文章编号　1004-924X(2026)03-0497-16

基于结构先验与多尺度残差聚合的大面积破损图像修复

魏　赟[1]，贾白雪[1*]，冯丹丹[1]，张治瑞[1]，辛子昊[1]，邬侗辰[2]

(1. 兰州交通大学 电子与信息工程学院，甘肃 兰州 730070；
2. 兰州交通大学 光电技术与智能控制教育部重点实验室，甘肃 兰州 730070)

摘要：针对大面积破损图像在人脸、街景、建筑等修复工程中存在的纹理模糊、结构扭曲与边界伪影的问题，提出一种基于结构先验与多尺度残差聚合(Structural Priors and Multi-scale Residual Aggregation，SPMRA-Net)的大面积破损图像修复方法。首先，设计动态自适应多尺度(Dynamic Adaptive Multi-scale，DAM)模块，通过空洞卷积与残差结构增强模型的上下文感知能力。其次，构建由 Transformer路径与一致语义注意力路径组成的瓶颈模块，并通过交叉注意力机制将 Transformer路径输出的全局语义信息与自注意力路径输出的局部纹理信息融合，以抑制大面积缺损场景中的结构扭曲。然后，为缓解传统跳跃连接因简单拼接导致的语义鸿沟问题，引入自适应多尺度聚合(Adaptive Multi-scale Aggregation，AMSA)模块提升深层特征与浅层特征的交互能力，保证修复图像的边界连续性。最后，引入双判别器对修复后的结果与原始图像进行一致性评估，以提升生成图像在视觉感知上的真实性。实验结果表明，该模型在CelebA 数据集上的PSNR 提升 3.06 dB，SSIM 提升 0.087，LPIPS 降低 0.078。在 Paris StreetView，Places2 数据集上的主观评价以及客观评价指标均优于对比算法。本文方法在结构一致性与感知层面上均取得一定提升，验证了其在大面积破损图像修复任务中的有效性。

关　键　词：深度学习；图像修复；生成对抗网络；多尺度残差；结构先验；Transformer

中图分类号：TP391.4　　**文献标识码**：A

doi：10.37188/OPE.20263403.0497　　**CSTR**：32169.14.OPE.20263403.0497

Inpainting of large-area damaged images based on structural priors and multi-scale residual aggregation

WEI Yun[1]，JIA Baixue[1*]，FENG Dandan[1]，ZHANG Zhirui[1]，XIN Zihao[1]，WU Tongchen[2]

(1. *School of Electronic and Information Engineering，Lanzhou Jiaotong University，Lanzhou* 730070，*China*；
2. *Key Laboratory of Optical Technology and Intelligent Control of Ministry of Education，Lanzhou Jiaotong University，Lanzhou* 730070，*China*)
* *Corresponding author，E-mail*：3489395305@*qq.com*

Abstract：To address texture blurring，structural distortion，and boundary artifacts in the restoration of images with large-area damage (e.g.，faces，street scenes，and buildings)，a large-area damaged image

收稿日期：2025-10-30；**修订日期**：2025-11-19.

基金项目：甘肃省教育厅高校教师创新基金资助项目(No. 2025A-054)；甘肃省联合基金资助项目(No. 25JRRA1160)；甘肃省自然科学基金资助项目(No. 25JRRA177)；中央引导地方科技发展资金项目(No. 25ZYJF001)

restoration network, termed SPMRA-Net, is proposed. First, a Dynamic Adaptive Multi-scale (DAM) module is designed to enhance contextual modeling through dilated convolutions and a residual architecture. Subsequently, a bottleneck module consisting of a Transformer branch and a consistent semantic attention branch is constructed; global semantic information from the Transformer branch is fused with local texture information from the self-attention branch via a cross-attention mechanism, effectively suppressing structural distortion in severely corrupted regions. Moreover, to mitigate the semantic gap introduced by naive concatenation in conventional skip connections, an Adaptive Multi-scale Aggregation (AMSA) module is introduced to strengthen interactions between deep and shallow features and to preserve boundary continuity in restored images. Finally, a dual-discriminator framework is employed to assess consistency between restored results and original images, thereby improving the perceptual realism of generated outputs. Experiments demonstrate that, on the CelebA dataset, the proposed method improves PSNR by 3.06 dB and SSIM by 0.087, while reducing LPIPS by 0.078. Subjective evaluations on the Paris StreetView and Places2 datasets, together with the aforementioned objective metrics, consistently outperform comparative methods. These results indicate notable gains in both structural consistency and perceptual quality, validating the effectiveness of the proposed approach for large-area damaged image restoration.

Key words: deep learning; image inpainting; generative adversarial network; multi-scale residuals; structural prior; Transformer

1 引　言

图像修复是一项基于已知区域信息还原图像中缺失或破损区域的视觉任务,该任务在数字摄影修复、视频修复和医学影像处理等领域[1]有重要应用价值。传统修复方法主要分为基于扩散[2-4]的方法和基于补丁[5-7]的方法。基于扩散的方法通过将边界信息逐步传播至缺失区域内部修复图像,但此类方法缺乏全局上下文理解,在处理大面积缺失时效果不佳。基于补丁的方法则通过在已知区域搜索最相似的图块填充未知区域,但依赖于低层次像素相似性,难以适应复杂场景[8]。

随着深度学习技术的崛起,Ronneberger等[9]提出的U-Net以及Goodfellow等[10]提出的生成对抗网络对受损图像的低频和高频特征信息进行联合建模,改善传统方法在纹理、结构与语义方面的不匹配问题。基于此类端到端的学习框架,研究人员探索不同网络结构在图像修复任务中的作用机理。Chen等[11]提出的DNNAM将多尺度通道注意力机制与深度神经网络相结合,采用残差结构扩展U-Net感受野生成合理结构。雷帅等[12]利用深度可分离卷积提取特征以缓解修复结果中的纹理模糊问题。陈清江等[13]通过构建多尺度交互调制模块,从低分辨率图像中提取多尺度特征,以增强信息流的丰富性。但上述方法在融合多尺度信息时往往采用静态、非自适应策略,缺乏对局部图像内容的调整能力。但卷积神经网络的感受野有限,难以充分捕获全局结构信息。近年来,Transformer架构凭借其建模长距离依赖的能力广泛应用于图像修复任务。Chen等[14]提出的SWMH Transformer将CNN与Transformer相结合,利用Transformer捕捉图像像素间的长距离依赖关系,提升修复结果的结构一致性。Huang等[15]提出稀疏注意力(Spa-former),实现大感受野下的长距离建模,以处理复杂场景和大面积破损图像。Liu等[16]通过整合CNN的局部建模能力与Transformer的长距离依赖建模能力,从而实现局部特征与全局语义的协同建模。但Transformer与CNN的融合方式仍停留在浅层特征层面,难以实现全局语义与局部纹理间的信息交互。基于此,本文将CNN与Transformer相结合以共同处理全局结构与局部纹理,改善大面积破损图像下结构扭曲以及图像伪影问题。此外,传统跳跃连接虽能将编码器浅层特征传递至解码器,以保留边缘、纹理等高频

细节。但此类特征与解码器的深层特征在语义层面存在差异，二者简单拼接易导致边界伪影与高频细节的丢失。为此，Wang 等[17]针对 U-Net 中的语义鸿沟引入可学习的注意力模块，以实现更精细的跳跃连接融合。Ates G C 等[18]提出双重交叉注意力模块（Dual Cross-Attention，DCA）增强跳跃连接，以捕捉多尺度上下文信息。Sun 等[19]通过在跳跃连接层中引入 DA-Block，增强模型的特征提取能力并优化编码器-解码器效率。然而，现有方法在实现基于局部内容特征的精细化自适应融合方面仍存在不足。因此，本文通过使用动态加权的多尺度空洞卷积减小浅层与深层特征间的语义差距，显著改善修复结果的边界连续性与结构完整性。

根据上述分析，本文提出一种结构先验与多尺度残差聚合的大面积破损图像修复方法。首先，设计 DAM 模块，通过深度可分离卷积与残差结构增强上下文感知能力，从而有效改善纹理模糊问题。其次，提出结构感知双路径瓶颈（Structure-Aware Dual-Path Bottleneck，SADP-Bottleneck）模块，该瓶颈层通过交叉注意力机制，将 Transformer 路径捕捉的全局语义与一致语义注意力路径提供的稀疏局部纹理先验进行融合，缓解大面积缺损造成的结构不一致与细节失真问题。最后，提出 AMSA 模块，将编码器浅层特征与解码器深层特征进行交互，利用多尺度感知机制确保生成边界连续、结构完整的修复图像。

2　SPMRA-Net 算法

2.1　整体架构

本文模型的整体框架如图 1 所示，由生成器和双判别器系统构成。其中，生成器由编码器、瓶颈模块及解码器构成，并通过跳跃连接实现特征的信息交互与细节保留。判别器由补丁判别器（Patch discriminator[20]）和特征判别器（Feature discriminator[21]）组成双判别器系统，对输出图像进行对抗性监督，从而确保生成结果具有视觉感知层面的真实性与语义理解层面的一致性。

2.1.1　生成器

生成器采用编码器-解码器架构，对输入图像 I_{in} 进行特征压缩、深度推理与重建[22]，映射为高维初始特征表示。大核卷积能够提供更大的感受野，使模型在早期阶段即可感知更广泛的空间上下文信息，从而有效缓解修复边界处的伪影。随后，初始特征依次通过 4 个 DAM 模块，利用不同空洞率的深度可分离卷积与残差聚合机制，增强上下文感知能力，实现局部纹理与全局

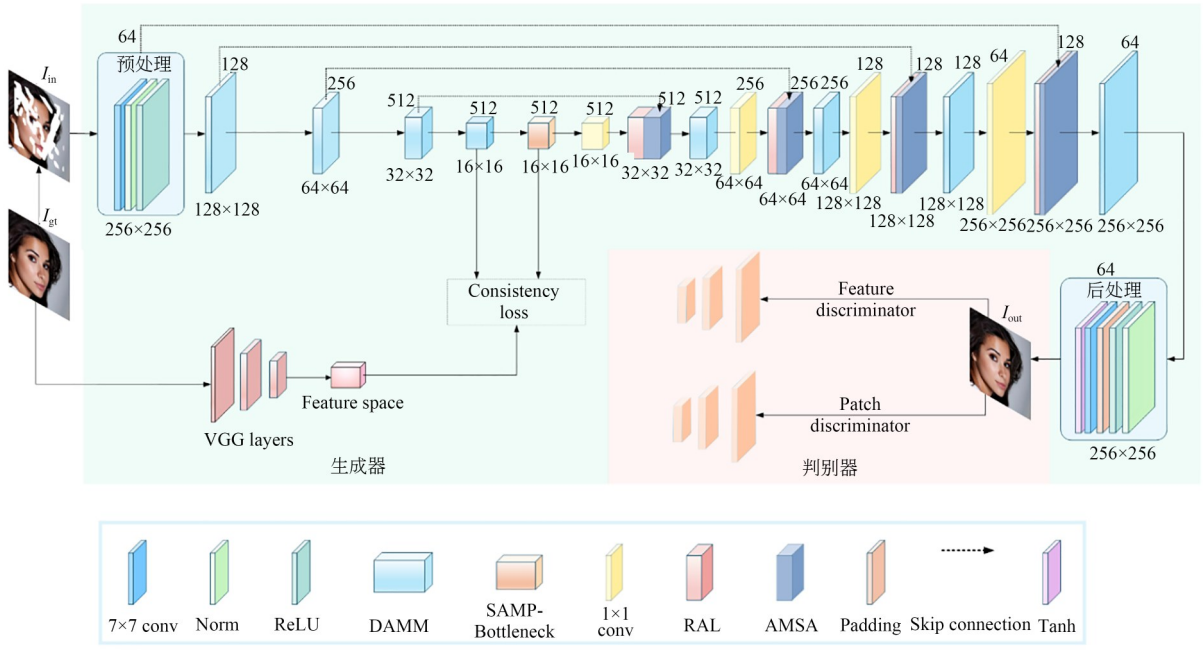

图 1　基于结构先验与多尺度残差聚合网络的整体框架

Fig. 1　Overall framework of structural priors and multi-scale residual aggregation net

结构的多尺度建模。同时，编码器通过跳跃连接将浅层的高频细节特征传递至解码器，从而改善纹理模糊问题。在瓶颈层设计Transformer路径与一致语义注意力路径的双路径并行机制，有效建模长距离依赖关系，输出全局语义信息与局部纹理信息，进一步改善大面积缺失场景下的结构扭曲与全局不一致问题。在解码阶段，解码器将瓶颈层输出的特征信息进行逐层上采样恢复图像分辨率。为解决传统简单拼接导致的语义鸿沟问题，解码器首先通过1×1卷积进行通道对齐并结合区域亲和度学习(Region Affinity Learning, RAL)模块进行初步内容填充。随后，利用AMSA模块动态融合来自解码器的深层上采样特征信息与编码器的浅层跳跃连接特征信息，显著提升修复图像的边界连续性与结构合理性。最终，经后处理生成纹理清晰、结构合理的修复结果I_{out}。

2.1.2　判别器

I_{out}在训练过程中受到双判别器系统的约束。其中，Patch discriminator作用于图像空间，通过对局部图像块进行判别，来监督生成内容局部纹理的真实性；Feature discriminator基于预训练的深度特征空间，通过判别特征分布，弥补Patch discriminator因感受野受限而导致的全局结构一致性不足，从语义层面对生成图像的结构合理性与全局一致性进行评估。两者共同驱动生成器在视觉感知与语义理解层面实现协同优化，从而生成具有结构完整性与细节保真度的高质量修复结果。

2.2　编码器

2.2.1　动态自适应多尺度模块

传统基于标准卷积的残差块存在感受野受限与特征融合方式较为静态两大局限性，导致修复结果出现纹理模糊与上下文不协调等问题，多尺度模块通过在不同感受野下并行提取与自适应融合特征，有效增强模型的上下文建模能力[23-24]。AOT-GAN[25]模块的多尺度上下文分支与门控融合机制，能够一定程度上处理结构缺失区域的长距离依赖问题。然而，其上下文建模过程受限于卷积的固有感受野，缺乏对结构信息的建模。基于此，本文提出DAM模块，增强上下文感知能力，实现全局一致性与局部细节表达。如图2所示，该模块由两条并行路径构成：用于建模复杂上下文特征的特征变换路径(Feature Transformation Path, FTP)和用于保持原始结构与细节信息的信息保留路径(Information Preservation Path, IPP)。整体设计上采用了多级残差结构，以维持特征在深层传递过程中的稳定性。包括3类：主干输入残差，多尺度空洞卷积内部残差和门控路径加权残差。

在FTP中，输入特征X经部分归一化和ReLU激活后得到X_{act}，X_{act}在Layer1中并行进入N个具有不同空洞率的深度可分离卷积分支中。本文采用的空洞卷积模块由一组具有递增膨胀率(Dilation rate=1，2，4，8)的并行卷积分支构成，给模型注入不同尺度的结构先验。在不显著增加参数量的情况下，通过扩大感受野捕获不同尺度下的上下文语义信息。随着网络层级的加深，模块逐步减少并行分支数量(N=4～1)，从而使模型在不同尺度下能够聚焦于不同范围的上下文建模需求。每个空洞卷积分支均采用ReLU激活并与前一层输出以残差形式相连接，从而缓解深层堆叠带来的梯度衰减问题，并保持特征表达的稳定性。分支内部依次包括反射填充(ReflectionPad2d，Pad)、3×3深度卷积(DW-Conv)、1×1逐点卷积(PW-Conv)及ReLU激活函数。融合后的特征首先输入至通道注意力模块(Channel Attention)[26]中，生成相应的动态权重W_i，随后，各分支的特征输出按照其权重进行加权融合，从而获得最终的多尺度特征表示$F(X)$，如式(1)所示：

$$F(X)=\sum_{i=1}^{N} w_i \cdot \mathrm{Branch}_i(X_{act}), \tag{1}$$

其中：N是并行分支总数，w_i是第i个分支的权重，$\mathrm{Branch}_i(\cdot)$是第i个多尺度处理分支的函数。该机制使模型能够根据图像内容自适应分配各尺度特征权重，从而实现多尺度特征的有效融合并增强上下文感知，进而缓解纹理模糊问题。

与此同时，X_{act}进入图2第二层门控路径(Gating path)中，经过深度可分离卷积、层归一化(Layer Normalization，Norm)与Sigmoid激活函数，生成自适应门控掩码M。该掩码自适应调节变换特征$F(X)$与原始特征X_{act}的融合比例，从而在特征更新与信息保留之间实现动态平衡，获得

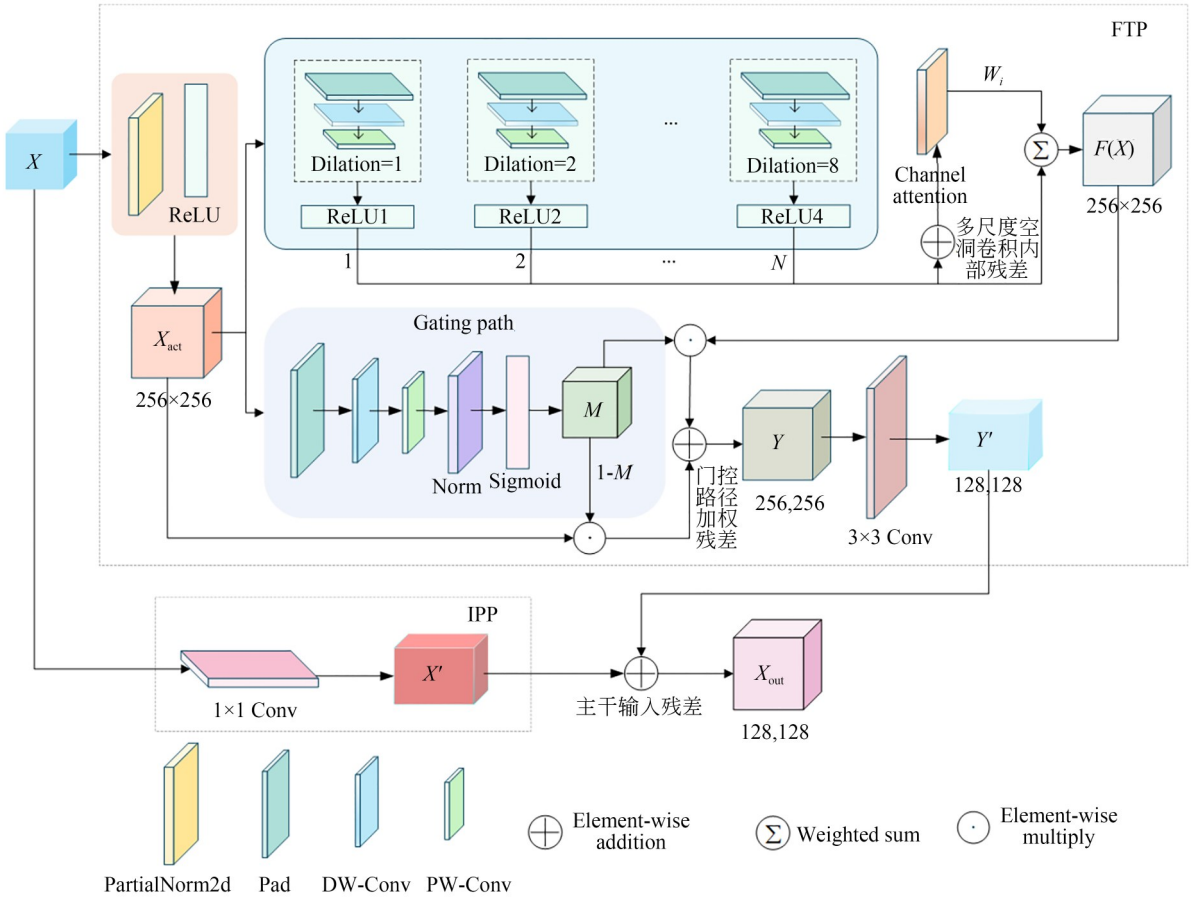

图 2　动态自适应多尺度模块

Fig. 2　Dynamic adaptive multi-scale block

融合特征 Y，如式(2)所示：

$$Y = M \odot F(X) + (1 - M) \odot X_{act}. \quad (2)$$

Y 通过 3×3 卷积层进行进一步特征提取，得到主路径输出 Y'。

在 IPP 中，为有效缓解梯度消失并维持结构完整性与边缘信息的传递，原始输入特征 X 通过 1×1 卷积进行投影，得到与 FPT 输出维度完全匹配的表示 X'。

最终，X' 与 Y' 通过逐元素相加实现残差融合，得到模块的最终输出 X_{out}。

2.2.2　结构感知双路径瓶颈模块

在大面积不规则破损图像修复任务中，网络瓶颈层往往面临全局结构建模与局部纹理生成之间的矛盾[27]。然而，传统瓶颈结构难以同时兼顾这两类特征的差异，易导致修复结果出现结构扭曲、纹理模糊及边界不连续等问题。李雄心等[28]提出混合特征提取块 HFEB，将 ResNet 残差块和 Transformer 与双重注意力结合的并行特征提取模块 PFEM 串联起来，提取图像局部特征，同时捕捉全局上下文信息，有助于更全面地理解图像内容。受此启发，本文提出 SADP-Bottleneck 模块。如图 3 所示，该模块采用双路径并行机制：Transformer 路径通过建模长距离依赖关系，增强模型对整体结构的感知能力，以确保修复区域的结构一致性；一致语义注意力路径从已知区域中获取纹理细节，以强化局部纹理的表达。两条路径通过交叉注意力融合进行信息交互，实现全局结构与局部纹理的协同建模。首先，输入特征图 F_{enc} 被展平并转置形成特征序列 X_{seq}，并依次输入由两个 Transformer Block 组成的全局语义路径。在每个 Transformer Block 内部，X_{seq} 经 RMS Norm 归一化后，通过独立的线性映射层 $\boldsymbol{Q}'$，$\boldsymbol{K}'$ 和 $\boldsymbol{V}'$ 分别生成查询($\boldsymbol{Q}_1$)、键($\boldsymbol{K}_1$)和值($\boldsymbol{V}_1$)三个张量。为保留空间位置信息，$\boldsymbol{Q}_1$，$\boldsymbol{K}_1$ 与 $\boldsymbol{V}_1$ 经 rearrange 函数重塑后，引入二维旋转位置编码(2D-RoPE)进行变换。随后，经过位置编

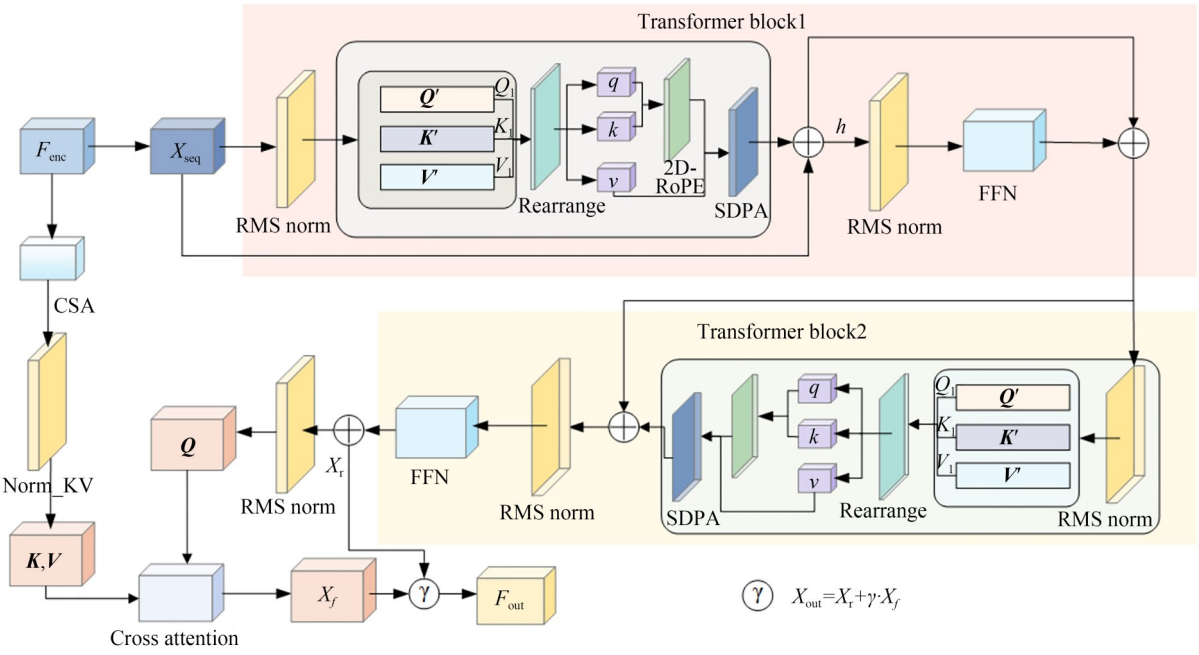

图 3　结构感知双路径瓶颈模块

Fig. 3　Structure-aware dual-path bottleneck block

码的 q,k 与重塑后的 v 一起通过缩放点积注意力(Scaled Dot-Product Attention, SDPA)完成全局信息聚合,从而克服传统卷积的局部感受野限制。聚合结果与输入序列 $\boldsymbol{X}_{\text{seq}}$ 通过残差连接形成中间特征 h,为后续的语义交互与特征融合提供全局依赖支撑,如式(3)所示:

$$h=\boldsymbol{X}_{\text{seq}}+A(\text{RMSNorm}(\boldsymbol{X}_{\text{seq}})),\tag{3}$$

其中,$A(\cdot)$ 表示注意力函数,计算公式如式(4)所示:

$$A(\boldsymbol{X}')=\boldsymbol{P}(\text{SDPA}(\boldsymbol{Q}_r,\boldsymbol{K}_r,\boldsymbol{V})),\tag{4}$$

其中:$\boldsymbol{Q}_r$,$\boldsymbol{K}_r$ 是 2D-RoPE 变换后的结果,$\boldsymbol{P}(\cdot)$ 是映射函数。

最后,经过 RMSNorm 归一化与前馈网络(Feed-Forward Network, FFN)变换后,特征 h 在 Transformer2 中重复执行与 Transformer1 相同的处理流程,输出特征 $\boldsymbol{X}_r$,以进一步编码长距离依赖关系的语义信息。经过归一化后得到的查询矩阵 $\boldsymbol{Q}$ 可有效捕获全局语义依赖,从而增强模型的全局上下文建模能力。与此同时,在一致语义注意力路径中,$\boldsymbol{F}_{\text{enc}}$ 先进入 CSA[29](Coherent Semantic Attention)模块,该模块通过引入图像有效区域的纹理和语义先验,实现对破损区域的合理填充,缓解纹理模糊的问题。接着经 RMSNorm 归一化后形成键($\boldsymbol{K}$)和值($\boldsymbol{V}$)。最终,两条路径分别得到的 $\boldsymbol{Q}$ 与 $\boldsymbol{K}$,$\boldsymbol{V}$ 在交叉注意力融合(CrossAttention)模块融合。融合后的特征通过带有可学习缩放因子 γ 的残差连接,与 Transformer 路径的输出 $\boldsymbol{X}_r$ 相结合,得到最终的瓶颈层输出 F_{out}。

2.2.3　自适应多尺度聚合模块

AMSA 模块的整体架构如图 4 所示,其接收输入特征图 $\boldsymbol{P}\in\mathbf{R}^{B\times C\times W}$,在 Path A 中,输入特征 P 首先经由 3×3 卷积层提取局部信息,并通过 ReLU 函数激活,随后经过 1×1 卷积保持通道数。通过 Softmax 函数在尺度维度上进行归一化,生成空间位置自适应权重图 $\boldsymbol{w}_i\in\mathbf{R}^{B\times K\times H\times W}$。与此同时,输入特征 P 被送入 Path B,该路径由 k 个带有不同空洞率 $r_1\sim r_k$ 的 3×3 空洞卷积与 ReLU 激活函数组成,以捕获不同感受野的上下文信息,从而生成一组多尺度特征图 $F_1\sim F_k$,如式(5)所示:

$$F_i=\text{ReLU}\left(\text{Conv}_{3\times 3}^{r=r_i}(x)\right),i=1,\cdots,k,\tag{5}$$

其中,$F_i\in\mathbf{R}^{B\times C\times H\times W}$ 表示空洞率为 r_i 的卷积分支输出。

最终,权重图 WeightMap 的第 i 个通道 $\boldsymbol{W}_i$ 与对应的特征图 F_i 进行逐元素相乘,利用自适应权重对不同尺度的特征进行调整。加权特征图最终进行逐元素求和,生成输出特征图 $\boldsymbol{P}_{\text{out}}$,如式

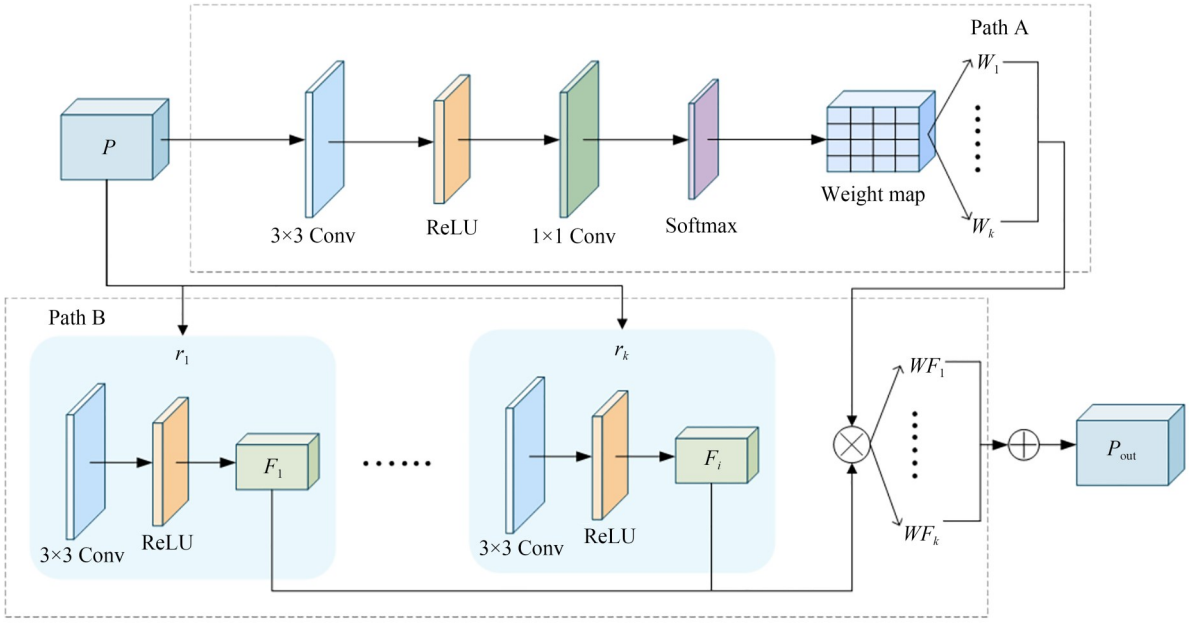

图 4　自适应多尺度聚合模块

Fig. 4　Adaptive multi-scale aggregation block

(6)所示：

$$P_{out}=\sum_{i=1}^{K}W_i\odot F_i, \tag{6}$$

其中：⊙表示逐元素相乘。

2.2.4　判别器

本文采用 Patch discriminator 与 Feature discriminator 双判别器系统，以实现对修复图像从像素级到语义级的多维度监督。两类判别器均与生成器采用对抗训练框架，Patch discriminator 作用于图像空间，通过对局部图像块的判别，监督生成内容的局部纹理真实性；Feature discriminator 则对深层特征进行语义级判别。通过对特征分布的约束提升生成结果的结构一致性。为在训练过程中兼顾局部细节与高层语义特征的对齐，将图像域对抗损失的权重设置为 $\lambda_{patch}=1.0$，特征域对抗损失的权重设置为 $\lambda_{feat}=0.5$。该配置经实验验证能够在保证局部纹理真实性的同时，有效引导高层特征对齐，从而实现生成结果在局部细节与全局结构上的均衡优化。

2.3　损失函数

为确保修复结果在纹理细节、结构一致性及整体视觉质量方面的优化，在训练过程中联合多种损失函数，具体包括重建损失、一致性损失以及对抗损失[29]。最终的目标函数定义如式(7)所示：

$$\mathcal{L}=\lambda_r\mathcal{L}_{re}+\lambda_c\mathcal{L}_c+\lambda_d\mathcal{L}_{adv}, \tag{7}$$

其中：$\mathcal{L}_{re}$，$\mathcal{L}_c$，$\mathcal{L}_{adv}$ 分别表示重建损失、一致性损失、对抗损失，λ_r，λ_c，λ_d 分别为 3 类损失的权衡系数，$\lambda_r=1$，$\lambda_c=0.01$，$\lambda_d=0.002$。

2.3.1　重建损失

为约束修复图像与真实图像之间的像素级差异，采用 L1 范数作为重建损失，如式(8)所示：

$$\mathcal{L}_{re}=\|I_{out}-I_{gt}\|_1, \tag{8}$$

其中：I_{out} 表示修复结果，I_{gt} 表示真实图像，该损失有助于恢复整体结构与基础纹理信息。

2.3.2　一致性损失

传统的感知损失虽能提升语义层次的感知能力，但无法直接约束 CSA 层及其对应解码层的特征一致性，可能导致训练过程不稳定。为此，引入一致性损失，通过对齐 CSA 模块与解码器对应层的特征表征，实现特征层级的有效约束。

具体而言，利用 ImageNet 预训练的 VGG-16 提取第 4-3 层的高层特征作为监督信号，对于缺损区域 M，一致性损失定义如式(9)所示：

$$\mathcal{L}_c=\sum_{y\in M}(\|\mathrm{CSA}(I_{in})_y-\Phi_n(I_{gt})_y\|_2^2+\|\mathrm{CSA}_d(I_{in})_y-\Phi_n(I_{gt})_y\|_2^2), \tag{9}$$

其中：I_{in} 表示带掩码的输入图像，$\Phi_n(\cdot)$ 表示 VGG-

16第n层的激活特征，CSA(·)与CSA_d(·)分别表示CSA模块与解码器的特征输出。该损失在保证语义一致性的同时，有效减少结构扭曲与细节缺失问题。

2.3.3　对抗损失

为提升修复图像在真实感与细节还原方面的表现，引入双判别器机制。首先，Feature discriminator基于VGG-16 pool 3层特征图，通过下采样获取Markovian patches[30]的统计特征，并在特征空间计算对抗损失，从而有效增强全局一致性与语义合理性。其次，Patch discriminator直接在像素空间对比修复结果I_{out}与真实图像I_{gt}，以确保局部纹理的精细度与真实性。

对抗损失采用相对平均最小二乘损失，分别定义生成器与判别器的目标函数。如式(10)和式(11)所示：

$$\mathcal{L}_G=-\boldsymbol{E}_{I_{gt}}[D(I_{gt},I_r)^2]-\boldsymbol{E}_{I_r}[(1-D(I_r,I_{gt}))^2], \tag{10}$$

$$\mathcal{L}_D=-\boldsymbol{E}_{I_{gt}}[(1-D(I_{gt},I_r))^2]-\boldsymbol{E}_{I_r}[D(I_r,I_{gt})^2], \tag{11}$$

其中：$D(\cdot)$表示判别器，$\boldsymbol{E}_{I_{gt}}$，$\boldsymbol{E}_{I_r}$分别表示对真实样本与生成样本的期望。

3　实验结果与分析

3.1　实验设置

实验环境在Window11系统，python3.11和pytorch2.1.2环境下进行，硬件环境为Intel® Core™i5-12600KF中央处理器，16G的随机访问存储器，采用NVIDIA GeForce RTX 4060Ti图形处理器，所有的对比实验均在相同环境下进行。在模型训练阶段，采用Adam优化器对网络参数进行优化，批次大小为2，训练周期为200轮，前100轮保持学习率不变，后100轮学习率线性衰减至趋于0。评价指标采用峰值信噪比(PSNR[31])、结构相似性(SSIM[32])和学习感知图像块相似度(LPIPS[33])3种客观评价指标对修复结果进行分析。为全面评估所提出方法在不同应用场景中的泛化能力与修复效果，采用CelebA、Places2及Paris StreetView 3个公开数据集验证模型性能。从每个数据集中选取8 000张图片作为训练集，1 000张图片作为测试集，输入图片大小为256×256。本模型总参数量约58.5 M，在训练阶段峰值显存约13.6 GB，推理速度为0.01 s/张。该模型计算效率可控、对多类破损图像具有良好泛化性，能够在实际场景中保持稳定可靠的修复性能。此外，为更真实地模拟自然图像中常见的不规则遮挡场景，统一采用标准的不规则掩码集合进行训练与测试。

为增强主观评价的完整性与说服力，展示前期工作中修复效果欠佳的典型案例，涵盖人脸、街景、建筑及敦煌壁画4种场景。如图5所示，人脸的眉毛和眼睛修复的结构边界不清晰；街景图像中窗口区域与树叶部分出现模糊现象；在建筑场景中，屋顶及窗户区域的修复结果不完整；在敦煌壁画的修复示例中，人脸部位未能成功修复。通过对比分析模型在复杂纹理、弱结构边界、大面积破损条件下的表现，为论文改进方向提供依据。

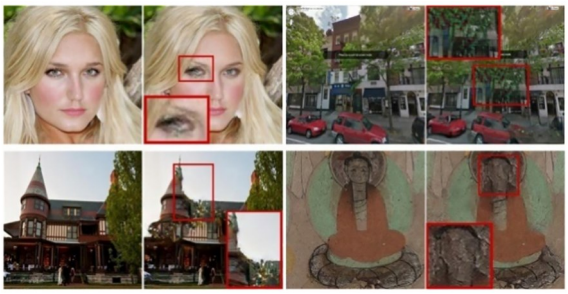

图5　典型案例对比

Fig. 5　Comparison of typical cases

3.2　对比实验

将本文方法与SWMH Transformer，AOT-GAN，ARPG[34]，DNNAM这4种方法在CelebA，Paris StreetView和Places2 3个数据集上进行全面的定性和定量对比，实验结果表明，本文算法在不同数据集上均有明显的性能提升，在PSNR，SSIM，LPIPS评价指标上，本文算法表现出明显优势。这3项指标的提升表明，模型在缓解局部伪影与纹理模糊的同时，在整体结构一致性与视觉自然度上也得到改善。由于PSNR，SSIM偏向低层结构保真，而LPIPS关注高层感知质量，它们之间的互补关系进一步验证方法在多层次质量维度上的有效性，从而增强实验结果的解释性与可信度。

SWMH采用条带式窗口自注意力机制，但条带划分会限制局部区域的特征表达能力，使模型难以充分捕获细节信息；AOT-GAN依赖多尺度上下文聚合来推断缺失区域的整体语义，但其特征建模更侧重结构一致性；ARPG采用随机顺序的自回归生成机制，过度依赖全局上下文而缺乏对局部空间连续性的约束；DNNAM的全局注意力在大面积缺失区域中过度依赖语义推断而缺乏对局部高频纹理的建模。本文模型通过对全局结构一致性与局部高频纹理的协同建模，从而在复杂纹理修复与边缘连续性方面表现出更优的修复质量，在整体结构保持与细节上均取得了更为突出的修复效果。

图6为在CelebA数据集中各个模型的修复对比图，掩码几乎覆盖人脸五官。SWMH算法在处理简单纹理特征时效果较好，但在处理复杂纹理时存在模糊问题，如图6第二列瞳孔部位存在模糊现象；AOT-GAN算法虽能生成合理的逻辑结构，但对于边缘区域的修复不够清晰，如图6第三列牙齿和眉毛部分出现结构不清晰和断裂；ARPG算法对全局上下文的依赖性较强，对于复杂纹理的修复效果欠佳，如图6第四列牙齿在视觉上较为突兀；DNNAM算法虽然能够保持整体结构的合理性，但在局部高频纹理恢复上不够精细，如图6第五列出现上嘴唇不自然且眼睛模糊，影响整体真实感。相较于前面提到的4种对比算法，本文算法在修复过程中更注重局部和全局信息的融合，具有更好的修复效果，在整体结构和局部细节上均展现出较好的修复结果。

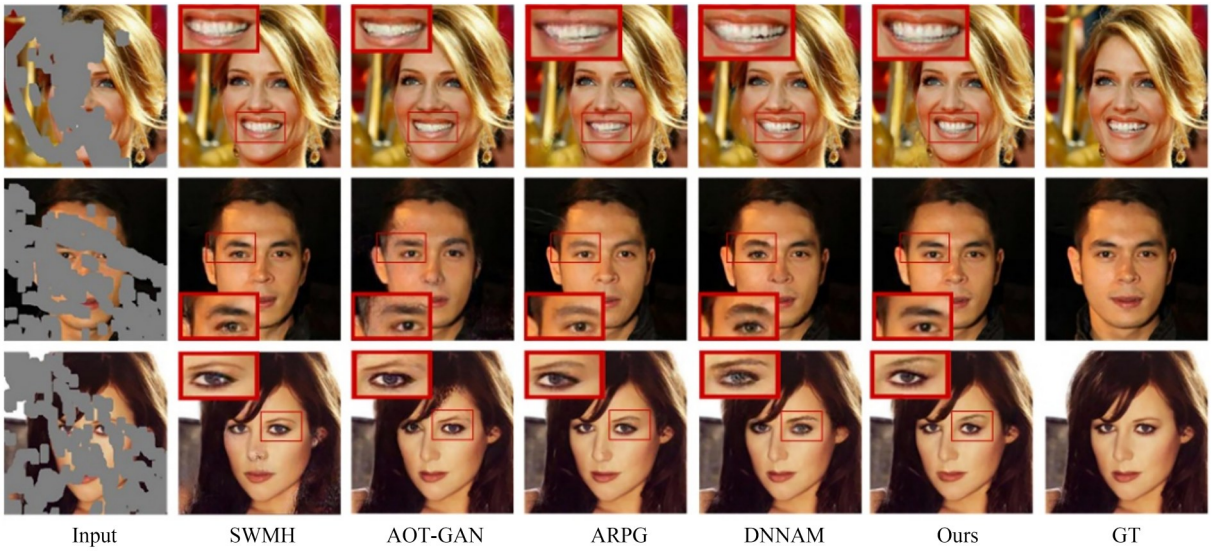

图6　CelebA数据集修复对比

Fig. 6　Repair comparison on CelebA dataset

表1为本文算法与其他4种算法在CelebA数据集上的定量对比结果。其中最优结果使用粗体表示，↑表示数值越大效果越好，↓表示数值越小效果越好。

表1　CelebA数据集定量分析

Tab. 1　Quantitative analysis of CelebA dataset

修复算法	PSNR/dB(↑)	SSIM(↑)	LPIPS(↓)
SWMH	22.95	0.852	0.097
AOT-GAN	23.57	0.861	0.104
ARPG	24.23	0.843	0.098
DNNAM	24.65	0.857	0.087
Ours	**25.82**	**0.885**	**0.079**

图7为在Paris StreetView数据集中各个模型的修复对比图。SWMH算法能将整体结构修复出来，但特征提取能力差，造成修复结果模糊。AOT-GAN算法对结构边缘特征提取能力较差，修复结果会出现伪影，如图7第三列窗口。ARPG算法对局部语义理解不到位，面对复杂结构时会与真实图像出现差异。DNNAM算法面对复杂背景修复结果出现混乱，如图7第五列楼梯

图 7　Paris StreetView 数据集修复对比

Fig. 7　Repair comparison on Paris StreetView dataset

与窗口。相比之下，本文算法在结构、纹理、语义和边界处理上都能达到更好的效果。表 2 为本文算法与其他 4 种算法在 Paris StreetView 数据集上的定量对比结果。

图 8 为在 Places2 数据集中各个算法的修复对比图。SWMH 算法在修复大面积缺失区域时，对纹理结构的修复效果较差，如图 8 第二列窗口和屋顶部分。AOT-GAN 算法、ARPG 算法对局部上下文不敏感，对细节处的几何结构修复能

表 2　Paris StreetView 数据集定量分析

Tab. 2　Quantitative analysis of Paris StreetView dataset

修复算法	PSNR/dB(↑)	SSIM(↑)	LPIPS(↓)
SWMH	22.13	0.781	0.163
AOT-GAN	21.56	0.795	0.198
ARPG	22.89	0.772	0.135
DNNAM	23.31	0.788	0.129
Ours	**24.97**	**0.821**	**0.114**

图 8　Places2 数据集修复对比

Fig. 8　Repair comparison on Places2 dataset

力较差。本文算法的修复效果显然优于其他算法。表 3 为本文算法与其他 4 种算法在 Places2 数据集上的定量对比结果。

表 3　Places2 数据集定量分析

Tab. 3　Quantitative analysis of Places2 dataset

修复算法	PSNR/dB(↑)	SSIM(↑)	LPIPS(↓)
SWMH	22.05	0.763	0.197
AOT-GAN	21.33	0.772	0.161
ARPG	21.81	0.755	0.145
DNNAM	22.69	0.768	0.153
Ours	**23.85**	**0.794**	**0.137**

为进一步评估模型在真实应用场景中的表现，引入敦煌壁画的实际破损图像开展补充实验，从艺术遗产保护的视角验证模型在真实场景下的泛化能力与稳健性。实验结果表明：SWMH 算法的修复结果不完整、AOT-GAN 算法对缺失区域修复后出现局部特征模糊，如图 9 第二列、第三列；ARPG 与 DNNAM 算法在修复区域出现色彩不均、结构扭曲问题，如图 9 第四列、第五列。相比之下，本文算法在大面积剥落区域中能够修复壁画的纹理细节，并在结构一致性方面呈现出更高稳定性。表 4 为本文算法与其他 4 种算法在真实破损壁画上的定量对比结果。

表 4　真实破损壁画修复实验的定量分析

Tab. 4　Quantitative analysis of authentic damaged mural restoration experiments

修复算法	PSNR/dB(↑)	SSIM(↑)	LPIPS(↓)
SWMH	21.86	0.658	0.183
AOT-GAN	22.79	0.713	0.166
ARPG	23.14	0.737	0.152
DNNAM	23.69	0.762	0.146
Ours	**24.17**	**0.804**	**0.139**

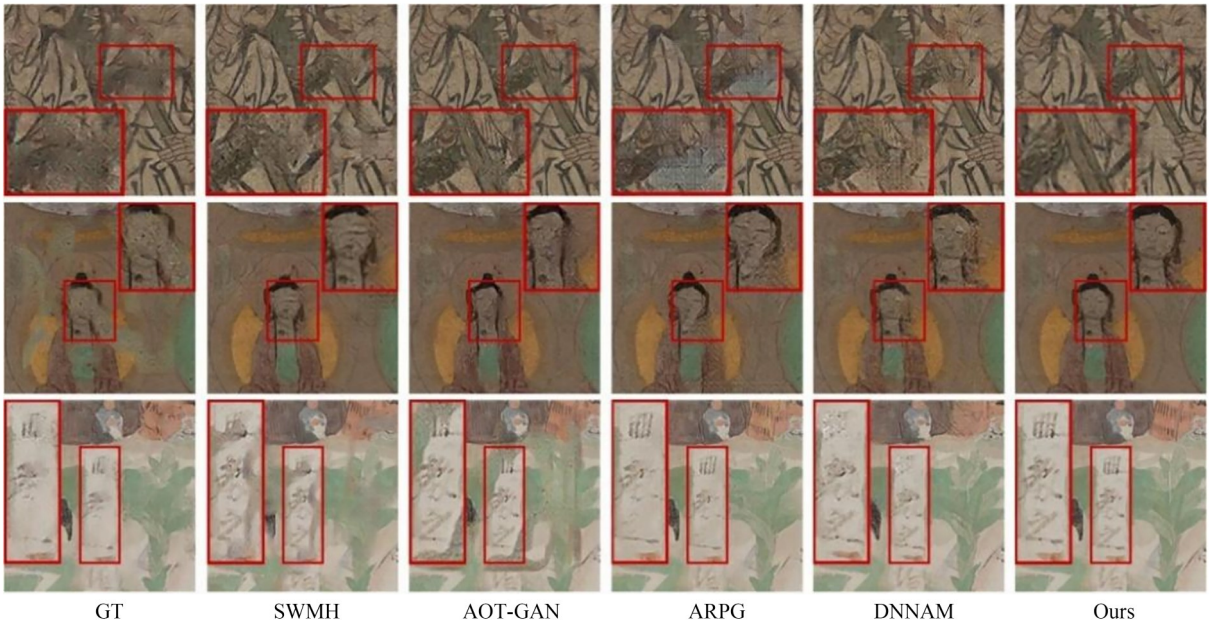

图 9　真实破损壁画修复结果对比

Fig. 9　Comparison of restoration results for authentic damaged murals

3.3　消融实验

为系统性验证本文提出的关键模块 DAM，SADP-Bottleneck 以及 AMSA 在整体修复性能中的有效性，采取并展示最为典型的 CelebA 数据集上的结果。通过逐步引入各个模块的核心组成部分，以验证各组成部分对整体修复性能的贡献。如表 5 所示，基线网络采用编码器-解码器框架，并在编码器与解码器之间设置跳跃连接，以实现高低层特征的融合，共包含 8 个残差块。

表 5　基线模型

Tab. 5　Baseline model

序号	基线模型结构	输入通道数	输出通道数	对应残差
①	3×3 Conv	64	64	⑯
②	4×4 Dilated Conv＋3×3 Conv	64	128	⑮
③	4×4 Dilated Conv＋3×3 Conv	128	256	⑭
④	4×4 Dilated Conv＋3×3 Conv＋CSA	256	512	⑬
⑤	4×4 Dilated Conv＋3×3 Conv	512	512	⑫
⑥	4×4 Dilated Conv＋3×3 Conv	512	512	⑪
⑦	4×4 Dilated Conv＋3×3 Conv	512	512	⑩
⑧	4×4 Dilated Conv＋3×3 Conv	512	512	⑨
⑨	4×4 Conv＋3×3 Conv＋3×3 Conv	512	512	⑧
⑩	4×4 Conv＋3×3 Conv＋3×3 Conv	512	512	⑦
⑪	4×4 Conv＋3×3 Conv＋3×3 Conv	512	512	⑥
⑫	4×4 Conv＋3×3 Conv＋3×3 Conv	512	512	⑤
⑬	4×4 Conv＋3×3 Conv＋3×3 Conv	512	512	④
⑭	4×4 Conv＋3×3 Conv＋3×3 Conv	512	256	③
⑮	4×4 Conv＋3×3 Conv＋3×3 Conv	256	128	②
⑯	4×4 Conv＋3×3 Conv	128	64	①

注：①～⑧为编码器，⑨～⑯为解码器

3.3.1　动态自适应多尺度模块

基线模型在缺失区域的修复结果中因其有限的感受野与静态特征融合方式，常表现出纹理模糊和上下文不协调问题，导致修复结果不理想。基于此，本研究在基线模型上逐步引入多尺度建模与动态融合，M1通过引入多尺度空洞卷积分支扩展有效感受野，使模型能够在不同尺度下并行建模上下文信息，因此在PSNR与SSIM上较基线模型有一定提升。然而，由于多尺度特征采用静态融合方式，各分支贡献无法根据内容自适应调节，且并行分支带来一定参数冗余，其纹理细节表达仍受限制。在此基础上，M2引入通道注意力对多尺度分支进行加权，并采用深度可分离卷积降低计算冗余。结果显示，M2在参数量降低约28%的情况下，性能与M1基本持平。在M2基础上，通过对各路径进行动态融合，构成完整的DAM模块。如表6所示，M3在各项指标上均取得最优结果，显著提升了修复质量与模型稳定性。

表 6　动态自适应多尺度模块分析

Tab. 6　Adaptive multi-scale aggregation block analysis

模块	描述	PSNR/dB(↑)	SSIM(↑)	LPIPS(↓)
基线模型	ResBlock U-Net	22.76	0.798	0.157
M1	Baseline＋多尺度分支	23.57	0.813	0.129
M2	M1＋深度可分离	23.98	0.815	0.126
M3	M2＋动态融合	**24.16**	**0.839**	**0.105**

本文DAM模块采用递增空洞率(d=1,2,4,8)的深度可分离卷积分支扩展有效感受野，使各分支分别聚焦于局部细粒度纹理(d=1)、中尺度结构关系(d=2,4)及全局轮廓与语义依赖(d=8)。该局部到全局的分层感受野布局可稳定覆盖大尺度破损场景中的不同范围上下文，从而抑制纹理模糊并提升结构连贯性；反之，仅采用低膨胀率会削弱长距离结构恢复，而仅采用高膨胀率则易丢失局部细节，递增空洞率的组合实现了局部细节与全局语义的多尺度特征建模平衡。

3.3.2　结构感知双路径瓶颈模块

尽管DAM模块显著增强了编码器的多尺度上下文感知能力，但在处理长距离依赖关系时仍存在全局结构不一致与远距离伪影等问题。为此，对瓶颈层结构进行改进，分析不同路径的提升效果。首先，在M3基础上将瓶颈层替换为仅包含Transformer路径的模块，单独使用Transformer路径可以较好地恢复全局结构与语义一致性。如表7所示，单路径Transformer的PSNR达到24.81，但在细节纹理与破损边界处常出现纹理模糊。随后，构建仅包含一致语义注意力（CSA）路径的瓶颈结构。该模块通过上下文自注意力实现非局部稀疏特征匹配，从而增强局部纹理重建能力。在细节纹理恢复和边缘对齐方面表现更为突出，其PSNR达到24.69，如表7所示。但在大面积结构缺失的场景中，由于局部匹配依赖性增强，易出现语义重建不足、结构扭曲或因错误匹配带来的伪影现象。最后，验证完整SADP-Bottleneck模块。保留Transformer的全局语义推理能力并借助CSA提供的局部纹理信息，从而在结构完整性与边界细节上达到更好的修复效果，在所有定量指标上均优于任意单一路径。

表7　结构感知双路径瓶颈模块分析

Tab. 7　Structure-aware dual-path bottleneck block analysis

模块	描述	PSNR/dB（↑）	SSIM（↑）	LPIPS（↓）
M4	M3＋单路径Transformer	24.81	0.831	0.097
M5	M3＋单路径一致语义注意力	24.69	0.802	0.099
M6	M3＋SADP-Bottleneck	**25.43**	**0.843**	**0.083**

3.3.3　自适应多尺度聚合模块

M6为传统U-Net将前一层输出特征与跳跃连接特征直接拼接，导致修复结果出现边界伪影。为此，M7引入AMSA模块对跳跃连接特征进行自适应融合。其通过多空洞率卷积分支并行建模不同感受野的上下文信息，并利用Softmax归一化生成空间位置对应的自适应权重，对多尺度特征进行动态重加权，实现对不同空间位置尺度的自适应加权。该机制有效缓解了特征融合阶段的尺度冲突与边界不稳定问题，使解码阶段的特征过渡更加平滑自然。如表8所示，各项指标均得到优化。

表8　自适应多尺度聚合模块分析

Tab. 8　Adaptive multi-scale aggregation block analysis

模块	描述	PSNR/dB（↑）	SSIM（↑）	LPIPS（↓）
M6	直接拼接	25.43	0.843	0.083
M7	M6＋AMSA	**25.82**	**0.885**	**0.079**

4　结　论

本文提出一种基于结构先验融合与多尺度残差传递的大面积破损图像修复模型，以解决大面积不规则缺失场景下的纹理失真、结构扭曲与语义鸿沟问题。在编码器阶段，设计DAM模块以提升上下文感知能力，有效缓解纹理模糊问题；瓶颈层引入Transformer路径与一致语义注意力路径，并通过交叉注意力机制实现全局语义信息与局部纹理信息的融合，从而缓解结构扭曲问题。在解码阶段，构建AMSA模块，提升细节重建质量。经实验评估，本文算法在视觉效果和定量指标上均优于现有方法，在CelebA数据集上PSNR提升3.06 dB，说明模型有效减少了像素级误差；SSIM提升0.087，表明结构一致性得到增强；LPIPS降低0.078则反映生成结果在深层特征空间中生成的内容与真实图像更加接近。

生成结果在结构一致性与纹理真实感方面均具显著优势。尽管模型在性能上表现突出，但在处理复杂随机纹理及降低计算复杂度方面仍存在挑战。未来工作将重点探索高分辨率场景下的图像修复任务及模型轻量化设计，以平衡模型的修复性能与效率。

作者贡献声明：

魏赟：论文写作指导，论文框架提供；

贾白雪：研究思路确定与方法框架搭建，实验设计与论文撰写；

冯丹丹：数据预处理与参数调优；

张治瑞：实验结果分析与可视化；

辛子昊：理论推导与模型性能验证；

邬侗辰：论文审阅与编辑。

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 W X，TIAN Q CH，LIAN L，*et al*. Research progress of image inpainting methods based on deep learning[J]. *Computer Engineering and Applications*，2024，60(22)：58-73.（in Chinese）

[2] YU J H，LIN Z，YANG J M，*et al*. Generative image inpainting with contextual attention[C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition. June* 18-23，2018，*Salt Lake City*，*UT*，*USA*. IEEE，2018：5505-5514.

[3] CORNEANU C，GADDE R，MARTINEZ A M. LatentPaint：image inpainting in latent space with diffusion models[C]. 2024 *IEEE/CVF Winter Conference on Applications of Computer Vision*（*WACV*）. *January* 3-8，2024，*Waikoloa*，*HI*，*USA*. IEEE，2024：4322-4331.

[4] LIU H P，WANG Y，QIAN B，*et al*. Structure matters：tackling the semantic discrepancy in diffusion models for image inpainting[C]. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*（*CVPR*）. *June* 16-22，2024，*Seattle*，*WA*，*USA*. IEEE，2024：8038-8047.

[5] GUO X F，YANG H Y，HUANG D. Image inpainting *via* conditional texture and structure dual generation[C]. 2021 *IEEE/CVF International Conference on Computer Vision*（*ICCV*）. *October* 10-17，2021，*Montreal*，*QC*，*Canada*. IEEE，2022：14114-14123.

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

[7] WANG W L，JIA Y J. Damaged region filling and evaluation by symmetrical exemplar-based image inpainting for Thangka[J]. *EURASIP Journal on Image and Video Processing*，2017，2017(1)：38.

[8] 李雪涛，王耀雄，高放. 图像修复方法综述[J]. 激光与光电子学进展，2023，60(2)：0200002.
LI X T，WANG Y X，GAO F. Review of image inpainting methods[J]. *Laser & Optoelectronics Progress*，2023，60(2)：0200002.（in Chinese）

[9] RONNEBERGER O，FISCHER P，BROX T. *U-Net*：*Convolutional Networks for Biomedical image Segmentation*[M]. Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Cham：Springer International Publishing，2015：234-241.

[10] GOODFELLOW I J，POUGET-ABADIE J，MIRZA M，*et al*. Generative adversarial nets[C]. *NIPS'14*：*Proceedings of the* 28*th International Conference on Neural Information Processing Systems*. 2014：2672- 2680.

[11] CHEN Y T，XIA R L，YANG K，*et al*. DNNAM：Image inpainting algorithm *via* deep neural networks and attention mechanism[J]. *Applied Soft Computing*，2024，154：111392.

[12] 雷帅，仇明鑫，柳先辉，等. 基于多尺度深度可分离ResNet的废弃家电回收图像分类模型[J]. 计算机科学，2025，52(S1)：389-395.
LEI SH，QIU M X，LIU X H，*et al*. Image classification model for waste household appliance recycling based on multi-scale depthwise separable ResNet[J]. *Computer Science*，2025，52(S1)：389-395.（in Chinese）

[13] 陈清江，陈鹏民. 多维度聚合Transformer的图像超分辨率重建[J]. 光学 精密工程，2025，33(12)：1955-1970.
CHEN Q J，CHEN P M. MDAT：Multi-dimensional aggregation transformer for image super-resolution reconstruction[J]. *Opt. Precision Eng.*，2025，33(12)：1955-1970.（in Chinese）

[14] CHEN B W，LIU T J，LIU K H. Lightweight image inpainting by stripe window transformer with joint attention to CNN[C]. 2023 *IEEE* 33*rd International Workshop on Machine Learning for Signal Processing*（*MLSP*）. *September* 17-20，2023，*Rome*，*Italy*. IEEE，2023：1-6.

[15] HUANG W L，DENG Y，HUI S Q，*et al*. Sparse self-attention transformer for image inpainting[J]. *Pattern Recognition*，2024，145：109897.

[16] LIU J L，GONG M G，GAO Y，*et al*. Bidirectional interaction of CNN and Transformer for image inpainting[J]. *Knowledge-Based Systems*，2024，299：112046.

[17] WANG H N, CAO P, YANG J Z, *et al*. Narrowing the semantic gaps in U-Net with learnable skip connections: The case of medical image segmentation[J]. *Neural Networks*, 2024, 178: 106546.

[18] ATES G C, MOHAN P, CELIK E. Dual Cross-Attention for medical image segmentation[J]. *Engineering Applications of Artificial Intelligence*, 2023, 126: 107139.

[19] SUN G Q, PAN Y Z, KONG W K, *et al*. DA-TransUNet: integrating spatial and channel dual attention with transformer U-Net for medical image segmentation[J]. *Frontiers in Bioengineering and Biotechnology*, 2024, 12: 1398237.

[20] 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*). *July* 21-26, 2017, *Honolulu*, *HI*, *USA*. IEEE, 2017: 5967-5976.

[21] PARK S J, SON H, CHO S, *et al*. SRFeat: single image super-resolution with feature discrimination[C]. *Computer Vision-ECCV* 2018. Cham: Springer, 2018: 455-471.

[22] CELARD P, IGLESIAS E L, SORRIBES-FDEZ J M, *et al*. A survey on deep learning applied to medical images: from simple artificial neural networks to generative models[J]. *Neural Computing and Applications*, 2023, 35(3): 2291-2323.

[23] 邬开俊, 张治瑞, 汪滢, 等. 全局感知与多尺度特征融合的城市道路语义分割[J]. 光学 精密工程, 2025, 33(14): 2262-2277.
WU K J, ZHANG ZH R, WANG Y, *et al*. Urban road semantic segmentation with global awareness and multi-scale feature fusion[J]. *Opt. Precision Eng.*, 2025, 33(14): 2262-2277. (in Chinese)

[24] 谢欣丹, 李晓艳, 王鹏, 等. 基于多尺度残差特征融合的图像去雾算法[J]. 液晶与显示, 2024, 39(6): 822-832.
XIE X D, LI X Y, WANG P, *et al*. Image defogging algorithm based on multi-scale residual feature fusion[J]. *Chinese Journal of Liquid Crystals and Displays*, 2024, 39(6): 822-832. (in Chinese)

[25] ZENG Y H, FU J L, CHAO H Y, *et al*. Aggregated contextual transformations for high-resolution image inpainting[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2023, 29(7): 3266-3280.

[26] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition. June* 18-23, 2018, *Salt Lake City*, *UT*, *USA*. IEEE, 2018: 7132-7141.

[27] 张家骏, 廉敬, 刘冀钊, 等. 利用图像平滑结构信息指导图像修复[J]. 光学 精密工程, 2024, 32(4): 549-564.
ZHANG J J, LIAN J, LIU J ZH, *et al*. Using image smoothing structure information to guide image inpainting[J]. *Opt. Precision Eng.*, 2024, 32(4): 549-564. (in Chinese)

[28] 李雄心, 夏丰领, 张靠民, 等. 强化边缘特征提取的双分支融合真实图像去雾[J]. 光学 精密工程, 2025, 33(2): 247-261.
LI X X, XIA F L, ZHANG K M, *et al*. Enhanced edge feature extraction dual branch fusion network for real image dehazing[J]. *Opt. Precision Eng.*, 2025, 33(2): 247-261. (in Chinese)

[29] LIU H Y, JIANG B, XIAO Y, *et al*. Coherent semantic attention for image inpainting[C]. 2019 *IEEE/CVF International Conference on Computer Vision* (*ICCV*). *October* 27-*November* 2, 2019. *Seoul*, *Korea*. IEEE, 2019: 4169-4178.

[30] LI C, WAND M. Precomputed real-time texture synthesis with Markovian generative adversarial networks[C]. *Computer Vision-ECCV* 2016. Cham: Springer, 2016: 702-716.

[31] HORÉ A, ZIOU D. Image quality metrics: PSNR *vs*. SSIM[C]. 2010 20*th International Conference on Pattern Recognition. August* 23-26, 2010, *Istanbul*, *Turkey*. IEEE, 2010: 2366-2369.

[32] 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.

[33] ZHANG R, ISOLA P, EFROS A A, *et al*. The unreasonable effectiveness of deep features as a per-

ceptual metric[C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition. June* 18-23, 2018, *Salt Lake City*, *UT*, *USA*. IEEE, 2018: 586-595.

[34] LI H P, YANG J Y, LI G Q, *et al*. Autoregressive image generation with randomized parallel decoding [EB/OL]. 2025: *arXiv*: 2503.10568. https://arxiv.org/abs/2503.10568.

作者简介：

魏　赟(1970—),女,河南鹤壁人,博士,副教授,主要从事计算机视觉,车联网,智能交通工作的研究。E-mail: weiyun@mail.lzjtu.cn

通讯作者：

贾白雪(2001—),女,河南周口人,硕士研究生,主要从事计算机视觉、图像修复的研究。E-mail: 3489395305@qq.com