计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
2025,34(4):104−114 [doi: 10.15888/j.cnki.csa.009849] [CSTR: 32024.14.csa.009849]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

融合像素差分卷积与 Transformer 的岩石 CT 图像超分辨率重建①

张　威[1,2], 尹　祎[1,2], 林钰斌[1,2]

[1](武汉科技大学 计算机科学与技术学院, 武汉 430081)
[2](武汉科技大学 智能信息处理与实时工业系统湖北省重点实验室, 武汉 430081)
通信作者: 张　威, E-mail: 202213704129@wust.edu.cn

摘　要: 针对岩石 CT 图像超分辨率重建中纹理和边缘细节恢复不佳, 以及传统 Transformer 模型资源消耗大的问题, 本文提出了一种轻量级混合架构 PDCLT 模型. 该模型结合了基于像素差分卷积的细节强化 CNN 模块和轻量级 Transformer 模块, 以实现对局部与全局特征的高效提取. 具体而言, 首先提出细节强化模块, 融合了像素差分卷积和残差增强注意力, 并提出了自适应路径权重缩放方法, 以动态调整特征提取路径的权重, 增强了对细微结构和关键特征的捕捉. 其次, 轻量级 Transformer 模块集成高效多头注意力和多尺度特征融合网络, 在降低 GPU 内存需求的同时提取全局和多尺度特征. 最后, 在损失函数中加入孔隙度损失以优化孔隙结构的保留. 实验结果显示, PDCLT 模型在重建质量和细节还原方面表现出色, 显著提升了岩石 CT 图像的超分辨率重建质量.

关键词: 岩石 CT 图像; 超分辨率重建; 像素差分卷积; 残差增强注意力; 高效多头注意力; 多尺度特征融合网络

引用格式: 张威,尹祎,林钰斌.融合像素差分卷积与 Transformer 的岩石 CT 图像超分辨率重建.计算机系统应用,2025,34(4):104–114. http://www.c-s-a.org.cn/1003-3254/9849.html

Super-resolution Reconstruction of Rock CT Images by Fusing Pixel Difference Convolution and Transformer

ZHANG Wei[1,2], YIN Yi[1,2], LIN Yu-Bin[1,2]

[1](College of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430081, China)
[2](Hubei Provincial Key Laboratory of Intelligent Information Processing and Real-time Industrial Systems, Wuhan University of Science and Technology, Wuhan 430081, China)

Abstract: To address the inadequate restoration of textures and edge details in super-resolution reconstruction of rock CT images, along with the high resource consumption of traditional Transformer models, this study proposes a lightweight hybrid architecture, the pixel difference convolution and lightweight Transformer (PDCLT) model. The model integrates a detail-enhancement convolutional neural network (CNN) module based on pixel difference convolution and a lightweight Transformer module to efficiently extract both local and global features. Specifically, the model first introduces a detail enhancement module that combines pixel difference convolution with residual enhanced attention. It also proposes an adaptive path weight scaling method to dynamically adjust the weights of feature extraction paths, which enhances the capture of subtle structures and key features. Secondly, the lightweight Transformer module incorporates efficient multi-head self-attention and a multi-scale feature fusion network to reduce GPU memory demands while extracting global and multi-scale features. Finally, porosity loss is added to the loss function to optimize the preservation of pore structures. Experimental results show that the PDCLT model excels in reconstruction quality and detail restoration, significantly improving the super-resolution reconstruction quality of rock CT images.

① 基金项目: 省部共建耐火材料与冶金国家重点实验室开放基金 (G202410)
收稿时间: 2024-10-13; 修改时间: 2024-10-30, 2024-11-29; 采用时间: 2024-12-06; csa 在线出版时间: 2025-03-04
CNKI 网络首发时间: 2025-03-06

Key words: rock CT image; super-resolution reconstruction; pixel differential convolution; residual enhanced attention; efficient multi-head attention; multi-scale feature fusion network

1 引言

超分辨率图像处理技术在数字图像领域占据了重要地位, 尤其在地质科学中, 如岩石 CT 图像的分析中, 起到了关键作用. 该技术通过从低分辨率图像中恢复高分辨率细节, 显著提高了图像的视觉质量, 从而为更深入的图像分析和处理提供了强有力的数据支持.

在地质学研究和油气资源的开发过程中, 准确分析岩石的成分和孔隙结构至关重要. 然而, 岩石的复杂性使得这一过程极具挑战. 此外, 现场环境中的设备通常受到资源和计算能力的限制, 使得开发资源高效的图像处理方法尤为必要. 当前的 X 射线微计算机断层扫描 (X-ray micro-CT) 技术虽然能够提供岩石特征的高分辨率观察, 但在分辨率和视野范围之间存在权衡, 限制了图像的清晰度, 导致岩石微小结构的边界和纹理上可能出现模糊不清的问题, 这使得在后续岩石分割等处理中可能出现矿物边界模糊等问题[1,2]. 目前, 许多超分辨率模型在处理岩石的显微图像时往往难以恢复这些关键细节, 而这些细节对于岩石的科学分析和资源评估至关重要. 因此, 为解决 X 射线扫描图像中边界模糊和纹理细节缺失以及设备资源限制的问题, 有必要开发一种高效、资源友好的超分辨率重建方法, 以在资源受限的设备上提升岩石 CT 图像的清晰度和纹理细节.

传统的插值方法 (如双线性和双三次插值) 在简单场景下表现良好, 但在应对复杂纹理和细节丰富的岩石 CT 图像时, 往往无法有效恢复高频细节, 限制了其应用范围[3]. 目前, 随着机器学习和深度学习的迅速发展, 基于学习的超分辨率方法成为图像处理领域的研究热点[4–6]. 早期模型如 SRCNN[7]通过卷积神经网络对低分辨率图像进行处理, 显著提升了图像质量, 开创了深度学习在超分辨率中的应用.

在岩石 CT 图像超分辨率中, 准确捕捉细粒度纹理和边缘特征至关重要. 传统卷积神经网络 (CNN) 虽然能有效提取高层次语义特征, 但其固定网格卷积核在处理局部细节时存在不足, 尤其在纹理和边缘提取上. 传统 CNN 通过局部感受野进行平滑操作, 这会模糊细节, 难以捕捉微小的灰度变化和局部特征, 尤其在复杂纹理或模糊边缘区域. 相比之下, 像素差分卷积 (pixel difference convolution, PDC)[8] 通过直接计算相邻像素的灰度差异, 能够更精确地捕捉局部变化. 与传统卷积核的平滑处理不同, PDC 强调局部像素差异, 突出图像细节, 在处理高纹理和复杂边缘时, 能够敏感地捕捉微小特征变化, 保留岩石 CT 图像中的精细结构与纹理特征, 提升高分辨率复现效果. 相比传统 CNN, PDC 在局部细节恢复和纹理增强上具有明显优势.

近年来, Transformer 模型因其在自然语言处理中的卓越表现, 逐渐引起了图像处理领域的广泛关注. Transformer[9]的核心自注意力机制能够在较大范围内捕捉图像特征间的长距离依赖关系, 能有效提升图像的细节恢复能力. 例如, ViT[10]和 SwinIR[6]等模型在图像超分辨率任务中展示了 Transformer 在捕捉细节和重建纹理方面的强大潜力, 显著超越了传统卷积网络的表现. 然而, 这些模型对计算资源的高需求限制了其在低算力设备上的应用场景, 因此, 有必要引入轻量化的 Transformer 架构, 以在满足性能需求的同时优化资源利用.

为应对岩石 CT 图像超分辨率重建的挑战, 本文提出一种融合细节强化 CNN 模块与轻量级 Transformer 模块的混合架构 PDCLT 模型. 在该模型中, 首先设计了细节强化模块 (DIB), 融合像素差分卷积 (PDC) 和残差增强注意力 (REA), 并设计自适应路径权重缩放 (APWS) 方法, 使模型能够动态调整特征提取路径的权重. PDC 有效捕捉了图像中的细微结构和局部变化, REA 则确保模型专注于关键特征, 从而提升了复杂纹理的细节重建效果. 在此基础上, 轻量级 Transformer 模块 (LTB) 负责提取全局信息, 集成了高效多头注意力 (*EMHA*) 和多尺度特征融合网络 (*MFFN*), *EMHA* 在保持计算效率的同时捕捉长距离依赖, *MFFN* 通过并行卷积核提取多尺度的局部和空间特征. 最后, 针对岩石 CT 图像的特点, 本文在损失函数中引入了孔隙度损失, 使模型在重构时能更好地保留孔隙结构. 得益于这些设计, PDCLT 在纹理细节保留和整体重建效果上表现出色, 实现了岩石 CT 图像的高质量超分辨率重建.

2 相关工作

2.1 基于卷积神经网络的超分辨率方法

图像超分辨率技术取得的显著进步, 主要得益于深度学习技术的飞速发展, 特别是卷积神经网络的应用. 2014 年, Dong 等人[7]提出了 SRCNN 模型, 首次将深度学习用于图像超分辨率. SRCNN 通过 3 层卷积网络学习低分辨率到高分辨率图像的映射, 效果显著优于传统方法. 随后, VDSR[11]和 DRCN[12]通过引入残差学习来增加网络深度, 有效缓解了深层网络中的梯度消失问题, 进一步提升了超分辨率性能. EDSR[13]则通过优化残差块堆叠, 显著提高了重建质量.

在提升特征利用效率方面, MemNet[14]和 RDN[5]引入了密集连接架构, 通过每层输出与后续层的有效连接, 优化了特征的综合利用, 尽管这些改进增加了计算成本和内存消耗. 为了应对计算资源受限的环境, 研究者们提出了多种轻量级模型. 信息蒸馏网络 IDN[15]和信息多蒸馏网络 IMDN[16]通过蒸馏块逐步提取层次特征, 以实现单图像超分辨率的高效重建. BSRN[17]利用可分离卷积减少冗余.

尽管这些 CNN 方法在局部特征提取上表现出色, 但它们在全局特征的捕捉和细节恢复方面仍然存在局限性, 特别是在重建图像纹理的精细化方面表现不足.

2.2 边缘检测方法

自 Julesz 于 1959 年完成其工作以来[18], 边缘检测一直是计算机视觉的重要研究方向. 边缘检测方法可大致分为边缘微分算子、传统机器学习和深度学习方法. 早期的边缘微分算子, 如 Canny[19]和 Laplace[20]是通过物体边缘与背景像素之间的灰度梯度信息实现检测, Canny 算子还引入了非极大值抑制, 成为经典算法. 然而, 这些算子设计依赖手动调参, 难以处理复杂图像.

随着机器学习的发展, 传统方法逐步引入了手工设计特征以监督学习. Dollár 等人[21]提出了快速随机结构森林方法, 利用局部模板预测边缘; Martin 等人[22]提出的 Pb 算法结合亮度、纹理和颜色特征进行边缘检测, 但尽管这些方法设计了复杂的特征表示, 边缘信息提取效果有限. 随着深度学习迅速发展, 卷积神经网络 (CNN) 广泛应用于边缘检测任务. N^4-field[23]等早期方法采用逐块或逐像素策略, 一定程度上改善了边缘检测效果, 但削弱了整体预测准确性. 后续的 RCF[24]模型通过结构改进增强了检测能力, 但这类深度模型结构复杂、参数量大, 难以满足计算资源有限设备的需求.

在此背景下, 像素差分卷积 (pixel difference convolution, PDC)[8]被提出. PDC 通过直接计算相邻像素差异, 能够高效捕捉局部差分信息, 突出图像中局部的变化区域. 与传统边缘检测方法侧重通过全局梯度信息来识别图像中的显著边界和物体轮廓来提取图像的结构性信息不同, PDC 更关注图像的局部细节, 通过计算相邻像素之间的灰度差异, 突出细微的变化和局部特征, 这对于超分辨率任务中细节恢复和纹理增强尤为重要. 因此本文借鉴 PDC 的优势, 将其用于岩石 CT 图像超分辨率重建, 以捕捉微小边缘和复杂纹理, 在资源受限环境下实现高质量细节复原.

2.3 基于 Transformer 的超分辨率方法

自从 Transformer[9]在自然语言处理领域取得成功后, 其强大的长期依赖建模能力逐渐被应用于图像超分辨率任务中. Transformer 通过自注意力机制捕捉全局信息, 使其在图像细节重建方面优于传统 CNN 方法.

SwinIR[6]模型基于 Swin Transformer[25]构建, 采用移位窗口机制在捕捉局部和全局特征的同时有效降低了计算复杂度, 显著提升了超分辨率任务的效果. HAT[26]通过结合通道注意力和自注意力机制来增强模型的特征提取能力. 通道注意力强调特征通道间的重要性, 自注意力则捕捉全局依赖, 从而在高频细节的重建上表现突出.

尽管基于 Transformer 的模型在性能上取得了显著进展, 但其高计算成本和内存需求仍是主要瓶颈. 为应对这些挑战, ESRT[27]模型结合了轻量级 CNN 与高效 Transformer 的混合架构, 采用特征分离策略和高效多头注意力机制, 减少了 GPU 内存占用和参数量, 同时保持了较高的图像重建质量. LBNet[28]模型则通过递归 Transformer 与对称 CNN 结构的结合来优化效率和性能, 递归结构共享参数, 有效降低了内存需求, 并在复杂图像的细节处理上保持较高的精度.

这些轻量化设计为实现更高效的 Transformer 架构提供了可能性, 使得在资源受限的环境中也能实现优异的图像超分辨率效果. 尽管已有显著进步, 进一步提升 Transformer 架构的效率, 以在资源消耗与重建质量间取得更好的平衡, 依然是图像超分辨率重建研究的重要方向.

3 模型介绍

3.1 网络架构

本文提出的方法的整体网络架构如图 1 所示. 该模型由浅层特征提取、深层特征提取和高分辨率图像重建 3 个部分组成, 适用于高效的岩石 CT 图像超分辨率重建任务.

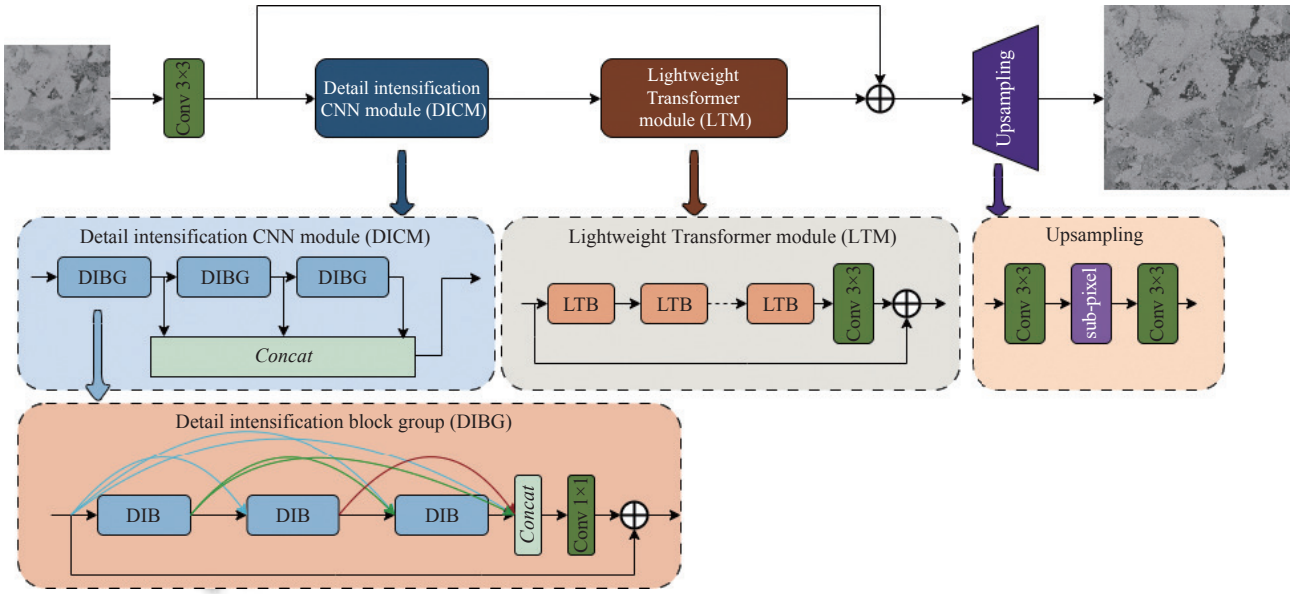

图 1 整体网络架构

3.1.1 浅层特征提取

首先, 输入的低分辨率图像$I_{\mathrm{LR}} \in \mathbb{R}^{H\times W\times 3}$通过浅层特征提取层, 该层包含一个标准的$3\times 3$卷积层, 用于提取初步的粗略特征$F_0 \in \mathbb{R}^{H\times W\times C}$. 其中, $H\times W$ 表示图像的空间分辨率, C 代表中间特征的通道数. 特征提取的过程可通过以下公式表示:

$$F_0 = W(I_{\mathrm{LR}}) \tag{1}$$

其中, $W(\cdot)$表示浅层特征提取层.

3.1.2 深层特征提取

在深层特征提取阶段, 模型包含细节强化 CNN 模块 (DICM) 和轻量级 Transformer 模块 (LTM). 首先, DICM 用于提取图像的局部特征, 并将这些细节特征整合后作为 LTM 的输入, LTM 进一步处理这些特征以捕捉图像的全局信息. DICM 和 LTM 模块构建了一个双尺度特征融合网络结构, 在局部特征的基础上有效结合全局特征, 以实现更优的岩石 CT 图像纹理恢复效果.

具体来说, 浅层特征F_0经过 DICM, 并通过多级细节强化模块组 (DIBG) 处理, 提取出一系列深层的局部特征F_n. 具体表述如下:

$$F_n = \delta^n\left(\delta^{n-1}\left(\cdots\left(\delta^1(F_0)\right)\right)\right) \tag{2}$$

其中, $\delta^n(\cdot)$表示第 n 个 DIBG 的特征映射, F_n表示第 n 个 DIBG 所提取出的特征.

每个 DIBG 由多个具有残差密集连接的细节强化模块 (DIB) 组成, 以充分捕捉细节信息. 所有 DIBG 的输出特征沿通道维度拼接为整体, 形成融合后的局部特征F_{DICM}. 随后, 这些局部特征F_{DICM}被输入至 LTM. LTM 由多个轻量级 Transformer 块 (LTB) 组成, 进一步处理 DICM 提取的特征, 以捕捉全局信息. 通过这种方式, LTM 对 DICM 输出的中间特征进行增强, 整合了局部和全局信息. 最终, LTM 和 DICM 的特征通过残差连接进行融合, 将 LTM 提取的全局特征与 DICM 提取的局部特征整合, 实现了特征融合. 该过程可表示为:

$$F_{\mathrm{DICM}} = Concat(F_1, F_2, \cdots, F_n) \tag{3}$$

$$F_{\mathrm{LTM}} = \phi^m\left(\phi^{m-1}\left(\cdots\left(\phi^1(F_{\mathrm{DICM}})\right)\right)\right) + F_{\mathrm{DICM}} \tag{4}$$

其中, F_{DICM}表示 DICM 的输出特征, F_{LTM}表示 LTM 的输出特征, $\phi(\cdot)$是 LTB 的操作, $Concat(\cdot)$是通道拼接操作. 为简洁起见, 卷积层操作已省略.

3.1.3 高分辨率图像重建

通过上采样器将融合后的深层特征放大到原始高分辨率尺寸. 上采样器由两个卷积层和一个亚像素卷

积层组成, 输出通道为$3\times s^2$, 其中 s 是放大因子. 该过程可表示为:

$$I_{\mathrm{SR}} = H_{\mathrm{up}}(F_{\mathrm{LTM}} + F_0) \tag{5}$$

3.1.4 损失函数

在岩石领域的研究中, 孔隙度是评估岩石相关属性的重要指标. 对于计算机 CT 扫描获得的岩石 CT 图像, 可通过图像边缘检测将岩石 CT 图像分割为固体部分和孔隙部分. 根据孔隙像素与固体像素的比例, 可以计算岩石的孔隙度. 由于岩石成分不同, 分割阈值往往不同, 因此采用 Otsu 算法来确定每张岩石 CT 图像的最佳分割阈值.

设p_{pore}和p_{solid}分别表示岩石孔隙部分和固体部分的像素数. 对于第 n 张岩石 CT 图像I, 其孔隙度ϕ可表示为:

$$\phi_I = p_{\mathrm{pore}}(I) / \left(p_{\mathrm{pore}}(I) + p_{\mathrm{solid}}(I)\right) \tag{6}$$

在超分辨率处理岩石 CT 图像时, 其孔隙度保持不变. 为此, 将孔隙度作为损失函数的一部分引入, 总损失函数由$L1$损失L_{L1}与孔隙度损失L_{porosity}组合优化, 以加速网络收敛, 并用于与其他超分辨率方法的公平比较. 损失函数定义如下:

$$L = L_{L1} + \lambda L_{\mathrm{porosity}} \tag{7}$$

其中, L_{L1}定义为:

$$L_{L1} = \frac{1}{N}\sum_{i=1}^{N} I_{\mathrm{SR}}^{i} - I_{\mathrm{HR}}^{i} \tag{8}$$

孔隙度损失L_{porosity}定义为:

$$L_{\mathrm{porosity}} = \left(\phi\left(I_{\mathrm{SR}}^{i}\right) - \phi\left(I_{\mathrm{HR}}^{i}\right)\right)^2 \tag{9}$$

其中, λ是权重系数, I_{SR}和I_{HR}分别表示超分辨率图像和高分辨率图像.

3.2 细节强化模块

在图像超分辨率领域, 传统方法通常使用普通卷积块进行特征提取. 然而, 常规卷积不仅计算成本高、运行效率低, 而且其固定网格结构在庞大的解空间中缺乏针对细节信息的约束, 限制了对图像边缘和纹理的精确捕捉能力. 考虑到边缘先验知识可以帮助恢复更清晰的图像轮廓, 本文借鉴了像素差分卷积 (pixel difference convolution, PDC)[8]的思想, 提出了如图 2 左侧所示的细节强化模块 (DIB) 用于更有效的深层特征提取. 具体而言, 该模块包含基于中心差分的像素差分卷积 (CPDC)、基于角度差分的像素差分卷积 (APDC)、基于径向差分的像素差分卷积 (RPDC)、常规卷积以及残差增强注意力 (REA). 不同于原始像素差分卷积, 本方法使用常规卷积替代了深度可分离卷积. 为避免参数量急剧增加, 本文在平衡参数量和性能的基础上, 将通过像素差分卷积的通道数减半, 随后再通过常规卷积恢复.

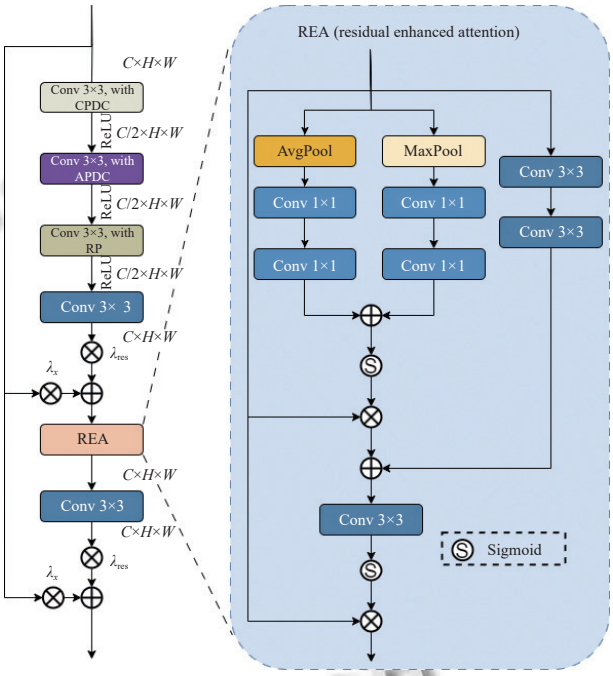

图 2 细节强化模块

在像素差分卷积之后, 本文设计的残差增强注意力 (REA) 被用于进一步提升特征表达能力, 其结构如图 2 右侧所示. 在图像超分辨率重建任务中, 注意力机制已被广泛应用, 然而直接将通道注意力机制嵌入主干网络未必总能显著提升性能, 因为这可能削弱网络对重要特征的提取能力, 影响图像纹理和细节的保留. 为解决这一问题, REA 结合了传统通道注意力块和残差块, 将特征提取与通道加权有效融合, 使网络在关注关键通道信息的同时, 保留更多有价值的纹理和细节. 同时, 受 ReZero[29]启发, 提出了一种带有自适应路径权重缩放 (APWS) 方法的残差连接, 通过可学习的权重系数来动态调整残差路径与恒等路径的重要性. 与传统残差连接中的固定权重相比, APWS 可以提高模型的稳定性, 并加快模型收敛. 整个细节强化模块的过程可表示为:

$$F_{\mathrm{PDC}} = \lambda_x \cdot x + \lambda_{\mathrm{res}} \cdot (W_1(f_{\mathrm{RPDC}}(f_{\mathrm{APDC}}(f_{\mathrm{CPDC}}(x))))) \tag{10}$$

$$F_{\text{out}} = \lambda_x \cdot x + \lambda_{\text{res}} \cdot (W_2(f_{\text{REA}}(F_{\text{PDC}}))) \quad (11)$$

其中, $f_{\text{CPDC}}(\cdot)$、$f_{\text{APDC}}(\cdot)$和$f_{\text{RPDC}}(\cdot)$分别表示基于中心差分的像素差分卷积操作、基于角度差分的像素差分卷积操作和基于径向差分的像素差分卷积操作. $W_1(\cdot)$和$W_2(\cdot)$分别表示3×3卷积操作, λ_x和λ_{res}为自适应路径权重系数. $f_{\text{REA}}(\cdot)$是残差增强注意力, F_{out}是输出特征.

3.3 轻量级 Transformer 模块

在图像超分辨率任务中, 大多数基于 Transformer 的方法通常需要大量计算资源, 并占用较多 GPU 内存. 为此, 本文提出了一种轻量级 Transformer 模块 (LTB), 用于建模图像的长距离依赖关系并提取全局特征, 如图 3 所示. 本文针对岩石 CT 图像超分辨率任务优化了传统 Transformer 块[9], 以提升计算效率并适应 GPU 内存受限的情况. 受 ESRT[27]启发, LTB 引入了预处理和后处理模块. 具体而言, LTB 采用展开技术, 将输入特征图分割成补丁. 展开操作描述如下: 对于 Transformer 块的输入特征$F_{\text{ori}} \in \mathbb{R}^{C\times H\times W}$, 通过$k\times k$的卷积核将其展开为$N$个补丁, 每个补丁特征为$F_{pi} \in \mathbb{R}^{k^2\times C}$, 其中$N = H\times W$. 展开操作自动嵌入每个补丁的位置信息, 无需额外的位置编码. 随后, 补丁F_p直接输入至 LTB, 且 LTB 的输入和输出形状一致, 最终通过折叠操重建特征图.

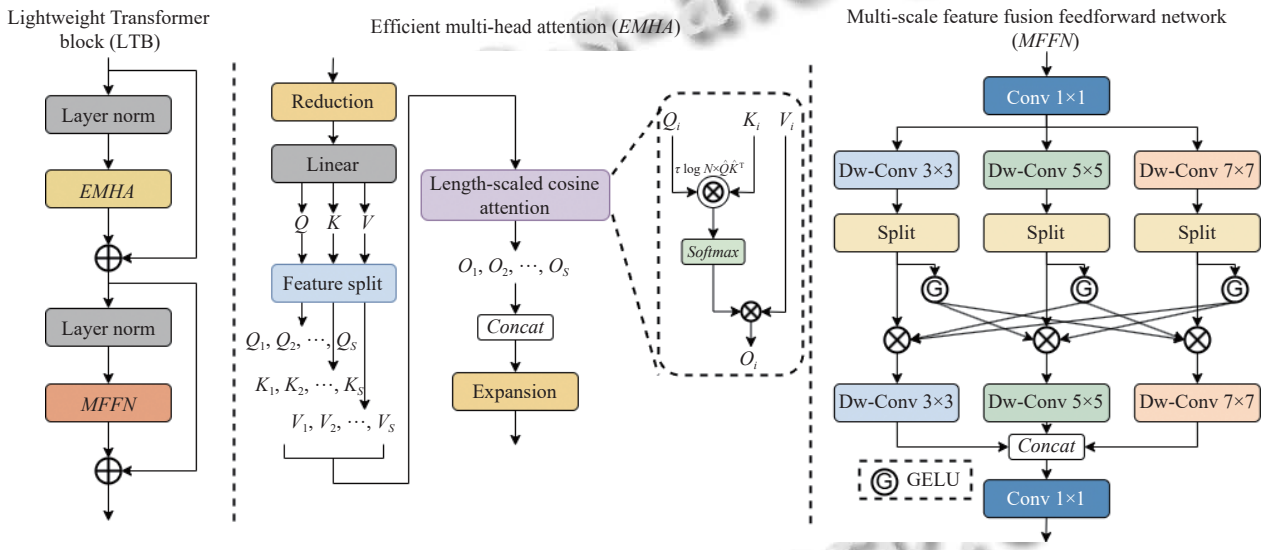

图 3 轻量级 Transformer 模块

近期研究[30]表明, 随着输入序列长度的增加, 注意力输出的置信度会降低, 这主要是由于缩放因子与输入序列长度的关联性. 为解决这个问题, 本文引入 Trans-NeXt[31]中的长度缩放余弦注意力 (LSCA). 如图 3 所示, LSCA 用于替代查询和键值之间的原始缩放点积注意力 (SDPA). LSCA 使用余弦相似度, 被观察到能生成更平滑的注意力权重, 有效提高大型视觉模型的训练稳定性. 同时, LSCA 中的缩放因子能够更好适应长序列输入, 有助于稳定模型在未知长度上的泛化表现. LSCA 的公式如下:

$$Attention(Q,K,V) = Softmax(\tau \log N \times \hat{Q}\hat{K}^{\mathrm{T}})V \quad (12)$$

其中, $\hat{Q}$, $\hat{K}$是进行l_2归一化的查询和键矩阵, V是值矩阵, τ是一个可学习参数, 初始值为1/0.24. N表示每个查询和有效键的交互次数.

为进一步节省计算开销, 类似于 ESRT, 查询矩阵Q、键矩阵K和值矩阵V采用了特征分割策略, 将输入序列分成S段以降低 GPU 内存需求. 每段特征分别表示为$Q_1, Q_2, \cdots, Q_S$, $K_1, K_2, \cdots, K_S$和$V_1, V_2, \cdots, V_S$, 并依次输入 LSCA 操作, 省略了 mask 处理. 得到的输出$O_1, O_2, \cdots, O_S$再连接形成最终输出, 如图 3 左侧所示.

考虑到传统 MLP 在捕捉多尺度特征方面的局限性, 尤其是在岩石 CT 图像中对局部和空间特征的捕捉能力不足且参数量较大, 本文设计了一种多尺度特征融合网络 (*MFFN*) 来解决该问题. 如图 3 右侧所示, 特征E_m经过1×1卷积降维, 然后被输入到 3 条包含不同卷积核大小 (3×3, 5×5, 7×7) 的深度可分离卷积 (Dw-Conv) 分支中, 以提取多尺度特征, 丰富特征表达. 为增强各尺度特征之间的交互, 沿通道维度将特征分为两

部分, 其中一部分经过 GELU 激活与另一部分通过逐元素相乘进行结合. 整个过程可以表示为:

$$\begin{cases} E_3 = f_{3\times3}^{\mathrm{dwc}}\left(f_{1\times1}\left(E_m\right)\right) \\ E_5 = f_{5\times5}^{\mathrm{dwc}}\left(f_{1\times1}\left(E_m\right)\right) \\ E_7 = f_{7\times7}^{\mathrm{dwc}}\left(f_{1\times1}\left(E_m\right)\right) \end{cases} \tag{13}$$

$$\begin{cases} \overline{E}_3 = f_{3\times3}^{\mathrm{dwc}}\left(E_3^{P_1}\cdot E_5^{P_2}\cdot E_7^{P_2}\right) \\ \overline{E}_5 = f_{5\times5}^{\mathrm{dwc}}\left(E_3^{P_2}\cdot E_5^{P_1}\cdot E_7^{P_2}\right) \\ \overline{E}_7 = f_{7\times7}^{\mathrm{dwc}}\left(E_3^{P_2}\cdot E_5^{P_2}\cdot E_7^{P_1}\right) \\ \overline{E} = f_{1\times1}\left[\overline{E}_3, \overline{E}_5, \overline{E}_7\right] \end{cases} \tag{14}$$

其中, E_m 表示经高效多头注意力的输出, $f_{1\times1}(\cdot)$表示 1×1卷积, $f_{n\times n}^{\mathrm{dwc}}(\cdot)$表示卷积核大小为 n 的深度卷积, X^{P_1} 和X^{P_2} 分别表示未经过 GELU 激活和激活的特征, $[\cdot]$表示通道级联.

如图 3 所示, 本文在高效多头注意力 (*EMHA*) 中引入了 LSCA, 在 *EMHA* 和 *MFFN* 之前应用了层归一化. 整个 LTB 的计算流程概括如下:

$$\begin{cases} E_m = EMHA\left(LN\left(E_i\right)\right) + E_i \\ E_{\mathrm{out}} = MFFN\left(LN\left(E_m\right)\right) + E_m \end{cases} \tag{15}$$

其中, E_i 表示输入, $LN(\cdot)$表示层归一化操作, $EMHA(\cdot)$和$MFFN(\cdot)$分别表示高效多头注意力操作和多尺度特征融合网络操作.

4 实验

4.1 数据集和评价指标

本研究采用了两种常用的公开超分辨率岩石 CT 图像数据集进行实验: DRSRD1_2D[32]和 DeepRock-SR_2D[33]. DRSRD1_2D 数据集中包含 3 种不同类型的岩石 CT 图像: 碳酸盐 (carbonate)、砂岩 (sandstone) 和二者的混合 (shuffled). 本实验选择了其中的混合图像数据集 (shuffled), 该数据集包含 2 000 张图像, 其中 1 600 张用于训练, 200 张用于验证, 200 张用于测试. 另一数据集 DeepRock-SR_2D 包含 4 种类型的岩石 CT 图像: 碳酸盐 (carbonate)、砂岩 (sandstone)、煤岩 (coal2D) 以及三者的混合 (shuffled). 同样地, 本实验选择了其中的混合图像数据集 (shuffled), 该数据集包含 12 000 张图像, 其中 9 600 张用于训练, 1 200 张用于验证, 1 200 张用于测试.

为了定量评估模型性能, 本文采用了峰值信噪比 (PSNR) 和结构相似性 (SSIM)[34]两种广泛用于图像质量评价的指标. PSNR 用于衡量重建图像与原始图像之间的差异, 其值越高, 说明重建图像越接近原始图像, 代表重建质量越好. SSIM 则从亮度、对比度和结构 3 个方面衡量图像的结构信息, 值的范围在 [0, 1] 之间, SSIM 越接近 1, 表示重建图像与真实图像的相似性越高.

4.2 实验细节

实验在配置为 GPU V100 (4 核, 16 GB 显存) 的计算环境中进行, 使用了 PyTorch 1.7 和 CUDA 11.0 框架进行模型训练.

在训练过程中, 我们采用随机剪裁方法, 每个训练时期随机裁剪尺寸为 48×48 的低分辨率图像块作为输入, 以提高数据多样性、减少过拟合并增强模型的鲁棒性. 同时, 数据增强还包括随机水平翻转和 90°旋转. 在网络的学习阶段, 设置的初始学习率设置为 2×10^{-4}, 并使用 Adam 优化器 ($\beta_1 = 0.9, \beta_2 = 0.999$) 进行 200 个迭代的训练. 此外, 为了获得更锐利的重建效果, 用 $L1$ 损失代替了 $L2$ 损失, 并引入了孔隙度损失函数来增强模型对岩石 CT 图像特性的捕捉. 图 4 展示了在 DRSRD1_2D 数据集中的×4 倍训练数据集上训练过程中的损失函数收敛情况. 从图中可以看出, 损失函数随训练进展迅速下降, 且曲线较为平滑. 这一现象与使用像素差分卷积 (PDC) 密切相关, PDC 通过捕捉局部像素灰度差异, 提供了强有力的先验知识, 使得模型在初期就能快速优化损失, 导致收敛速度较快. 并且 PDC 局部特征学习方式避免了传统卷积方法中的平滑作用, 能够减少图像细节的丢失, 使得损失优化更加稳定. 图 5 则显示了在 DRSRD1_2D 数据集中的×4 倍验证数据集上, 验证过程中 PSNR/SSIM 值的变化趋势. 随着训练次数的增加, PSNR/SSIM 值逐渐上升, 并在第 200 个 epoch 左右模型达到收敛状态.

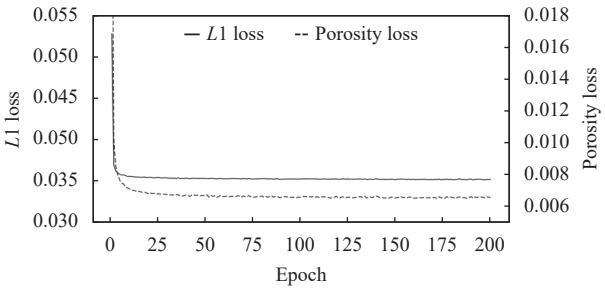

图 4 损失函数收敛图

对于轻量级混合架构 PDCLT 模型的配置细节, 细节强化 CNN 模块 (DICM) 由 3 个细节强化模块组

(DIBG) 组成, 每个 DIBG 的输入通道数设置为 32, 并包含 3 个细节强化模块 (DIB). 在每个 DIB 中, 自适应路径权重系数λ_x和λ_{res}的初始值均设为为 1. 同时, 轻量级 Transformer 模块 (LTM) 中包含一个轻量级 Transformer 块 (LTB), 以节省 GPU 内存. 损失函数中的权重参数λ设置为 0.1.

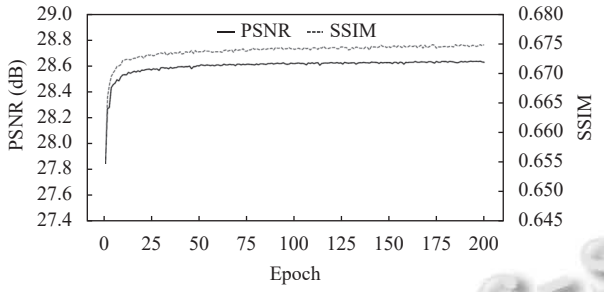

图 5 DRSRD1_2D 数据集上 PSNR/SSIM 变化图

4.3 消融实验

4.3.1 细节强化模块中的不同卷积

在细节强化模块中, 主分支通过像素差分卷积进行特征提取. 为评估像素差分卷积不同配置的有效性, 本文进行了消融实验, 重点对比了使用深度可分离卷积 (DSConv) 和常规卷积 (Conv) 的性能差异, 实验结果汇总如表 1 所示. 研究发现, 尽管常规卷积使模型参数量增加了 113k, 但在图像质量指标上带来了显著提升: PSNR 提高了 0.05 dB, SSIM 增加了 0.000 8. 相比深度可分离卷积, 常规卷积能更充分捕捉细节信息, 增强了模型的特征学习能力.

表 1 像素差分卷积的不同配置评估

方法	参数 (k)	PSNR (dB)/SSIM
DSConv 3×3	480	28.21/0.662 1
Conv 3×3	593	**28.26/0.662 9**

从性能与模型复杂度的平衡角度看, 结果表明, 尽管常规卷积增加了计算开销, 但在图像超分辨率重建效果上带来了显著改善.

4.3.2 细节强化模块中的残差增强注意力效果

在细节强化模块中, 我们设计并应用了残差增强注意力 (REA) 模块. 该模块结合了通道注意力机制与残差结构, 有效增强了对关键通道信息的捕捉, 并保留了更多重要的纹理和细节信息. 通过特征提取与注意力加权的融合, REA 显著提升了模型的细节感知能力.

如表 2 所示, 尽管 REA 模块仅引入了少量额外参数, 模型在 PSNR 和 SSIM 等图像重建指标上依然有所提升.

表 2 细节强化模块中 REA 的消融研究

方法	参数 (k)	PSNR (dB)/SSIM
w/o REA	590	28.23/0.662 2
w/ REA	593	**28.26/0.662 9**

4.3.3 多尺度特征融合网络的效果

在提出的轻量级 Transformer 模块中, 设计并引入了多尺度特征融合网络 (*MFFN*). 相比传统 MLP, *MFFN* 结合了多尺度卷积操作, 能够有效捕捉图像的多尺度特征, 并通过特征分支和融合机制增强各尺度间的交互, 从而更好地挖掘相关对象的潜在尺度关系. 为验证 *MFFN* 的有效性, 本文将 *MFFN* 与 MLP 进行了对比实验, 展示了其在多尺度特征提取和复杂图像处理方面的明显优势.

在保证模型性能的同时, 本文通过减少 *MFFN* 的输入通道数, 进一步优化了计算效率和参数规模. 实验结果如表 3 所示, *MFFN* 在捕捉复杂多尺度特征的过程中显著减少了参数量, 并在 PSNR 和 SSIM 等图像重建指标上优于传统 MLP. 这表明, *MFFN* 在处理岩石 CT 图像的多尺度特征方面表现出色, 达到了性能与参数量的良好平衡.

表 3 MLP 和 *MFFN* 的消融研究

方法	参数 (k)	PSNR (dB)/SSIM
MLP	652	28.24/0.662 5
MFFN	593	**28.26/0.662 9**

4.4 实验结果和分析

4.4.1 孔隙度损失函数对结果的影响

本文将岩石孔隙度作为一种重要的约束条件. 为了评估这一条件的效果, 实验针对 DRSRD1_2D 数据集×4 倍的图像进行了客观指标的对比分析. 在相同实验条件下, 比较了引入孔隙度损失函数$L_{porosity}$的 PDCLT 模型和未加入该约束条件的 PDCLT 模型的 孔隙度.

通过对测试集所有图像在不同条件下的孔隙度进行分析, 发现加入孔隙度损失函数后的 PDCLT 相较于 PDCLT 更接近原图 (HR) 的孔隙度值. 表 4 具体展示了分别对碳酸盐和砂岩随机采样得到的 10 张测试图像孔隙度.

4.4.2 与其他轻量级单图像超分辨率模型比较

在表 5 中, PDCLT 与其他轻量级单图像超分辨率

模型分别在 DRSRD1_2D 和 DeepRock-SR_2D 数据集上进行了比较, 包括 SRCNN[7], FSRCNN[35], SAFMN[36], BSRN[17], IMDN[16], ESRT[27]和 CRAFT[37]. 结果显示, 提出的 PDCLT 在整体性能上优于这些轻量级模型. 具体而言, 针对×4 倍的岩石 CT 图像超分辨率任务, PDCLT 仅需 593k 参数量, 便实现了性能与计算复杂度之间的良好平衡. 在 DRSRD1_2D 数据集上, 尽管 ESRT 和 CRAFT 在×2 倍放大下的性能接近 PDCLT, 但它们的参数量显著高于 PDCLT. 而 BSRN 和 SAFMN 虽然参数量少于 PDCLT, 但其性能明显不及 PDCLT.

表 4 HR, PDCLT 和加入 $L_{porosity}$ 的 PDCLT 的孔隙度对比

岩石种类	图片编号	孔隙度 (%)		
		HR	PDCLT	PDCLT-$L_{porosity}$
carbonate	1814	10.49	11.19	10.87
	1847	9.41	9.91	9.70
	1891	11.14	12.46	11.89
	1923	9.19	9.94	9.70
	1975	9.61	10.26	9.99
sandstone	1809	12.95	13.51	13.01
	1851	11.74	12.53	11.62
	1888	9.79	11.27	10.39
	1930	12.05	11.15	11.45
	1976	12.67	13.45	13.11

此外, 图 6 和图 7 分别展示了 PDCLT 与其他轻量级 SISR 模型分别在 DRSRD1_2D 测试数据集中的碳酸盐和砂岩上×4 倍放大的视觉对比. 从图中可以看出, PDCLT 重建的超分辨率图像在细节上更加精确, 尤其是在边缘和线条区域, 展现了更为清晰和准确的纹理细节.

表 5 DRSRD1_2D 和 DeepRock-SR_2D 测试集的对比实验

倍率	模型	参数 (k)	DRSRD1_2D	DeepRock-SR_2D
			PSNR (dB)/SSIM	PSNR (dB)/SSIM
×2	Bicubic	—	31.29/0.8525	36.03/0.8875
	SRCNN[7]	8	32.59/0.8705	37.90/0.9072
	FSRCNN[35]	13	32.92/0.8811	38.06/0.9097
	SAFMN[36]	228	33.47/0.8852	38.85/0.9125
	BSRN[17]	332	33.48/0.8852	38.82/0.9123
	IMDN[16]	694	33.48/0.8851	38.77/0.9122
	ESRT[27]	677	33.49/0.8854	38.85//0.9124
	CRAFT[37]	737	33.51/0.8858	**38.93/0.9128**
	PDCLT	519	**33.52/0.8859**	38.88/0.9126
×4	Bicubic	—	24.72/0.5721	30.04/0.6966
	SRCNN[7]	8	26.42/0.6120	31.95/0.7289
	FSRCNN[35]	13	27.26/0.6375	32.22/0.7354
	SAFMN[36]	240	28.20/0.6607	33.69/0.7596
	BSRN[17]	352	28.20/0.6610	33.68/0.7594
	IMDN[16]	715	28.23/0.6622	33.73/0.7601
	ESRT[27]	751	28.21/0.6611	33.72/0.7599
	CRAFT[37]	753	28.19/0.6608	**33.80/0.7606**
	PDCLT	593	**28.26/0.6629**	33.76/0.7604

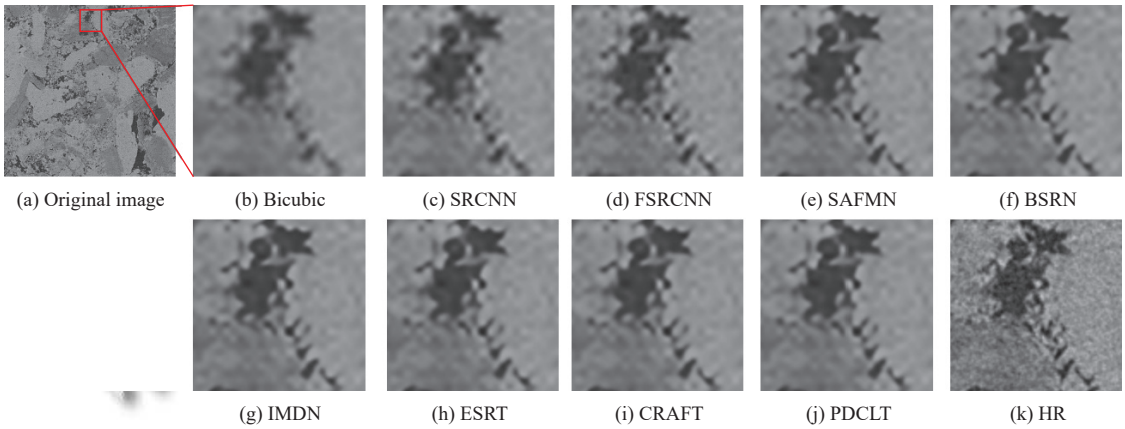

图 6 编号为 1818 的碳酸盐图像×4 倍上重建效果比较

5 结论

本文提出了一种轻量级混合架构模型 PDCLT, 模型结合了细节增强的 CNN 模块和轻量级 Transformer 模块. 其中, 细节增强模块通过像素差分卷积提取岩石 CT 图像中的细节, 并采用残差增强注意力机制捕捉关键特征, 提升局部细节表达能力. 轻量级 Transformer 模块则利用高效的多头注意力处理长距离依赖问题, 捕获全局特征, 并用多尺度融合网络对特征进行多尺度融合, 加强尺度间的交互, 充分利用空间信息, 提升超分辨率重建效果. 在 DRSRD1_2D 和 DeepRock-SR_2D 数据集上的实验验证表明, PDCLT 模型在岩石 CT 图像重建质量上优于主流图像超分辨率模型, 并在性能和参数规模之间实现了良好平衡. 未来研究将聚焦于×2 倍放大下的模型优化及模型泛化能力的验证.

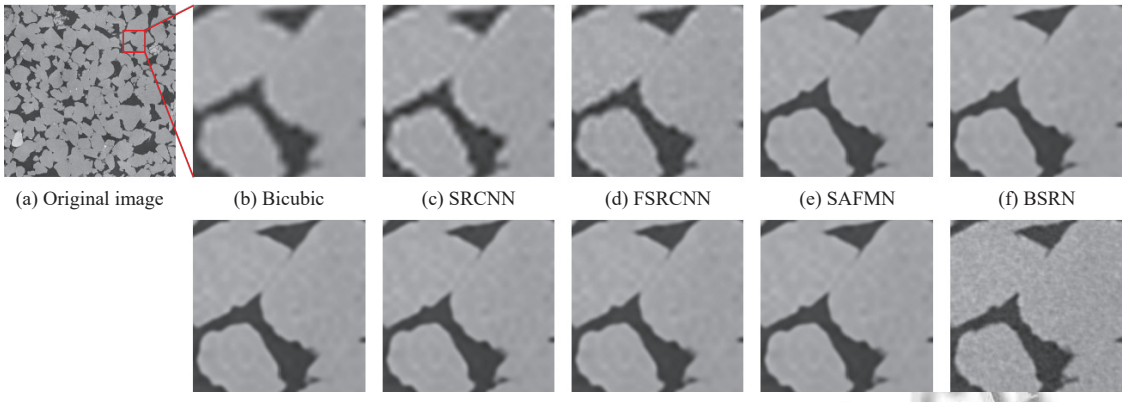

图 7 编号为 1897 的砂岩图像×4 倍上重建效果比较

参考文献

1 Ketcham RA, Carlson WD. Acquisition, optimization and interpretation of X-ray computed tomographic imagery: Applications to the geosciences. Computers & Geosciences, 2001, 27(4): 381–400.

2 Cnudde V, Boone MN. High-resolution X-ray computed tomography in geosciences: A review of the current technology and applications. Earth-science Reviews, 2013, 123: 1–17. [doi: 10.1016/j.earscirev.2013.04.003]

3 Anwar S, Khan S, Barnes N. A deep journey into super-resolution: A survey. ACM Computing Surveys, 2021, 53(3): 60.

4 Ledig C, Theis L, Huszár F, *et al*. Photo-realistic single image super-resolution using a generative adversarial network. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 105–114.

5 Zhang YL, Tian YP, Kong Y, *et al*. Residual dense network for image super-resolution. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 2472–2481.

6 Liang JY, Cao JZ, Sun GL, *et al*. SwinIR: Image restoration using swin Transformer. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Montreal: IEEE, 2021. 1833–1844.

7 Dong C, Loy CC, He KM, *et al*. Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(2): 295–307. [doi: 10.1109/TPAMI.2015.2439281]

8 Su Z, Liu WZ, Yu ZT, *et al*. Pixel difference networks for efficient edge detection. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2021. 5097–5107.

9 Vaswani A, Shazeer N, Parmar N, *et al*. Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.

10 Dosovitskiy A. An image is worth 16x16 words: Transformers for image recognition at scale. Proceedings of the 9th International Conference on Learning Representations. OpenReview.net, 2021. 1–21.

11 Kim J, Lee JK, Lee KM. Accurate image super-resolution using very deep convolutional networks. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 1646–1654.

12 Kim J, Lee JK, Lee KM. Deeply-recursive convolutional network for image super-resolution. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 1637–1645.

13 Lim B, Son S, Kim H, *et al*. Enhanced deep residual networks for single image super-resolution. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu: IEEE, 2017. 1132–1140.

14 Tai Y, Yang J, Liu XM, *et al*. MemNet: A persistent memory network for image restoration. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 4549–4557.

15 Hui Z, Wang XM, Gao XB. Fast and accurate single image super-resolution via information distillation network. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 723–731.

16 Hui Z, Gao XB, Yang YC, *et al*. Lightweight image super-resolution with information multi-distillation network. Proceedings of the 27th ACM International Conference on Multimedia. Nice: ACM, 2019. 2024–2032.

17 Li ZY, Liu YQ, Chen XY, *et al*. Blueprint separable residual network for efficient image super-resolution. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). New Orleans: IEEE, 2022. 832–842.

18 Julesz B. A method of coding television signals based on edge detection. The Bell System Technical Journal, 1959, 38(4): 1001–1020. [doi: 10.1002/j.1538-7305.1959.tb01586.x]

19 Canny J. A Computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986, PAMI-8(6): 679–698. [doi: 10.1109/tpami.1986.4767851]

20 van Vliet LJ, Young IT, Beckers GL. A nonlinear laplace operator as edge detector in noisy images. Computer Vision, Graphics, and Image Processing, 1989, 45(2): 167–195. [doi: 10.1016/0734-189x(89)90131-x]

21 Dollár P, Zitnick CL. Fast edge detection using structured forests. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(8): 1558–1570. [doi: 10.1109/TPAMI.2014.2377715]

22 Martin DR, Fowlkes CC, Malik J. Learning to detect natural image boundaries using local brightness, color, and texture cues. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(5): 530–549. [doi: 10.1109/tpami.2004.1273918]

23 Ganin Y, Lempitsky V. N^4-fields: Neural network nearest neighbor fields for image transforms. Proceedings of the 12th Asian Conference on Computer Vision (ACCV 2014). Singapore: Springer, 2015. 536–551. [doi: 10.1007/978-3-319-16808-1_36]

24 Liu Y, Cheng MM, Hu XW, *et al*. Richer convolutional features for edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(8): 1939–1946. [doi: 10.1109/tpami.2018.2878849]

25 Liu Z, Lin YT, Cao Y, *et al*. Swin Transformer: Hierarchical vision Transformer using shifted windows. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 9992–10002.

26 Chen XY, Wang XT, Zhou JT, *et al*. Activating more pixels in image super-resolution Transformer. Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 22367–22377.

27 Lu ZS, Li JC, Liu H, *et al*. Transformer for single image super-resolution. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). New Orleans: IEEE, 2022. 456–465.

28 Gao GW, Wang ZX, Li JC, *et al*. Lightweight bimodal network for single-image super-resolution via symmetric CNN and recursive Transformer. Proceedings of the 31st International Joint Conference on Artificial Intelligence. Vienna: ijcai.org, 2022. 913–919.

29 Bachlechner T, Majumder BP, Mao H, *et al*. ReZero is all you need: Fast convergence at large depth. Proceedings of the 37th Conference on Uncertainty in Artificial Intelligence. AUAI Press, 2021. 1352–1361.

30 Chiang D, Cholak P. Overcoming a theoretical limitation of self-attention. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Dublin: Association for Computational Linguistics, 2022. 7654–7664. [doi: 10.18653/v1/2022.acl-long.527]

31 Shi D. TransNeXt: Robust foveal visual perception for vision Transformers. Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024. 17773–17783.

32 Wang YD, Armstrong RT, Mostaghimi P. A super resolution dataset of digital rocks (DRSRD1): Sandstone and carbonate [Technical Report]. Digital Rocks Portal, 2019. https://www.digitalrocksportal.org/projects/211 [doi: 10.17612/P7D38H]

33 Wang YD, Armstrong RT, Mostaghimi P. A diverse super resolution dataset of digital rocks (DeepRock-SR): Sandstone, carbonate, and coal [Technical Report]. Digital Rocks Portal, 2019. https://www.digitalrocksportal.org/projects/215 [doi: 10.17612/s3m9-e024]

34 Wang Z, Bovik AC, Sheikh HR, *et al*. Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing, 2004, 13(4): 600–612. [doi: 10.1109/TIP.2003.819861]

35 Dong C, Loy CC, Tang XO. Accelerating the super-resolution convolutional neural network. Proceedings of the 14th European Conference on Computer Vision (ECCV 2016). Amsterdam: Springer, 2016. 391–407.

36 Sun L, Dong JX, Tang JH, *et al*. Spatially-adaptive feature modulation for efficient image super-resolution. Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision. Paris: IEEE, 2023. 13144–13153.

37 Li A, Zhang L, Liu Y, *et al*. Feature modulation Transformer: Cross-refinement of global representation via high-frequency prior for image super-resolution. Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision. Paris: IEEE, 2023. 12480–12490.

(校对责编: 张重毅)