

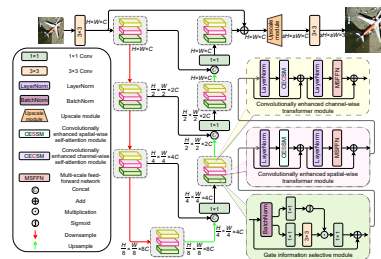
基于自适应双域注意力网络的遥感图像重建

吴菲¹, 陈嘉诚¹, 杨俊^{1*}, 王万良², 李国庆³

¹ 嘉兴大学信息科学与工程学院, 浙江 嘉兴 314000;

² 浙江工业大学计算机科学与技术学院, 浙江 杭州 310000;

³ 宁波大学信息科学与工程学院, 浙江 宁波 315000



摘要: 随着卷积神经网络 (convolutional neural networks, CNN) 和 Transformer 模型的快速发展, 它们在遥感图像超分辨率 (remote-sensing image super-resolution, RSISR) 重建任务中取得了显著进展。然而, 现有方法在处理不同尺度物体特征时表现不足, 同时未能充分挖掘通道与空间维度间的隐性关联, 限制了重建性能的进一步提升。针对上述问题, 本文提出了一种自适应双域注意力网络 (adaptive dual-domain attention network, ADAN)。该网络通过融合通道域与空间域的自注意力信息, 增强了特征提取能力; 设计的多尺度前馈网络 (multi-scale feed-forward network, MSFFN) 能够捕捉丰富的多尺度特征; 结合新颖的门控卷积模块, 进一步提升了局部特征表达能力。基于 U 型结构的网络骨干设计, 实现了高效的多层次特征融合。在多个公开遥感数据集上的实验结果表明, 所提出的 ADAN 方法在定量指标 (如 PSNR 和 SSIM) 以及视觉质量方面均显著优于现有的先进算法, 充分验证了其有效性与先进性, 为遥感图像超分辨率重建提供了新的研究思路和技术路径。

关键词: 双域注意力; Transformer; 注意力机制; 遥感图像; 超分辨率

中图分类号: TP751.1

文献标志码: A

吴菲, 陈嘉诚, 杨俊, 等. 基于自适应双域注意力网络的遥感图像重建 [J]. 光电工程, 2025, 52(4): 240297

Wu F, Chen J C, Yang J, et al. Remote-sensing images reconstruction based on adaptive dual-domain attention network[J].

Opto-Electron Eng, 2025, 52(4): 240297

Remote-sensing images reconstruction based on adaptive dual-domain attention network

Wu Fei¹, Chen Jiacheng¹, Yang Jun^{1*}, Wang Wanliang², Li Guoqing³

¹ College of Information Science and Engineering, Jiaxing University, Jiaxing, Zhejiang 314000, China;

² College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, Zhejiang 310000, China;

³ College of Information Science and Engineering, Ningbo University, Ningbo, Zhejiang 315000, China

Abstract: With the rapid development of convolutional neural networks (CNNs) and Transformer models, significant progress has been made in remote sensing image super-resolution (RSSR) reconstruction tasks. However, existing methods have limitations in effectively handling multi-scale object features and fail to fully explore the implicit correlations between channel and spatial dimensions, thus restricting further improvements in

收稿日期: 2024-12-17; 修回日期: 2025-02-14; 录用日期: 2025-02-17

基金项目: 国家自然科学基金 (61873240, 62302197); 浙江省自然科学基金 (LQN25F020024, LQ23F020006, LQ23F030007); 嘉兴市科技计划基金 (2024AD10045)

*通信作者: 杨俊, juneryoung@zjxu.edu.cn。

版权所有©2025 中国科学院光电技术研究所

reconstruction performance. To address these issues, this paper proposes an adaptive dual-domain attention network (ADAN). The network integrates self-attention information from both channel and spatial domains to enhance feature extraction capabilities. A multi-scale feed-forward network (MSFFN) is designed to capture rich multi-scale features. At the same time, an innovative gated convolutional module is introduced to further enhance the representation of local features. The network adopts a U-shaped backbone structure, enabling efficient multi-level feature fusion. Experimental results on multiple publicly available remote sensing datasets show that the proposed ADAN method significantly outperforms state-of-the-art approaches in terms of quantitative metrics (e.g., PSNR and SSIM) and visual quality. These results validate the effectiveness and superiority of ADAN, providing novel insights and technical approaches for remote sensing image super-resolution reconstruction.

Keywords: dual-domain attention; transformer; attention mechanism; remote sensing images; super-resolution

1 引言

在大数据时代背景下, 遥感技术对高分辨率 (high-resolution, HR) 图像^[1-2]的需求不断增加, HR 图像在土地覆盖分割^[3]、遥感图像分类^[4-5]、地面目标检测^[6]和变化检测^[7]等研究领域中具有重要作用。然而, 由于成像设备与目标物体距离较远, 目标物体的分辨率较低, 影响了后续高层任务的性能。尽管通过升级物理设备可以获取更清晰的 HR 图像, 但成本高昂, 难以广泛应用。而基于图像超分辨率 (single-image super-resolution, SISR) 技术^[8-9]的方法无需依赖昂贵的硬件设备, 为提高遥感图像分辨率提供了一种灵活且经济的解决方案, 因而成为当前研究的重点方向。该技术旨在通过仅利用退化的低分辨率 (low-resolution, LR) 图像生成自然且真实的 HR 纹理。然而, 由于多个不同的 HR 图像可能映射为同一 LR 图像, SISR 属于典型的病态逆问题, 给重建任务带来巨大挑战。尽管提出了插值法、重建法和基于实例的方法等多种方案, 但这些方法往往难以避免伪影问题, 且优化过程复杂且耗时。

近年来, 深度学习推动了 SISR 方法的进展, 尤其是基于 CNN 的方法在特征提取能力上优于传统方法。SRCNN^[10]首次将 CNN 应用于 SISR, 之后基于 CNN 的方法陆续被研究。NLSA^[11]通过稀疏非局部注意力提升性能, Kim 等^[12]则引入残差学习, 构建了 20 层深度网络。这些方法普遍采用 CNN 的局部连通性和平移不变性, 具有一定的高效性和良好的泛化能力, 但也存在两个问题: 1) 感受野有限, 难以捕捉远距离像素依赖; 2) 卷积滤波器静态偏置, 难以灵活适应输入内容。

基于此, Transformer^[13]因其自注意力 (self-

attention, SA) 机制在图像超分辨率中的应用受到关注, 能有效建立全局依赖并弥补 CNN 的不足。结合 CNN 和 Transformer 的混合网络^[14-16], 如 SwinIR^[17]和 ELAN^[18], 取得了显著进展。然而, 这些方法忽视了通道与空间的交互, 并且前馈神经网络 (feed-forward network, FFN) 受通道冗余限制, 未充分考虑多尺度特征与相似物体的尺度关系。

为了解决上述问题, 本文提出了基于自适应双域注意力网络 (adaptive dual-domain attention network, ADAN), 该方法结合了通道域和空间域的自注意力机制, 以提取更可靠的综合深度特征。具体来讲, 在 Transformer 架构^[19]的基础上引入了卷积增强空间维度 Transformer 模块 (convolutionally enhanced spatial-wise transformer module, CESTM) 以及卷积增强通道维度模块 (convolutionally enhanced channel-wise transformer module, CECTM), 两个注意力模块有效地结合了通道和空间之间的相互作用和依赖关系。同时, 为了更好地保存多尺度信息, 提出了一种多尺度前馈神经网络 (multi-scale feed-forward network, MSFFN) 来探索相似物体的潜在尺度关系, 丰富多尺度特征。此外, 考虑到基于自注意力机制的 Transformer 架构缺乏局部特征提取的能力和归纳偏置特性, 因此, 本文进一步引入了一种新颖的门控信息选择模块 (gate information selective module, GISM) 以提升模型挖掘局部细节特征的能力。通过无缝且灵活的集成上述所提出的三种基础模块, 本文构建了一种能够高效重建遥感图像的超分辨率图像网络。

本文的主要贡献包括以下四点:

1) 提出了一种适用于遥感图像超分辨率 (RSSR) 任务的全新方法——基于自适应双域注意力网络 (ADAN)。

2) 设计了并行的通道、空间特征提取模块及门控卷积模块, 全面挖掘通道、空间和卷积维度特征, 增强特征表示能力。

3) 提出了多尺度前馈神经网络 (MSFFN), 有效探索潜在尺度关系, 增强全局表示能力。

4) 在多个数据集上的实验结果表明, ADAN 的重建性能显著优于现有遥感图像 SR 方法。

2 相关工作

2.1 基于 CNN 的自然图像超分辨率重建算法

近年来, 基于 CNN 的技术在单图像超分辨率重建任务中快速发展。Dong 等^[10]提出了第一个基于 CNN 的 SISR 方法 (SRCNN)。受 SRCNN 的启发, 大量基于 CNN 的 SR 方法被用于开发, 在图像 SR 任务中取得了一定进展^[20-23]。He 等^[24]从加深网络深度的角度出发, 利用残差连接建立了一个有 20 层卷积层的深度模型 VDSR^[12], 其性能明显优于浅层网络 SRCNN。这也意味着在一定程度上加深网络深度, 可以提高网络性能。鉴于此发现, 研究人员开始致力于开发设计更深、更宽、更复杂的网络。随后, Lim 等^[25]通过去除传统残差网络中不必要的模块 (比如批归一化^[26]) 以优化整体网络结果和减少模型计算复杂度, 设计了一种约 50 层的增强型深度超分辨率 (enhanced deep super-resolution, EDSR) 网络。但是该方法平等地处理低分辨率特征并且忽略了全局依赖, 导致细节检索效率低下。因此, 注意力机制被引入到基于 CNN 的超分辨率模型中^[27], 例如通道注意力 (channel attention, CA) 机制^[28]和整体注意力 (holistic attention, HA) 机制^[29]等, 以重新平衡各种元素的相关性。Zhang 等^[30]利用残差结构组中的残差块构建了一个超过 400 层的网络, 通过多个残差连接直接传输丰富的低频信息, 使网络专注于学习丢失的高频信息, 从而减轻了模型的训练负担和拟合难度。此外, 引入了 CA 机制, 以考虑通道间的相互依存关系, 自适应地调整通道特征。实验结果表明, 其可以大幅度提高重建性能。为了提取全局先验知识 (例如自相似性), 一些学者提出了自相似注意力机制^[31]和非局部稀疏注意力模块。例如: Mei 等^[11]将非局部操作和稀疏表示结合到 SISR 任务中, 并提出了一种非局部稀疏注意力, 以缓解非局部操作所需的大量计算资源。Zhang 等^[32]针对深度卷积神经网络 (deep convolutional neural networks, DCNN) 没有充分利用特征图、通道

和像素之间的互补优势的缺点, 提出了三个注意力模块: 可训练的特征注意力模块 (feature attention module, FAM)、通道注意力模块 (channel attention module, CAM) 以及像素注意力模块 (pixel attention module)。尽管这些方法在一定程度上增强了模型的全局特征表示能力, 但密集的非局部建模带来了巨大的计算复杂性和训练成本。上述问题, 在空间分辨率较大的遥感图像 SR 任务上尤其突出。

2.2 遥感图像的超分辨率重建算法

遥感图像^[33-34]在地理、环境监测等领域有重要应用, 但由于卫星和传感器的限制, 常常需要对低分辨率图像进行超分辨率处理以获得更高的细节和准确性。遥感图像超分辨率重建在图像领域逐渐得到广泛的关注, 近年来众多学者已经在遥感图像超分辨率重建的方法研究上取得了卓越的成果^[35]。其中基于深度学习的图像超分辨率方法相较于早期的传统方法其高效性, 准确性与鲁棒性都展示出一定的优势。LGCNet^[20]是第一个基于 CNN 的遥感图像超分辨率模型, 使用局部和全局表示来学习重建的 SR 图像。局部模块用于捕捉图像的细节信息, 全局模块用于捕捉图像的整体结构和上下文信息。Dong 等^[36]提出了 SMSR, 该网络通过密集采样策略, 能够更好地捕捉图像中的细节信息和复杂的特征。它利用一阶和二阶学习机制聚合了不同的多尺度特征。近年来, 注意力机制在遥感图像超分辨率重建任务中得到了广泛的应用。例如: 在 HSENet 中, Lei 等^[31]引入了一个单尺度自相似性模块 (single-scale self-similarity exploitation module, SSEM) 来计算同一尺度图像内的特征相关性。此外, 利用了一个跨尺度连接结构 (cross-scale connection structure, CCS) 来捕获跨不同尺度的相似性。Xiao 等^[37]为了降低自注意力机制中的计算冗余, 提出了一种自适应选择 Top-*k* Token 的 Transformer 模型 (Top-*k* token selective transformer, TTST), 其中还包含了多尺度特征聚合以及全局上下文注意力机制, 以提高全局特征以及局部特征的融合。

3 方法描述

3.1 主体结构

本文的主要目标是开发一个高效的 Transformer 模型, 可以处理高分辨率图像的重建任务。所提出的 ADAN 由浅层特征提取、深度特征提取和图像重建三个模块组成, 如图 1 所示, 给定一个低分辨率输

入图像 $I_{LR} \in \mathbb{R}^{H \times W \times 3}$, 浅层特征提取阶段采用了一个 3×3 卷积从低分辨率遥感图像中提取浅层特征 $F_s \in \mathbb{R}^{H \times W \times C}$, 其中 H 和 W 分别表示输入图像的高和宽, C 则表示特征通道的数量。之后, 通过 Transformer 的核心组件: 门控信息选择模块 (gate information selective module, GISM)、卷积增强空间维度 Transformer 模块 (CESTM) 以及卷积增强通道维度 Transformer 模块 (CECTM) 作为深度特征提取模块, 以编解码的方式提取深度特征。其中, 为了能够高效的提取低质图像特征, 门控信息选择模块被用来对浅层特征做预处理, 以便后续的多尺度特征提取。CESTM 以及 CECTM 则用来提取空间以及通道维度的双重信息。具体来讲, 浅层特征 F_s 通过四层相对应的编解码器后得到深度特征 $F_d \in \mathbb{R}^{H \times W \times C}$ 。每个编码器-解码器的层都包含多个 Transformer 块, 从顶部到底部层逐渐增加以保持效率。从高分辨率输入开始, 编码器逐层减少空间尺寸, 同时扩展通道容量。解码器以低分辨率的潜在特征 $F_l \in \mathbb{R}^{\frac{H}{8} \times \frac{W}{8} \times 8C}$ 作为输入, 逐步恢复高分辨率的表示。对于特征下采样和上采样, 分别应用步

长卷积和反卷积实现。为了辅助恢复过程, 编码器的特征通过恒等跳跃连接与解码器的特征拼接。之后通过 1×1 的卷积来降低通道的数量。最后利用上采样模块对高分辨率图像 $F_{HR} \in \mathbb{R}^{H \times W \times C}$ 进行图像重建获得输出图像 $I \in \mathbb{R}^{sH \times sW \times 3}$ 。其中 sH 和 sW 分别表示输出重建图像的高和宽。

3.2 门控信息选择模块

门控信息选择模块 (GISM) 是基于 gMLP^[38] 以及 GLU^[39] 改进的, 其目的是进行高维特征映射以挖掘更精确的信息。批归一化 (BatchNorm) 在每个小批量数据中标准化网络的输入, 使其具有零均值和单位方差。它可以与相邻的线性层合并, 更符合轻量级网络的需求。此外, DehazeFormer^[40] 指出 BatchNorm^[26] 具有维持空间相关性的优点, 不具有 LayerNorm^[41] 打破空间相关性的缺点, 因此本文采用 BatchNorm 来归一化输入数据。具体来说, 令 x 为特征图, 首先采用 BatchNorm 对 x 做归一化处理 $\hat{x} = \text{BatchNorm}(x)$ 。之后对 $\hat{x}$ 进行特征选择以控制信息传输, 具体过程如下:

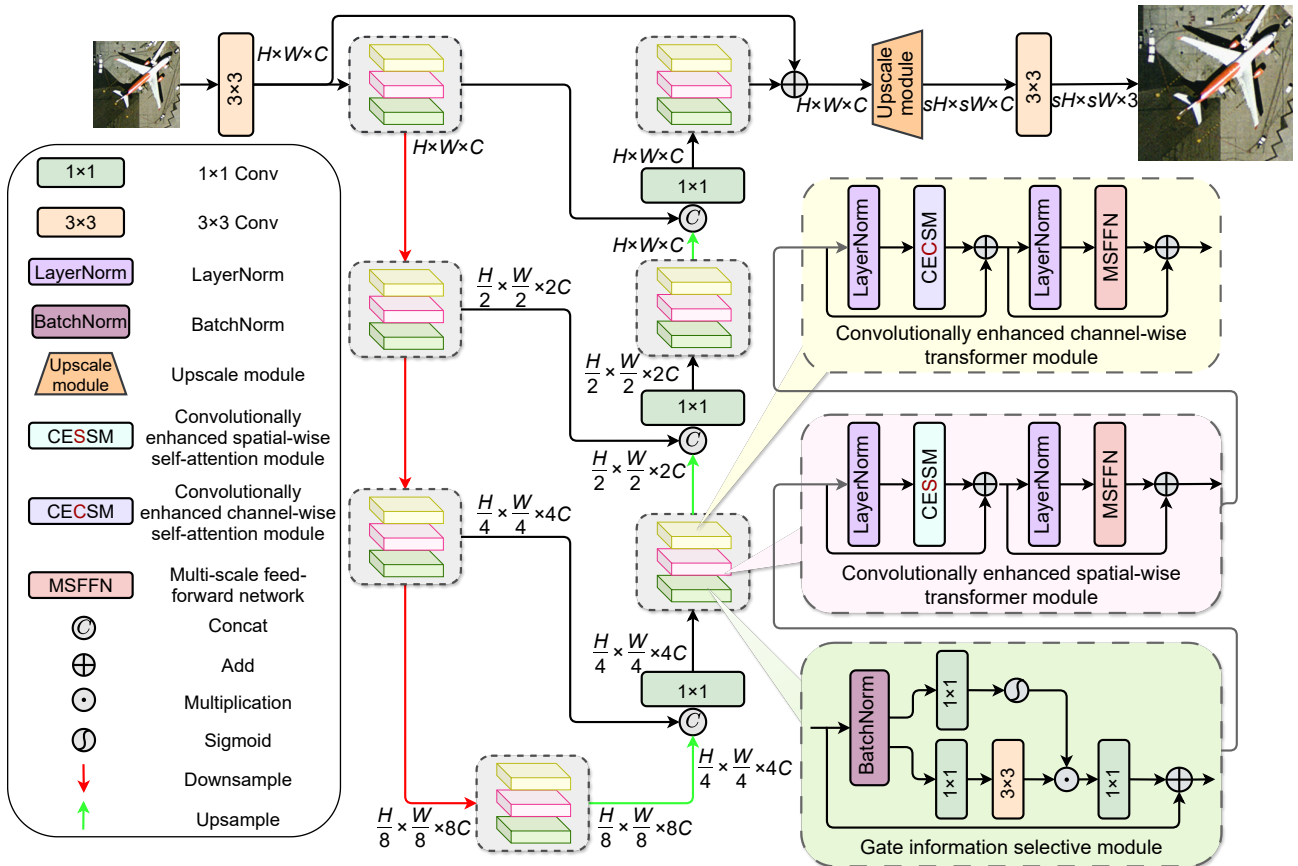


图 1 ADAN 的总体架构和各模块结构

Fig. 1 Overall architecture and module structures of ADAN

$$\begin{cases} x_1 = \text{Sigmoid}(H_{1 \times 1}(\hat{x})) \\ x_2 = H_{\text{DW}3 \times 3}(H_{1 \times 1}(\hat{x})) \end{cases}, \quad (1)$$

式中: $H_{1 \times 1}(\cdot)$ 是卷积核为 1×1 的卷积层; $H_{\text{DW}3 \times 3}(\cdot)$ 是卷积核为 3×3 的深度可分离卷积层。然后使用 x_1 作为 x_2 的门控信号, 接着对 x_2 使用另一个 1×1 卷积进行投影, 并且输出与恒等跳跃连接 x 相加, 可以表示为

$$x_{\text{GISM}} = x + H_{1 \times 1}(x_1 \cdot x_2). \quad (2)$$

3.3 卷积增强空间维度 Transformer 模块

卷积增强空间维度 Transformer 模块 (CESTM) 是 ADAN 的重要组成模块之一。该模块由卷积增强空间维度自注意力模块 (convolutionally enhanced spatial-wise self-attention module, CESSM) 以及多尺度前馈神经网络 (MSFFN) 组成。给定输入 $x_{\text{GISM}} \in \mathbb{R}^{H \times W \times C}$, 该模块被定义为

$$\begin{cases} x' = \text{CESSM}(\text{LN}(x_{\text{GISM}})) + x_{\text{GISM}} \\ x_{\text{CESTM}} = \text{MSFFN}(\text{LN}(x')) + x' \end{cases}, \quad (3)$$

式中: x_{CESTM} 为 CESTM 的输出结果; $\text{LN}(\cdot)$ 为层归一化操作。

3.3.1 卷积增强空间维度自注意力模块

如图 2 所示, CESSM 为了降低计算成本并且不牺牲全局特征学习能力, 本文使用了 CESTM, 而不是常见的自注意力机制。具体来说, 将层归一化 (LN) 后的图像输入 $\hat{x}_{\text{GISM}}$, 划分成大小为 $(HW/N_w) \times N_w \times C$ 的互不重叠的窗口。随后, 对每个窗口 $Z_i (0 \leq i \leq N_w)$ 分别进行自注意力计算。在计算相似性时, 在注意模块中加入一个相对位置偏差 B :

$$\begin{cases} Q_i, K_i, V_i = H_{\text{DW}3 \times 3}(H_{1 \times 1}(Z_i)) \\ x_i^{\text{WSA}} = \text{Softmax}(Q_i \cdot K_i^T / \sqrt{d} + B) \cdot V_i \end{cases}, \quad (4)$$

式中: Q 、 K 、 V 分别为查询 (query)、键 (key) 和值

(value), $Q, K, V \in \mathbb{R}^{\frac{HW}{N_w} \times N_w \times C}$ 。最后将 N_w 个窗口进行拼接得到了完整的 x^{WSA} 。

由于 Transformer 能够高效地捕获非局部依赖, 倾向于提取低频信息, 往往会忽视高频局部细节信息, 而这些高频信息往往对重建任务至关重要。卷积操作能够通过内置的平移等变约束对观察到的图像中的潜在语义模式进行编码, 换言之卷积能够有效地捕捉图像的细致的局部结构。因此为了克服 Transformer 模型的固有缺陷, 本文在 CESTM 中添加了卷积补充模块以进一步提取高频细节信息。因此卷积补充模块可以被表示为

$$x_{\text{Conv}}^{\text{WSA}} = H_{\text{SE}}(H_{3 \times 3}(\delta(H_{3 \times 3}(\hat{x}_{\text{GISM}}))))), \quad (5)$$

式中: $\delta(\cdot)$ 为 ReLU 激活函数; $H_{\text{SE}}(\cdot)$ 的具体实现请参考文献 [30]。最后 CESSM 表示为

$$x_{\text{CESSM}} = x^{\text{WSA}} + x_{\text{Conv}}^{\text{WSA}}. \quad (6)$$

3.4 卷积增强通道维度 Transformer 模块

卷积增强通道维度 Transformer 模块 (CECTM) 与卷积增强空间维度 Transformer 模块 (CESTM) 类似, 不同的是 CECTM 则通过卷积增强通道维度自注意力模块 (convolutionally enhanced channel-wise self-attention module, CECSM) 更加关注通道维度的信息。给定输入 $x_{\text{CESTM}} \in \mathbb{R}^{H \times W \times C}$, CECTM 则被定义为

$$\begin{cases} x' = \text{CECSM}(\text{LN}(x_{\text{CESTM}})) + x_{\text{CESTM}} \\ x_{\text{CECTM}} = \text{MSFFN}(\text{LN}(x')) + x' \end{cases}. \quad (7)$$

3.4.1 卷积增强通道维度自注意力模块

CECSM 用于通道重校准和丰富通道的底层特征的信息交互。该模块与 CESSM 相结合, 有效地结合通道和空间位置之间的相互作用和依赖关系。如图 3 所示, 具体来说, 给定输入 $\hat{x}_{\text{CESTM}} \in \mathbb{R}^{H \times W \times C}$, 利用一

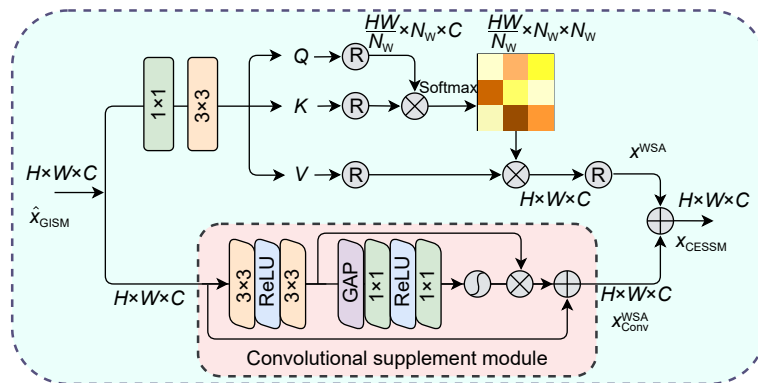


图 2 卷积增强空间维度自注意力模块

Fig. 2 Convolution-enhanced spatial-wise self-attention module

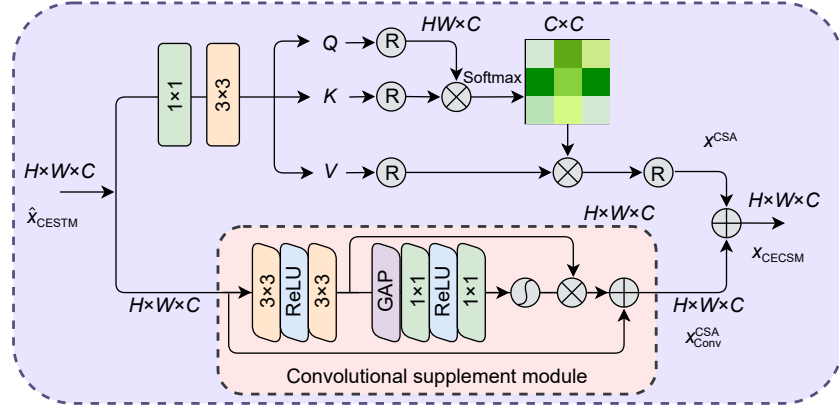


图3 卷积增强通道维度自注意力模块

Fig. 3 Convolution-enhanced channel-wise self-attention module

个 1×1 和一个 3×3 卷积得到查询 (Q)、键 (K) 和值 (V), $Q, K, V \in \mathbb{R}^{H \times W \times C}$ 。然后进行自注意力计算:

$$x^{\text{CSA}} = \text{Softmax}(Q^T \cdot K / \tau) \cdot V, \quad (8)$$

式中: τ 为归一化 $Q^T \cdot K$ 的结果。卷积补充模块与 3.3.1 节中所提方法一致, 最后 CECSM 可以被表示为

$$\begin{cases} x_{\text{Conv}}^{\text{CSA}} = H_{\text{SE}}(H_{3 \times 3}(\delta(H_{3 \times 3}(\hat{X}_{\text{CESTM}})))) \\ x_{\text{CECSM}} = x^{\text{CSA}} + x_{\text{Conv}}^{\text{CSA}} \end{cases} \quad (9)$$

3.5 多尺度前馈神经网络

原始的 MLP 层采用线性投影进行特征传播, 没有明确考虑多尺度特征。例如, TransENet 方法未使用多尺度前馈神经网络, 缺乏多尺度建模能力, 导致性能受限。此外, 自然图像领域中的一些多尺度前馈神经网络虽然通过不同尺度的卷积核提取特征, 但缺乏不同感受野特征的交互。然而, 在遥感图像中通常会出现多尺度特性, 例如一张图中有多个大小不一的同类物体, 不同尺寸的同类物体之间信息可以得到互补。因此, 本文设计了一个高效的多尺度前馈神经网络 (MSFFN) 来生成一组丰富的特征, 有助于将更多的多尺度线索聚集到全局表示中。

如图 4 所示, 在经过一个层归一化之后, 得到归一化特征 $X_1 = \text{LN}(x)$, 然后将 X_1 输入划分成三个并行分支, 分别以 3×3 、 5×5 和 7×7 卷积探索多尺度的表征, 表达式为

$$\begin{cases} X_3 = H_{\text{DW}3 \times 3}(X_1) \\ X_5 = H_{\text{DW}5 \times 5}(X_1) \\ X_7 = H_{\text{DW}7 \times 7}(X_1) \end{cases} \quad (10)$$

为了增强多尺度之间的信息交互, 使用拆分操作沿着通道维度将多尺度特征划分为三个部分 (例如: 以 X_3 为例, 三个部分分别记为: $X_3^{\text{p}_1}$ 、 $X_3^{\text{p}_2}$ 、 $X_3^{\text{p}_3}$), 并且在激活函数 ReLU 之后拼接不同尺度的特征信息。

该过程数学表示为

$$\begin{cases} \bar{X}_3 = \delta(H_{\text{DW}3 \times 3}[X_3^{\text{p}_1}, X_5^{\text{p}_1}, X_7^{\text{p}_1}]) \\ \bar{X}_5 = \delta(H_{\text{DW}5 \times 5}[X_3^{\text{p}_2}, X_5^{\text{p}_2}, X_7^{\text{p}_2}]) \\ \bar{X}_7 = \delta(H_{\text{DW}7 \times 7}[X_3^{\text{p}_3}, X_5^{\text{p}_3}, X_7^{\text{p}_3}]) \\ x_{\text{MSFFN}} = H_{1 \times 1}[\bar{X}_3, \bar{X}_5, \bar{X}_7] + X_1 \end{cases} \quad (11)$$

式中: $[\cdot]$ 表示通道维度的拼接; $H_{1 \times 1}$ 表示 1×1 卷积; x_{MSFFN} 表示多尺度前馈神经网络的输出。

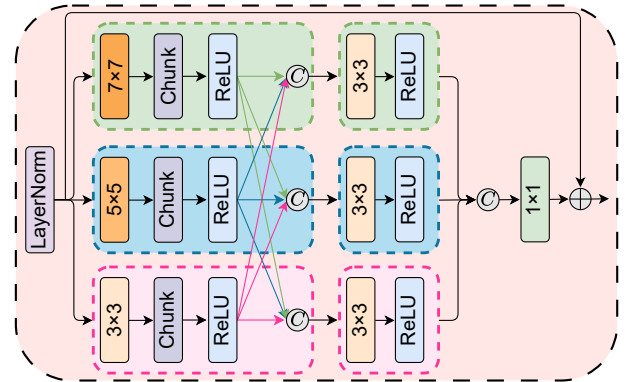


图4 多尺度前馈神经网络

Fig. 4 Multi-scale feedforward neural network

4 实验验证

4.1 实验设置

4.1.1 数据集说明

为了验证方法的有效性, 本文采用了两种公开数据集进行测试 (UCMerced LandUse^[42] 数据集和 AID 数据集^[43])。两种数据集在遥感图像超分辨率任务中得到了广泛的应用。在 MATLAB 设置中使用双三次插值运算对 HR 图像进行相应比例的降采样, 以生成 LR 图像。

UCMerced LandUse 数据集: 该数据集包含了 21 种类别图像, 其中包含了农业、飞机、棒球场以及建筑物等, 共计 2100 张图片, 每个类别分别为 100 张图片, 每张图片的大小是 $256 \text{ pixel} \times 256 \text{ pixel}$, 分辨率为 0.3 m/pixel 。该数据集被分为两部分: 训练集和测试集, 其中 20% 的训练集作为验证集。

AID 数据集: 该数据集主要用于遥感图像分类、目标检测和分割等任务。它由多种遥感图像获取途径汇集而成, 通常包括卫星图像和无人机图像等。该数据集包含了来自 30 种不同类型的遥感场景的 10000 张照片, 如机场、建筑物、汽车、住宅区域等。图片大小为 $600 \text{ pixel} \times 600 \text{ pixel}$, 空间分辨率为 0.5 m/pixel 。根据 TransNet^[44], 随机选择整个数据集的 80% 数据作为训练集, 其余图像则作为 AID 数据集的测试集。此外, 本文在 30 个类别中选取 5 张图片作为验证图像, 因此验证集共计 150 张图片。

4.1.2 评估指标

本文采用了图像超分辨率领域的常用评估指标, 即峰值信噪比 (peak signal-to-noise ratio, PSNR) 和结构相似性指数度量 (structure similarity index measurement, SSIM)^[45], 作为本文图像重建质量的评估指标。所有的超分辨率结果都在 RGB 空间上进行评估。

给定一个超分辨率重建图像 X_{SR} 和一个相对应的高分辨率图像 X_{HR} , PSNR 指标表示为

$$\text{PSNR}(X_{\text{SR}}, X_{\text{HR}}) = 10 \cdot \log_{10} \frac{\max(X_{\text{HR}})^2}{\|X_{\text{SR}} - X_{\text{HR}}\|_2^2}, \quad (12)$$

式中: $\max(X_{\text{HR}})$ 为图像像素最大值; $\|X_{\text{SR}} - X_{\text{HR}}\|_2^2$ 为均方误差。PSNR 值被广泛用于测量 X_{SR} 和 X_{HR} 之间的距离, 反映了绝对重建精度。PSNR 值越大, 表明恢复的图像质量越高。

此外, SSIM 指标评估图像的结构相似性, 其表达式为

$$\text{SSIM}(X_{\text{SR}}, X_{\text{HR}}) = \frac{(2 \cdot \mu_{\text{SR}} \cdot \mu_{\text{HR}} + C_1) \cdot (2 \cdot \sigma_{\text{SR-HR}} + C_2)}{(\mu_{\text{SR}}^2 + \mu_{\text{HR}}^2 + C_1) \cdot (\sigma_{\text{SR}}^2 + \sigma_{\text{HR}}^2 + C_2)}, \quad (13)$$

式中: μ_{SR} 、 σ_{SR} 分别为超分辨率重建图像 X_{SR} 的平均值和标准差; μ_{HR} 、 σ_{HR} 分别为对应的高分辨率图像 X_{HR} 的平均值和标准差; $\sigma_{\text{SR-HR}}$ 为 X_{SR} 和 X_{HR} 的协方差; C_1 和 C_2 为正值常数。SSIM 值越大, 图像纹理的重构效果越好。

本文还进一步引入了学习到的感知图像块相似

度 (learned perceptual image patch similarity, LPIPS)^[46] 来评估竞争方法的重建质量。LPIPS 值越低, 表示感知质量越高。与此同时, 本文还分析了模型的浮点运算 (FLOPS) 和运行时间, 其中 FLOPS 是根据 $48 \text{ pixel} \times 48 \text{ pixel}$ 图像所计算的。

4.1.3 实验环境以及参数设定

为了获得更好的泛化性能, 本文使用了数据增强技术, 其中包括 90° 的随机旋转、随机水平旋转以及垂直翻转。本文采用了 Pytorch 框架来实现和训练所提出的 ADAN, 并使用一台具有一张 NVIDIA GeForce GTX 3090 显卡的服务器对模型进行训练与测试。通过随机初始化, 训练不同的模型来重建放大因子分别为 $\times 2$ 、 $\times 3$ 和 $\times 4$ 的遥感图像。使用 ADAM^[47] 优化器训练模型, 参数设置为 $\beta_1 = 0.900$ 和 $\beta_2 = 0.999$ 。学习速率初始化为 2×10^{-4} , 每 400 个周期减半。在训练过程中, LR 图像被随机裁剪为 $48 \text{ pixel} \times 48 \text{ pixel}$ 输入到网络中, 批大小为 16, 且 HR 图像大小与放大因子相对应。在 ADAN 中, 除 1×1 卷积层外, 所有卷积层通道数为 64。本文选择了 L_1 作为损失函数, 该损失函数通过计算模型输出图像与真实高分辨率图像之间的损失值来训练网络模型中的参数, 模型训练目标为最小化损失函数, 表示为

$$L_1(\theta) = \frac{1}{N} \sum_{i=1}^N \|H_{\text{Model}}(I_{\text{LR}}^i) - I_{\text{LR}}^i\|_1, \quad (14)$$

式中: θ 表示整个网络中需要学习的所有参数; N 表示每个 batch 中的图像数目。

4.2 实验结果与分析

4.2.1 定量结果

为了验证所提出的算法 ADAN 的有效性, 本文与 8 种当下流行的超分辨率算法进行了比较, 8 种算法分别为 Bicubic、SRCNN^[10]、FSRCNN^[48]、VDSR^[12]、LGCNet^[20]、DCM^[49]、HSENet^[31] 和 TransENet^[44]。其中 SECANN、FSRCNN 以及 VDSR 是针对自然图像所设计的超分辨率网络模型, LGCNN、DCM、HSENet 和 TransENet 是针对遥感图像所设计的更加新颖的超分辨率网络模型。除此之外, 本文在 UCMerced LandUse 数据集上进行了定量以及视觉质量比较。为了不失一般性, 所有对比算法都采用了公开的源代码, 并且在相同的环境进行了训练以及评估。

实验在两个不同的数据集上对三种不同的放大因子 ($\times 2$, $\times 3$, $\times 4$) 下的图像进行了超分辨率重建。黑色

粗体数字和下划线数字分别表示最优和次优结果。[表 1](#) 则展示了本文所提出的 ADAN 算法与 8 种对比算法在 UCMerced LandUse 测试数据集上不同缩放尺度下的评估指标的平均值。从[表 1](#) 不难看出 ADAN 的性能优于其他方法, 并且不论是 $\times 2$ 、 $\times 3$ 还是 $\times 4$, ADAN 的结果都表现为最优, 其中 PSNR 单位为 dB。相较于次优方法 TransENet, 在三种尺度上分别进步了 0.19 dB、0.07 dB 和 0.10 dB。此外, 在 SSIM 度量上, 本文所提出的 ADAN 比 TransENet 分别高出了 0.0362、0.0285 和 0.0309。与 UCMerced Landuse 数据集相比, AID 数据集包含更多的类别和更高的差异, 因此本文使用该数据集来进一步评估通用性和泛化性能。

如[表 2](#) 所示, ADAN 在 AID 数据集上同样也表现出了优越性, 具体来讲, 相较于次优算法 TransENet, 随着放大因子的增大, ADAN 的性能逐渐与 TransENet 拉开了差距, 其中 PSNR 单位为 dB。例如, 放大因子为 $\times 2$ 时, ADAN 的 PSNR (SSIM) 从 35.28 (0.9374) 提高到 37.04 (0.9617)。放大因子为 $\times 3$ 时, ADAN 的 PSNR(SSIM) 从 31.45 (0.8595) 提高到 32.96 (0.889)。而放大因子为 $\times 4$ 时, ADAN 的 PSNR (SSIM) 从 29.38 (0.7909) 提高到 29.99 (0.8177)。结果表明, 在大多数情况下, 所设计的 ADAN 都超过了对比方法, 因此, ADAN 具有更强的泛化能力。此外, 为了更好地验证 ADAN 的性能, [表 3](#) 展示了所有算法在 AID 数据集中 30 种场景类别在放大因子为 $\times 4$ 时的 PSNR 结果。从[表 3](#) 不难看出, 在大部分类别中

本文所提出的 ADAN 都获得了最高的 PSNR 值, 与其相比最有竞争力的是 TransENet, 然而 ADAN 的结果也是紧随其后, 表现也良好。值得一提的是, 对于总体的 PSNR 值, ADAN 比 TransENet 高 0.61 dB。为了进一步证明所提出的方法的优越性, 实验还采用了 LPIPS 度量。图像质量越低, LPIPS 值就越高。如[表 4](#) 所示, ADAN 的性能显著优于其他方法。具体来说, 当放大因子为 $\times 2$ 时, ADAN 比 HSENet 算法低 0.001; 而当放大因子为 $\times 3$ 和 $\times 4$ 时, ADAN 则比具有强竞争力的 TransENet 低 0.0008。最后, 如[图 5](#) 所示, 不难看出 ADAN 不仅能够取得更快的收敛速度, 而且能够收敛到更优的局部解, 进一步突出了该模型的有效性与优越性。此外, 如[图 6](#) 所示, ADAN 能够取得更好的重建精度, 而且当训练次数大于 400 时, 模型性能增益逐渐趋于饱和。

4.2.2 视觉对比

为了进一步评估 ADAN 相较于目前领先的方法的视觉质量, 以进一步验证其有效性。从[图 7](#) ~ [图 9](#) 的视觉结果可知, ADAN 取得了更好的重建效果。从[图 7](#) 中不难看出, HSENet 和 TransENet 产生了显著的棋盘效应和伪影, 而 ADAN 更加接近真值。这可能是由于这两种方法受到噪声和退化的影响, 聚集了不正确的信息。从[图 8](#) 中可以看出, ADAN 在 $\times 3$ 放大因子下产生最清晰的田野图像, 而其他方法产生不同程度的模糊和扭曲, 这进一步证明了 ADAN 的优越性。如[图 9](#) 所示, 其他方法在最具挑战性的放大比例 $\times 4$ 时, 其他方法几乎无法重建路面的信息, 而

表 1 在 UCMerced LandUse 数据集 ($\times 2$ 、 $\times 3$ 和 $\times 4$) 上的 PSNR/SSIM 结果

Table 1 PSNR/SSIM results on the UCMerced LandUse dataset ($\times 2$, $\times 3$, and $\times 4$)

Scale	PSNR/SSIM								
	Bicubic	SRCNN	FSRCNN	VDSR	LGCNet	DCM	HSENet	TransENet	Ours
2	30.76/0.8789	32.84/0.9152	33.18/0.9196	33.38/0.9220	33.48/0.9235	33.65/0.9274	34.22/0.9327	35.43/0.9355	35.62/0.9717
3	27.46/0.7631	28.66/0.8038	29.09/0.8167	29.28/0.8232	29.28/0.8238	29.52/0.8349	30.00/0.8420	31.03/0.8526	31.10/0.8811
4	25.65/0.6725	26.78/0.7219	26.93/0.7267	26.85/0.7317	27.02/0.7333	27.22/0.7528	27.73/0.7623	28.74/0.7694	28.84/0.8003

表 2 在 AID 数据集 ($\times 2$ 、 $\times 3$ 和 $\times 4$) 上的 PSNR/SSIM 结果

Table 2 PSNR/SSIM results on the AID dataset ($\times 2$, $\times 3$, and $\times 4$)

Scale	PSNR/SSIM								
	Bicubic	SRCNN	FSRCNN	VDSR	LGCNet	DCM	HSENet	TransENet	Ours
2	32.39/0.8906	34.49/0.9286	34.73/0.9331	35.05/0.9346	34.80/0.9320	35.21/0.9366	35.24/0.9368	35.28/0.9374	36.93/0.9617
3	29.08/0.7863	30.55/0.8372	30.98/0.8401	31.15/0.8522	30.73/0.8417	31.31/0.8561	31.39/0.8572	31.45/0.8595	32.96/0.8889
4	27.30/0.7036	28.40/0.7561	28.77/0.7729	28.99/0.7753	28.61/0.7626	29.17/0.7824	29.21/0.7850	29.38/0.7909	29.99/0.8177

表 3 AID 数据集中放大因子为×4 时每个类别的平均 PSNR

Table 3 Average PSNR for each category with an upscaling factor of ×4 on the AID dataset

Class	PSNR/dB							
	Bicubic	SRCNN	LGCNet	VDSR	DCM	HSENet	TransENet	Ours
Airport	27.03	28.17	28.39	28.82	28.99	29.03	29.23	29.31
Bareland	34.88	35.63	35.78	35.98	36.17	36.21	36.20	36.42
Baseball field	29.06	30.51	30.75	31.18	31.36	31.23	31.59	31.28
Beach	31.07	31.92	32.08	32.29	32.45	32.76	32.55	33.51
Bridge	28.98	30.41	30.67	31.19	31.39	31.30	31.63	30.83
Center	25.26	26.59	26.92	27.48	27.72	27.84	28.03	27.44
Church	22.15	23.41	23.68	24.12	24.29	24.39	24.51	24.62
Commercial	25.83	27.05	27.24	27.62	27.78	27.99	27.97	28.39
Dense residential	23.05	24.13	24.33	24.70	24.87	24.44	25.13	24.62
Desert	38.49	38.84	39.06	39.13	39.27	39.37	39.31	38.99
Farmland	32.30	33.48	33.77	34.20	34.42	33.90	34.58	34.19
Forest	27.39	28.15	28.20	28.36	28.47	38.31	28.56	28.37
Industrial	24.75	26.00	26.24	26.72	26.92	26.99	27.21	27.30
Meadow	32.06	32.57	32.65	32.77	32.88	32.74	32.94	33.30
Medium residential	26.09	27.37	27.63	28.06	28.25	28.11	28.45	26.94
Mountain	28.04	28.90	28.97	29.11	29.18	29.26	29.28	28.89
Park	26.23	27.25	27.37	27.69	27.82	28.23	28.01	28.11
Parking	22.33	24.01	24.40	25.21	25.74	26.17	26.40	26.01
Playground	27.27	28.72	29.04	29.62	29.92	31.18	30.30	32.00
Pond	28.94	29.85	30.00	30.26	30.39	30.40	30.53	30.33
Port	24.69	25.82	26.02	26.43	26.62	26.92	26.91	27.47
Railway station	26.31	27.55	27.76	28.19	28.38	28.47	28.61	28.42
Resort	25.98	27.12	27.32	27.71	27.88	27.99	28.08	27.66
River	29.61	30.48	30.60	30.82	30.91	30.88	31.00	30.28
School	24.91	26.13	26.34	26.78	26.94	27.51	27.22	27.52
Sparse residential	25.41	26.16	26.27	26.46	26.53	26.44	26.43	26.58
Square	26.75	28.13	28.39	28.91	29.13	29.05	29.39	28.79
Stadium	24.81	26.10	26.37	26.88	27.10	27.28	27.41	28.01
Storage tanks	24.18	25.27	25.48	25.86	26.00	26.07	26.20	26.80
Viaduct	25.86	27.03	27.26	27.74	27.93	28.12	28.21	28.01
AVG	27.30	28.40	28.61	28.99	29.17	29.21	29.38	29.99

表 4 LPIPS 在尺度为×2、×3 和×4 时的 UCMerced LandUse 数据集上的结果

Table 4 LPIPS results on the UCMerced LandUse dataset with scaling factors of ×2, ×3, and ×4

Scale	LPIPS								
	Bicubic	SRCNN	FSRCNN	VDSR	LGCNet	DCM	HSENet	TransENet	Ours
2	0.0721	0.0444	0.0471	0.0287	0.0293	0.0284	0.0266	0.0279	0.0256
3	0.1281	0.0945	0.1062	0.0801	0.0752	0.0698	0.0654	0.0649	0.0641
4	0.1650	0.1260	0.1395	0.1102	0.1093	0.1046	0.1081	0.1030	0.1022

ADAN 却产生了良好的视觉效果。具体来讲, ADAN 能够更好地修复线等细节。通过以上分析, 可以得出

结论, ADAN 可以产生视觉上令人满意的 HR 图像, 它具有丰富和真实的纹理, 清晰的边缘和界线。如

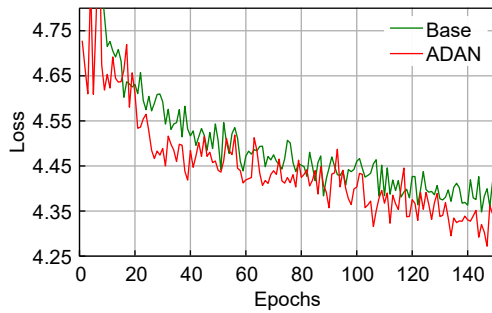


图 5 放大因子为 $\times 2$ 时 UCMerced Landuse 的损失函数分析结果

Fig. 5 Loss function analysis results on UCMerced LandUse with upscale factor of $\times 2$

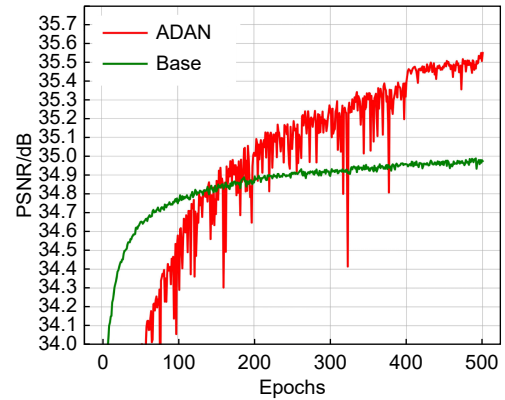


图 6 放大因子为 $\times 2$ 时 UCMerced Landuse 的 PSNR 分析结果

Fig. 6 PSNR analysis results on UCMerced LandUse with upscale factor of $\times 2$

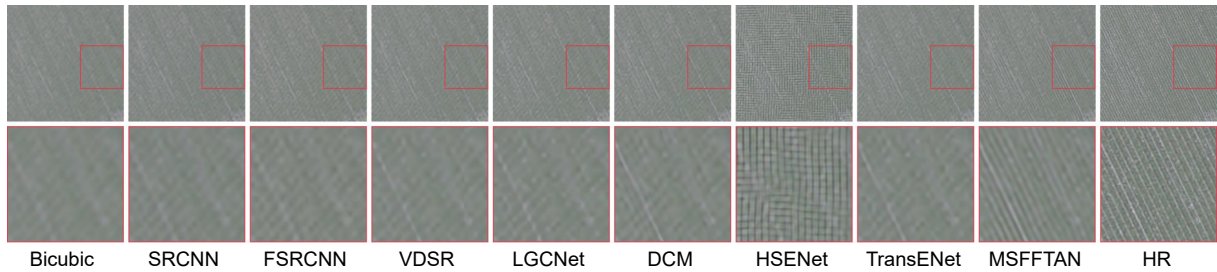


图 7 放大因子为 $\times 2$ 时 UCMerced Landuse 的视觉比较结果

Fig. 7 Visual comparison on UCMerced LandUse with an upscaling factor of $\times 2$

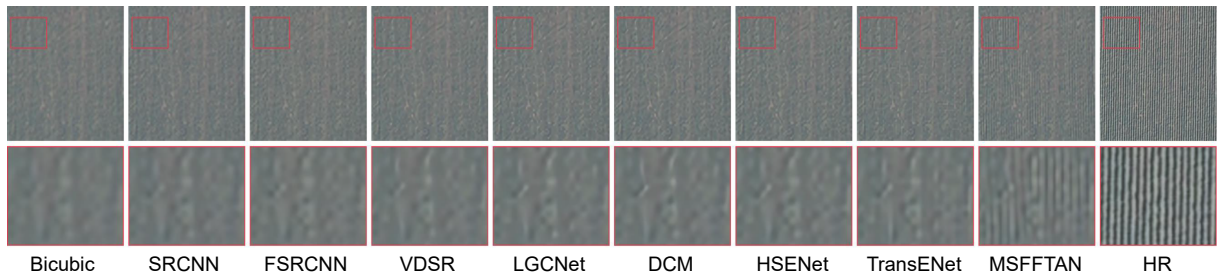


图 8 放大因子为 $\times 3$ 时 UCMerced Landuse 的视觉比较结果

Fig. 8 Visual comparison on UCMerced LandUse with an upscaling factor of $\times 3$

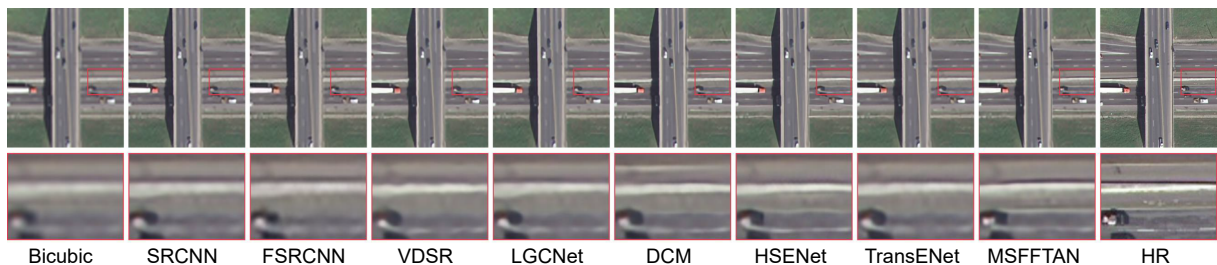


图 9 放大因子为 $\times 4$ 时 UCMerced Landuse 的视觉比较结果

Fig. 9 Visual comparison on UCMerced LandUse with an upscaling factor of $\times 4$

图 10 所示, 根据残差图可以看出 ADAN 显示出了最小的误差。

4.3 消融实验

为了进一步验证 ADAN 性能, 本文进行了广泛

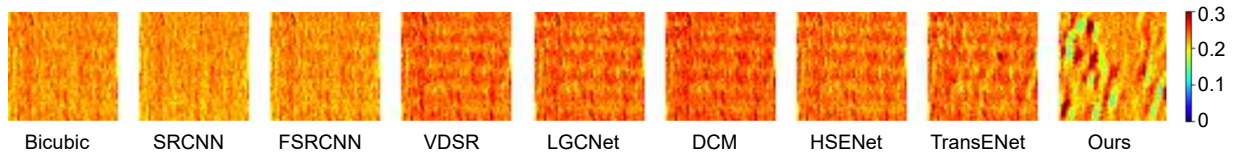


图 10 放大因子为 $\times 2$ 时对 UCMerced Landuse 进行残差对比
Fig. 10 Residual comparison on UCMerced Landuse with an upscaling factor of $\times 2$

的消融实验, 深入评估每个单独模块。本章节所有消融实验都在 UCMerced LandUse 数据集上进行了 500 代的放大因子为 $\times 2$ 的训练。

4.3.1 各模块的有效性消融研究

ADAN 中各模块的有效性验证则通过逐步增加本文精选设计的模块检验。从表 5 可以看出, 与基线模型 (Model 0) 相比, 使用 GISM (Model 1) 可以提高 0.05 dB。GISM 模块对高维特征映射, 相较于基线模型 GISM 挖掘了更精确的特征信息。而在此基础上增加了 CESSM 后, PSNR 以及 SSIM 则分别上涨了 0.04 dB 和 0.0002。CESSM 则提取了空间维度的更深层次特征, 这对超分辨率重建任务至关重要。此外通道维度的信息也是在空间维度特征信息提取的基础之上提取了全局信息, 因此, 最终 ADAN 相较于基线模型 PSNR 提升了 0.12 dB, SSIM 则提升了 0.0008。综上所述, 当合并所提出的所有模块后, 网络的整体性能提升显著, 从而证明了本文所提出的模块的有效性。

4.3.2 多尺度前馈神经网络模块的消融研究

为了探讨多尺度的有效性, 实验中令 MSFFN 单独采用不同卷积核大小的深度可分离卷积 (3×3 、 5×5 、 7×7)。本文发现, 虽然在前馈过程中使用了各种单尺度的深度可分离卷积, 但是这些都不能同时

获取多尺度的信息。相比之下, MSFFN 在前馈过程中探索并整合了多尺度特征, 从而使得网络得到了显著的性能改进。具体来说, 从表 6 可以看出, 相较于单独使用 3×3 、 5×5 、 7×7 深度可分离卷积核, MSFFN 分别取得了 0.02 dB、0.05 dB、0.03 dB 的 PSNR 性能提升, 这可以充分证明本文所提多尺度策略的有效性。

此外, 为了进一步评估 MSFFN 的有效性, 本文将所提出的 MSFFN 与常见的 Transformer 中采用的前馈神经网络层进行比较。表 7 列出了在 MSFFN 与 MLP、Conv-FFN 和 GDFN 之间 PSNR 性能和模型效率分析比较, 结果显示 MSFFN 在几乎不牺牲计算高效性的情况下, 取得了最优异的重建结果, 进一步证明了本文所提 MSFFN 的高效性和优越性。

4.3.3 基于 CNN-Transformer 模型的消融研究

当前遥感图像领域的 CNN-Transformer 混合架构 (如 TransENet 模型) 通常将 CNN 与 Transformer 割裂开来, 先通过 CNN 模块提取局部和细节特征, 再通过 Transformer 模块聚合全局和非局部依赖。然而, 这种串行结构难以充分利用 CNN 的归纳偏置能力和 Transformer 的全局理解能力。此外, 卷积特征与 Transformer 特征的异质性可能导致特征冲突, 降低

表 5 模块结构的消融结果
Table 5 Ablation results of module structures

Model	GISM	CESSM	CECSM	PSNR/dB	SSIM
Model 0	×	×	×	36.81	0.9609
Model 1	√	×	×	36.86	0.9613
Model 2	√	√	×	36.90	0.9615
Model 3 (ours)	√	√	√	36.93	0.9617

表 6 多尺度前馈神经网络 (MSFFN) 消融结果
Table 6 Ablation results of the multi-scale feedforward neural network (MSFFN)

Method	Params/M	FLOPs/G	PSNR/dB
3×3	1.98	126	36.91
5×5	2.08	140	36.88
7×7	2.35	157	36.90
MSFFN (ours)	2.13	147	36.93

表 7 多尺度前馈神经网络 (MSFFN) 与其他代表性前馈神经网络的对比分析

Table 7 Comparison analysis of the multi-scale feedforward neural network (MSFFN) with other representative feedforward neural networks

Method	Param/M	FLOPs/G	PSNR/dB
MLP	1.96	120	36.85
Conv-FFN	2.02	131	36.88
GDFN	2.09	142	36.89
MSFFN (ours)	2.13	147	36.93

模型性能。为此, 本文提出了一种新颖的 CNN-Transformer 并行架构, 能够充分融合局部特征和非局部关系, 同时自适应解决特征混乱问题。另一方面, 现有混合架构多基于空间维度的自注意力机制, 忽视了遥感图像中频谱相似性和通道全局依赖。为此, 本文联合建模空间和频谱维度的全局关系, 显著提升了重建性能。如表 8 所示, 所提算法在性能上优于 TransENet 及单纯使用空间或频谱自注意力的模型。

表 8 ADAN 与其他代表性 CNN-Transformer 架构的对比分析

Table 8 Comparison analysis of ADAN with other representative CNN-Transformer architectures

Method	PSNR/dB	SSIM
TransENet	35.43	0.9355
Spatial dimension Transformer	35.52	0.9521
Frequency dimension Transformer	35.56	0.9602
ADAN(ours)	35.62	0.9719

4.4 模型复杂性分析

表 9 展示了 UCMerced LandUse 数据集中放大因子为 $\times 2$ 的性能和复杂性分析, 其中 FLOPs 代表浮点运算, 它被定义为计算的数量, 可以用来度量模型的复杂性。从表格中不难看出 ADAN 的 FLOPs 的结果优于多数模型, 仅仅比 LGCNet 高了 0.05 G。而在 PSNR 的结果上 ADAN 比 LGCNet 高出 2.14 dB, 这表明 ADAN 可以在模型的复杂性和性能之间达到合

表 9 模型复杂性分析

Table 9 Model complexity analysis

Method	Param/M	Flops/G	PSNR/dB
LGCNet	0.193	7.11	33.48
DCM	2.180	7.32	33.65
HSENet	5.400	10.80	34.22
TransENet	37.800	9.32	35.43
ADAN(ours)	4.120	7.16	35.62

理的平衡。综上所述, ADAN 比其他方法参数量更少, 但结果更佳, 更好地平衡了网络复杂度和图像超分辨率质量之间关系。

5 结 论

本文提出了一种并行的通道-空间 Transformer 方法, 用于高效的图像超分辨率。通过通道注意力和空间注意力信息的交互作用, 从而增强模型的综合特征表示能力。同时, 为了考虑遥感图像中相似地面物体的尺度多样性, 本文设计了一个多尺度前馈层神经网络来生成一组丰富的多尺度特征。此外, 引入了一种新颖的门控信息选择模块, 增强了模型挖掘局部高频特征的能力, 以弥补 Transformer 模型倾向于学习低频全局特征的缺陷。在多个标准测试集上, ADAN 的性能都优于先进的 SR 模型。尽管取得了不错的性能, 遥感图像总是遭受各种错综复杂的退化, 使得 ADAN 在现实场景中性能急剧下降, 限制了 RSSR 模型的实际价值。因此, 下一步的工作将聚焦于真实场景下的遥感图像超分辨率重建算法的构建, 从而增强模型应对真实复杂降质的能力。

利益冲突: 所有作者声明无利益冲突

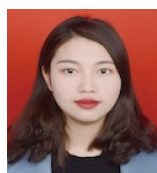
参考文献

- [1] Chen M H, Lu Y Q, Yang W Y, et al. Super-resolution reconstruction of retinal OCT image using multi-teacher knowledge distillation network[J]. *Opto-Electron Eng*, 2024, **51**(7): 240114.
陈明惠, 芦焱琦, 杨文逸, 等. OCT 图像多教师知识蒸馏超分辨率重建[J]. *光电工程*, 2024, **51**(7): 240114.
- [2] Xiao Z J, Zhang J H, Lin B H. Feature coordination and fine-grained perception of small targets in remote sensing images[J]. *Opto-Electron Eng*, 2024, **51**(6): 240066.
肖振久, 张杰浩, 林瀚翰. 特征协同与细粒度感知的遥感图像小目标检测[J]. *光电工程*, 2024, **51**(6): 240066.
- [3] Wu L S, Fang L Y, Yue J, et al. Deep bilateral filtering network for point-supervised semantic segmentation in remote sensing images[J]. *IEEE Trans Image Process*, 2022, **31**: 7419–7434.
- [4] Yang J Q, Du B, Xu Y H, et al. Can spectral information work while extracting spatial distribution?—An online spectral

- information compensation network for HSI classification[J]. *IEEE Trans Image Process*, 2023, **32**: 2360–2373.
- [5] Sitaula C, Kc S, Aryal J. Enhanced multi-level features for very high resolution remote sensing scene classification[J]. *Neural Comput Appl*, 2024, **36**(13): 7071–7083.
- [6] Zhang C, Lam K M, Liu T S, et al. Structured adversarial self-supervised learning for robust object detection in remote sensing images[J]. *IEEE Trans Geosci Remote Sens*, 2024, **62**: 5613720.
- [7] Yu W Q, Cheng G, Wang M J, et al. MAR20: a benchmark for military aircraft recognition in remote sensing images[J]. *Natl Remote Sens Bull*, 2023, **27**(12): 2688–2696.
禹文奇, 程攀, 王美君, 等. MAR20: 遥感图像军用飞机目标识别数据集[J]. *遥感学报*, 2023, **27**(12): 2688–2696.
- [8] Zhang W X, Luo Y H, Liu Y Q, et al. Image super-resolution reconstruction based on active displacement imaging[J]. *Opto-Electron Eng*, 2024, **51**(1): 230290.
张文雪, 罗一涵, 刘雅卿, 等. 基于主动位移成像的图像超分辨率重建[J]. *光电工程*, 2024, **51**(1): 230290.
- [9] Cheng D Q, Ma X, Kou Q Q, et al. Lightweight image super-resolution reconstruction algorithm based on multi-path feature calibration[J]. *J Comput-Aided Des Comput Graph*, 2025, **36**(12): 241211.
程德强, 马祥, 寇旗旗, 等. 基于多路特征校准的轻量级图像超分辨率重建算法[J]. *计算机辅助设计与图形学学报*, 2025, **36**(12): 241211.
- [10] Dong C, Loy C C, He K M, et al. Learning a deep convolutional network for image super-resolution[C]//*Proceedings of the 13th European Conference on Computer Vision*, 2014: 184–199. https://doi.org/10.1007/978-3-319-10593-2_13.
- [11] Mei Y Q, Fan Y C, Zhou Y Q. Image super-resolution with non-local sparse attention[C]//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 3517–3526. <https://doi.org/10.1109/CVPR46437.2021.00352>.
- [12] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 1646–1654. <https://doi.org/10.1109/CVPR.2016.182>.
- [13] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017: 6000–6010.
- [14] Chen X Y, Wang X T, Zhou J T, et al. Activating more pixels in image super-resolution Transformer[C]//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 22367–22377. <https://doi.org/10.1109/CVPR52729.2023.02142>.
- [15] Yang F Z, Yang H, Fu J L, et al. Learning texture Transformer network for image super-resolution[C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 5791–5800. <https://doi.org/10.1109/CVPR42600.2020.00583>.
- [16] Lu Z S, Li J C, Liu H, et al. Transformer for single image super-resolution[C]//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2022: 457–466. <https://doi.org/10.1109/CVPRW56347.2022.00061>.
- [17] Liang J Y, Cao J Z, Sun G L, et al. SwinIR: image restoration using Swin Transformer[C]//*Proceedings of 2021 IEEE/CVF International Conference on Computer Vision Workshops*, 2021: 1833–1844. <https://doi.org/10.1109/ICCVW54120.2021.00210>.
- [18] Zhang X D, Zeng H, Guo S, et al. Efficient long-range attention network for image super-resolution[C]//*Proceedings of the 17th European Conference on Computer Vision*, 2022: 649–667. https://doi.org/10.1007/978-3-031-19790-1_39.
- [19] Wang Z H, Qian Y T. Super-resolution reconstruction of spatiotemporal fusion for dual-stream remote sensing images based on Swin Transformer[J]. *Comput Eng*, 2024, **50**(9): 33–45.
王志浩, 钱云海. 基于 Swin Transformer 的双流遥感图像时空融合超分辨率重建[J]. *计算机工程*, 2024, **50**(9): 33–45.
- [20] Lei S, Shi Z W, Zou Z X. Super-resolution for remote sensing images via local-global combined network[J]. *IEEE Geosci Remote Sens Lett*, 2017, **14**(8): 1243–1247.
- [21] Jiang K, Wang Z Y, Yi P, et al. Deep distillation recursive network for remote sensing imagery super-resolution[J]. *Remote Sens*, 2018, **10**(11): 1700.
- [22] Jiang K, Wang Z Y, Yi P, et al. Edge-enhanced GAN for remote sensing image superresolution[J]. *IEEE Trans Geosci Remote Sens*, 2019, **57**(8): 5799–5812.
- [23] Jiang K, Wang Z Y, Yi P, et al. Hierarchical dense recursive network for image super-resolution[J]. *Pattern Recognit*, 2020, **107**: 107475.
- [24] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 770–778. <https://doi.org/10.1109/CVPR.2016.90>.
- [25] Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017: 136–144. <https://doi.org/10.1109/CVPRW.2017.151>.
- [26] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]//*Proceedings of the 32nd International Conference on Machine Learning*, 2015: 448–456.
- [27] Dong C, Loy C C, He K M, et al. Image super-resolution using deep convolutional networks[J]. *IEEE Trans Pattern Anal Mach Intell*, 2016, **38**(2): 295–307.
- [28] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 7132–7141. <https://doi.org/10.1109/CVPR.2018.00745>.
- [29] Niu B, Wen W L, Ren W Q, et al. Single image super-resolution via a holistic attention network[C]//*Proceedings of the 16th European Conference on Computer Vision*, 2020: 191–207. https://doi.org/10.1007/978-3-030-58610-2_12.
- [30] Zhang Y L, Li K P, Li K, et al. Image super-resolution using very deep residual channel attention networks[C]//*Proceedings of the 15th European Conference on Computer Vision*, 2018: 286–301. https://doi.org/10.1007/978-3-030-01234-2_18.
- [31] Lei S, Shi Z W. Hybrid-scale self-similarity exploitation for remote sensing image super-resolution[J]. *IEEE Trans Geosci Remote Sens*, 2022, **60**: 5401410.
- [32] Zhang W D, Zhao W Y, Li J, et al. CVANet: cascaded visual attention network for single image super-resolution[J]. *Neural Netw*, 2024, **170**: 622–634.
- [33] Zhu B Y, Chen Z H, Sheng B. Remote sensing image detection based on perceptually enhanced Swin Transformer[J]. *Comput Eng*, 2024, **50**(1): 216–223.
祝冰艳, 陈志华, 盛斌. 基于感知增强 Swin Transformer 的遥感图像检测[J]. *计算机工程*, 2024, **50**(1): 216–223.

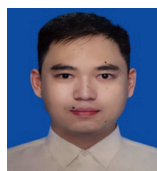
- [34] Zhang Y Y, Zhang B H, Zhao Y F, et al. Remote sensing image classification based on dual-channel deep dense feature fusion[J]. *Laser Technol*, 2021, **45**(1): 73–79.
张艳月, 张宝华, 赵云飞, 等. 基于双通道深度密集特征融合的遥感影像分类[J]. *激光技术*, 2021, **45**(1): 73–79.
- [35] Salvetti F, Mazzia V, Khaliq A, et al. Multi-image super resolution of remotely sensed images using residual attention deep neural networks[J]. *Remote Sens*, 2020, **12**(14): 2207.
- [36] Dong X Y, Sun X, Jia X P, et al. Remote sensing image super-resolution using novel dense-sampling networks[J]. *IEEE Trans Geosci Remote Sens*, 2021, **59**(2): 1618–1633.
- [37] Xiao Y, Yuan Q Q, Jiang K, et al. TTST: a top-k token selective transformer for remote sensing image super-resolution[J]. *IEEE Trans Image Process*, 2024, **33**: 738–752.
- [38] Liu H X, Dai Z H, So D R, et al. Pay attention to MLPs[C]// *Proceedings of the 35th International Conference on Neural Information Processing Systems*, 2021: 704.
- [39] Dauphin Y N, Fan A, Auli M, et al. Language modeling with gated convolutional networks[C]// *Proceedings of the 34th International Conference on Machine Learning*, 2017: 933–941.
- [40] Song Y D, He Z Q, Qian H, et al. Vision transformers for single image dehazing[J]. *IEEE Trans Image Process*, 2023, **32**: 1927–1941.
- [41] Ba J L, Kiros J R, Hinton G E. Layer normalization[Z]. arXiv: 1607.06450, 2016. <https://doi.org/10.48550/arXiv.1607.06450>.
- [42] Yang Y, Newsam S. Bag-of-visual-words and spatial extensions for land-use classification[C]// *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2010: 270–279.
<https://doi.org/10.1145/1869790.1869829>.
- [43] Xia G S, Hu J W, Hu F, et al. AID: a benchmark data set for performance evaluation of aerial scene classification[J]. *IEEE Trans Geosci Remote Sens*, 2017, **55**(7): 3965–3981.
- [44] Lei S, Shi Z W, Mo W J. Transformer-based multistage enhancement for remote sensing image super-resolution[J]. *IEEE Trans Geosci Remote Sens*, 2022, **60**: 5615611.
- [45] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[J]. *IEEE Trans Image Process*, 2004, **13**(4): 600–612.
- [46] Zhang R, Isola P, Efros A A, et al. The unreasonable effectiveness of deep features as a perceptual metric[C]// *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 586–595.
<https://doi.org/10.1109/CVPR.2018.00068>.
- [47] Kingma D P, Ba J. Adam: a method for stochastic optimization[Z]. arXiv: 1412.6980, 2014.
<https://doi.org/10.48550/arXiv.1412.6980>.
- [48] Dong C, Loy C C, Tang X O. Accelerating the super-resolution convolutional neural network[C]// *Proceedings of the 14th European Conference on Computer Vision*, 2016: 391–407.
https://doi.org/10.1007/978-3-319-46475-6_25.
- [49] Haut J M, Paoletti M E, Fernández-Beltrán R, et al. Remote sensing single-image superresolution based on a deep compendium model[J]. *IEEE Geosci Remote Sens Lett*, 2019, **16**(9): 1432–1436.

作者简介



吴菲 (1995-), 女, 博士, 嘉兴大学讲师, 主要从事多目标优化、图像超分辨率重建等方面研究。

E-mail: WFMOOK@163.com



陈嘉诚 (1997-), 男, 博士, 嘉兴大学讲师, 主要从事人工智能、图像处理等方面研究。

E-mail: cjcjdx@zjxu.edu.cn



【通信作者】杨俊 (1978-), 男, 博士, 嘉兴大学信息科学与工程学院副教授, 主要从事智能多媒体信息分析与处理、机器学习等方面的研究。

E-mail: juneryoung@zjxu.edu.cn



王万良 (1957-), 男, 博士, 浙江工业大学计算机科学与技术学院教授, 博士生导师, 主要从事控制理论、大数据、人工智能等方面研究。

E-mail: zjutwll@zjut.edu.cn



李国庆 (1994-), 男, 博士, 宁波大学信息科学与工程学院讲师, 主要研究方向为人工智能、智能计算等方面研究。

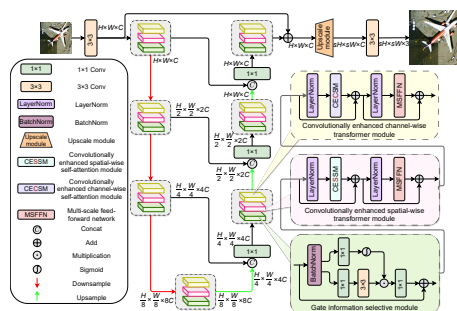
E-mail: li241700@126.com



扫描二维码, 获取PDF全文

Remote-sensing images reconstruction based on adaptive dual-domain attention network

Wu Fei¹, Chen Jiacheng¹, Yang Jun^{1*}, Wang Wanliang², Li Guoqing³



Overall architecture and module structures of ADAN

Overview: With the rapid development of convolutional neural networks (CNNs) and Transformer models, significant progress has been made in the task of remote sensing image super-resolution reconstruction (RSISR). However, existing methods have limitations in handling features of objects at different scales and fail to fully exploit the implicit relationships between channel and spatial dimensions, which restricts further improvement in reconstruction performance. To address these issues, an adaptive dual-domain attention network (ADAN) is proposed, aiming to enhance feature extraction capabilities by integrating self-attention information from both channel and spatial domains. Additionally, it combines multi-scale feature mining and local feature representation to improve the performance of remote sensing image super-resolution reconstruction.

The research aims to address the shortcomings of existing methods in multi-scale feature extraction and insufficient exploration of channel-spatial relationships in remote sensing image super-resolution tasks. To this end, the ADAN network designs a multi-scale feed-forward network (MSFFN) to capture rich multi-scale features and incorporates a novel gate information selective module (GISM) to enhance local feature representation. Furthermore, the network adopts a U-shaped architecture to achieve efficient multi-level feature fusion. Specifically, ADAN introduces a convolutionally enhanced spatial-wise transformer module (CESTM) and a convolutionally enhanced channel-wise transformer module (CECTM) to extract channel and spatial features in parallel, comprehensively exploring the interactions and dependencies between features.

Experimental results demonstrate that ADAN significantly outperforms state-of-the-art algorithms on multiple public remote sensing datasets in terms of quantitative metrics (e.g., PSNR and SSIM) and visual quality, validating its effectiveness and superiority. The main contributions are as follows: 1) Proposing a novel method, ADAN, tailored for remote sensing image super-resolution tasks; 2) Designing parallel channel and spatial feature extraction modules along with a gated convolution module to comprehensively explore features across channel, spatial, and convolutional dimensions; 3) Introducing a multi-scale feed-forward network (MSFFN) to effectively explore potential scale relationships and enhance global representation capabilities; 4) Experimentally validating the superior performance of ADAN in remote sensing image super-resolution reconstruction. This research provides new insights and technical pathways for remote sensing image super-resolution reconstruction.

Wu F, Chen J C, Yang J, et al. Remote-sensing images reconstruction based on adaptive dual-domain attention network[J]. *Opto-Electron Eng*, 2025, 52(4): 240297; DOI: 10.12086/oee.2025.240297

Foundation item: National Natural Science Foundation of China (61873240, 62302197), the Natural Science Foundation of Zhejiang Province (LQN25F020024, LQ23F020006, LQ23F030007), and the Science and Technology Plan Project of Jiaxing City (2024AD10045)

¹College of Information Science and Engineering, Jiaxing University, Jiaxing, Zhejiang 314000, China; ²College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, Zhejiang 310000, China; ³College of Information Science and Engineering, Ningbo University, Ningbo, Zhejiang 315000, China

* E-mail: juneryoung@zjxu.edu.cn