



· 深度学习视觉处理与分析 ·

基于双域聚合 Transformer 的高光谱图像 超分辨重建方法

王海军¹, 王要威¹, 杨腾飞², 霍利美¹, 李培峦¹

1. 河南科技大学 数学与统计学院, 河南 洛阳 471023; 2. 中国科学院 空天信息创新研究院, 北京 100094

摘要 高光谱图像超分辨率旨在提升低空间分辨率高光谱图像的质量, 对遥感、精准农业等领域的应用至关重要。尽管现有深度学习方法已经取得显著进展, 但在刻画高光谱数据复杂的空间-光谱长程依赖关系以及充分挖掘空间与光谱特征互补性方面仍存在不足。为此, 提出一种双域空间-光谱聚合 Transformer 网络(dual-domain spatial-spectral aggregation Transformer, DSAT)用于单高光谱图像超分辨率重建。DSAT 围绕空间域与光谱域的双域建模与交互聚合展开设计: 一方面, 通过双向门控空间-光谱聚合模块(bidirectional gated spatial-spectral aggregation, BGSA)分别构建跨域空间自注意力和跨域光谱自注意力, 其中, 前者在局部窗口自注意力中引入全局上下文信息, 以增强远距离空间位置之间的依赖建模, 后者则聚焦光谱维度的全局关系, 以精准刻画不同光谱波段之间的内在相关性; 另一方面, 在 BGSA 中嵌入自适应交互模块, 从空间交互和通道交互 2 个层面对空谱特征进行自适应加权与融合, 充分挖掘空间与光谱信息的互补性。3 个基准数据集上的实验结果表明, 该文中所提方法在主客观质量方面均优于其他主流对比方法。

关键词 高光谱图像; 图像超分辨率重建; 深度学习; Transformer

中图分类号: TP391 DOI: 10.16152/j.cnki.xdxzbzr.2026-03-008

Dual-domain aggregation Transformer for hyperspectral image super-resolution

WANG Haijun¹, WANG Yaowei¹, YANG Tengfei², HUO Limei¹, LI Peiluan¹

1. School of Mathematics and Statistics, Henan University of Science and Technology, Luoyang 471023, China;

2. Aerospace Information Research Institute, Chinese Academy of Science, Beijing 100094, China

Abstract Hyperspectral image super-resolution aims to enhance the spatial resolution of low-resolution hyperspectral images and is critical for applications such as remote sensing and precision agriculture. Existing deep learning methods have made significant progress, but they still struggle to capture complex long-range spatial-spectral dependencies and to fully exploit complementary information between spatial and spectral features. To address these challenges, we propose a dual-domain spatial-spectral aggregation Transformer (DSAT) for single-image hyperspectral super-resolution. Central to DSAT is dual-domain modeling and interactive aggregation in the spatial and spectral domains. Concretely, a bidirectional gated spatial-spectral aggregation (BGSA)

收稿日期: 2026-01-20 修回日期: 2026-01-30

基金项目: 国家自然科学基金(T2541012, 12401116); 河南省科技攻关计划项目(252102210020, 252102240125); 河南省高等学校重点科研项目(25A520045)。

第一作者: 王海军, 男, 博士, 副教授, 硕士生导师, 从事机器学习、计算机视觉研究, hjwang@haust.edu.cn。

module constructs cross-domain spatial self-attention and cross-domain spectral self-attention. The spatial branch injects global context into local-window self-attention to strengthen long-range spatial dependency modeling, while the spectral branch focuses on global relationships along the spectral dimension to accurately capture intrinsic correlations among bands. Moreover, an adaptive interaction module is embedded within BGSA to perform adaptive weighting and fusion of spatial-spectral features via spatial interaction and channel interaction, thereby exploiting complementarity between spatial and spectral information. Experiments on three benchmark datasets show that the proposed method outperforms other competitors in both objective metrics and visual quality.

Keywords hyperspectral image; image super-resolution; deep learning; Transformer

高光谱图像(hyperspectral images, HSI)记录物体在多个连续窄波段的反射信息,提供识别物体属性所需的丰富光谱细节,使其在目标检测^[1]、环境监测^[2]、农业^[3]和医学^[4]等领域展现出巨大应用潜力。然而,受硬件或成本制约,大量高光谱图像分辨率不高,限制了其在下游任务的应用。

高光谱图像超分辨率重建主要分为 2 类:基于融合的方法和单高光谱图像超分辨率方法。前者融合空间分辨率低但光谱分辨率高的 HSI 与空间分辨率高但光谱分辨率低的辅助图像(如 RGB)实现高分辨率重建,但其面临辅助数据构建与精确配准的挑战。后者则能够直接从单幅低分辨率高光谱图像中重建高分辨率图像,因其限制较少而受到广泛关注。

单高光谱图像超分辨率重建方法主要分为传统方法和基于深度学习的方法。传统方法如基于低秩约束^[5]、稀疏表示^[6]和字典学习^[7]等,因其对先验知识和参数敏感,且难以捕捉复杂的非线性关系,重建性能并不理想。随着深度学习的快速发展,卷积神经网络(convolutional neural network, CNN)在图像超分辨率领域表现出色。早期研究中,Mei 等利用 3D 卷积提取空间光谱信息,但参数和计算量大^[8];为解决此问题, Li 等提出结合 2D 和可分离 3D 卷积的混合卷积神经网络^[9];而 Wang 等则设计并行的 2D 与可分离 3D 卷积结构,并引入频域损失函数^[10]。此外,基于分组策略的方法也受到广泛关注。Li 等人将分组递归模块嵌入全局残差结构,提出分组深度递归残差网络^[11]; Jiang 等则将渐进式上采样策略引入分组卷积以学习空间光谱先验^[12]; Wang 等基于序列建模和循环反馈有效挖掘高光谱的光谱序列相关性^[13];针对邻近波段空间信息提取易破坏光谱结构以及上下文利用不足的问题, Wang 等

提出兼顾相邻光谱保持与上下文关联增强的双阶段网络^[14]。然而,这些基于 CNN 的方法主要依赖局部特征提取,在建模复杂空间-光谱长程依赖方面仍显不足。

近年来,Transformer 模型凭借自注意力机制,有效弥补 CNN 在捕获长距离依赖方面的不足,并在计算机视觉领域展现出卓越性能。在高光谱图像超分辨率领域, Liu 等设计并行分支网络,结合 Transformer 与 3D 卷积提取谱带相关性^[15]; Zhang 等引入高效 SCC 核自注意力机制,以较低计算成本提取多尺度空间-光谱特征^[16]; Chen 等融合多尺度光谱注意力与可变形卷积 Transformer 模块,实现局部与全局空间-光谱信息的高效建模^[17],并在此基础上 Chen 等发展跨域空间-光谱模块,将全局信息引入矩形窗口自注意力,进一步整合局部与全局信息^[18]; Hu 等则将超分辨率问题转化至丰度域,并结合空间保持分解和空间-光谱注意力机制,实现高效特征挖掘^[19]。

然而,现有 Transformer 方法多侧重光谱自注意力,往往忽略长距离空间依赖。部分工作虽然同时引入光谱与空间自注意力,但是未能充分挖掘二者的互补性和相关性。为此,本文提出一种双域空间-光谱聚合 Transformer 网络(dual-domain spatial-spectral aggregation Transformer, DSAT)。DSAT 基于 Transformer 设计双向门控空间-光谱聚合模块(bidirectional gated spatial-spectral aggregation, BGSA),实现空间和光谱维度的并行自注意力计算。此外, BGSA 内部集成自适应交互模块(adaptive interaction module, AIM),进一步优化空间和光谱特征的自适应加权融合,提升空谱特征利用效率。

1 网络模型设计

本文网络由浅层特征提取、深层特征提取和

重建 3 部分组成,具体流程如图 1(a) 所示。设输入低分辨率图像张量为 $I_{LR} \in \mathbf{R}^{H \times W \times B}$, 经过双 3 次插值得到的高分辨率图像张量为 $I_{BI} \in \mathbf{R}^{sH \times sW \times B}$, 网络输出的超分辨率图像张量为 $I_{SR} \in \mathbf{R}^{sH \times sW \times B}$ 。其中: $H \times W$ 是图像的空间分辨率; B 是光谱波段

数; s 是尺度因子。首先, I_{LR} 通过卷积层进行浅层特征提取; 随后, 将提取的浅层特征送入一系列 BGSA 残差组(每个含多个 BGSA 模块和残差连接)以提取深层特征; 最后, 将深层特征上采样并与 I_{BI} 融合, 得到超分辨率图像张量 I_{SR} 。

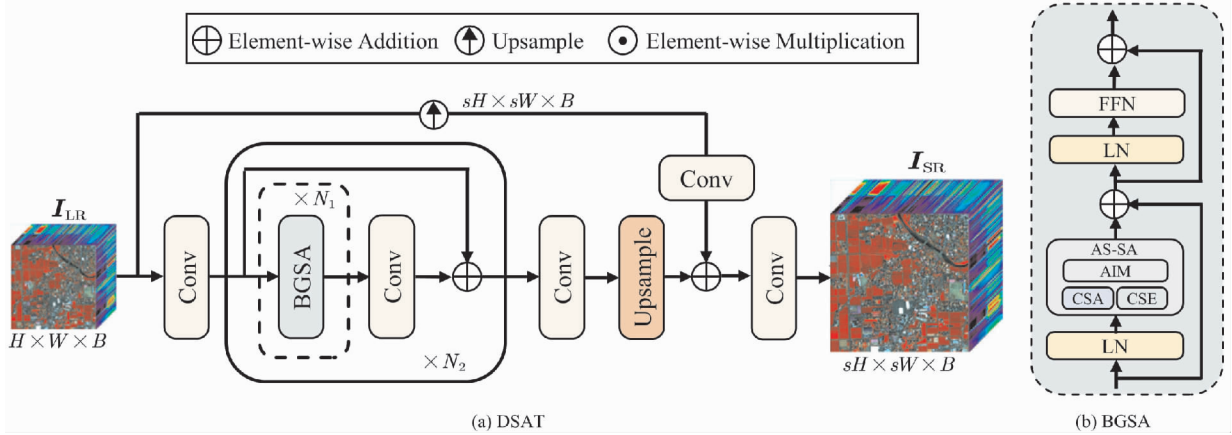
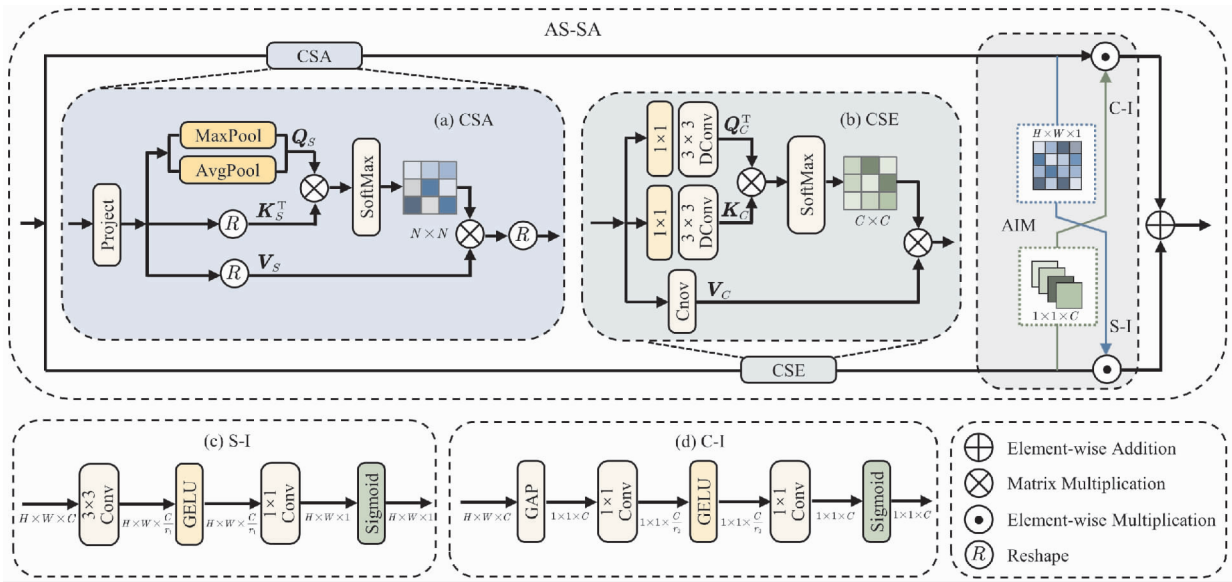


图 1 DSAT 总体结构
Fig. 1 The overall architecture of DSAT

1.1 双向门控的空间 - 光谱聚合模块

为了深度挖掘空间 - 光谱特征, 本文基于 Transformer 架构设计 BGSA。如图 1(b) 所示, BGSA 包含 2 个层归一化 (layer norm, LN) 层、1 个自适应空间 - 光谱自注意力 (adaptive spatial-spectral self-attention, AS-SA) 模块 (见图 2) 和 1 个前馈神经网络 (feed-forward neural network, FFN)。AS-SA 模块引入跨域空间自注意力

(cross-scope spatial self-attention, CSA) 和跨域光谱自注意力 (cross-scope spectral self-attention, CSE) 机制, 实现空间与光谱特征并行提取。考虑到简单特征叠加可能导致信息融合不足, 受 Chen 等^[20] 的启发, AS-SA 模块融入 AIM, 对提取的空间和光谱特征进行自适应加权处理, 实现空谱特征的有效融合。



(a) 跨域空间自注意力 CSA; (b) 跨域通道自注意力 CSE; (c) 空间交互 S-I; (d) 通道交互 C-I。

图 2 自适应空间光谱自注意力模块

Fig. 2 Adaptive spatial-spectral self-attention module

1.2 自适应空间光谱自注意力

1.2.1 跨域空间自注意力

传统的基于窗口自注意力机制虽然计算开销较低,但其局部性限制了全局空间依赖建模。本文引入 CSA,旨在将全局上下文信息融入局部窗口,在保证计算效率的同时,实现局部与全局信息有效交互。如图 2(a)和图 2(b)所示,将层归一化后的输入特征 $\mathbf{X} \in \mathbf{R}^{H \times W \times C}$ 送入 CSA 和 CSE 分别提取空间特征和光谱特征。

基于 CSA, $\mathbf{X}$ 首先通过线性映射生成查询 $\mathbf{Q}$ 、键 $\mathbf{K}$ 、值 $\mathbf{V}$ 。对 $\mathbf{Q}$ 进行自适应平均池化和最大池化得到 $\mathbf{Q}_s$,从而捕获长距离的空间依赖;同时,对 $\mathbf{K}$ 和 $\mathbf{V}$ 进行窗口划分和重塑。之后,CSA 并行计算多头注意力,并在通道维度上进行划分与拼接,最终输出 $\mathbf{Y}_s$ 。以单头为例,主要计算过程为

$$\mathbf{Q} = \mathbf{X}L_Q, \mathbf{K}_s = \mathbf{X}L_K, \mathbf{V}_s = \mathbf{X}L_V \quad (1)$$

$$\mathbf{Q}_s = \text{CAT}(H_{\text{Avgpool}}(\mathbf{Q}), H_{\text{Maxpool}}(\mathbf{Q})) \quad (2)$$

$$\mathbf{Y}_s = \text{Softmax}(\mathbf{Q}_s \mathbf{K}_s^T / \sqrt{d} + P) \mathbf{V}_s \quad (3)$$

式中: L_Q, L_K, L_V 是可学习参数; $H_{\text{Avgpool}}(\cdot)$ 和 $H_{\text{Maxpool}}(\cdot)$ 表示自适应平均池化和最大池化; $\text{CAT}(\cdot)$ 表示通道维度拼接; d 是特征维度; P 是相对位置编码。

1.2.2 跨域光谱自注意力

传统的 3D 卷积侧重局部光谱特征,难以捕获长距离依赖。部分基于 Transformer 的方法虽然结合光谱自注意力与 3D 卷积改进了上述问题,却仍存在高维特征带来的内存占用与计算开销过大问题。本文所提框架引入 CSE,旨在有效捕获特征的空间-光谱相关性和全局交互。如图 2(b)所示, CSE 通过逐点卷积和深度卷积提取输入特征 $\mathbf{X}$ 的局部空间-光谱特征,得到查询 $\mathbf{Q}$ 和键 $\mathbf{K}$,同时通过 1×1 卷积生成值 $\mathbf{V}$ 。随后,将 $\mathbf{Q}$ 、 $\mathbf{K}$ 、 $\mathbf{V}$ 重塑为多头注意力所需尺寸,进行自注意力计算。以单头为例,主要计算过程如下:

$$\mathbf{Q}_c = W_p^1(W_d^1(\mathbf{X})), \mathbf{K}_c = W_p^2(W_d^2(\mathbf{X})), \quad \mathbf{V}_c = W_v(\mathbf{X}) \quad (4)$$

$$\mathbf{Y}_c = \mathbf{V}_c \cdot \text{Softmax}(\mathbf{Q}_c \mathbf{K}_c^T / \sqrt{d}) \quad (5)$$

式中: $W_p^1(\cdot)$ 和 $W_p^2(\cdot)$ 表示逐点卷积; $W_d^1(\cdot)$ 和 $W_d^2(\cdot)$ 表示深度卷积; W_v 表示标准的 1×1 卷积; d 是特征维度。

1.2.3 自适应交互模块

CSA 与 CSE 可并行捕捉长距离的空间与光谱依赖,但简单地将两者相加往往导致跨模态信

息融合不足。为此本文引入 AIM,对来自空间和光谱分支的特征进行重新加权与交互融合。AIM 包含 2 类互补的交互机制:空间交互 (spatial-interaction, S-I) 和通道交互 (channel-interaction, C-I)。

S-I 模块侧重空间表征建模,旨在构建位置敏感的注意力权重。如图 2(c)所示, CSA 的输出特征 $\mathbf{Y}_s$ 经 3×3 卷积捕获局部空间上下文并压缩通道,再经过 GELU 激活函数与 1×1 卷积,生成像素级的空间注意力图 (spatial attention map, S-Map)。这一过程可表示为

$$\text{S-Map}(\mathbf{Y}_s) = f(F_p^1(\sigma(F(\mathbf{Y}_s)))) \quad (6)$$

式中: $F(\cdot)$ 表示通道压缩比例 r_1 的标准卷积; $F_p^1(\cdot)$ 表示通道压缩比例 $\frac{C}{r_1}$ 的点卷积; $\sigma(\cdot)$ 表示 GELU 激活函数; $f(\cdot)$ 表示 Sigmoid 函数。

C-I 模块探索跨通道相关性,旨在自适应学习通道间的权重分配。如图 2(d)所示, CSE 的输出特征 $\mathbf{Y}_c$ 首先通过自适应全局平均池化汇聚空间信息至通道级统计量,随后通过 2 层 1×1 卷积构建瓶颈式映射学习通道权重,并结合 GELU 激活函数提升非线性表达,最终生成通道注意力图 (channel attention map, C-Map)。这一过程可以表示为

$$\text{C-Map}(\mathbf{Y}_c) = f(F_p^3(\sigma(F_p^2(H_{\text{GP}}(\mathbf{Y}_c))))) \quad (7)$$

式中: $H_{\text{GP}}(\cdot)$ 表示自适应全局平均池化; $F_p^2(\cdot)$ 和 $F_p^3(\cdot)$ 分别表示通道压缩比例 r_2 和通道激励比例 r_2 的点卷积; $\sigma(\cdot)$ 表示 GELU 激活函数; $f(\cdot)$ 表示 Sigmoid 函数。

基于 AIM,进一步构建 AS-SA 模块,在 CSA 分支引入空间交互,在 CSE 分支引入通道交互,能够更好地融合空间和光谱特征,提升模型表示能力。交互公式为

$$\text{S-I}(\mathbf{Y}_c, \mathbf{Y}_s) = \mathbf{Y}_c \odot \text{S-Map}(\mathbf{Y}_s) \quad (8)$$

$$\text{C-I}(\mathbf{Y}_s, \mathbf{Y}_c) = \mathbf{Y}_s \odot \text{C-Map}(\mathbf{Y}_c) \quad (9)$$

式中: $\odot$ 表示逐元素相乘。

值得注意的是, AIM 主要从 2 个方面提升特征利用效率:一方面, S-I 生成空间注意力图,在空间维度上自适应增强边缘与纹理等高频区域的响应,有效突出对结构恢复更为关键的空间细节特征;另一方面, C-I 生成通道注意力图,在光谱维度上自适应调整各特征通道的重要性,抑制冗余通道,强化与重建任务高度相关的光谱响应。

1.3 损失函数

本文基于 L1 损失、光谱角映射(spectral angle mapper, SAM)损失以及空间域和光谱域的梯度损失复合而成损失函数。L1 损失用于衡量重建图像和原始图像之间的像素级误差,具有良好收敛性。

$$L_1 = \frac{1}{N} \sum_{n=1}^N \| \mathbf{I}_{\text{HR}}^n - \mathbf{I}_{\text{SR}}^n \|_1 \quad (10)$$

式中: N 表示批次数量; $\mathbf{I}_{\text{HR}}^n$ 和 $\mathbf{I}_{\text{SR}}^n$ 分别表示第 n 个原始图像和重建图像。为保留光谱特征之间的相关性,引入 SAM 损失:

$$L_{\text{sam}} = \frac{1}{N} \sum_{n=1}^N \frac{1}{\pi} \arccos \left(\frac{\mathbf{I}_{\text{HR}}^n \cdot \mathbf{I}_{\text{SR}}^n}{\| \mathbf{I}_{\text{HR}}^n \|_2 \cdot \| \mathbf{I}_{\text{SR}}^n \|_2} \right) \quad (11)$$

另外,为了提升感知质量,引入梯度损失:

$$L_{\text{grad}} = \frac{1}{N} \sum_{n=1}^N \| \nabla(\mathbf{I}_{\text{HR}}^n) - \nabla(\mathbf{I}_{\text{SR}}^n) \|_1 \quad (12)$$

式中: $\nabla(\cdot)$ 表示沿水平、垂直和光谱维度的像元梯度。最终损失函数定义为

$$L_{\text{total}} = L_1 + \lambda_1 L_{\text{sam}} + \lambda_2 L_{\text{grad}} \quad (13)$$

式中: λ_1 和 λ_2 分别表示光谱损失和梯度损失的权重系数,本文 λ_1 和 λ_2 的取值为 0.5 和 0.1。

2 实验结果与分析

2.1 数据集

实验采用 3 个基准数据集,分别为 Pavia_Centre^[21]、CAVE^[22] 和 XiongAn^[23]。

1) Pavia_Centre: 由 ROSIS 传感器采集自意大利帕维亚中心区域。原始图像尺寸为 $1\,096 \times 1\,096$ 像素,经裁剪后的尺寸为 $1\,096 \times 715$ 像素,空间分辨率 1.3 m,光谱范围 470 ~ 838 nm,共 102 个波段。Pavia_Centre 测试集由图像前 256 列中的 2 块区域组成;验证集裁剪 257 ~ 320 列间的图像块;训练集则取自 320 列后带 32 像素重叠的图像块。

2) CAVE: 由哥伦比亚大学发布,包含 32 张室内场景的高光谱图像。图像的尺寸为 512×512 像素,光谱范围 400 ~ 700 nm,共 31 个波段。本文随机选择其中 80% 用于训练,测试集和验证集各占 10%。

3) XiongAn: 由上海技术物理研究所开发的全光谱多模态成像光谱仪采集。经处理得到的图像尺寸为 $1\,580 \times 3\,750$ 像素,空间分辨率 0.5 m,光

谱范围 400 ~ 1 000 nm,共 256 个波段。这里参照 Pavia_Centre 数据集的数据划分方式,构建测试集、验证集和训练集。具体地,从前 512 列裁剪出图像块构建测试集;513 ~ 768 列裁剪出图像块构建验证集;剩余列则裁剪出重叠 32 像素的图像块构建训练集。

低分辨率高光谱图像基于对应高分辨率图像通过双三次插值下采样获得。

2.2 实验设置

本文所提网络中 BGSA 个数设置为 6, BGSA 残差组个数为 4,中间特征图通道数 C 为 180。空间交互的通道压缩比 r_1 为 16,通道交互的通道压缩比 r_2 为 8。上采样阶段采用 PixelShuffle^[24]。训练采用 Adam 优化器,初始学习率为 0.000 1,每隔 30 个 epoch 衰减为原来的 10%。总训练轮数为 150,批量大小为 16。所有实验均在 PyTorch 框架下,基于 NVIDIA RTX A6000 GPU(48 GB 显存)运行,并采用峰值信噪比(peak signal-to-noise ratio, PSNR)、结构相似性指数(structural similarity index, SSIM)、光谱角映射(spectral angle mapper, SAM)、相对全局均方根误差(erreur relative globale adimensionnelle de synthèse, ERGAS)和相关系数(cross correlation, CC)等指标评价重建性能。

2.3 实验结果分析

为了客观评价 DSAT 方法的有效性,本文选取 GDRRN^[11]、MCNet^[9]、SSPSR^[12]、PDE-Net^[25]、GELIN^[10]、MSDformer^[17]、AS³ITransUnet^[26] 和 SNLSR^[19] 等主流方法在以上 3 个数据集进行对比实验。

2.3.1 Pavia_Centre 数据集

表 1 为 Pavia_Centre 数据集上尺度因子为 4 的性能对比结果。该数据集规模较小,故对基于深度学习的方法构成巨大挑战。本文所提 DSAT 方法在所有指标上均优于其他对比方法。具体而言,DSAT 在 PSNR、SSIM 和 CC 指标上分别较次优方法提升了 0.781 2 dB、0.020 5 和 0.011 2。同时,SAM 和 ERGAS 指标则分别降低了 0.943 4 和 0.615 8,表明 DSAT 方法的有效性。

图 3 进一步展示重建图像的主观质量对比结果,由第 32、21 和 11 波段合成伪彩色图。可以看出,DSAT 重建的图像展现出更为丰富的细节和更锐利的边缘。分析原因,主要得益于设计的 BGSA 模块能够高效并行提取空谱特征,并对其

进行自适应加权融合。

表 1 Pavia_Centre 数据集上的定量比较结果

Tab. 1 Quantitative comparison results on the Pavia_Centre dataset

Method	PSNR ↑ /dB	SSIM ↑	SAM ↓	ERGAS ↓	CC ↑
GDRRN	29.609 2	0.801 8	13.564 7	7.981 4	0.895 7
SSPSR	30.141 3	0.830 3	12.644 2	7.606 1	0.908 2
MCNet	29.912 3	0.824 3	12.958 0	7.733 9	0.903 5
PDE-Net	30.043 7	0.829 7	12.816 0	7.764 6	0.904 0
GELIN	30.135 8	0.833 0	12.625 7	7.596 7	0.907 9
MSDformer	<u>30.206 5</u>	<u>0.838 9</u>	<u>12.492 3</u>	<u>7.468 7</u>	<u>0.912 5</u>
AS ³ ITransUnet	29.910 7	0.824 1	12.962 1	7.820 4	0.903 4
SNLSR	29.787 0	0.822 2	13.150 2	7.885 0	0.899 9
DSAT	30.987 7	0.859 4	11.548 9	6.852 9	0.923 7

注:加黑数值表示最优结果;下划线数值表示次优结果。

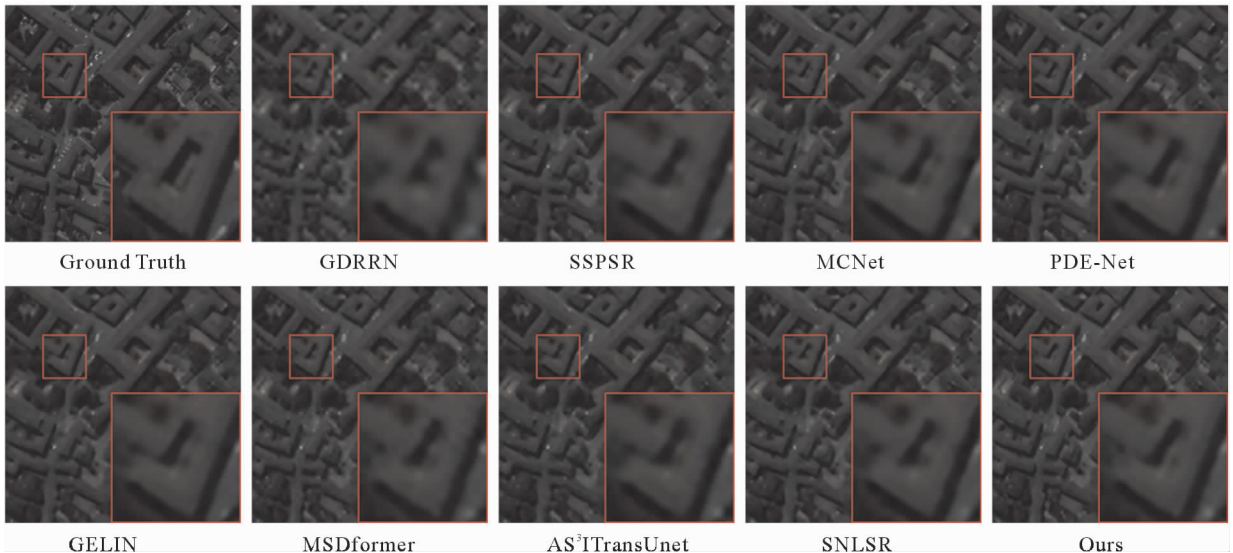


图 3 不同方法在 Pavia_Centre 数据集上的可视化对比结果

Fig. 3 Visual comparison results of different methods on the Pavia_Centre dataset

2.3.2 CAVE 数据集

表 2 展示了 CAVE 数据集上尺度因子为 4 的客客观质量比较结果。DSAT 方法相较于次优方法，在 PSNR、SSIM 和 CC 指标上分别取得了 0.848 8 dB、0.003 5 和 0.002 4 的提升,同时,SAM 和 ERGAS 指标分别降低 0.024 7 和 0.707 5。

表 2 CAVE 数据集上的定量比较结果

Tab. 2 Quantitative comparison results on the CAVE dataset

Method	PSNR ↑ /dB	SSIM ↑	SAM ↓	ERGAS ↓	CC ↑
GDRRN	33.784 5	0.933 4	9.419 4	9.350 7	0.978 1
SSPSR	35.503 4	0.951 2	7.733 6	<u>7.379 3</u>	<u>0.985 8</u>
MCNet	34.170 4	0.947 5	8.506 6	9.348 8	0.977 7
PDE-Net	<u>35.506 4</u>	0.953 4	7.747 4	7.470 6	0.985 1
GELIN	35.463 4	<u>0.955 0</u>	<u>7.203 1</u>	7.545 1	0.985 0
MSDformer	35.330 1	0.954 0	8.054 1	7.622 1	0.985 1
AS ³ ITransUnet	35.004 2	0.951 7	7.953 6	8.038 2	0.983 7
SNLSR	34.858 0	0.950 8	8.233 5	8.053 6	0.983 2
DSAT	36.355 2	0.958 5	7.178 4	6.671 8	0.988 2

注:加黑数值表示最优结果;下划线数值表示次优结果。

图4展示测试图像重建结果的主观质量对比,以第30波段合成灰度图。与其他方法相比,DSAT方法可恢复出更为逼真的细节。

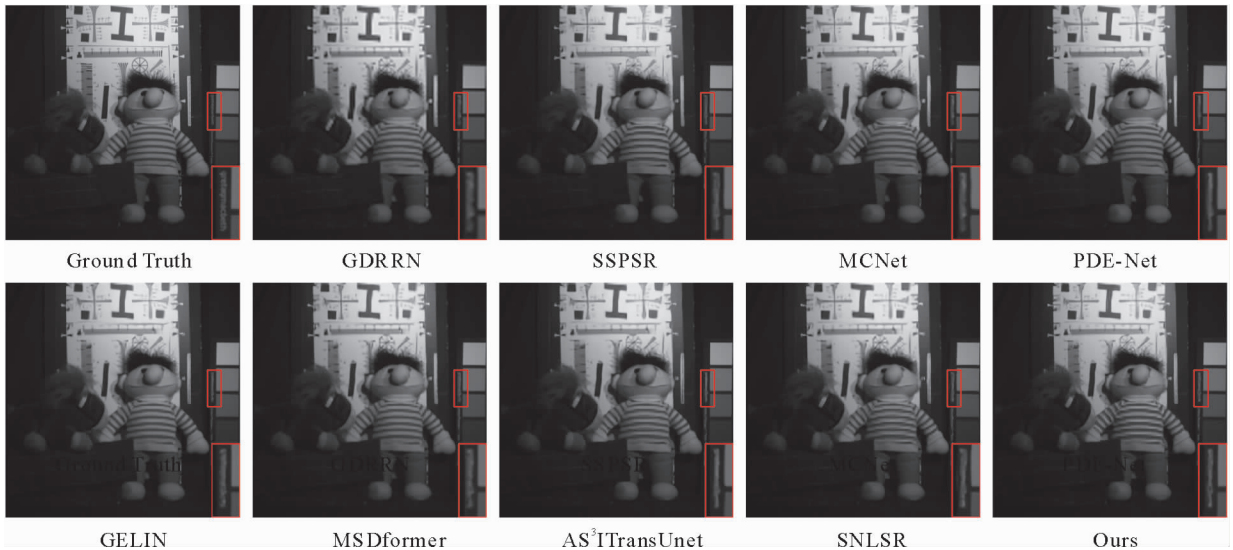


图4 不同方法在CAVE数据集上的可视化对比结果

Fig.4 Visual comparison results of different methods on the CAVE dataset

2.3.3 XiongAn数据集

表3列出XiongAn数据集上尺度因子为4的质量对比结果。鉴于该数据集规模较大且场景相对简单,所有方法均取得了良好的表现。但DSAT方法在所有指标上依然取得最佳。具体而言,DSAT在PSNR、SSIM和CC指标上分别较次优方法高出0.2456 dB、0.0031和0.0033。同

时,在SAM和ERGAS指标上则分别较次优方法降低了0.0736和0.0266。此外,图5展示了重建结果的主观视觉对比,由第120、72和36波段合成伪彩色图。可以看出,在边缘和纹理区域,DSAT展现出更卓越的性能,提供了更优的视觉质量。

表3 XiongAn数据集上的定量比较结果

Tab.3 Quantitative comparison results on the XiongAn dataset

Method	PSNR ↑ /dB	SSIM ↑	SAM ↓	ERGAS ↓	CC ↑
GDRRN	41.787 5	0.932 8	2.468 2	1.236 9	0.935 9
SSPSR	42.372 1	0.939 4	2.327 0	1.171 2	0.915 6
MCNet	42.307 2	0.938 7	2.336 3	1.182 9	0.929 8
PDE-Net	<u>42.436 4</u>	0.939 9	<u>2.308 9</u>	<u>1.146 8</u>	0.929 0
GELIN	42.298 0	0.939 8	2.329 9	1.174 3	<u>0.932 0</u>
MSDformer	42.287 6	<u>0.941 7</u>	2.339 8	1.149 9	0.928 4
AS ³ ITransUnet	42.072 4	0.936 4	2.394 7	1.220 0	0.930 8
SNLSR	42.165 7	0.937 9	2.367 2	1.190 2	0.925 4
DSAT	42.682 0	0.944 8	2.235 3	1.120 2	0.935 3

注:加黑数值表示最优结果;下划线数值表示次优结果。

2.4 消融实验

为了评估DSAT中模块的有效性,本文在Pavia_Centre数据集上对其进行消融实验。其中,训练阶段使用的低分辨率图像块大小为16×16×102像素,尺度因子为4,测试阶段使用的低

分辨率图像块大小为64×64×102像素,使用PSNR、SSIM和SAM3个主要评价指标进行评估。

为了验证BGSA的有效性,本文在保留其他网络结构不变的前提下,使用3个卷积层替换BGSA,构建对照模型,记为“Ours w/o BGSA”。

如表 4 所示,相比 Ours w/o BGSA,完整模型 Ours 在 PSNR 提高 0.884 8 dB,SSIM 提高 0.030 1, SAM 下降 1.159 0,3 个评价指标上均取得明显优势,验证了所提 BGSA 模块的有效性。



图 5 不同方法在 XiongAn 数据集上的可视化对比结果
Fig. 5 Visual comparison results of different methods on the XiongAn dataset

表 4 在不同变体上的消融实验结果

Tab. 4 Results of the ablation study on different variants

Variant	Params/ 10^6	FLOPs/ 10^9	PSNR $\uparrow$ /dB	SSIM $\uparrow$	SAM $\downarrow$
Ours	11.676 1	5.113 9	30.987 7	0.859 4	11.548 9
Ours w/o BGSA	6.625 7	3.861 6	30.102 9	0.829 3	12.707 9
Ours w/o AIM	11.259 6	5.039 8	29.728 9	0.814 1	13.311 9
Ours w/CSA only	10.751 3	5.742 8	30.384 6	0.842 0	12.300 9
Ours w/CSE only	9.675 8	5.196 7	30.696 3	0.846 8	11.894 2

注:加黑数值表示最优结果。

为了进一步分析 BGSA 内部组件对性能的影响,本文在 DSAT 的基础上设计多种变体。首先,通过移除 AIM 得到对照模型,记为 Ours w/o AIM。如表 4 所示,与完整模型 Ours 相比,移除 AIM 模块后,PSNR 下降 1.258 8 dB,SSIM 下降 0.045 3,SAM 则增加 1.763 0,表明 AIM 模块能够有效提高空谱特征利用效率,显著提升重建精度与光谱保持能力。其次,为了研究不同维度自注意力机制的作用,本文构造了 2 种单一类型注意力主导的 DSAT 变体:在 Ours w/CSA only 中,BGSA 的空间分支与光谱分支均采用 CSA 结构,执行空间自注意力;在 Ours w/CSE only 中,BGSA 的 2 个分支则均采用 CSE 结构,执行光谱自注意力。如表 4 所示,这 2 种变体的性能均劣于完整模型 Ours。其中,Ours w/CSA only 相比 Ours 的 PSNR 降低 0.603 1 dB,SSIM 降低 0.017 4,SAM

增加 0.752 0,说明仅依赖空间自注意力,难以兼顾光谱一致性;Ours w/CSE only 虽然在 SAM 指标上接近 Ours,但 PSNR 和 SSIM 仍存在明显差距,表明仅使用光谱自注意力在边缘与纹理细节恢复方面仍有不足。

为了研究损失函数中各项权重系数对模型性能的影响,本文对损失函数超参数进行了系统的消融实验。具体来说,将 L1 损失视为主损失项,其系数固定为 1,光谱损失和梯度损失的权重系数 λ_1 和 λ_2 分别取 $\{0,0.1,0.3,0.5,0.7\}$ 中的不同组合进行实验。表 5 给出了部分代表性结果。可以看到,适当引入光谱和梯度约束能在一定程度上提高 PSNR 和 SSIM 并降低 SAM。当 λ_1 和 λ_2 分别取 0.5 和 0.1 时,模型在 3 项指标上均取得最优表现,继续增大其中任一权重,模型性能反而下降。

表5 损失函数超参数的消融研究

Tab.5 Ablation study of loss function hyperparameters

λ_1	λ_2	PSNR $\uparrow$ /dB	SSIM $\uparrow$	SAM $\downarrow$
0.0	0.1	30.824 3	0.853 9	11.768 0
0.1	0.1	30.874 1	0.855 2	11.697 0
0.3	0.1	30.925 4	0.856 2	11.647 4
0.5	0.1	30.987 7	0.859 4	11.548 9
0.7	0.1	30.639 9	0.849 7	11.971 0
0.5	0.0	30.684 7	0.851 5	11.904 8
0.5	0.3	30.796 5	0.853 5	11.805 9
0.5	0.5	30.761 2	0.854 5	11.843 8
0.5	0.7	30.745 8	0.853 2	11.883 1

注:加黑数值表示最优结果。

3 结语

本文针对单高光谱图像超分辨率中难以有效捕获复杂的空间-光谱长程依赖的关键挑战,提出一种双域空间-光谱聚合变压器(DSAT)网络。DSAT通过引入BGSA,实现空间和光谱维度的并行高效自注意力计算,显著增强模型捕获全局依赖信息的能力。此外,AIM的引入,进一步优化空间和光谱特征的融合策略,提升模型的特征表达能力。在3个基准数据集上的实验结果表明,与现有主流方法相比,DSAT在5个客观质量评价指标以及主观质量对比上均取得显著提升,验证了其在高光谱图像超分辨率重建领域的有效性。DSAT在捕获长程依赖方面性能显著,但鉴于Transformer架构固有的计算与参数开销,如何在保持超分效果的同时实现模型轻量化是未来的研究重点。

参考文献

[1] XU Y C, ZHANG L F, DU B, et al. Hyperspectral anomaly detection based on machine learning: An overview[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2022, 15: 3351-3364.

[2] ZHANG Z, HUANG L H, WANG Q W, et al. UAV hyperspectral remote sensing image classification: A systematic review[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2025, 18: 3099-3124.

[3] XU X, LI J, HUANG X, et al. Multiple morphological component analysis based decomposition for remote sensing image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(5): 3083-3102.

[4] LU G L, FEI B W. Medical hyperspectral imaging: A review[J]. Journal of Biomedical Optics, 2014, 19(1): 010901.

[5] DIAN R W, FANG L Y, LI S T. Hyperspectral image super-resolution via non-local sparse tensor factorization[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI: IEEE, 2017: 3862-3871.

[6] XU Y, WU Z B, CHANUSSOT J, et al. Nonlocal patch tensor sparse representation for hyperspectral image super-resolution [J]. IEEE Transactions on Image Processing, 2019, 28(6): 3034-3047.

[7] YANG J C, WRIGHT J, HUANG T S, et al. Image super-resolution via sparse representation [J]. IEEE Transactions on Image Processing, 2010, 19(11): 2861-2873.

[8] MEI S H, YUAN X, JI J Y, et al. Hyperspectral image super-resolution via convolutional neural network [C]//2017 IEEE International Conference on Image Processing (ICIP). Beijing: IEEE, 2017: 4297-4301.

[9] LI Q, WANG Q, LI X L. Exploring the relationship between 2D/3D convolution for hyperspectral image super-resolution[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(10): 8693-8703.

[10] WANG X Y, HU Q, JIANG J J, et al. A group-based embedding learning and integration network for hyper-

- spectral image super-resolution[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5541416.
- [11] LI Y, ZHANG L, DING L C, et al. Single hyperspectral image super-resolution with grouped deep recursive residual network[C]//2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM). Xi'an: IEEE, 2018: 1-4.
- [12] JIANG J J, SUN H, LIU X M, et al. Learning spatial-spectral prior for super-resolution of hyperspectral imagery[J]. IEEE Transactions on Computational Imaging, 2020, 6: 1082-1096.
- [13] WANG X Y, MA J Y, JIANG J J. Hyperspectral image super-resolution via recurrent feedback embedding and spatial-spectral consistency regularization [J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-13.
- [14] WANG H, WANG C, YUAN Y. Neighbor spectra maintenance and context affinity enhancement for single hyperspectral image super-resolution[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 5515315.
- [15] LIU Y T, HU J W, KANG X D, et al. Interactformer: Interactive transformer and CNN for hyperspectral image super-resolution[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5531715.
- [16] ZHANG M J, ZHANG C, ZHANG Q M, et al. ES-SAformer: Efficient Transformer for hyperspectral image super-resolution [C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris: IEEE, 2023: 23016-23027.
- [17] CHEN S, ZHANG L F, ZHANG L P. MSDformer: Multiscale deformable transformer for hyperspectral image super-resolution[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 5525614.
- [18] CHEN S, ZHANG L F, ZHANG L P. Cross-scope spatial-spectral information aggregation for hyperspectral image super-resolution[J]. IEEE Transactions on Image Processing, 2024, 33: 5878-5891.
- [19] HU Q, WANG X Y, JIANG J J, et al. Exploring the spectral prior for hyperspectral image super-resolution [J]. IEEE Transactions on Image Processing, 2024, 33: 5260-5272.
- [20] CHEN Z, ZHANG Y L, GU J J, et al. Dual aggregation transformer for image super-resolution[C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris: IEEE, 2023: 12278-12287.
- [21] GAMBA P. A collection of data for urban area characterization [C]//2004 IEEE International Geoscience and Remote Sensing Symposium. Anchorage: IEEE, 2004: 69-72.
- [22] YASUMA F, MITSUNAGA T, ISO D, et al. Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum[J]. IEEE Transactions on Image Processing, 2010, 19(9): 2241-2253.
- [23] 岑奕, 张立福, 张霞, 等. 雄安新区马蹄湾村航空高光谱遥感影像分类数据集[J]. 遥感学报, 2020, 24(11): 1299-1306.
- CEN Y, ZHANG L F, ZHANG X, et al. Aerial hyperspectral remote sensing classification dataset of Xiongan New Area (Mativan Village) [J]. Journal of Remote Sensing (Chinese), 2020, 24(11): 1299-1306.
- [24] SHI W Z, CABALLERO J, HUSZÁR F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 1874-1883.
- [25] HOU J H, ZHU Z Y, HCU J H, et al. Deep posterior distribution-based embedding for hyperspectral image super-resolution[J]. IEEE Transactions on Image Processing, 2022, 31: 5720-5732.
- [26] XU Q, LIU S J, WANG J H, et al. AS³ITransUNet: Spatial-spectral interactive transformer U-Net with alternating sampling for hyperspectral image super-resolution[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-13.

(编辑 李 静)