Donoho D. Image decomposition via the combination of sparse representations and variational approach[J]. *IEEE Transactions Image Processing*, 2005, 14(10): 1570–1582
- [23] Elad M, Stark J, Querre P, et al. Simultaneous cartoon and texture image inpainting using morphological component analysis[J]. *Applied & Computational Harmonic Analysis*, 2005, 19(3): 340–358
- [24] Mairal J, Elad M, Sapiro G. Sparse representation for color image restoration[J]. *IEEE Transactions on Image Processing*, 2008, 17(1): 53–69
- [25] Mairal J, Bach F, Ponce J, et al. Online learning for matrix factorization and sparse coding[J]. *Journal of Machine Learning Research*, 2010, 11(1): 19–60
- [26] Shen Bin, Hu Wei, Zhang Yimin, et al. Image inpainting via sparse representation [C] //Proc of the 7th IEEE Int Conf on Acoustics, Speech and Signal Processing (ICASSP). Piscataway, NJ: IEEE, 2009: 697–700
- [27] Li Xin. Image recovery via hybrid sparse representations: A deterministic annealing approach[J]. *IEEE Journal of Selected Topics Signal Processing*, 2011, 5(5): 953–962
- [28] Gabriel P. Sparse modeling of textures[J]. *Journal of Mathematical Imaging & Vision*, 2009, 34(1): 17–31
- [29] Zhang Jian, Zhao Debin, Xiong Ruiqin, et al. Image restoration using joint statistical modeling in space-transform domain[J]. *IEEE Transactions on Circuits System and Video Technology*, 2014, 24(6): 915–928
- [30] Zhang Jian, Zhao Debin, Gao Wen. Group-based sparse representation for image restoration[J]. *IEEE Transactions on Image Processing*, 2014, 23(8): 3336–3351
- [31] Zhang Lin, Kang Bin, Liu Bin, et al. A new inpainting method for object removal based on patch local feature and sparse representation[J]. *International Journal of Innovative Computing Information and Control*, 2016, 12(1): 113–124
- [32] Goodfellow I, Pouget J, Mirza M, et al. Generative adversarial nets [C] //Proc of the 27th Advances in Neural Information Processing Systems (NIPS). La Jolla, CA: NIPS, 2014: 2672–2680
- [33] Yang Chao, Lu Xin, Lin Zhe, et al. High-resolution image inpainting using multi-scale neural patch synthesis [C] //Proc of the 11th IEEE Int Conf on Computer Vision (ICCV). Piscataway, NJ: IEEE, 2017: 1033–1038
- [34] Zeng Yanhong, Fu, Jialong, Chao Hongyang, et al. Learning pyramid

- context encoder network for high-quality image inpainting [C] // Proc of the 32nd IEEE Conf on Computer Vision and Pattern Recognition (CVPR). Los Alamitos, CA: IEEE Computer Society, 2019: 1486–1494
- [35] Ren Yurui, Yu Xiaoming, Zhang Ruonan, et al. Structureflow: Image inpainting via structure-aware appearance flow [C] // Proc of the 13th IEEE Int Conf on Computer Vision (ICCV). Piscataway, NJ: IEEE, 2019: 181–190
- [36] Xiong Wei, Yu Jiahui, Lin Zhe, et al. Foreground-aware image inpainting [C] // Proc of the 32nd IEEE Conf on Computer Vision and Pattern Recognition (CVPR). Los Alamitos, CA: IEEE Computer Society, 2019: 5840–5848
- [37] Zheng Chuanxia, Cham T, Cai Jiafei. Pluralistic image completion [C] // Proc of the 32nd IEEE Conf on Computer Vision and Pattern Recognition (CVPR). Los Alamitos, CA: IEEE Computer Society, 2019: 1438–1447
- [38] Zhao Lei, Mo Qihang, Lin Sihuan, et al. UCTGAN: Diverse image inpainting based on unsupervised cross-space translation [C] // Proc of the 33rd IEEE Conf on Computer Vision and Pattern Recognition (CVPR). Los Alamitos, CA: IEEE Computer Society, 2020: 5741–5750
- [39] Liao Liang, Xiao Jing, Wang Zheng, et al. Image inpainting guided by coherence priors of semantics and textures [C] // Proc of the 34th IEEE Conf on Computer Vision and Pattern Recognition (CVPR). Los Alamitos, CA: IEEE Computer Society, 2021: 6539–6548
- [40] Liu Hongyu, Jiang Bin, Song Yibing, et al. Rethinking image inpainting via a mutual encoder decoder with feature equalizations [C] // Proc of the 18th European Conf on Computer Vision (ECCV). Berlin: Springer, 2020: 725–741
- [41] Nazeri K, Ng E, Joseph T, et al. EdgeConnect: Generative image inpainting with adversarial edge learning [C] // Proc of the 11th IEEE Int Conf on Computer Vision (ICCV). Piscataway, NJ: IEEE, 2019: 472–480
- [42] Yu Jiahui, Lin Zhe, Yang Jimei, et al. Free-form image inpainting with gated convolution [C] // Proc of the 11th IEEE Int Conf on Computer Vision (ICCV). Piscataway, NJ: IEEE, 2019: 4471–4480
- [43] Yu Fang, Koltun V, Funkhouser T. Dilated residual networks [C] // Proc of the 31st IEEE Conf on Computer Vision and Pattern Recognition (CVPR). Los Alamitos, CA: IEEE Computer Society, 2017: 472–480
- [44] Clevert D, Unterthiner T, Hochreiter S. Fast and accurate deep network learning by exponential linear units (Elus) [J]. arXiv preprint, arXiv: 1511.07289, 2015
- [45] Zhang Han, Goodfellow I, Metaxas D, et al. Self-attention generative adversarial networks [C] // Proc of the 36th Int Conf on Machine Learning (ICML). New York: ACM, 2019: 7354–7363
- [46] Gulrajani I, Ahmed F, Arjovsky M, et al. Improved training of wasserstein GANs [C] // Proc of the 30th Advances in Neural Information Processing Systems (NIPS). La Jolla, CA: NIPS, 2017: 5767–5777
- [47] Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks [C] // Proc of the 30th Advances in Neural Information Processing Systems (NIPS). La Jolla, CA: NIPS, 2017: 214–223
- [48] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint, arXiv: 1409.1556, 2014
- [49] Feng Shun. On the art of "leaving blank" in Chinese painting [J]. Grand Stage, 2013(6): 90–91 (in Chinese)  
(封顺. 论中国画中的“留白”艺术 [J]. 大舞台, 2013(6): 90–91)
- [50] Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks [C] // Proc of the 34th Int Conf on Machine Learning (ICML). New York: ACM, 2017: 214–223
- [51] Yu Jiahui, Lin Zhe, Yang Jimeiecc, et al. Generative image inpainting with contextual attention [C] // Proc of the 31st IEEE Conf on Computer Vision and Pattern Recognition (CVPR). Los Alamitos, CA: IEEE Computer Society, 2018: 5505–5514
- [52] Liu Guilin, Reda F A, Shih K J, et al. Image inpainting for irregular holes using partial convolutions [C] // Proc of the 16th European Conf on Computer Vision (ECCV). Berlin: Springer, 2018: 85–100



**Zhao Lei**, born in 1975. PhD, associate professor. Member of CCF. His main research interests include image restoration, deep learning, and intelligent image generation.

赵磊, 1975年生. 博士, 副研究员. CCF会员. 主要研究方向为图像修复、深度学习、图像智能生成.



**Ji Boyan**, born in 2000. Master candidate. His main research interest includes image cross domain translation. (qinglanwuji@zju.edu.cn)

吉柏言, 2000年生. 硕士研究生. 主要研究方向为图像跨域迁移.



**Xing Wei**, born in 1967. PhD, associate professor. His main research interest includes image intelligent processing.

邢卫, 1967年生. 博士, 副教授. 主要研究方向为图像智能处理.



**Lin Huaizhong**, born in 1970. PhD, associate professor. His main research interest includes image intelligent processing.

林怀忠, 1970年生. 博士, 副教授. 主要研究方向为图像智能处理.



**Lin Zhijie**, born in 1980. PhD, associate professor. His main research interest includes image intelligent processing. (bytelin@qq.com)

林志洁, 1980年生. 博士, 副教授. 主要研究方向为图像智能处理.