计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
2025,34(2):225−236 [doi: 10.15888/j.cnki.csa.009747] [CSTR: 32024.14.csa.009747]

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

基于多重蒸馏与 Transformer 的遥感图像超分辨率重建①

王　军[1,2], 陈莹莹[1], 程　勇[1,2]

[1](南京信息工程大学 软件学院, 南京 210044)
[2](南京信息工程大学 科技产业处, 南京 210044)
通信作者: 陈莹莹, E-mail: 202212490365@nuist.edu.cn

摘　要: 现有的基于卷积神经网络的超分辨率重建方法由于感受野限制, 难以充分利用遥感图像丰富的上下文信息和自相关性, 导致重建效果不佳. 针对该问题, 本文提出了一种基于多重蒸馏与 Transformer 的遥感图像超分辨率 (remote sensing image super-resolution based on multi-distillation and Transformer, MDT) 重建方法. 首先结合多重蒸馏和双注意力机制, 逐步提取低分辨率图像中的多尺度特征, 以减少特征丢失. 接着, 构建一种卷积调制 Transformer 来提取图像的全局信息, 恢复更多复杂的纹理细节, 从而提升重建图像的视觉效果. 最后, 在上采样过程中添加全局残差路径, 提高特征在网络中的传播效率, 有效减少了图像的失真与伪影问题. 在 AID 和 UCMerced 两个数据集上的进行实验, 结果表明, 本文方法在放大至 4 倍超分辨率任务上的峰值信噪比和结构相似度分别最高达到了 29.10 dB 和 0.780 7, 重建图像质量明显提高, 并且在细节保留方面达到了更好的视觉效果.

关键词: 超分辨率重建; 多重蒸馏; Transformer; 双注意力机制; 遥感图像

引用格式: 王军,陈莹莹,程勇.基于多重蒸馏与 Transformer 的遥感图像超分辨率重建.计算机系统应用,2025,34(2):225–236. http://www.c-s-a.org.cn/1003-3254/9747.html

Remote Sensing Image Super-resolution Reconstruction Based on Multi-distillation and Transformer

WANG Jun[1,2], CHEN Ying-Ying[1], CHENG Yong[1,2]

[1](School of Software, Nanjing University of Information Science & Technology, Nanjing 210044, China)
[2](Science and Technology Industries Division, Nanjing University of Information Science & Technology, Nanjing 210044, China)

Abstract: Existing super-resolution reconstruction methods based on convolutional neural networks are limited by their receptive fields, which makes it difficult to fully utilize the rich contextual information and auto-correlation in remote sensing images, resulting in suboptimal reconstruction performance. To address this issue, this study proposes a novel network, termed MDT, a remote sensing image super-resolution rebuilding method based on multi-distillation and Transformer. Firstly, the network combines multiple distillations with a dual attention mechanism to progressively extract multi-scale features from low-resolution images, thereby reducing feature loss. Next, a convolutional modulation-based Transformer is constructed to capture global information in the images, recovering more complex texture details and enhancing the visual quality of the reconstructed images. Finally, a global residual path is added during upsampling to improve the propagation efficiency of features within the network, effectively reducing image distortion and artifacts. Experiments conducted on the AID and UCMerced datasets demonstrate that the proposed method achieves a peak signal-to-noise ratio (PSNR) and a peak structural similarity index (SSIM) of 29.10 dB and 0.780 7, respectively, on ×4 super-resolution tasks. The quality of the reconstructed images is significantly improved, with better visual effects in terms of

① 基金项目: 国家自然科学基金 (41975183)
收稿时间: 2024-07-02; 修改时间: 2024-07-25; 采用时间: 2024-08-01; csa 在线出版时间: 2024-12-06
CNKI 网络首发时间: 2024-12-06

detail preservation.

Key words: super-resolution reconstruction; multi-distillation; Transformer; dual attention mechanism; remote sensing image

在计算机视觉领域中, 图像超分辨率 (super-resolution, SR) 重建被视为一项非常重要和极具挑战的任务. 单图像超分辨率 (single image super resolution, SISR) 重建旨在从退化的低分辨率 (low resolution, LR) 图像中恢复出高分辨率 (high resolution, HR) 图像. 近年来, 高分辨率遥感图像被广泛用于气象观测[1]、资源测绘[2]、灾害监测[3]等多个重要领域. 然而, 受硬件和环境因素影响, 采集到的遥感图像通常分辨率较低, 存在细节丢失的问题. 鉴于从硬件方面提升遥感图像质量的成本较高, 图像超分辨率重建技术已经成为提高遥感图像分辨率和质量的有效解决方案.

随着深度学习的发展, 卷积神经网络 (convolutional neural network, CNN) 因其强大的特征提取能力在图像超分辨率重建领域展现出巨大的潜力, 许多基于 CNN 的 SISR 方法相继被提出. 然而, CNN 在捕捉图像的长距离依赖关系方面存在不足且网络结构通常较深, 无法很好地提取遥感图像中的全局特征, 容易导致图像细节损失或产生伪影. 此外, 遥感图像存在高空间分布的特点[4], 且地物大小和形状呈现多样性, 使其超分辨率重建任务面临独特的挑战. 同时, 遥感图像还具有高度自相似性, 即单个图像中存在相似斑块重复出现的情况, 如建筑物的屋顶、道路斑马线等. 因此, 相比于 CNN, Transformer 是一种适合处理序列数据的神经网络, 可以有效地对序列之间的依赖关系进行建模. 在遥感图像超分辨率任务中, Transformer 可以帮助模型学习整个图像中相似斑块的边缘和纹理, 对全局特征进行远距离建模从而提升图像重建性能. 但是 Transformer 也存在以下两个缺点: 首先基于 Transformer 的网络更加庞大, 需要消耗更多的计算资源, 推理速度也更加缓慢. 其次, Transformer 缺乏在局部区域内进行信息交换的机制, 忽略了图像的局部细节信息.

针对上述问题, 本文充分考虑 CNN 和 Transformer 的互补性, 提出基于多重蒸馏与 Transformer 的遥感图像超分辨率 (multi-distillation with Transformer for remote sensing super-resolution, MDT) 重建方法. 本文的贡献可归纳如下.

1) 提出一种新的融合架构——MDT. 该架构采用串联方式结合 CNN 和 Transformer. 其中, CNN 用于提取 LR 图像的浅层局部特征, Transformer 则在深层上提取全局特征. 通过结合 CNN 与 Transformer 的优势有效增大网络的感受野, 从而充分提取遥感图像的多尺度特征.

2) 提出多重蒸馏双注意力模块 (multi-distillation dual-attention block, MDDAB), 利用蒸馏机制和双注意力机制逐层提取 LR 图像的局部特征, 从而减少低级特征的丢失.

3) 提出卷积调制 Transformer (convolutional modulation Transformer, CMT), 引入卷积调制层获取全局特征, 并设计空间仿射前馈网络模块 (spatial-affine feed-forward network, SAFN), 有效整合局部和全局特征, 提升模型的重建效果.

4) 在 UCMerced 和 AID 两个遥感数据集上进行大量实验并进行数据和可视化分析, 证明本文方法在平衡参数量、计算复杂度和重构图像质量 3 个方面具有显著效果.

1 相关工作

目前, 图像超分辨率重建技术主要分为 3 类[5]: 基于插值的方法、基于重建的方法和基于学习的方法. 基于插值的方法[6]利用不同的插值操作来估计未知像素值, 这些方法简单易实现, 但重建的图像缺乏细节. 基于重建的方法[7]通过将图像的先验信息作为约束合并到高分辨率图像中来提高图像质量, 然而这些方法需要大量计算成本, 难以在实际应用中推广. 基于学习的方法[8]通过学习 LR 图像和 HR 图像之间的映射关系来指导图像重建. 与上述两种方法相比, 基于学习的方法取得了更好的性能并成为该领域的主流方法.

1.1 基于卷积神经网络的超分辨率重建

2016 年, Dong 等人[9]首次引入了深度学习模型来解决单图像超分辨率任务, 该模型通过 3 层 CNN 实现了低分辨率图像与高分辨率图像之间的端到端特征映射. 2016 年, 为扩大网络的感受野, Kim 等人[10]将递归

神经网络应用于 SR 任务, 利用递归监督策略和残差学习来解决模型梯度爆炸问题. 2017 年, Lim 等人[11]为克服网络深度增加后的优化难题, 提出增强深度超分辨率网络 EDSR. 随后, 为了恢复更多高频细节, Ledig 等人[12]通过引入对抗损失和感知损失, 提出基于生成对抗网络的超分辨率方法 SRGAN. 2018 年, Zhang 等人[13]提出了一个残差密集网络 RDN, 通过密集连接提取了大量的局部特征和层级特征. 紧接着, Ahn 等人[14]为进一步学习多级特征表征, 提出级联残差网络 CARN. 2019 年, 基于 IDN[15]的思想, Hui 等人[16]进一步提出了信息多蒸馏网络 IMDN, 通过提取层级特征来扩大网络感受野, 从而改善性能. Pan 等人[17]针对模型参数量大的问题, 提出了用于遥感图像 SR 的残差密集反投影网络, 该网络利用残差反投影模块简化网络并加速重建过程. 2020 年, Liu 等人[18]提出的 RFDN 模型改变了基于 IMDN 模块的通道分裂方法, 对残差块中的卷积层采用跳跃式连接, 优化了网络性能. 2021 年, Zhang 等人[19]设计了一种基于混合高阶注意力网络的遥感图像 SR 网络, 采用高阶注意力模型进行遥感特征细节恢复, 并通过引入频率感知连接将特征提取网络和特征细化网络连接起来, 极大地提高了遥感图像的清晰度. 2023 年, Chang 等人[20]提出了一种基于 PEGAN 的超分辨率重建算法模型, 该模型仿照密集网络的连接方式, 利用全局网络的多径连接对网络中各个层间信息进行融合, 得到更丰富的细节信息. 为了轻化网络模型, Wang 等人[21]构建了一个轻量级 FeNet 网络, 在计算成本和重建精度之间取得了较好的平衡. 2023 年, Yi 等人[22]将对比学习融入盲遥感图像超分辨率任务中, 探索了未知退化的盲遥感图像超分辨率方法, 进一步推进了该领域的研究.

得益于 CNN 在局部特征提取方面的卓越性能, 这些网络在图像超分辨率重建领域中取得了显著成果. 然而, 在小感受野内, CNN 很难学习到图像的全局信息. 因此, CNN 在恢复图像纹理和边缘细节方面还存在很大的限制.

1.2 基于 Transformer 的超分辨率重建

Transformer 模型最开始由 Vaswani 等人[23]提出, 通过在视觉任务中采用 Transformer, 它已成功应用于图像分类[24]、目标检测[25]和低级图像处理[26]. 2020 年, Yang 等人[27]将 Transformer 应用于图像 SR 重建, 提出了基于参考图像的 SR 重建模型 TTSR. 2021 年, Liang 等人[28]在 Swin Transformer[29]的基础上提出了 SwinIR, 使用滑动窗口自注意力增强图像块间的信息联系. 2022 年, Wang 等人[30]提出了一种基于 Swin Transformer 融合注意力网络的遥感图像 SR 网络, 该网络提出具有融合注意力机制的 Swin Transformer 模块, 该模块的主要作用是提取高频信息, 再将梯度方法加入网络中, 有效增强了网络重建细节的能力. He 等人[31]提出了一种基于 ResNet 的密集光谱 Transformer 来实现多光谱遥感图像的光谱 SR, 该网络将 Transformer 与 ResNet 结合起来, 满足遥感图像学习远程关系的需求. Lu 等人[32]提出具有高效多头注意力机制的 ESRT 模型, 显著降低了模型的计算成本和内存占用. 2023 年, Choi 等人[33]提出一种高效的 SR 网络 NGSwin, 首次将 N-Gram 上下文引入到基于 Transformer 的低级视觉算法中, 通过扩展感受野来恢复更多的退化像素. Chen 等人[34]提出一种混合注意力 Transformer, 引入重叠交叉注意模块来增强相邻窗口特征之间的交互, 进一步地提升了超分辨率模型的性能. 由于多头自注意力在提取跨尺度信息方面能力有限, Yoo 等人[35]在 Transformer 分支中采用跨令牌注意力以充分利用不同尺度的信息. 2024 年, Xiao 等人[36]提出一种结合多尺度特征聚合和自适应令牌选择的 Transformer 方法, 在大面积地球观测场景的 SISR 任务中显著提升了性能.

总的来说, 尽管 CNN 在捕获局部相关性方面表现出色, Transformer 在处理全局依赖性方面具有优势, 但两者均在充分提取图像特征方面有一定的缺陷. 具体来说, CNN 网络无法充分利用遥感图像高度自相似性的特点, Transformer 无法有效提取 SISR 任务中不可或缺的局部信息. 因此, 本文将 CNN 和 Transformer 两者优势结合, 利用 CNN 部分的蒸馏机制和双注意力机制逐层获取低层次的局部特征, 使模型具备超分辨率重建的初步能力, 随后利用卷积调制 Transformer 中的卷积调制块建立全局特征依赖关系, 提取深层遥感图像特征.

2 本文方法

2.1 网络整体架构

MDT 的总体架构如图 1 所示, 主要由 4 个部分组成: 浅层特征提取模块、多重蒸馏 CNN 模块、卷积调制 Transformer 模块和图像重建模块. 本文将低分辨率图像 I_{LR} 和超分辨率重建图像 I_{SR} 分别定义为网络的输入和输出, 高分辨率图像 I_{HR} 定义为理想图像.

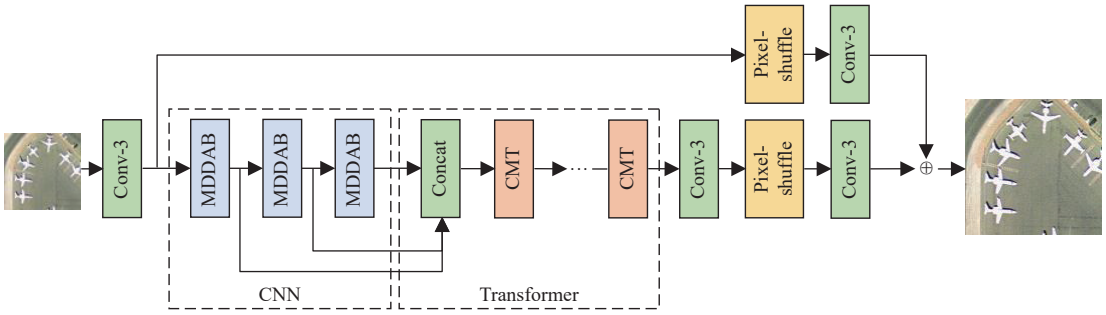

图 1 MDT 网络的总体架构

首先, 我们使用一个 3×3 卷积层从I_{LR}中提取浅层特征, 具体公式为:

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

其中, F_{conv}表示浅层特征提取模块, F_0表示提取的浅层特征. 随后, 将浅层特征F_0传递给 CNN 和 Transformer 部分, 以提取更深层次的特征. 其中, CNN 部分主要由 MDDAB 构成, Transformer 部分主要由 CMT 构成. 假设 MDDAB 的数量为N, 则第n $(1 \leqslant n \leqslant N)$个 MDDAB 的输出可表示为:

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

其中, ζ^n表示第n个 MDDAB 模块, F_n表示第n个 MDDAB 模块的输出. 此外, 每个 MDDAB 模块的输出都会被合并发送到 Transformer 中用于提取更高级、更深层的特征. 假设 CMT 的数量为M, 则第m $(1 \leqslant m \leqslant M)$个 CMT 的输出可表示为:

$$F_d = \phi^m\left(\phi^{m-1}\left(\cdots\left(\phi^1([F_1, F_2, \cdots, F_m])\right)\right)\right) \tag{3}$$

其中, ϕ^m表示第m个 CMT 模块, F_d表示第m个 CMT 的输出. 最后, 将F_0和F_d同时送入重建模块, 得到高质量图像. 超分辨率重建图像I_{SR}重构过程可表示为:

$$I_{\mathrm{SR}} = F_{\mathrm{conv}}\left(F_p(f(F_d))\right) + F_{\mathrm{conv}}\left(F_p(F_0)\right) \tag{4}$$

其中, F_{conv}表示 3×3 卷积层, F_p表示亚像素卷积层.

2.2 多重蒸馏双注意力模块

多重蒸馏 CNN 模块的作用是提取低分辨率图像中的潜在的局部特征, 使模型具有初步的超分辨率重建能力.

研究表明目前大多数 SISR 模型在处理过程中通常保持特征图的空间分辨率不变, 这导致模型通常具有庞大的计算成本. 为解决这个问题, 本文基于增强残差特征注意块 (enhanced residual feature attention block, ERFAB) 设计了一种多重蒸馏双注意力模块 MDDAB. MDDAB 的结构如图 2 所示, 首先使用 ERFAB 提取输入特征, 标记为P_{high}. 随后减小特征图的大小, 将下采样的特征图表示为F'_{n-1}, 通过级联多个 ERFAB 模块来获取 LR 图像中的潜在局部特征, 这些 ERFAB 共享权重以减少参数. 特征提取后, 通过双线性插值将F'_{n-1}上采样到原始大小, 得到的特征仍标记为F'_{n-1}. 同时, 使用单个 ERFAB 模块来处理P_{high}得到P'_{high}, 使其与F'_{n-1}的特征空间对齐. 最后, 将F'_{n-1}与P'_{high}拼接以获得特征F''_{n-1}. 该过程可表示为:

$$F''_{n-1} = cat\left(f_a\left(P_{\mathrm{high}}\right), \uparrow f_a^5\left(\downarrow F'_{n-1}\right)\right) \tag{5}$$

其中, ↑和↓分别表示上采样和下采样操作, f_a表示 ERFAB 的操作, f_a^5表示采用了 5 个级联的 ERFAB 操作, $cat(\cdot)$表示拼接操作.

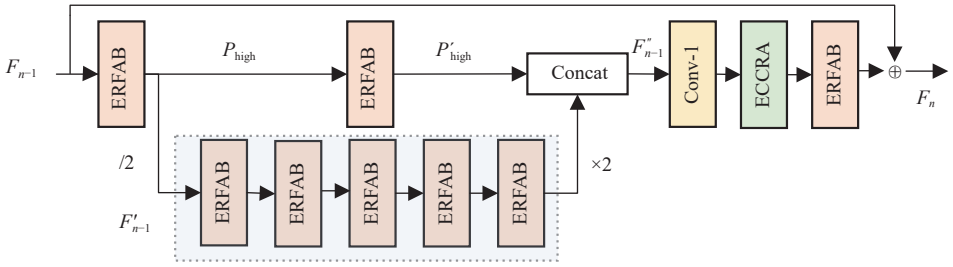

图 2 多重蒸馏双注意力块 (MDDAB)

由于F''_{n-1}是F'_{n-1}与P'_{high}两个特征连接而成, 因此使用一个 1×1 的卷积层来减少通道数. 然而, 特征图通道的减小不可避免地会导致图像细节的丢失. 为解决该问题, 在 MDDAB 模块的最后采用增强型对比通道残差注意力 (enhanced contrast residual attention, ECCRA) 模块[37]和 ERFAB 模块来提取通道方向上的增强特征, 从而缓解了生成的高分辨率图像不自然的问题. ECCRA 的结构如图 3 所示, 与传统的注意力模块不同, ECCRA 利用标准差与均值之和的对比信息来计算通道注意力权重, 并通过在模块内部嵌套残差结构, 解决了特征信息在传递过程中消失的问题. 此外, 由于多个级联的网络容易出现梯度爆炸, 因此 ResNet[38]的思想被引入 MDDAB 中, 添加全局残差连接来稳定训练.

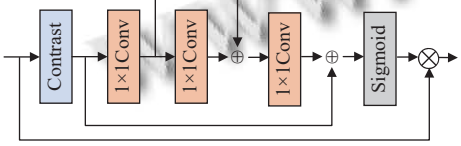

图 3 增强型对比通道残差注意力 (ECCRA) 模块

受 RFDN[39]启发, 信息蒸馏方法能够通过通道分裂操作有效提取蒸馏特征, 同时保证模型更加轻量和灵活. 因此, 本文通过改进 RFDB 模块, 设计增强残差特征注意块 ERFAB 作为 CNN 部分的基本特征提取块, 其结构与 RFDB 类似, 但在保持轻量的同时效率更高.

ERFAB 的结构如图 4 所示, ERFAB 由 3 个阶段组成: 特征蒸馏、特征压缩和特征增强. 在第 1 阶段, 对于输入特征, 模型通过连续的 1×1 卷积层和浅残差块 (shallow residual block, SRB) 提取蒸馏特征. 在特征压缩阶段, 将得到的蒸馏特征拼接在一起并利用 1×1 卷积减少特征的维度, 在减少计算复杂度的同时保留关键的特征信息. 在特征增强阶段, 为在保持效率的同时增强模型的表示能力, 我们引入增强型空间残差注意力 (enhanced spatial residual attention, ESRA)[37], 通过细化特征将遥感图像特征集中在空间尺度上, 达到减少提取冗余遥感特征的目的.

基于蒸馏的方法由于在单个基本块中逐步提取特征的通道分割设计存在冗余参数, 且对空间和通道的建模能力相对较弱. 因此, 对于整个 MDDAB 模块, 采用双注意力机制在通道尺度和空间尺度上提取增强的遥感多尺度特征, 分别对通道和空间维度的依赖关系进行建模. 具体来说, 利用 ERFAB 模块中的增强型空间残差注意力探索低分辨率图像中有价值的空间特征, 借助 MDDAB 模块中的增强型对比通道残差注意力以提取通道方向上的增强特征, 这样使模型兼具空间建模和通道建模能力, 从而提高图像重建性能.

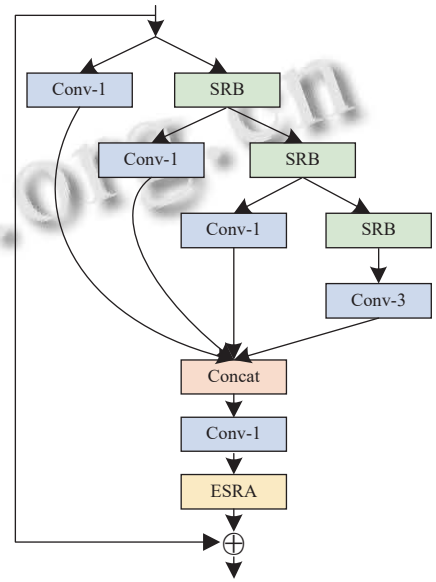

图 4 增强残差特征注意块 (ERFAB)

2.3 卷积调制 Transformer 模块

Transformer 通常由结构相似的编码器和解码器组成. 在 SISR 任务中, Transformer 的编码器可以学习低分辨率图像的特征表示, 使模型能够在特征层面处理和恢复低分辨率图像, 而解码器在 SISR 任务中影响较小, 所以 CMT 模块中只采用了编码器结构. 最近的大多数基于 Transformer 的方法计算成本高且内存占用大. Conv2Former 块[40]最初被开发用于改进传统的自注意力机制, 仅利用卷积和 Hadamard 乘积简化了自注意力机制[41], 与 Transformer 中的自注意力机制相比, 复杂度是线性的. 因此, 本文提出了卷积调制 Transformer 模块以减小模型的复杂度.

CMT 的结构如图 5 所示, 主要由两个归一化层、一个卷积调制层和一个空间仿射前馈层 (*SAFN*) 组成. 假设给定的输入特征为X, 则 CMT 的输出特征Y可表示为:

$$X_{m1} = ConvMod(Norm(X)) + X \tag{6}$$

$$Y = SAFN(Norm(X_{m1})) + X_{m1} \tag{7}$$

其中, *ConvMod* 表示卷积调制操作, *Norm* 表示归一化操作, *SAFN* 表示空间仿射感知器, X_{m1} 表示中间特征.

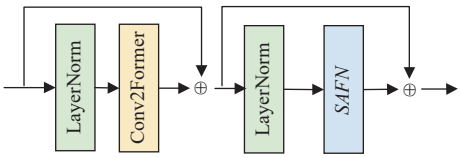

图 5 卷积调制 Transformer (CMT) 模块

前馈网络 (feed-forward network, FFN) 通过非线性激活层和两个线性投影层来提取特征, 但忽略了对空间信息建模, 并且通道中的冗余信息也妨碍了特征表达能力. 为此, 本文设计了空间仿射前馈网络模块 *SAFN*.

如图 6 所示, 输入特征通过仿射变换层进行数据预处理, 用于对数据进行初步标准化和调整. 数据随后通过一个线性层进行线性特征的提取并经过 *GELU* 激活函数处理, 通过引入非线性转换, 使模型能够处理更复杂的特征. 接下来, 特征图被分成两部分, 一部分通过深度卷积层进行处理, 有效地捕捉局部特征, 同时减少计算量. 另一部分特征图保持不变, 通过跳跃连接直接传递. 假设给定的输入特征为 X, 该操作可用公式表示为:

$$X_a, X_b = Split(GELU(Linear(Affine(X)))) \tag{8}$$

$$X_c = DW\text{-}Conv(X_b) \tag{9}$$

其中, *Affine* 表示仿射变换层, *Linear* 表示线性变换层, *DW-Conv* 表示深度卷积层, X_a, X_b 分别表示特征图被分割后的输出, X_c 表示深度卷积层的输出. 然后, 这两个分支的特征图通过逐元素乘法相乘, 结合两部分特征来增强特征表达能力. 最后, 数据再通过一个线性层和一个仿射变换层, 输出最终结果, 完成整个模块的处理过程. 这个设计能够有效地捕捉和整合局部和全局特征, 提高模型的表达能力和性能. *SAFN* 的输出 Y 可表示为式 (10).

$$Y = Affine(Linear(X_a \odot X_c)) \tag{10}$$

其中, $\odot$ 表示 Hadamard 乘积. 与 FFN 相比, *SAFN* 模块能够捕获非线性空间信息并缓解全连接层的通道冗余问题.

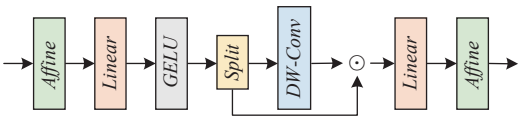

图 6 空间仿射前馈网络 (SAFN) 模块

3 实验结果与分析

3.1 数据集

本文使用两个公开的遥感数据集 UCMerced[42]和 AID[43]验证提出方法的有效性. UCMerced 数据集由 21 类遥感场景组成, 每类包含 100 张图像, 大小为 256×256 像素. 该数据集被平均分成两个不同的集合, 一半是训练集, 一半是测试集, 测试集的 10% 作为验证集. AID 数据集包含机场、教堂、森林、停车场等 30 个遥感场景类别, 共 10 000 张图像, 所有图像的像素大小为 600×600. 对于数据集的划分, 80% 用作训练集, 剩余的 20% 用作测试集, 此外, 在每个类别中随机挑选 5 张图像, 共 150 张图像用作验证集.

3.2 实验设置

本文对尺度因子为×2、×3、×4 的遥感影像数据进行实验. 实验通过对 HR 图像双三次下采样退化生成 LR 图像, 得到 HR-LR 图像对. 对于每次迭代, 网络输入 16 个大小为 48×48 的低分辨率块, 采用水平翻转和随机旋转 (90°、180°、270°) 来增强训练样本. MDDAB 的数量设置为 3, CMT 的数量设置为 2. 模型使用 Adam 优化器进行训练, 参数设置为 $\beta_1 = 0.9$, $\beta_2 = 0.99$. 初始学习率设置为 10^{-4}, 在每训练 400 个 epoch 后学习率下降一半, 批量大小设置为 16. 此外, 采用 L1 损失函数对模型进行共 800 次 epoch 训练. 最后, 训练和测试阶段均在 PyTorch 框架下进行, 使用 Python 3.8 和 PyTorch 1.12.0+CUDA 113 搭建模型, 在 RTX 3090 24 GB 上展开实验.

3.3 评价指标

实验使用峰值信噪比 (peak-signal-to-noise ratio, *PSNR*) 和结构相似度 (structural similarity, *SSIM*)[44]两种指标对实验进行客观评估.

PSNR 是指通过计算重建 HR 图像与真实 HR 图像对应像素点之间的误差, 从而客观地评估重建图像失真程度的指标. *PSNR* 值主要由均方误差 (mean-square error, *MSE*) 决定, *PSNR* 表达式如式 (11) 所示, 单位为分贝 (dB):

$$PSNR = 10 \times \log_{10}\left(\frac{MAX^2}{MSE}\right) = 20 \times \log_{10}\left(\frac{MAX}{\sqrt{MSE}}\right) \tag{11}$$

其中, *MAX* 为 I_{HR} 图像的最大像素值, 对于 8 比特精度的图像, *MAX* 取值为 255. *PSNR* 值的取值范围为 $[0, +\infty)$, 其值越大, 则表示重建 HR 图像与真实 HR 图

像之间的像素误差越小, 重建 HR 图像相对于真实 HR 图像的失真越少, 重建图像的质量越好.

SSIM 是从亮度、对比度和结构 3 个方面来衡量参考图像与失真图像之前结构相似性的方法. *SSIM* 的表达式如式 (12) 所示.

$$SSIM(x,y)=\left[\frac{2\mu_x\mu_y+C_1}{\mu_x^2+\mu_y^2+C_1}\right]\cdot\left[\frac{2\sigma_x\sigma_y+C_2}{\sigma_x^2+\sigma_y^2+C_2}\right]\cdot\left[\frac{\sigma_{xy}+\frac{C_2}{2}}{\sigma_x\sigma_y+\frac{C_2}{2}}\right] \tag{12}$$

其中, μ_x 与 μ_y 表示重建图像与真实图像的均值, σ_x 与 σ_y 表示重建图像与真实图像的标准差, σ_{xy} 表示重建图像与真实图像之间的协方差, C_1 与 C_2 为固定常数, 用来保证分母的稳定. *SSIM* 的取值范围为[−1, 1], *SSIM* 值越接近 1 表示 SR 图像与 HR 图像越接近.

3.4 实验结果及分析

3.4.1 与其他方法的定量分析

本文将提出的方法与 8 种主流算法分别在两个数据集上进行了比较, 包括: SRCNN[9]、FSRCNN[45]、VDSR[46]、LGCNet[47]、IMDN[16]、CTNet[48]、ESRT[32]、FeNet[21]. 表 1 展示了这些方法分别在 UCMerced 和 AID 测试集上进行放大至 2 倍、3 倍和 4 倍的比较结果, 其中加粗数值表示为最佳结果, 下划线数值表示次佳结果.

表 1 不同数据集在不同方法不同放大倍数下的平均 *PSNR* 及 *SSIM*

数据集	方法	2倍		3倍		4倍	
		PSNR (dB)	*SSIM*	*PSNR* (dB)	*SSIM*	*PSNR* (dB)	*SSIM*
UCMerced	SRCNN	32.84	0.9162	28.92	0.8116	26.79	0.7243
	FSRCNN	31.71	0.9003	28.21	0.7937	26.14	0.7002
	VDSR	33.43	0.9235	29.30	0.8260	27.12	0.7362
	LGCNet	33.44	0.9240	29.29	0.8240	27.05	0.7345
	IMDN	33.52	0.9243	29.42	0.8315	27.23	0.7432
	CTNet	33.54	0.9245	29.44	0.8317	27.28	0.7442
	ESRT	33.64	0.9257	29.50	**0.8331**	27.36	0.7468
	FeNet	33.62	**0.9264**	29.47	0.8320	27.30	0.7447
	MDT	**33.68**	0.9260	**29.56**	0.8323	**27.45**	**0.7516**
AID	SRCNN	34.50	0.9286	30.60	0.8380	28.45	0.7560
	FSRCNN	33.99	0.9212	30.19	0.8261	28.12	0.7435
	VDSR	35.03	0.9345	31.14	0.8518	28.97	0.7753
	LGCNet	34.81	0.9321	30.83	0.8443	28.62	0.7632
	IMDN	35.10	0.9346	31.14	0.8518	28.96	0.9698
	CTNet	35.11	0.9354	31.18	0.8524	29.00	0.7772
	ESRT	35.16	0.9356	31.23	0.8527	29.06	0.7801
	FeNet	35.15	0.9361	31.25	0.8549	29.08	0.7804
	MDT	**35.20**	**0.9366**	**31.27**	**0.8552**	**29.10**	**0.7807**

注: 加粗数据表示最优解,下划线数据表示次优解

通过表 1 能够发现, 在放大 4 倍的 SR 任务中, MDT 网络在 UCMerced 和 AID 两个测试集上 *PSNR* 和 *SSIM* 指标均取得了最佳表现, UCMerced 测试集的 *PSNR* 和 *SSIM* 分别为 27.45 dB 和 0.7516, AID 测试集的 *PSNR* 和 *SSIM* 分别为 29.10 dB 和 0.7807. 其中, MDT 网络在 AID 数据集上表现优于 UCMerced 数据集, 是因为 AID 数据集包含了更多样化和丰富的场景信息, 使 MDT 网络能够更好地分层聚合这些信息, 从而提取更具代表性的特征和细节.

在放大 2 倍和 3 倍的 SR 任务中, MDT 网络在 UCMerced 和 AID 测试集上的 *PSNR* 指标均获得最佳表现, *SSIM* 指标也非常接近最佳结果. 尽管 MDT 网络在放大 2 倍和 3 倍上的超分辨率重建表现不如在放大 4 倍任务中那样出色, 但综合衡量指标仍远优于之前的超分模型. 当进行放大 4 倍 SR 任务时, 信息丢失比放大 2 倍和 3 倍更严重, 这意味着从 LR 到 HR 的重建更具挑战性, 需要模型能够更好地推断丢失的细节. 综上所述, MDT 网络在处理复杂和多样化的图像数据时, 展现了卓越的适应能力和重建能力.

为进一步探索本文方法的优势, 表 2 列出了 UCMerced 数据集 21 个场景类别放大倍数为 3 时不同方法的对比结果, 其中加粗数值表示为最佳结果, 下划线数值表示次佳结果. 本文的 MDT 模型在 17 个场景类别上取得了最佳 *PSNR* 值, CTNet 模型在“农业”这一类别上取得了最佳 *PSNR* 值, ESRT 方法在大部分场景类别上取得了次优结果. 相比于 ESRT 方法, 本文方法在需要清晰边缘和纹理的场景中更加有效, 例如密集住宅、立交桥、网球场等.

表 2 UCMerced 数据集上放大倍数为 3 时每个类别的 *PSNR* 值 (dB)

类别	SRCNN	FSRCNN	VDSR	LGCNet	IMDN	CTNet	ESRT	FeNet	MDT
农业	27.44	27.26	27.82	27.64	28.56	**28.70**	28.20	28.45	27.72
飞机	28.65	27.60	28.58	29.23	29.22	29.26	29.40	29.35	**29.48**
棒球场	34.56	33.79	34.70	34.75	**34.94**	34.84	34.84	34.81	34.89
海滩	37.11	36.53	37.23	37.28	37.45	37.42	37.39	37.46	**37.51**
建筑物	27.22	26.23	27.90	27.88	27.97	27.99	28.14	28.10	**28.23**
灌木丛	26.18	26.05	26.36	26.35	26.42	26.43	26.41	26.39	**26.49**
密集住宅	27.77	26.83	28.31	28.29	28.43	28.44	28.51	28.42	**28.59**
森林	28.35	28.20	28.48	28.41	28.42	28.46	28.46	28.47	**28.50**
高速公路	28.89	28.16	**30.09**	29.55	29.54	29.55	29.88	29.75	29.90
高尔夫场	36.33	35.14	36.52	36.46	36.47	36.48	36.49	36.49	**36.53**
港口	23.09	22.46	23.62	23.61	23.84	23.83	23.93	23.77	**24.01**
十字路口	27.91	27.24	28.35	28.32	28.40	28.42	28.55	28.47	**28.58**
中心住宅	27.35	26.52	27.84	27.78	27.82	27.87	27.95	27.89	**28.01**
移动住房	24.23	23.29	24.68	24.70	24.89	24.90	24.96	24.88	**25.02**
立交桥	26.14	25.54	26.76	26.84	26.86	26.91	27.16	27.03	**27.31**
停车场	23.20	22.63	23.49	23.46	23.49	23.53	23.72	23.69	**23.83**
河流	29.03	28.79	29.03	29.13	29.10	**29.15**	29.12	29.11	29.14
跑道	29.99	29.01	30.55	30.58	30.68	30.71	30.78	30.79	**30.98**
稀疏住宅	30.88	30.00	31.15	31.17	31.28	31.26	31.33	31.29	**31.38**
储油罐	31.67	30.88	32.10	32.16	32.29	32.32	32.40	32.37	**32.50**
网球场	31.28	30.23	31.64	31.59	31.68	31.79	31.95	31.93	**32.12**
平均PSNR	28.92	28.21	29.30	29.29	29.42	29.44	29.50	29.47	**29.56**

注: 加粗数据表示最佳值, 下划线数据表示次优值

3.4.2 模型复杂度分析

表 3 是不同方法的参数量和乘加次数的比较, 输入图像大小为 96×96, 放大倍数为 2. 结果表明, 尽管 LGCNet 模型参数量和乘加次数最少, 但其重构效果并不理想; VDSR 模型的参数量大, 计算复杂度高, 重构效果较差. IMDN、CTNet、ESRT、FeNet 这 4 个模型的 *PSNR* 接近, 其中 ESRT 模型的参数量和计算量与本文方法接近, 但本文 MDT 重构效率更高; 虽然 MDT 方法的参数量和计算量略高于 FeNet, 但在重构效率上表现更为出色, 可见本文方法能够在可接受的计算量和内存成本下实现更佳的性能.

3.4.3 与其他方法的定性分析

为进一步验证本文方法, 图 7、图 8 分别给出了不同模型在 UCMerced 数据集放大倍数为 3 和 4 的情况下的主观视觉感知评价, 即不同网络所重建图像之间的对比图, 并对局部区域进行标注与放大.

表 3 放大倍数为 2 时的模型复杂度

指标	VDSR	LGCNet	IMDN	CTNet	ESRT	FeNet	MDT
参数量 (k)	670	193	694	401	677	350	676
乘加次数 (G)	6.15	1.77	6.35	2.37	12.77	5.46	6.16
在AID上的*PSNR* (dB)	35.03	34.81	35.10	35.11	35.16	35.15	35.20

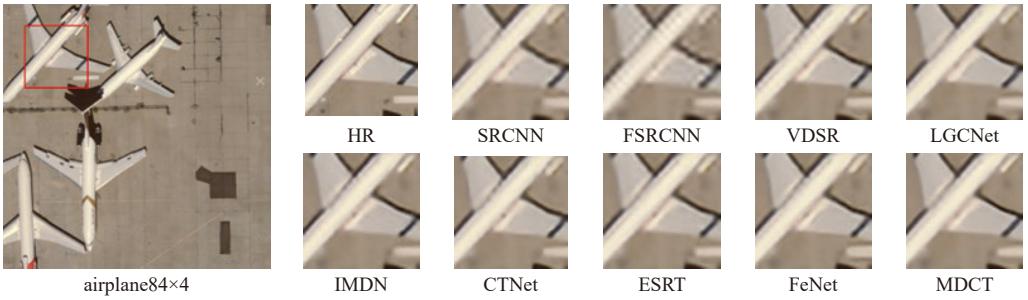

图 7 3 倍放大下不同算法对 UCMerced 数据集中 airplane84.tif 的重构效果

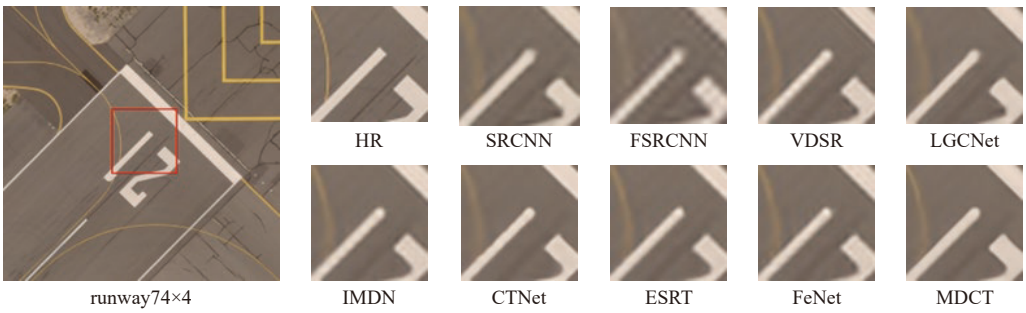

图 8　4 倍放大下不同算法对 UCMerced 数据集中 runway74.tif 的重构效果

图 7 展示了 UCMerced 数据集中的 airplane84.tif 图像在 3 倍 SR 任务中的测试结果. SRCNN 和 FSRCNN 模型重建得到的图像十分模糊, 图像的细节与边缘基本丢失. 尽管 VDSR、LGCNet、IMDN、CTNet 模型有所改进, 但边缘细节仍然模糊. ESRT 和 FeNet 模型基本可以重建图像内容, 但在清晰度方面还有提升空间. 综合来看, MDT 模型所重构的图像内容更清晰, 对飞机机翼线条细节的处理得更为流畅, 与 HR 图像更为接近.

图 8 给出了 UCMerced 数据集 runway74.tif 图像在 4 倍 SR 任务中的测试结果. FSRCNN 模型的重建结果十分模糊, 几乎无法呈现图像细节. 其他方法虽然可以恢复主要轮廓, 但是在纹理结构和边缘细节方面存在不同程度的扭曲. 相比之下, MDT 方法重建的图像轮廓更加自然与清晰. 综上, 本文方法相较于所列其他方法, 更加重视线条纹理与细节部分的恢复.

3.5　消融实验及分析

为验证 MDT 模型中所提出的模块的有效性, 在 UCMerced 数据集上对各结构进行了放大 4 倍超分辨率任务的消融实验, 并对实验结果进行对比与分析.

3.5.1　多重蒸馏双注意力块的有效性

本文方法中用于局部特征提取的卷积网络部分采用了双注意力机制, 为验证其有效性, 本节以仅使用 RFDB 模块作为 MDDAB 模块基础块的基线展开讨论. 双注意力机制是 MDT 模型的重要组成部分, 可以在保持模型参数量几乎不变的情况下, 有效提升模型的性能. 根据表 4 中的方案 1、2 和 3 可以看出, 分别引入 ESRA 模块和 ECCRA 模块后, 模型的 *PSNR* 指标比基线分别提高 0.05 dB 和 0.04 dB. 方案 1 和方案 4 的对比表明, 双注意力机制可以在平衡参数的同时将 *PSNR* 提高 0.08 dB, 这充分说明了在 MDDAB 模块中使用双注意力的必要性. 双注意力机制通过在通道和空间方向建模, 有效利用遥感低分辨率图像的多尺度信息, 因此表现出更好的重建性能.

表 4　多重蒸馏双注意力块的消融实验

方案	RFDB	ESRA	ECCRA	Params (k)	*PSNR* (dB)	*SSIM*
1	√	×	×	739	27.37	0.7474
2	√	√	×	741	27.42	0.7503
3	√	×	√	748	27.41	0.7488
4	√	√	√	750	**27.45**	**0.7516**

3.5.2　增强残差特征注意块的有效性

为验证所提出的 ERFAB 模块的有效性, 本节用 SISR 模型中一些常用的特征提取模块替换 ERFAB 模块进行实验对比. 表 5 展示了在 UCMerced 数据集下放大倍数为 4 时, 不同方案的参数量、乘加次数、*PSNR* 和 *SSIM* 指标的对比结果, 其中乘加次数计算时输入图像大小为 96×96. 从表 5 可以看出, 虽然 ERFAB 模块会带来一定的参数量和计算复杂度的增加, 但其在特征提取方面表现出了更强的能力, 有效提升了模型对高频细节和边缘信息的恢复能力.

表 5　ERFAB 与其他基本模块在放大 4 倍时的性能比较

模块	Params (k)	M-Adds (G)	*PSNR* (dB)	*SSIM*
RCAB[49]	504	6.78	27.32	0.7460
ESDB[50]	397	5.82	27.27	0.7436
SCPA[51]	391	5.85	27.36	0.7472
RFDB[37]	748	9.07	27.41	0.7488
IMDB[16]	594	7.80	27.31	0.7447
ERFAB	750	9.07	**27.45**	**0.7516**

3.5.3　卷积调制 Transformer 模块的数量

为了捕获图像的全局信息, 本文提出了卷积调制 Transformer 模块. 表 6 展示了去掉 Transformer 模块以及使用不同数量 Transformer 模块时, 各项指标的对比结果. 从表 6 第 1 行可以看出, 去掉 Transformer 模块后模型性能明显下降, 这表明 Transformer 模块能充分利用图像中相似图像块之间的信息并建立远程依赖.

表 6 不同数量 Transformer 模块的消融实验

数量	Params (k)	M-Adds (G)	*PSNR* (dB)	*SSIM*
w/o T	682	8.45	27.32	0.7449
1 T	716	8.76	27.39	0.7482
2 T	750	9.07	**27.45**	**0.7516**
3 T	784	9.38	27.42	0.7490

表 6 的第 1 行为没有 Transformer 模块情况下的结果. 第 2、3、4 行分别表示采用 1 个、2 个、3 个卷积调制 Transformer 模块时的实验结果. 随着 CMT 模块数量的增加, 模型的性能也有所提升, 同时模型的参数量和乘加次数也随之增加. 然而, 当 CMT 模块数量达到 3 个时, 模型性能反而有所下降. 这可能是由于过多的 Transformer 模块导致了冗余计算和过拟合现象, 从而影响了模型的整体表现. 为在模型大小和性能之间取得良好的平衡, 最终的 MDT 模型选择使用 2 个 CMT 模块. 使用 2 个 CMT 模块能够在保持较低计算复杂度和参数量的情况下, 显著提升模型对图像全局信息的捕获能力, 从而改善图像重建的清晰度和细节表现.

3.5.4 空间仿射前馈网络模块的有效性

本文提出一个空间仿射前馈网络 *SAFN* 来取代 Transformer 中的 MLP. 从表 7 中可以看出, 虽然 *SAFN* 会带来一定的参数量和计算复杂度的增加, 但其性能显著提升. 具体来说, MLP 方案的 *PSNR* 值为 27.36 dB, *SSIM* 值为 0.7480. 采用去掉门控机制的 *SAFN* 方案, 参数量和计算复杂度保持不变, *PSNR* 和 *SSIM* 值略微下降, 说明门控机制在保持计算量的情况下提升了模型性能. 在仅去掉仿射变换层的 *SAFN* 方案中, 虽然参数量和乘加次数较 MLP 有所增加, 但 *PSNR* 提升了 0.04 dB. 值得注意的是, 使用同时采用门控机制和仿射变换层的 *SAFN* 模块, 参数量和乘加次数变化不大, *PSNR* 显著提高至 27.45 dB, *SSIM* 也提升至 0.7516. 这表明, SGFN 模块在特征提取和归一化过程中, 比常用的 MLP 层表现出更强的能力, 能够更好地重建图像的质量.

表 7 *SAFN* 模块的消融实验

不同方案	Params (k)	M-Adds (G)	*PSNR* (dB)	*SSIM*
MLP	732	8.91	27.36	0.7480
SAFN w/o SG	732	8.91	27.34	0.7457
SAFN w/o Affine	750	9.07	27.40	0.7486
SAFN	750	9.07	**27.45**	**0.7516**

4 结论与展望

本文研究了一种结合多重蒸馏与 Transformer 的遥感图像超分辨率重建方法, 将卷积神经网络和 Transformer 相结合, 使网络同时具备强大的局部拟合能力和全局信息捕获能力, 从而有效提升模型对遥感图像边缘细节的恢复能力. 在两个公开的遥感数据集上进行了大量的实验, 结果证明本文方法在平衡参数量、计算复杂度和重构图像质量 3 个方面具有显著效果, 尤其遥感图像边缘细节恢复上表现出色. 此外, 本文提出的模型参数量只有 676k, 便于在资源有限的硬件环境中部署, 具有广泛的实际应用潜力.

通过将该模型应用于遥感图像处理, 可以显著提高相关下游任务的准确性. 例如, 在变化检测任务中, 模型能够更精准地识别出时间序列图像中的细微变化, 提升监测和管理的效率和可靠性. 在建筑物提取任务中, 模型可以更准确地分割和提取建筑物轮廓, 为城市规划、灾害评估和地图更新等领域提供更高精度的数据支持. 尽管 MDT 网络在 SR 任务中的表现相较主流网络有了明显的提升, 但在某些特定数据集上的表现仍显不足, 后续研究将通过分析这些数据集的特点, 对 MDT 网络进行改进, 以实现更具泛化性能的超分辨率模型.

参考文献

1 薛毅, 李博, 张广科. 浅谈我国卫星遥感应用现状与发展. 中国航天, 2020(4): 51–53.

2 Sun J, Zhu JJ, Tappen MF. Context-constrained hallucination for image super-resolution. Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco: IEEE, 2010. 231–238.

3 Wang YT, Li HF, Jia P, *et al*. Multi-scale densenets-based aircraft detection from remote sensing images. Sensors, 2019, 19(23): 5270. [doi: 10.3390/s19235270]

4 Li R, Zheng SY, Duan CX, *et al*. Land cover classification from remote sensing images based on multi-scale fully convolutional network. Geo-spatial Information Science, 2022, 25(2): 278–294. [doi: 10.1080/10095020.2021.2017237]

5 吴靖, 叶晓晶, 黄峰, 等. 基于深度学习的单帧图像超分辨率重建综述. 电子学报, 2022, 50(9): 2265–2294. [doi: 10.12263/DZXB.20220091]

6 Zhang L, Wu XL. An edge-guided image interpolation algorithm via directional filtering and data fusion. IEEE Transactions on Image Processing, 2006, 15(8): 2226–2238. [doi: 10.1109/TIP.2006.877407]

7 Li XL, Hu YT, Gao XB, *et al*. A multi-frame image super-resolution method. Signal Processing, 2010, 90(2): 405–414. [doi: 10.1016/j.sigpro.2009.05.028]

8 Zeng KL, Lu T, Liang XF, *et al*. Face super-resolution via bilayer contextual representation. Signal Processing: Image Communication, 2019, 75: 147–157. [doi: 10.1016/j.image.2019.03.019]

9 Dong C, Loy CC, He K, *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]

10 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. [doi: 10.1109/CVPR.2016.181]

11 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. 136–144.

12 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. 4681–4690. [doi: 10.1109/CVPR.2017.19]

13 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.

14 Ahn N, Kang B, Sohn KA. Fast, accurate, and lightweight super-resolution with cascading residual network. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 252–268.

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. [doi: 10.1109/CVPR.2018.00082]

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. [doi: 10.1145/3343031.3351084]

17 Pan ZX, Ma W, Guo JY, *et al*. Super-resolution of single remote sensing image based on residual dense backprojection networks. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(10): 7918–7933. [doi: 10.1109/TGRS.2019.2917427]

18 Liu J, Zhang WJ, Tang YT, *et al*. Residual feature aggregation network for image super-resolution. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 2359–2368. [doi: 10.1109/CVPR42600.2020.00243]

19 Zhang DY, Shao J, Li XY, *et al*. Remote sensing image super-resolution via mixed high-order attention network. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(6): 5183–5196. [doi: 10.1109/TGRS.2020.3009918]

20 Jing CW, Huang ZX, Ling ZY. An image super-resolution reconstruction method based on PEGAN. IEEE Access, 2023, 11: 102550–102561. [doi: 10.1109/ACCESS.2022.3142049]

21 Wang ZY, Li LL, Xue Y, *et al*. FeNet: Feature enhancement network for lightweight remote-sensing image super-resolution. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5622112.

22 Xiao Y, Yuan QQ, Jiang K, *et al*. From degrade to upgrade: Learning a self-supervised degradation guided adaptive network for blind remote sensing image super-resolution. Information Fusion, 2023, 96: 297–311. [doi: 10.1016/j.inffus.2023.03.021]

23 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.

24 Dihin RA, Al-Jawher WAM, AlShemmary EN. Diabetic retinopathy image classification using shift window transformer. International Journal of Innovative Computing, 2022, 13(1–2): 23–29. [doi: 10.11113/ijic.v13n1-2.415]

25 Chu XX, Tian Z, Wang YQ, *et al*. Twins: Revisiting the design of spatial attention in vision Transformers. Proceedings of the 35th International Conference on Neural Information Processing Systems. Curran Associates Inc., 2021. 716.

26 Chen HT, Wang YH, Guo TY, *et al*. Pre-trained image processing Transformer. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 12299–12310.

27 Yang FZ, Yang H, Fu JL, *et al*. Learning texture Transformer network for image super-resolution. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 5791–5800. [doi: 10.1109/CVPR42600.2020.00583]

28 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. Montreal: IEEE, 2021. 1833–1844. [doi: 10.1109/ICCVW54120.2021.00210]

29 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. 10012–10022. [doi: 10.1109/ICCV48922.2021.00986]

30 Wang ZL, Shang HX, Wang S. Super-resolution reconstruction of remote sensing images based on swin Transformer fusion attention network. Proceedings of the 2nd International Conference on Optics and Communication Technology. Hefei: SPIE, 2022. 173–181.
31 He J, Yuan QQ, Li J, *et al*. DsTer: A dense spectral Transformer for remote sensing spectral super-resolution. International Journal of Applied Earth Observation and Geoinformation, 2022, 109: 102773. [doi: 10.1016/j.jag.2022.102773]
32 Lu ZS, Li JC, Liu H, *et al*. Transformer for single image super-resolution. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. New Orleans: IEEE, 2022. 457–466.
33 Choi H, Lee J, Yang J. N-gram in swin Transformers for efficient lightweight image super-resolution. Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 2071–2081. [doi: 10.1109/CVPR52729.2023.00206]
34 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.
35 Yoo J, Kim T, Lee S, *et al*. Enriched CNN-Transformer feature aggregation networks for super-resolution. Proceedings of the 2023 IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2023. 4956–4965. [doi: 10.1109/WACV56688.2023.00493]
36 Xiao Y, Yuan QQ, Jiang K, *et al*. TTST: A top-*k* token selective Transformer for remote sensing image super-resolution. IEEE Transactions on Image Processing, 2024, 33: 738–752. [doi: 10.1109/TIP.2023.3349004]
37 He KM, Zhang XY, Ren SQ, *et al*. Deep residual learning for image recognition. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770–778.
38 Liu J, Tang J, Wu GS. Residual feature distillation network for lightweight image super-resolution. Proceedings of the 2020 European Conference on Computer Vision. Glasgow: Springer, 2020. 41–55.
39 Wang Y, Shao ZF, Lu T, *et al*. A lightweight distillation CNN-Transformer architecture for remote sensing image super-resolution. International Journal of Digital Earth, 2023, 16(1): 3560–3579. [doi: 10.1080/17538947.2023.2252393]
40 Hou QB, Lu CZ, Cheng MM, *et al*. Conv2Former: A simple Transformer-style ConvNet for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(12): 8274–8283. [doi: 10.1109/TPAMI.2024.3401450]
41 Luo JJ, Sun XF, Yiu ML, *et al*. Piecewise linear regression-based single image super-resolution via Hadamard transform. Information Sciences, 2018, 462: 315–330. [doi: 10.1016/j.ins.2018.06.030]
42 Yang Y, Newsam S. Bag-of-visual-words and spatial extensions for land-use classification. Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems. San Jose: ACM, 2010. 270–279.
43 Xia GS, Hu JW, Hu F, *et al*. AID: A benchmark data set for performance evaluation of aerial scene classification. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(7): 3965–3981. [doi: 10.1109/TGRS.2017.2685945]
44 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]
45 Dong C, Loy CC, Tang XO. Accelerating the super-resolution convolutional neural network. Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 391–407.
46 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. [doi: 10.1109/CVPR.2016.182]
47 Lei S, Shi ZW, Zou ZX. Super-resolution for remote sensing images via local-global combined network. IEEE Geoscience and Remote Sensing Letters, 2017, 14(8): 1243–1247. [doi: 10.1109/LGRS.2017.2704122]
48 Wang SZ, Zhou TF, Lu Y, *et al*. Contextual transformation network for lightweight remote-sensing image super-resolution. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5615313.
49 Zhang YL, Li KP, Li K, *et al*. Image super-resolution using very deep residual channel attention networks. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 286–301.
50 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. New Orleans: IEEE, 2022. 833–843. [doi: 10.1109/CVPRW56347.2022.00099]
51 Zou WB, Ye T, Zheng WX, *et al*. Self-calibrated efficient Transformer for lightweight super-resolution. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. New Orleans: IEEE, 2022. 930–939. [doi: 10.1109/CVPRW56347.2022.00107]

(校对责编: 张重毅)