doi：10. 12068/j. issn. 1005－3026. 2025. 20250072

基于双阶段深度学习的图像修复方法

王天琪[1]，张 迪[2]，张鑫宇[2]，张一鸣[3]
（1. 东北大学 档案馆，辽宁 沈阳 110819；2. 东北大学 信息科学与工程学院，辽宁 沈阳 110819；
3. 东北大学 机械工程与自动化学院，辽宁 沈阳 110819）

摘 要：针对现有图像修复方法在处理复杂损伤和依赖成对训练样本方面的不足，提出一种端到端的双阶段自监督图像修复方法．该方法包括退化模拟、边缘修复和色彩重建3个环节，并通过协同优化实现结构与色彩的协同还原．实验在档案馆历史图像集上进行，采用峰值信噪比(peak signal-to-noise ratio,PSNR)、感知相似度(learned perceptual image patch similarity,LPIPS)、弗雷歇起始距离(Fréchet inception distance,FID)和色彩丰富度得分(Colorfulness Score)4项指标进行评估．结果表明，该方法在图像还原精度、感知质量和色彩表现方面均优于主流方法，具备良好的实用价值和鲁棒性．

关 键 词：图像修复；自监督学习；图像退化模拟；协同优化；属性解耦；深度学习
中图分类号：TP 391 **文献标志码：**A **文章编号：**1005-3026(2025)11-0048-10

Dual-Stage Deep Learning-Based Image Inpainting Method

WANG Tian-qi[1], *ZHANG Di*[2], *ZHANG Xin-yu*[2], *ZHANG Yi-ming*[3]
(1. Archives, Northeastern University, Shenyang 110819, China; 2. School of Information Science & Engineering, Northeastern University, Shenyang 110819, China; 3. School of Mechanical Engineering & Automation, Northeastern University, Shenyang 110819, China. Corresponding author: ZHANG Di, E-mail: 531513256@qq.com)

Abstract: In view of the shortcomings of existing image inpainting methods in handling complex damage and relying on paired training samples, an end-to-end dual-stage self-supervised image inpainting method was proposed. The method included degradation simulation, edge restoration, and color reconstruction; synergistic restoration of structure and color was achieved through cooperative optimization. Experiments were conducted on a historical image dataset from archives, and evaluation was performed using four metrics: peak signal-to-noise ratio (PSNR), learned perceptual image patch similarity(LPIPS), Fréchet inception distarce(FID)and Colorfulness Scare. Experimental results demonstrate that the proposed method outperforms existing mainstream methods in terms of image restoration accuracy, perceptual quality, and color performance, exhibiting good practical value and robustness.

Key words: image inpainting; self-supervised learning; image degradation simulation; cooperative optimization; attribute disentanglement; deep learning

作为历史的“底片”和时代的“见证者”，大量珍贵的历史图像资料得以在档案馆保存和传承．然而，由于拍摄工艺和胶片老化等原因，这些档案历史图像常常伴随边缘模糊、细节缺失、划痕噪声等问题，严重影响展示效果和历史价值的再现．手工修复和上色不仅耗时费力，而且效果因人而异，难以满足档案馆大规模、批量化、数字化修复的需求．因此，图像修复技术在文化保护、艺术修复、数字内容生成等领域具有重要应用价值．传统的图像修复方法主要依赖于手工处理和基于规则的算法，如图像插值、滤波与频域分析等技术．这些方法在某些特定情况下能够取得一定的效果，但在处理复杂的退化模式和高层次语义信息时，往往效果不佳[1-2]．随着深度学习技术

收稿日期：2025-06-24
基金项目：辽宁省档案局科技项目(2024-B-17)；辽宁省教育科学规划重点项目(JG25DA009)．
作者简介：王天琪(1991—)，女，辽宁沈阳人，东北大学档案馆中级馆员．

的迅猛发展,研究者们开始探索利用深度神经网络,尤其是卷积神经网络(convolutional neural network, CNN)[3]和生成对抗网络(generative adversarial network, GAN)[4]等模型,来自动学习图像中的复杂特征,实现更为精确和高效的图像修复.

当前图像修复技术的研究主要集中在修复流程的多阶段设计、生成对抗网络与注意力机制的引入、Transformer[5]等新型网络架构的应用,以及多尺度特征的融合方法上.多阶段修复策略通过由粗到细的方式逐步生成缺失信息,如由粗到细的内容一致性图像修复方法,显著提升了修复结果的全局一致性和细节表现[6].Tai等[7]提出了一种持久记忆网络,通过引入记忆模块,有效捕捉了图像中的长期依赖信息,提升了修复质量.注意力机制与生成对抗网络的结合增强了特征提取和细节重建能力,例如周啟雪等[8]利用Involution级联注意力机制的古代壁画图像修复网络,有效修复了具有复杂结构和色彩变化的壁画内容.Transformer架构的应用为解决高频内容质量和非局部关系建模问题提供了新思路,如融合双向感知Transformer与频率分析策略的图像修复网络,显著提升了修复结果的语义一致性和细节表现[9].Liang等[10]利用Swin Transformer结构,在图像超分辨率、去噪和压缩伪影去除等任务中取得了优异的性能.基于生成对抗网络的图像修复方法[11-13]在训练时通常使用大量随机掩码,以对抗方式优化生成器,从而增强对上下文纹理的利用并保持结构一致性.但这类方法往往难以准确理解掩蔽区域的语义信息,生成结果在语义层面可能存在不合理之处.Gaa等[14]将扩散模型应用于图像修复任务,有效提高了生成图像的质量,但扩散模型通常需要数百到上千步的迭代采样过程,这导致推理速度较慢.

图像上色作为历史图像修复的重要组成部分[15-16],也得到了广泛关注.研究者们提出了多种基于深度学习的上色方法:DeOldify[17]通过自注意力机制实现语义感知色彩预测,但在历史场景中易出现时代错位(如将50年代中山装错误上色为现代西装颜色);DeepFill[18]采用门控卷积生成合理内容,但对结构性破损(如大面积面部缺失)处理效果不稳定.有国外研究团队提出了一种文本引导双重注意力修复网络(text-guided dual attention inpainting network, TDA-Net)[19],该方法通过双模态注意力机制从描述性文本中提取缺失区域的语义特征,并引入图像-文本匹配损失以提高修复结果与文本的一致性.此外,还有多模态特征融合方法(multimodal fusion learning, MMFL)[20]通过构建图像自适应词需求模块,合理过滤有效文本特征,使生成图像具有更精细的纹理.一些国内研究团队还提出了基于自适应特征融合与U-Net的双重退化网络(dual degradation network via adaptive feature fusion and U-Net, AFFU)[21],该网络使用单一网络结构同时解决图像降质问题,利用自引导模块融合多尺度图像信息,有效消除图像中的特定缺陷,并通过自适应多特征融合模块和信息转移机制链接这两个主要结构,自适应地选择和保留图像特征,防止有用信息的丢失.

在此背景下,本文提出了基于双阶段自监督框架的历史图像复原与上色算法(dual-stage self-supervised framework based historical image restoration and colorization algorithm, DSS-HIRC).该算法采用双阶段自监督学习框架,结合图像修复和上色任务,旨在实现对历史图像的高质量修复和真实感上色.通过引入自监督机制,能够在缺乏大规模标注数据的情况下,充分挖掘图像中的潜在信息,提升模型的泛化能力和修复效果.本研究旨在为历史图像修复领域提供一种高效、鲁棒的解决方案,通过系统性的修复,确保档案馆中保存的大量珍贵历史图像资料得以永久传承,促进文化遗产的数字化保护和传播.

1　算法架构

1.1　整体网络框架

本研究构建了一个端到端可训练的双阶段自监督图像修复框架,完整流程包括退化模拟、边缘修复和色彩恢复3个环节.通过自监督退化模拟构建训练样本,无需人工标注的破损-完整图像对,实现对高频边缘细节与低频色彩分布的协同重建.如算法1所示,首先,通过可控退化函数$D(\cdot)$在原始RGB(red green blue)或灰度图像上动态生成多种损伤模式,包括边缘断裂(随机剥离一定宽度的像素带)、模糊(可变高斯核卷积)与噪声(泊松噪声或高斯噪声混合),形成伪破损-完整数据对$(D(\boldsymbol{x}),\boldsymbol{x})$,并采用在线生成策略以覆盖更丰富的场景变化,有效避免模型对固定模式的过拟合.随后的训练阶段由两阶段组成.阶段一聚焦于边缘信息恢复:本研究基于经

典U-Net[9]架构引入残差连接和多尺度注意力模块,在网络输入端接收由退化图像输出的破损边缘图,通过卷积与下采样提取边缘特征,再经上采样与跳跃连接逐步还原细节,直至输出完整连贯的二值边缘轮廓.此阶段的自监督信号来自原始图像经Canny[22]算法提取的真实边缘图,采用二元交叉熵损失与结构相似性损失协同优化,使网络更好地聚焦于精细结构的连通性与准确性.阶段二则以色彩恢复为核心:将阶段一的修复边缘图作为输入,采用预训练的视觉Transformer(vision Transformer, ViT)作为全局语义编码器,从而捕捉图像中物体类别、空间布局等宏观信息;随后,融合了ViT特征的改进U-Net在局部结构建模上进一步发挥作用,通过跨层注意力机制将边缘与语义特征进行加权融合,生成初步的RGB重建图.为保证色彩的自然与真实,该阶段的损失函数包括:像素级均方误差(mean-square error, MSE),用于准确重构全局亮度与色彩分布;感知损失(基于VGG(visual geometry group)网络对高层语义特征的匹配)以增强图像的视觉连贯性与真实感;以及对抗损失(通过对局部斑块进行判别式训练)以进一步提升纹理细节与色彩过渡的精细度.两阶段网络在训练过程中采用协同优化策略,边缘修复与色彩恢复模块通过梯度反向传播共享部分特征提取器,从而实现相互补偿与协同增强.

算法1:基于双阶段深度学习的图像修复算法

输入:原始图像(RGB格式)$\boldsymbol{I}\in\mathbf{R}^{h\times w\times 3}$,训练轮次$T$,平衡系数$\lambda_1,\lambda_2$

输出:修复后图像$\boldsymbol{I}_{\text{out}}\in\mathbf{R}^{h\times w\times 3}$

初始化

1.边缘修复网络E_θ

2.色彩恢复网络C_ϕ

3.判别器D_ψ

4.优化器Adam(θ,ϕ,ψ)

5.可控退化函数$D(\cdot)$,边缘检测算子EdgeDetect$(\cdot)$

训练阶段

For t=1 to T do

第一阶段:边缘修复

1.生成自监督样本:

$E_{\text{gt}}\leftarrow$EdgeDetect$(\boldsymbol{I})$//Canny算法提取真实边缘

$E_{\text{deg}}\leftarrow D(E_{\text{gt}})$//退化模拟(掩膜+模糊+噪声)

2.边缘修复网络前向传播:

$E_{\text{pred}}\leftarrow E_\theta(E_{\text{deg}})$//输出概率图

3.计算边缘损失:

$L_{\text{edge}}\leftarrow\text{BCE}(E_{\text{pred}},E_{\text{gt}})+\alpha\text{SSIM}(E_{\text{pred}},E_{\text{gt}})$

第二阶段:色彩恢复

4.提取高维特征:

$E_{\text{vit}}\leftarrow$ViT-B/16$(\boldsymbol{I})[6,9,12]$//冻结预训练模型权重

$\boldsymbol{F}_{\text{vit}}\leftarrow$Interpolate$(\boldsymbol{F}_{\text{vit}})$//特征对齐

5.跨模态特征融合:

$\boldsymbol{F}_{\text{edge}}\leftarrow$ConvBlocks$(E_{\text{pred}})$//边缘特征编码

$\boldsymbol{F}_{\text{fusion}}\leftarrow$CrossAttention$(\boldsymbol{F}_{\text{edge}},\boldsymbol{F}_{\text{vit}})$//特征融合

6.色彩解码器生成:

$I_{\text{pred}}\leftarrow C_\phi(\boldsymbol{F}_{\text{fusion}})$//解码器解码

7.计算色彩损失:

$$\left.\begin{aligned}L_{\text{MSE}}&\leftarrow\text{MSE}(\boldsymbol{I}_{\text{pred}},\boldsymbol{I}_{\text{gt}})\\L_{\text{perc}}&\leftarrow\text{VGG19}(\boldsymbol{I}_{\text{pred}},\boldsymbol{I}_{\text{gt}})\\L_{\text{GAN}}&\leftarrow-\lg(D_\psi(\boldsymbol{I}_{\text{pred}}))\end{aligned}\right\}$$

$L_{\text{color}}\leftarrow L_{\text{MSE}}+\beta L_{\text{perc}}+\gamma L_{\text{GAN}}$

协同优化

8.总损失计算:

$L_{\text{total}}\leftarrow L_{\text{edge}}+\lambda' L_{\text{color}}$

9.参数更新:

更新$\theta,\phi,$Adam$(\nabla L_{\text{total}})$

End For

推理阶段

Function Pipeline$(\boldsymbol{I})$:

1.$E_{\text{deg}}\leftarrow$EdgeDetect$(\boldsymbol{I})+D(E_{\text{deg}})$

2.$E_{\text{pred}}\leftarrow E_0(E_{\text{deg}})$

3.$\boldsymbol{I}_{\text{out}}\leftarrow\text{C}_\phi(\text{Fusion}(E_{\text{pred}},\text{ViT}(\boldsymbol{I})))$

4.Return $\boldsymbol{I}_{\text{out}}$

其中:T为总训练轮数;t为当前训练轮次;EdgeDetect(·)为边缘检测算子表示从真实图像中提取真实边缘;$E_\theta(\cdot)$表示修复退化边缘并输出完整边缘图的过程;E_{gt}和E_{deg}分别是真实的边缘和退化边缘输入;E_{pred}为边缘修复网络输出的预测结果;α为二元交叉熵损失(L_{BCE})的平衡系数,用于平衡L_{BCE}和L_{SSIM}在边缘修复损失(L_{edge})中的重要性;β为感知损失(L_{perc})的平衡系数;γ为对抗损失(L_{GAN})的平衡系数;λ'为色彩恢复损失(L_{color})的平衡系数,用于平衡总损失(L_{total})中两个阶段损失的比重,即边缘修复损失(L_{edge})和色彩恢复损

失(L_{color}).ViT-B/16(·)用于从图像中提取全局语义特征;ConvBlocks(·)用于将边缘图编码为结构特征;CrossAttention(·)用于融合边缘特征与视觉特征;$\boldsymbol{F}_{fusion}$为融合后存储的特征;$C_{\phi}(\cdot)$指将融合特征解码为彩色图像的过程;$\boldsymbol{I}_{pred}$和$\boldsymbol{I}_{gt}$分别为生成的彩色图像以及真实的参考图像;$D_{\psi}(\cdot)$为判别输入图像真伪的判别器;$\boldsymbol{F}_{vit}$代表由预训练的ViT提取的全局语义特征图;$\boldsymbol{F}_{edge}$代表边缘特征(由E_{pred}经ConvBlocks(·)编码得到);L_{color}代表色彩恢复阶段的复合损失函数;L_{MSE}代表像素级均方误差损失(mean-square error loss);L_{GAN}代表对抗损失(generative adversarial network loss);L_{perc}代表感知损失(perceptual loss);L_{edge}代表边缘修复阶段的损失函数,而L_{total}则是总损失函数(边缘损失与色彩损失联合优化).MSE(·)和VGG19(·)分别计算像素级重建误差和感知特征差异.BCE(binary cross entropy,二元交叉熵)用于二分类任务的损失函数,在图像处理中当需要判断每个像素是否属于目标区域时,BCE对每个像素进行独立分类.SSIM(structural similarity index,结构相似性指数)是衡量两幅图像相似度的指标,通过模拟人类视觉系统特性,从亮度、对比度、结构3个维度综合评估图像质量.

模型整体结构如图1所示.

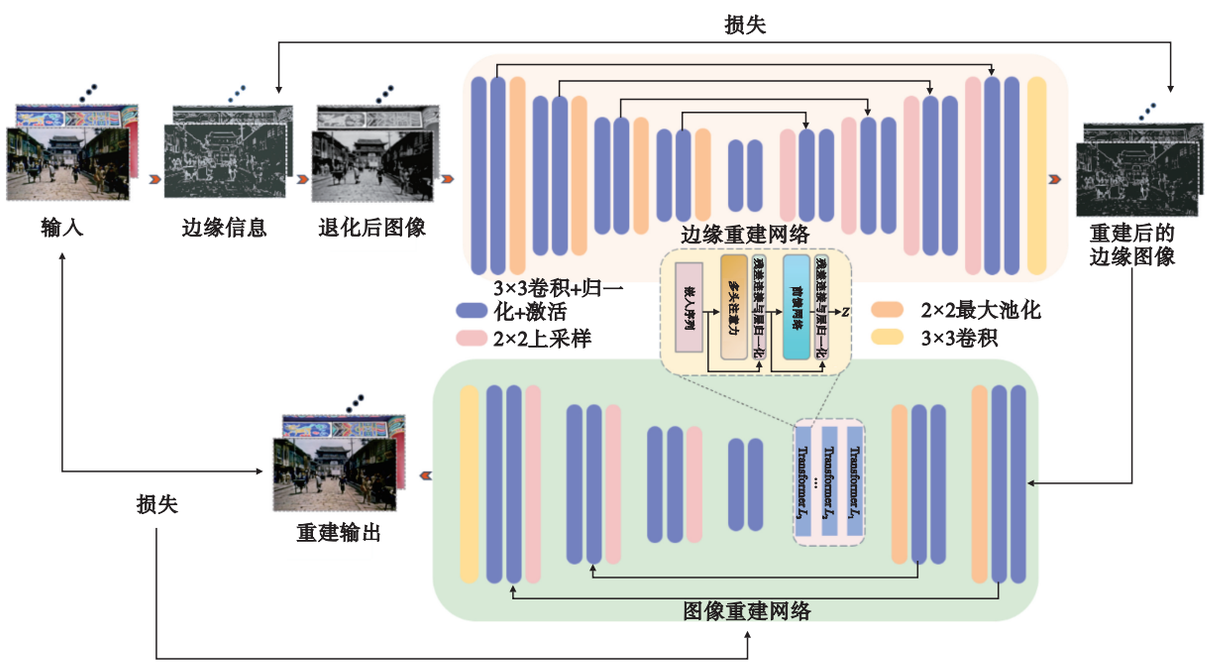

图1　用于历史图像修复与上色的双阶段自监督框架

Fig. 1　Dual-stage self-supervised framework for historical image restoration and colorization

1.2　边缘修复网络结构设置

1.2.1　退化过程模拟模块

本节提出了一种基于可控退化函数的自监督数据生成方案,其核心在于通过3步叠加模拟真实图像在采集或传输过程中可能出现的边缘断裂、模糊失真与噪声干扰,从而在线构建伪破损-完整训练对.具体来说,首先对输入图像$\boldsymbol{x}$施加边缘掩膜破坏,即生成一个二值掩膜,在掩膜覆盖率$p\in[0.1,0.3]$范围内随机引入3种破损形态:线状断裂(随机长度$l\in[5$像素$,20$像素$]$的直线段)、块状缺失(宽度$w\in[10$像素$,30$像素$]$、高度$h\in[10$像素$,30$像素$]$的矩形区域)和利用Perlin噪声生成的不规则孔洞,以模拟自然损伤结构;其次,将掩膜处理后的图像与均值为零、方差可控的高斯噪声$N(0,\delta_n)$(δ_n为方差,$\delta_n\in[0,25]$)相加,以复现传感器或环境噪声带来的随机干扰;最后,对上述结果应用高斯滤波器G_{σ},其标准差σ从均匀分布$U\in[0.5,3.0]$中采样,实现不同程度的模糊失真.三者按式(1)叠加生成多样化的退化图像$D(\boldsymbol{x})$.

$$D(\boldsymbol{x})=\mathrm{Blur}\big(\mathrm{Mask}(\boldsymbol{x})+\mathrm{Noise}(\boldsymbol{x})\big). \tag{1}$$

式中:Blur代表对得到的中间图像应用操作;Mask($\boldsymbol{x}$)代表掩膜(masking),即在原始图像$\boldsymbol{x}$上随机"擦除"或"遮盖"一部分像素;Noise($\boldsymbol{x}$)代表添加噪声.

为使自监督的边缘恢复阶段提供准确的"完整"目标,本方法在原始图像上采用Canny算法预处理提取真实边缘E_{gt},设置低阈值$T_{low}=50$、高阈

值$T_{high}=150$、高斯平滑核尺寸$k=5$、标准差$\sigma_{smooth}=1.0$，以精确捕捉图像中的物体轮廓.整个退化模拟过程支持在线生成，无需事先准备大规模标注数据，不仅大幅提高了样本多样性，还能有效避免模型对固定损伤模式的过拟合，同时为后续双阶段网络在自监督指引下专注于边缘结构与色彩细节的修复奠定坚实基础.最终输出图像如图2所示.

图2 退化过程图像输出

Fig. 2 Output images from degradation process

1.2.2 边缘重建网络模块

本节介绍了边缘修复阶段网络结构的核心改进，采用增强型U-Net架构，融合残差块、通道注意力机制与多尺度特征融合，以提升模型对复杂退化边缘结构的建模能力与鲁棒性.整个网络分为编码器与解码器两部分，通过对称结构与跳跃连接实现高效特征传递与信息恢复.首先，为防止在多次卷积与池化过程中“细节丢失”，编码器每层下采样前后都引入双卷积残差块(ResBlock)，其恒等映射路径保证了梯度在深层网络中的畅通并有效保留原始边缘的细微结构；在从输入的$h\times w$单通道退化边缘图到$\frac{h}{32}\times\frac{w}{32}\times 512$的五级下采样过程中，最大池化操作使网络获得足够大的感受野，以捕捉全局连通性.其次，为在跳跃连接阶段去除噪声响应强但结构信息弱的无效特征，解码器在融合编码器与上采样特征前，采用卷积块注意力模块(convolutional block attention module，CBAM)中的双重注意力机制，该模块依次通过通道注意力和空间注意力两个子模块对特征进行协同优化.在通道注意力阶段，首先对输入特征同时进行全局平均池化和全局最大池化操作，分别提取能够反映不同上下文信息的特征描述符；随后将这两个描述符输入一个权值共享的多层感知机中进行非线性变换与特征融合，并通过Sigmoid激活函数生成通道维度上的注意力权重图，从而自适应地校准各特征通道的重要性，有效抑制噪声通道并增强与边缘结构相关的特征响应.在空间注意力阶段，以前一阶段输出特征作为输入，首先在通道维度上分别执行平均池化与最大池化操作，将得到的两个二维特征图进行拼接，再通过一个7×7卷积层进行跨特征交互和信息压缩，生成空间维度的注意力分布图，并再次借助Sigmoid函数进行归一化.该方法还引入掩码调制机制，将生成的空间注意力图与标识图像破损区域的二值掩码进行点乘操作，以限制注意力仅作用于待修复区域，引导模型聚焦于断裂边缘及缺失结构的最可能位置.通过上述机制，CBAM模块在跳跃连接中显著提升了解码器所融合特征的质量：不仅抑制了噪声和伪边缘干扰，也增强了对关键细节的感知与重建能力，从而最终提升了图像修复结果的结构连贯性与视觉真实性.CBAM由两部分组成：首先是通道注意力模块(channel attention module)，通过对输入特征F进行全局平均池化与最大池化，再通过共享权重的全连接层生成通道级权重，对输入进行加权处理，公式为

$$\boldsymbol{F}'=\mathrm{ChannelAttention}(\boldsymbol{F})\otimes\boldsymbol{F}. \tag{2}$$

式中：$\boldsymbol{F}'$是经过通道注意力模块处理后的特征图；$\boldsymbol{F}$为输入特征.

其次是空间注意力模块(spatial attention module)，该模块通过通道聚合和卷积生成空间注意力图，从而进一步聚焦于结构显著区域：

$$\boldsymbol{F}''=\mathrm{SpatialAttention}(\boldsymbol{F}')\otimes\boldsymbol{F}'. \tag{3}$$

式中：$\boldsymbol{F}''$为经过空间注意力模块处理后的最终特征图.

最后，解码器通过5次转置卷积上采样重建至原始分辨率，每次上采样均与经过注意力筛选的编码器特征进行拼接，使得低层细节与高层语义互为补充，直至在末端用单通道Sigmoid激活输出像素级边缘置信图$E_{pred}\in[0,1]^{h\times w}$.得益于残差块的梯度直通与多尺度信息保留，模型能在深层次捕获全局连通模式；而基于CBAM的双重注意力策略，则实现了对噪声干扰与不规则破损区域

的精准区分与聚焦;再加上对上下文与空间语义的多次融合,上述三者相辅相成,既保证了复杂退化场景下高频边缘的连续性,也兼顾了局部微裂缝与断点的补全,从而显著提升了边缘恢复的完整性和稳定性.

1.3　色彩恢复网络结构设置

本节针对第一阶段所恢复的边缘轮廓仍缺乏色彩信息与纹理细节的问题,设计了一条“全局语义+局部结构”并行融合的色彩恢复通路.首先,为解决“单纯边缘无法提供物体类别与色彩先验”这一难题,引入了在 ImageNet-21k 上预训练的 ViT-B/16 作为全局语义编码器,分别抽取第 6,9,12 层的多头自注意力输出,通过双线性插值将其空间分辨率对齐至 $\frac{h}{16}\times\frac{w}{16}$,再用 1×1 卷积压缩通道至 256 维,形成包含多尺度、长程依赖的语义特征图 $\boldsymbol{F}_{\text{vit}}$.这种做法不仅克服了局部卷积对语境理解的局限,还为色彩还原提供了诸如“天空应为蓝色”“植物呈现绿色”等宏观指引.

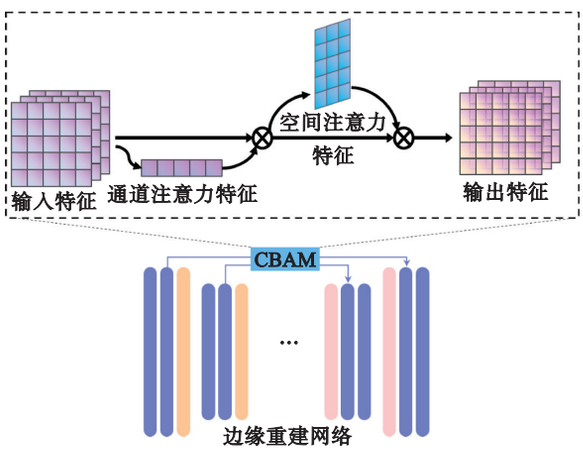

图 3　边缘重建网络结构

Fig. 3　Edge reconstruction network structure

与此同时,为解决“全局语义特征与局部边缘结构难以直接拼接导致信息冲突”的问题,在 U-Net 解码器前端增设了跨模态注意力融合模块:将第一阶段输出的恢复边缘图 E_{pred} 通过 3 层卷积与残差块提取出 $\frac{h}{8}\times\frac{w}{8}\times 128$ 的边缘特征 $\boldsymbol{F}_{\text{edge}}$,再以其为 Query,ViT 特征 $\boldsymbol{F}_{\text{vit}}$ 为 Key 和 Value,按照式(4)的交叉注意力机制,将全局语义动态注入本地结构中,生成同时兼具准确边界和语义一致性的融合特征 $\boldsymbol{F}_{\text{fusion}}$.这一策略有效避免了直接通道拼接带来的冗余信息、色彩错配与纹理失真.

$$\left.\begin{aligned}&\boldsymbol{Q}=\boldsymbol{W}_{\text{q}}\boldsymbol{F}_{\text{edge}},\boldsymbol{K}=\boldsymbol{W}_{\text{k}}\boldsymbol{F}_{\text{vit}},\boldsymbol{V}=\boldsymbol{W}_{\text{v}}\boldsymbol{F}_{\text{vit}};\\&\text{Attention}=\text{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\text{T}}}{\sqrt{d}}\right)\boldsymbol{V}.\end{aligned}\right\}\tag{4}$$

式中:$\boldsymbol{Q}$ 代表 Query(查询)张量矩阵;$\boldsymbol{K}$ 代表 Key(键)张量矩阵;$\boldsymbol{V}$ 代表 Value(值)张量矩阵;$\boldsymbol{W}_{\text{q}}$,$\boldsymbol{W}_{\text{k}}$,$\boldsymbol{W}_{\text{v}}$ 分别代表可学习的权重张量矩阵(用于特征线性变换);d 代表键($\boldsymbol{K}$)张量的维度(用于 Softmax 归一化)常量.

在融合特征的基础上,变体 U-Net 色彩解码器通过五级转置卷积上采样逐步恢复空间分辨率,每次上采样后都接入 CBAM 模块,通过通道注意力进一步剔除与色彩重建无关的特征、空间注意力聚焦于重要区域,不断细化色彩与纹理;最终以三通道 Tanh 激活输出归一化 RGB 图像 $\boldsymbol{I}_{\text{pred}}\in[-1,1]^{h\times w\times 3}$.该多任务解码路径在像素级 MSE、感知损失和对抗损失的联合约束下,不仅保障了色彩分布的全局正确性,还在局部细节与质感上实现了精细刻画.

1.4　损失函数

训练不稳定及其导致的生成图像失真是原始 GAN 面临的主要问题[23].为实现边缘结构准确、色彩自然且细节丰富的高质量图像修复,本节设计了双阶段的网络结构,并通过联合训练策略对其进行端到端优化.第一阶段聚焦于边缘信息的精确恢复,这是色彩还原的基础,若边缘模糊或失真将严重干扰后续的语义与纹理建模.为此,定义边缘修复阶段的损失函数为

$$L_{\text{edge}}=L_{\text{BCE}}\left(E_{\text{pred}},E_{\text{gt}}\right)+\alpha L_{\text{SSIM}}\left(E_{\text{pred}},E_{\text{gt}}\right).\tag{5}$$

其中,BCE(二元交叉熵)损失函数为

$$L_{\text{BCE}}=-\frac{1}{N}\sum_{n=1}^{N}[E_{\text{gt}}^{n}\lg E_{\text{pred}}^{n}+(1-E_{\text{gt}}^{n})\lg(1-E_{\text{pred}}^{n})].\tag{6}$$

式中:N 为样本总数;n 为每个样本.

用于监督像素级的边缘分类精度,而 SSIM(结构相似性)损失函数为

$$L_{\text{SSIM}}=1-\frac{(2\mu_{\text{pred}}\mu_{\text{gt}}+C_1)(2\sigma_{\text{pred,gt}}+C_2)}{(\mu_{\text{pred}}^2+\mu_{\text{gt}}^2+C_1)(\sigma_{\text{pred}}^2+\sigma_{\text{gt}}^2+C_2)}.\tag{7}$$

式中:μ_{pred} 代表预测边缘图;μ_{gt} 代表真实边缘图;σ_{pred}^2 代表预测边缘图局域块的方差;σ_{gt}^2 代表真实边缘图局域块的方差;$\sigma_{\text{pred,gt}}$ 代表 E_{pred} 和 E_{gt} 局域块的协方差;C_1 和 C_2 是避免分母为零的极小常数.

从结构一致性角度约束边缘图像的空间一致性,有助于生成更平滑、自然的边缘表示.本文设置 $\alpha=0.1$,以保证结构约束在损失函数中适度发挥作用而不主导训练过程.

第二阶段则面向色彩重建与纹理增强,其核

心目标是使生成的RGB图像在像素和语义层面同时贴近真实图像.为此,本文构建了复合损失函数:

$$L_{\text{color}}=L_{\text{MSE}}+\beta L_{\text{perc}}+\gamma L_{\text{GAN}}. \tag{8}$$

式中:L_{MSE}为保证生成图像在像素值上与真实图像尽可能接近的基础像素重建损失.

$$L_{\text{MSE}}=\frac{1}{3hw}\sum_{c=1}^{3}\sum_{i=1}^{h}\sum_{j=1}^{w}(\boldsymbol{I}_{\text{pred}}^{(c,i,j)}-\boldsymbol{I}_{\text{gt}}^{(c,i,j)})^2. \tag{9}$$

式中:$\boldsymbol{I}_{\text{pred}}^{(c,i,j)}$为模型预测的RGB图像像素值;$\boldsymbol{I}_{\text{gt}}^{(c,i,j)}$为真实参考图像像素值;$h$,$w$分别为图像的高度和宽度;$c$,$i$,$j$为图像的维度索引,$c$对应RGB图像的通道维度,$i$,$j$对应图像的空间维度($i$为高度方向的像素索引,$j$为宽度方向的像素索引).

为保证整体色彩的准确还原,并克服MSE在高频细节和语义层次上的不足,引入感知损失L_{perc}:

$$L_{\text{perc}}=\left\|\phi\left(\boldsymbol{I}_{\text{pred}}\right)-\phi\left(\boldsymbol{I}_{\text{gt}}\right)\right\|_2^2. \tag{10}$$

式中:$\phi(\cdot)$为特征提取函数,此处特指预训练的VGG-19网络的中层卷积特征.

该项通过VGG-19网络的中层特征度量图像在感知空间的相似性,使重建图像更具真实感.为提升纹理与细节表现,进一步引入对抗损失L_{GAN}:

$$L_{\text{GAN}}=-E_{\text{pred}}\left[\lg D'\left(\boldsymbol{I}_{\text{pred}}\right)\right]. \tag{11}$$

式中,$D'(\cdot)$为判别器网络(discriminator),用于判断输入的图像是真实或生成的.

基于PatchGAN判别器,强化图像的局部真实性表达.通过设置$\beta=0.5,\gamma=0.01$,在保证主监督信号稳定的前提下引入感知与对抗信息.最终,将两个阶段的损失函数协同优化,构成总损失函数:

$$L_{\text{total}}=L_{\text{edge}}+\lambda' L_{\text{color}}. \tag{12}$$

式中,λ'是平衡两个阶段贡献的超参数.

该联合策略体现了"结构优先-感知提升-细节增强"的设计思想,通过边缘结构引导语义重建,最终实现高质量图像修复的目标.

2 实验结果与分析

2.1 实验环境与参数设计

本研究在具备高性能计算能力的服务器上进行,硬件配置包括NVIDIA A100-SXM4-80GB GPU,配合AMD EPYC 7742 64核CPU和512 GB DDR4 ECC内存.软件环境基于Ubuntu 20.04 LTS操作系统,深度学习框架采用PyTorch 2.0.1(CUDA 11.8),所有的对比实验均在此统一环境下进行.数据加载使用多进程在线预处理与增强技术,所有输入图像统一缩放至256像素×256像素并归一化至[-1,1]区间.在模型训练阶段的关键超参数设置如下:$\alpha=0.1$,$\beta=0.5$,$\gamma=0.01$,$\lambda'=0.5$;采用Adam优化器(动量参数$\beta_1=0.9,\beta_2=0.999$)统一优化生成器、判别器及属性解耦模块,基础学习率设置为5×10^{-5}.整体分为两个阶段,并采用联合训练策略.边缘修复网络:参数量约28.7 M,批大小(Batch Size)=16×16,训练500轮次;色彩恢复网络:编码器冻结状态下,参数量约41.3 M(仅解码器部分),批大小=16×16,训练600轮次;在联合训练中,两个阶段网络在各自预热完成后共享损失函数和优化器状态,通过交替更新增强特征互补性与整体收敛性.采用PSNR和SSIM评估恢复结果的重建质量,同时计算LPIPS以量化图像主观视觉一致性.

2.2 实验数据集

本研究使用的档案馆图像修复数据集共包含2 495张历史照片,涵盖人物、建筑两大类典型场景.这些图像普遍存在划痕、褪色、模糊等多种退化形式,具有较强的现实代表性与挑战性.为合理构建训练与评估流程,将数据集按8∶1∶1的比例划分为训练集、验证集与测试集.划分过程中采用随机采样方式,同时确保各类别图像在不同子集中分布均衡.训练集用于模型参数学习,验证集用于调节超参数与早停策略,测试集则保持固定不变以用于最终性能评估与对比实验.

表1 图像修复数据集信息统计

Table 1 Image restoration dataset information statistics

类别	训练集	验证集	测试集	小计
人物	963	148	141	1 252
建筑	983	131	129	1 243
合计	1 946	279	270	2 495

2.3 结果分析

2.3.1 定量分析

本研究基于档案馆历史图像构建的复原任务数据集进行算法性能评估,涵盖人物与建筑两大类典型场景,采用PSNR,LPIPS,FID及Colorfulness Score等4项主流定量指标,全面衡量图像复原的像素精度、感知相似度、生成分布真实性与色彩还原质量.评价对比涵盖CycleGAN[24]、DeOldify与Restormer[25]等代表性方法.各项指标中,PSNR用于衡量图像的还原精度,LPIPS和FID分别反映人眼感知质量和生成图像的真实性分布,而

Colorfulness Score专注于评估复原图像的色彩丰富度,反映复原结果的视觉吸引力和自然度.

表2　以往算法在图像数据集上的性能表现
Table 2　Performance of previous algorithms on image datasets

模型	PSNR	LPIPS	FID	Colorfulness Score
CycleGAN	18.1	0.36	62.3	31.9
DeOldify	18.9	0.29	56.1	36.3
Restormer	18.7	0.23	48.0	38.3
DSS-HIRC	19.8	0.21	43.5	39.9

注:PSNR与Colorfulness Score数值越大越优,LPIPS与FID数值越小越优.

从定量结果来看,本文算法DSS-HIRC在整体性能与分场景效果上均表现最优.在整体测试中,PSNR达到19.8,较DeOldify提升4.8%,较Restormer提升5.9%;LPIPS值为0.21,较次优模型Restormer(0.23)降低8.7%;FID为43.5,较Restormer改善幅度达9.4%;Colorfulness Score达到39.9,较Restormer提升4.2%.在人物场景下,PSNR与FID分别为20.5与41.3,优于Restormer(20.3和42.9);Colorfulness Score为41.7,高于Restormer(41.1),呈现更优的感知与色彩质量.在建筑场景中,本文算法在PSNR(18.5)、FID(45.7)和Colorfulness Score(38.2)上全面领先,分别优于Restormer约8.2%,13.9%和7.9%,充分验证其在结构边缘细节与语义色彩建模方面的综合优势.

表3　以往算法在独立场景数据集上的性能表现
Table 3　Performance of previous algorithms on independent scene datasets

模型	场景	PSNR	LPIPS	FID	Colorfulness Score
CycleGAN	人物	18.5	0.42	61.2	32.1
	建筑	17.6	0.29	63.4	31.6
DeOldify	人物	19.6	0.33	55.3	36.6
	建筑	18.2	0.25	56.9	35.9
Restormer	人物	20.3	0.17	42.9	41.1
	建筑	17.1	0.29	53.1	35.4
DSS-HIRC	人物	20.5	0.17	41.3	41.7
	建筑	18.5	0.23	45.7	38.2

2.3.2　定性分析

图4展示了本文模型在不同类型图像上的视觉输出效果.从人像复原结果来看,本文算法相比CycleGAN与DeOldify在边缘轮廓等关键区域呈现出更完整清晰的线条,整体图像更具真实感与自然度;而Restormer生成的人像尽管色彩饱和度较高,但在细节还原和感知协调性方面存在欠缺,画面观感略显突兀.在建筑类图像复原中,受光照变化与结构遮挡等因素影响,Restormer与DeOldify存在一定程度的色彩偏差与结构扭曲;相比之下,本文模型在墙体边缘等区域的还原更准确,图像整体结构更连贯,细节纹理更丰富,进一步验证了其对不同场景的自适应能力与重建鲁棒性.

图4　模型输出效果图
Fig. 4　Model output samples

2.4　消融实验

为进一步验证所提双阶段自监督修复模型各组成模块的有效性与必要性,本研究设计并实施了一系列消融实验(如表4所示),围绕退化模拟策略、网络结构设计和训练方式3个维度展开.在退化策略方面,将完整模型与无自监督退化方案进行对比,评估随机退化生成机制对模型泛化能力的提升作用.其次,针对结构改进部分,分别移除CBAM注意力模块与将ViT替换为ResNet-50,以衡量结构感知能力对修复效果的影响.最后,比较联合训练与分阶段独立训练策略,评估两阶段协同优化在稳定性与性能上的综合优势.实验结果表明,自监督退化与协同训练对性能提升最为显著,CBAM模块在结构清晰度与色彩自然度方面亦发挥关键作用,模型在保持合理计算开销的前提下显著提升了修复质量,验证了设计方案的科学性与实用性.

同时,本研究在消融实验基础上,对网络结构、优化器设置等关键超参数进行了系统分析.通过在验证集上进行多组实验,确定了较优的参数组合,并总结其对性能提升的贡献.

表4 不同模型设置模型性能对比

Table 4 Comparison of model performance for different model settings

方法	PSNR	LPIPS	FID	训练时长/h
无自监督退化	17.3	0.36	58.8	18.3
使用ViT特征	18.3	0.25	45.2	19.1
使用ResNet-50特征	17.9	0.22	47.3	18.6
采用独立训练	18.9	0.22	45.4	25.2
DSS-HIRC	19.8	0.21	43.5	20.0

如表5所示,批大小为8×8的PSNR为18.9,低于16×16的19.3. PSNR衡量像素级别的重建准确度.PSNR分数降低表明8×8尽管捕捉的局部信息更细致,但在全局像素准确度上表现更差,这意味着更小的批大小难以稳定地将细粒度的局部信息融合到全局图像中,导致像素级重构质量下降,即稳定性不足.且8×8的LPIPS为0.24,高于16×16的0.23. LPIPS衡量人眼感知的主观相似度.LPIPS分数升高表明8×8生成的图像在视觉上与真实图像的差异更大.这可能是由于模型过度关注局部细节,却牺牲了全局纹理和色彩的连贯性与自然度,从而降低了感知真实感,进一步体现了整体性能的不稳定.以及8×8的FID为46.7,高于16×16的45.1. FID衡量生成图像的整体分布与真实图像分布的相似度.FID分数越高,表示生成图像的整体真实性越差.这有力地支持了"整体性能不及16×16稳定"的结论.因此综合来看较小的批大小(如8×8)虽然在局部细节捕捉上更加敏感,但由此带来的序列长度过长,模型训练的难度和计算开销增大,导致模型难以充分收敛或陷入局部最优,最终在PSNR,LPIPS,FID 3项核心指标上全面落后于16×16的配置,证明了其整体性能的不稳定性和次优性.在嵌入维度方面,512维的表示能力更强,能够更好地提升图像结构的表达效果,使得PSNR指标优于256维;同时,卷积核大小的设置对修复效果亦有明显影响,5×5卷积在处理边缘信息时更加鲁棒,尤其在人像与建筑类图像修复中表现出更自然和协调的效果.

表5 在不同网络相关超参数设置下的性能

Table 5 Performance under different network-related hyperparameter settings

批大小	嵌入维度	卷积核	PSNR	LPIPS	FID	训练时长/h
8×8	256	3×3	18.9	0.24	46.7	19.3
16×16	256	3×3	19.3	0.23	45.1	19.8
16×16	256	5×5	19.6	0.22	44.3	20.2
16×16	512	3×3	19.7	0.21	43.9	20.5
16×16	512	5×5	19.8	0.21	45.4	20.0

在超参数设置方面,本研究主要考察了学习率、批大小和权重衰减对模型性能的影响(如表6所示).学习率过大容易导致训练过程振荡甚至不收敛,而过小则可能造成收敛速度过慢和模型陷入局部最优.实验结果表明,当学习率设为时5×10^{-5},模型在PSNR与FID上取得最佳平衡.批大小则影响了梯度估计的稳定性与训练效率,过小的批大小(如8×8)虽然对显存友好,但易造成模型在不同批次间波动明显;而过大的批大小(如32×32)会降低对细节的捕捉能力.实验验证批大小=16×16时,模型在稳定性与性能上取得了较优表现.权重衰减的选择则主要影响模型的正则化能力,当不使用权重衰减时,模型容易过拟合,特别是在校史馆数据集有限规模的条件下;而当权重衰减值过大时,模型表达能力受到抑制,导致修复效果欠佳.最终实验结果显示,当权重衰减设定为1×10^{-4}时,能够有效平衡模型的泛化能力与生成质量,保证修复结果更加自然稳定.

表6 不同超参数设置下的模型性能对比

Table 6 Comparison of model performance under different hyperparameter settings

参数设置	PSNR	LPIPS	FID	训练时长/h
学习率=5e-4	18.5	0.28	50.6	17.8
学习率=5e-5	19.8	0.21	43.5	20.0
批大小=8×8	19.0	0.22	47.3	20.4
批大小=16×16	19.6	0.21	43.6	20.1
批大小=32×32	18.8	0.22	44.0	19.2
权重衰减=0	19.1	0.24	47.9	19.6
权重衰减=1e-4	19.8	0.21	43.5	20.0

3 结　论

本研究提出了一种基于双阶段自监督学习的老旧图像修复与上色框架,针对档案历史图像退化模式复杂、成对训练数据稀缺的核心挑战,通过解耦边缘修复与色彩恢复任务,结合自监督退化模拟与跨模态特征融合,实现了高效、高质量的历史图像复原.在档案馆自建数据集上的实验验证表明,该方法在定量指标与主观评估中均优于对比模型.本研究为老旧图像修复提供了一种高效、低依赖的解决方案,在历史档案保护、文化遗产数字化等领域具有实践意义.

参考文献:

[1] 陈文祥,田启川,廉露,等.基于深度学习的图像修复方法研究进展[J].计算机工程与应用,2024,60(22):58-73.
(Chen Wen-xiang, Tian Qi-chuan, Lian Lu, et al. Research progress of image inpainting methods based on deep learning [J]. *Computer Engineering and Applications*, 2024, 60(22):58-73.)

[2] Anwar S, Tahir M, Li C Y, et al. Image colorization: a survey and dataset [J]. *Information Fusion*, 2025, 114: 102720.

[3] Chua L O. CNN: a vision of complexity [J]. *International Journal of Bifurcation and Chaos*, 1997, 7(10): 2219-2425.

[4] Isola P, Zhu J Y, Zhou T H, et al. Image-to-image translation with conditional adversarial networks [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, 2017: 5967-5976.

[5] Parmar N, Vaswani A, Uszkoreit J, et al. Image Transformer [C]//Proceedings of the 35th International Conference on Machine Learning (ICML). Stockholm, 2018: 4055-4064.

[6] 魏赟,王璐璐,辛子昊,等.由粗到细的内容一致性图像修复方法[J].华中科技大学学报(自然科学版),2025,53(5):178-184.
(Wei Yun, Wang Lu-lu, Xin Zi-hao, et al. Coarse to fine approach to content-consistent image inpainting [J]. *Journal of Huazhong University of Science and Technology* (*Natural Science Edition*), 2025, 53(5): 178-184.)

[7] Tai Y, Yang J, Liu X M, et al. MemNet: a persistent memory network for image restoration [C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice, 2017: 4549-4557.

[8] 周啟雪,余映,胡家绿.利用Involution级联注意力机制的古代壁画图像修复网络[J].计算机科学,2025,52(12):158-165.
(Zhou Qi-xue, Yu Ying, Hu Jia-lyu). Ancient mural image restoration network using Involution cascade attention mechanism [J]. *Computer Science*, 2025, 52(12): 158-165.)

[9] Wang Z D, Cun X D, Bao J M, et al. Uformer: a general U-shaped Transformer for image restoration [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, 2022: 17662-17672.

[10] Liang J Y, Cao J Z, Sun G L, et al. SwinIR: image restoration using swin transformer [C]//2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Montreal, 2021: 1833-1844.

[11] Karras T, Laine S, Aittala M, et al. Analyzing and improving the image quality of StyleGAN [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, 2020: 8107-8116.

[12] Vitoria P, Raad L, Ballester C. ChromaGAN: adversarial picture colorization with semantic class distribution [C]//2020 IEEE Winter Conference on Applications of Computer Vision (WACV). Snowmass Village, 2020: 2434-2443.

[13] Guo Y, Gao Y, Lu Y X, et al. OneRestore: a universal restoration framework forComposite degradation [C]//Computer Vision-ECCV 2024. Cham: Springer, 2025: 255-272.

[14] Gaa D, Chizhov V, Peter P, et al. Connecting image inpainting with denoising in the homogeneous diffusion setting [J]. *Advances in Continuous and Discrete Models*, 2025, 2025(1): 74.

[15] Kim G, Kang K, Kim S, et al. BigColor: colorization using a generative color prior for natural images [M]//Computer Vision-ECCV 2022. Cham: Springer Nature Switzerland, 2022: 350-366.

[16] Kang X Y, Yang T, Ouyang W Q, et al. DDColor: towards photo-realistic image colorization *via* dual decoders [C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, 2024: 328-338.

[17] Salmona A, Bouza L, Delon J. DeOldify: a review and implementation of an automatic colorization method [J]. *Image Processing on Line*, 2022, 12: 347-368.

[18] Cui M Z, Jiang H, Li C Z. Progressive-augmented-based DeepFill for high-resolution image inpainting [J]. *Information*, 2023, 14(9): 512.

[19] Zhang H M, Wang M C, Zhang Y X, et al. TDA-Net: a novel transfer deep attention network for rapid response to building damage discovery [J]. *Remote Sensing*, 2022, 14(15): 3687.

[20] Yang F, Ning B, Li H Q. An overview of multimodal fusion learning [C]//Mobile Multimedia Communications. Cham: Springer, 2022: 259-268.

[21] Chen Y T, Xia R L, Yang K, et al. Dual degradation image inpainting method via adaptive feature fusion and U-Net network [J]. *Applied Soft Computing*, 2025, 174: 113010.

[22] Rong W B, Li Z J, Zhang W, et al. An improved Canny edge detection algorithm [C]//2014 IEEE International Conference on Mechatronics and Automation. Tianjin, 2014: 577-582.

[23] 杨晓雨,王爱侠,杨钢,等.基于生成对抗网络的人脸年龄渐进合成算法[J].东北大学学报(自然科学版),2024,45(7):944-952.
(Yang Xiao-yu, Wang Ai-xia, Yang Gang, et al. Progressive face age synthesis algorithm based on generative adversarial network [J]. *Journal of Northeastern University* (*Natural Science*), 2024, 45(7): 944-952.)

[24] Almahairi A, Rajeswar S, Sordoni A, et al. Augmented CycleGAN: learning many-to-many mappings from unpaired data [C]//Proceedings of the 35th International Conference on Machine Learning (ICML). Stockholm, 2018: 195-204.

[25] Zamir S W, Arora A, Khan S, et al. Restormer: efficient transformer for high-resolution image restoration [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, 2022: 5718-5729.