

图像代理注意力机制在遥感图像锐化中的研究

张绪杰

青岛大学计算机科学技术学院, 中国山东省青岛市

发布日期: 2025年5月30日

摘要

近年来, 注意力机制作为一种模拟人类视觉系统选择性聚焦能力的技术, 在图像处理领域引起了广泛关注。在图像处理领域中, 基于自注意力机制的Transformer架构具有广阔的应用前景, 能够有效解决长程依赖和全局特征建模问题。然而, 自注意力机制的高计算复杂度限制了模型的应用。本文提出一种新型的注意力机制——图像代理注意力机制, 应用于在遥感图像锐化模型HyperTransformer, 在公开数据集Pavia Center上进行实验, 分别从计算效率和图像锐化性能方面对改进模型进行量化评估。结果显示与基于自注意力、轴注意力的模型相比, 图像代理注意力模型在参数量和浮点运算次数上明显降低。与自注意力模型相比, 参数量减少了21.0%, 浮点运算量减少了22.6%。与轴注意力模型相比, 参数量减少了13.4%, 浮点运算量减少了15.7%。同时图像锐化质量评价指标也取得了较好的结果, 如相对全局合成误差降到了0.48。证明了图像代理注意力机制在遥感图像锐化中的适用性。

关键词

遥感图像锐化; 自注意力; 轴注意力; 图像代理注意力

A Study on Remote Sensing Image Sharpening Using Graph Proxy Attention

Xujie Zhang

School of Computer Science and Technology, Qingdao University, Qingdao 266071, China

Abstract

In recent years, as a technology that simulates the selective focusing ability of the human visual system, attention mechanisms have gained significant attention in the field of image processing. Within this field, Transformer architectures based on self-attention mechanisms show great potential for effectively addressing issues related to long-range dependencies and global feature modeling. Nevertheless, the high computational complexity associated with self-attention mechanisms has restricted their application. This paper introduces a novel attention mechanism—Graph Proxy Attention. It is applied in the remote sensing image sharpening model called HyperTransformer. Experiments are carried out on the public Pavia Center dataset. The

文章引用: 张绪杰. (2025). 图像代理注意力机制在遥感图像锐化中的研究[J]. 计算机科学辑要, 1(1), 164 - 173.

ISSN Print: 3079-5362; ISSN Online: 3079-5370.

DOI: 10.63313/CS.8016

improved model is evaluated in terms of computational efficiency and image sharpening performance. The results indicate that, compared to models based on self-attention and axial attention, the Graph Proxy Attention model significantly reduces the number of parameters and floating-point operations. Specifically, when compared to self-attention models, the parameter amount is reduced by 21.0%, and the floating-point operations are reduced by 22.6%. In comparison to axial attention models, the parameter amount is reduced by 13.4%, and the floating-point operations are reduced by 15.7%. At the same time, the evaluation indicators for image sharpening quality have also achieved favorable results. For example, the relative global synthetic error has dropped to 0.48. These findings demonstrate the applicability of the Graph Proxy Attention mechanism in remote sensing image sharpening.

Keywords

Remote Sensing Image Sharpening; Self-Attention; Axial Attention; Graph Proxy Attention

Copyright © 2025 by author(s) and Erytis Publishing Limited.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



1. 引言

随着人工智能技术的快速发展,图像处理已成为计算机视觉领域的核心研究方向之一[1]。从图像识别、图像分类到遥感图像锐化,图像处理技术的应用场景日益广泛。然而,传统图像处理方法在面对复杂场景时往往面临诸多挑战,例如计算复杂度高、对多尺度信息处理能力不足、对背景噪声的鲁棒性差等。这些问题限制了计算机视觉系统在实际应用中的性能和效率[2]。传统的遥感图像锐化方法,如分量替换(Component Substitution, CS)[3]、多分辨率分析(Multiresolution Analysis, MRA)[4]和贝叶斯(Bayesian Approach, BA)[5]方法,虽然在一定程度上提高了图像的空间分辨率,但往往导致光谱失真和空间扭曲,难以在提升空间细节的同时保持光谱信息的完整性。近年来,基于自注意力机制的遥感图像锐化研究带来了新的突破。例如,基于自注意力的融合跨阶段局部网络(CSPNet)与无参数注意力(SimAM)的遥感图像锐化方法[6],用于解决遥感图像全色锐化中存在的光谱分布不均和空间细节缺失问题。该方法通过引入CSPNet替代特征提取中的残差块,缓解梯度冗余问题,并利用SimAM块提取更深层次的特征信息。此外,设计了可学习作差参数以突出融合图像的边缘信息,有效改善了模型的梯度冗余,提升了融合图像的空谱分辨率;多尺度无参数自注意力机制增强网络[7],用于获取更高频信息和纹理细节的遥感图像。该网络通过多尺度特征提取和参数自由注意力机制抑制冗余信息,显著提升了图像的重建性能,包括峰值信噪比和结构相似性。现有基于注意力机制的遥感图像锐化方法在提升锐化效果方面虽有显著进展,但也存在多方面问题。如计算复杂度偏高,处理大规模遥感图像数据时计算量极大,导致实际应用效率低下且可扩展性受限。过拟合风险也不容忽视,这类方法通常参数量庞大,容易过拟合,致使模型在新数据上的泛化能力欠佳。为应对高计算复杂度与过拟合风险,往往需要更强的硬件支持及更长的训练时间,这无疑增加了总体成本。针对基于遥感图像锐化任务中深度学习方法的注意力机制计算复杂度高的问题,本文提出了一种新型的图像代理注意力。通过将图像代理注意力机制应用于遥感影像锐化模型HyperTransformer[8]中,在公开数据集Pavia Center[9]上的实验,分别从计算效率,如模型参数量Params、浮点运算量FLOPs,和影像锐化性能方面,如交叉相关(Cross-Correlation, CC)[10]、光谱角度映射(Spectral Angle Mapper, SAM)[11]、均方根误差(Root Mean Square Error, RMSE)[12]、全

局相对误差 (Erreur Relative Globale Adimensionnelle Desynthese, ERGAS) [13]、峰值信噪比 (Peak Signal to Noise Ratio, PSNR) [14] 多个维度对改进模型进行量化评估, 并做了图像代理向量数量消融实验验证了模型的有效性。

2. 图像代理注意力机制

2.1. 图像代理注意力的的基本思想

图像作为高度结构化的数据形式, 其本质是具备显著共性特征的空间分布表征。在图像域中, 有效的视觉表征应满足特定统计规律 (如局部连续性、空间相关性等), 而这些共性特征可被抽象为低维代理向量 (Proxy)。通过将传统注意力机制中的像素级查询替换为代理查询, 能够将高维图像特征间的复杂交互转化为代理与图像特征之间的相似性度量: 代理向量通过学习捕捉图像全局共性, 而图像特征则通过对比代理的共性特征与其自身特性之间的差异, 动态生成适配于当前任务的自适应表征。由于代理维度 (如代理数量或通道数) 远低于原始图像的空间维度 (如像素数或区域数), 该机制将降低注意力计算的时间复杂度, 从而在保留关键语义信息的前提下显著提升计算效率。

对于一张维度为 $w \times h$ 图像的所有像素, 共有 $(w \times h)^2$ 种像素排列组合方式, 但是这些组合中绝大部分都是不能被使用的无效图像, 如图 1 (b) 中的图像所示, 而可以被使用的图像只是所有组合方式中的一个子集, 如图 1 (a) 中的图像所示, 它们都有像素平滑等共有特征, 在图像代理注意力机制中, 引入少量的图像代理向量 A 来学习图像的这种共有特征, 由学习后的图像代理向量 A 来得到查询矩阵 Q , 进而和键矩阵 K 结合来计算注意力得分, 从而减少计算复杂度。

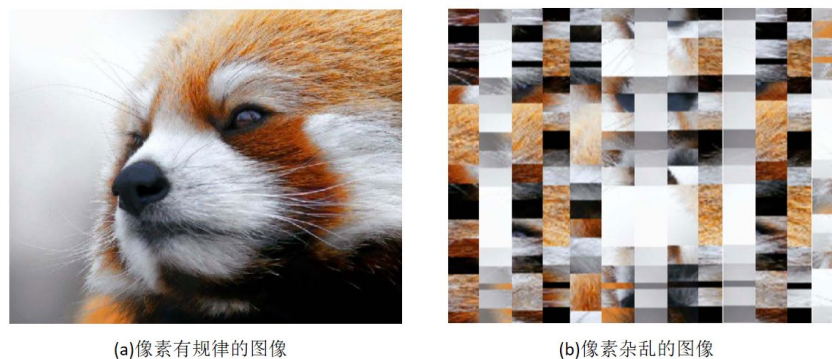


Figure 1. Comparison of Images with Regular and Irregular Pixel Patterns
图 1. 图像像素规律与杂乱的图像对比图

2.2. 图像代理注意力的架构

首先图像代理向量 A 通过学习获得图像的共有特征, 在查询、键和值矩阵的生成过程中, 通过 1×1 的卷积操作来保持通道数不变, 从而生成相应的矩阵。接着, 利用矩阵乘法计算注意力得分, 并通过缩放点积来稳定训练过程。然后添加相对位置偏移进一步增强模型对局部位置信息的捕捉能力, 使得模型能够更好地理解特征图中像素之间的相对关系。

最后, 通过矩阵乘法将注意力得分应用于值矩阵, 实现值的聚合, 并将宽度和高度的结果相加, 得到最终的输出特征图。这一过程不仅保留了局部信息的交互, 还显著减少了计算复杂度, 使得图像代理注意力机制在处理大规模特征图时具有较高的效率和效果, 具体来说, 图像代理注意力机制通过引入图像代理向量 A , 学习图像的固有特征, 减少了查询和键之间的直接接触, 从而减少部分计算量。同时,

相对位置偏移的引入使得模型能够更好地捕捉局部位置信息，保持全局感受野。图像代理注意力架构如图 2 所示。

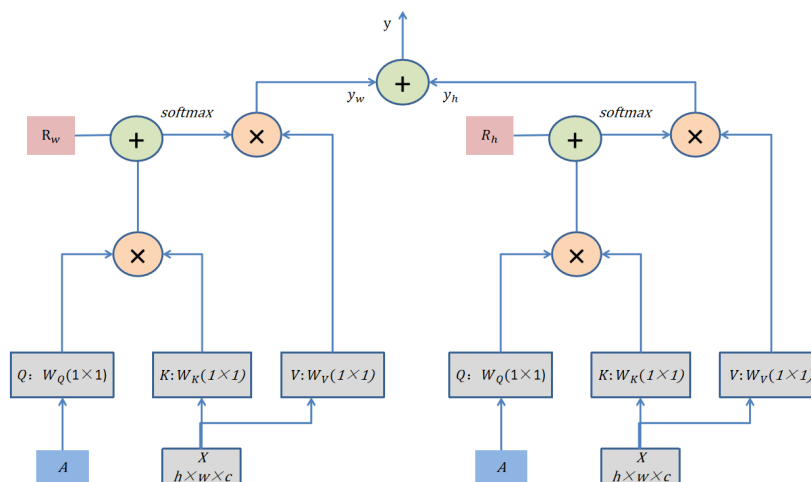


Figure 2. Graph Proxy Attention Mechanism Architecture
图 2. 图像代理注意力机制架构

3. HyperTransformer 模型

3.1. 基于图像代理注意力机制的 HyperTransformer 模型架构

HyperTransformer 是一种用于高光谱图像 Pansharpening 的深度学习模型，它通过结合 PAN 图像和 LR-HSI 图像来生成具有高光谱和空间分辨率的增强型 HSI。图 3 展示了 HyperTransformer 模型的总体结构。 p 为高分辨率的全色图像 PAN， $p \downarrow$ 为先下采样再上采样得到的 PAN 图像，其与 LR-HSI 图像在空间分辨率上对齐。 $y \uparrow$ 为上采样的 LR-HSI 图像，提供光谱信息。

FE-PAN 特征提取器用于提取 PAN 图像的特征，将 PAN 图像转换为特征空间，输出特征图 K (Key) 和 V (Value)，其中 K 作为键，用于与 LR-HSI 特征进行关联，V 作为值，用于提供 PAN 的纹理信息。FE-HSI 特征提取器用于提取 LR-HSI 图像的特征。它将上采样的 LR-HSI 图像转换为特征空间，由可学习的图像代理向量 A 来学习图像的共有特征，将图像共有特征作为查询矩阵 Q，用于与 PAN 特征进行关联。首先计算 Q 和 K 之间的相关性并应用 Softmax 函数来确定注意力权重，这些权重通过加权 V 中的特征来强调重要的特征。随后，这些加权的特征通过交叉相关性嵌入 (Cross-correlation Embedding) 模块进一步处理，以捕获 PAN 和 HSI 特征之间的相关性。接下来，通过包含图像代理注意力机制的特征软注意力模块 (Feature Soft-Attention) 处理不同特征子空间中捕获的信息，经过线性变换 (linear) 转换到适合融合的特征空间后，处理后的特征 (T) 被进一步送入纹理-光谱特征融合 (Texture & Spectral Feature Fusion) 模块，与从主干网络提取的光谱特征 (F) 相结合，其中 F 表示的是低分辨率高光谱图像 LR-HSI 的特征，它是通过主干网络对上采样的 LR-HSI 进行处理得到的特征表示。这些特征表示包含了 LR-HSI 的光谱信息，但在空间分辨率上较低。最后，这些融合的特征通过线性层进一步处理，以产生最终的增强型 HSI 输出 ($\hat{F}$)，这个输出具有高光谱和空间分辨率，有效地结合了 PAN 的高分辨率和 HSI 的丰富光谱信息，从而提高了锐化图像的空间质量和光谱质量。

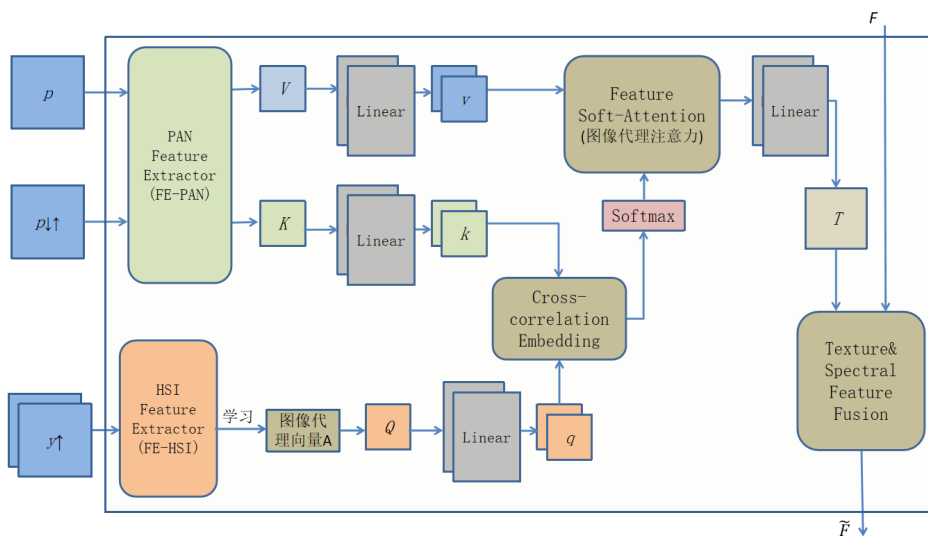


Figure 3. Based on the HyperTransformer Architecture with Graph Proxy Attention Mechanism
图 3. 基于图像代理注意力机制 HyperTransformer 架构

3.2. HyperTransformer 模型损失函数

本实验使用的 HyperTransformer 模型的损失函数在原模型 L_1 损失、VGG 感知损失 $L_{\text{vgg-per}}$ 和转移感知损失 $L_{\text{t-per}}$ 的基础上,增加了重建损失 L_{nrec} 和图像质量损失 L_{nrf} ,最终的损失函数表述为:

$$\mathcal{L}_{\text{overall}} = \lambda_1 \mathcal{L}_1 + \lambda_{\text{vgg-per}} \mathcal{L}_{\text{vgg-per}} + \lambda_{\text{t-per}} \mathcal{L}_{\text{t-per}} + \lambda_{\text{nrec}} \mathcal{L}_{\text{nrec}} + \lambda_{\text{nrf}} \mathcal{L}_{\text{nrf}} \quad (1)$$

其中 λ_1 、 $\lambda_{\text{vgg-per}}$ 、 $\lambda_{\text{t-per}}$ 、 λ_{nrec} 和 λ_{nrf} 是系数。实验中设置为 $\lambda_1 = 0.5$, $\lambda_{\text{vgg-per}} = 0.1$, $\lambda_{\text{t-per}} = 0.05$, $\lambda_{\text{nrec}} = 0.5$, $\lambda_{\text{nrf}} = 0.05$ 。

(1) L_1 损失

L_1 损失计算的是参考 HR-HSI 图像和预测 HR-HSI 图像之间的差距。

$$\mathcal{L}_{\text{rec}} = \frac{1}{\text{CWH}} \|\mathbf{x}_{\text{ref}} - \mathbf{x}\|_1 \quad (2)$$

其中, $\mathcal{L}_{\text{rec}}$ 是 L_1 损失, $\mathbf{x}_{\text{ref}}$ 是参考 HR-HSI 图像, $\mathbf{x}$ 是预测 HR-HSI 图像, (C,W,H)是高分辨率高光谱图像的尺寸。使用 L_1 损失,因为它在高光谱图像锐化任务中比 L_2 损失表现更好。

(2)VGG 感知损失函数

VGG 感知损失函数已被证明在 RGB 域的超分辨率任务中可以用来提升视觉质量[20]。VGG 感知损失主要是增强预测图与目标图特征空间的相似性,首先在预训练 VGG 网络上得到预测图,然后合成 RGB 图像,再计算感知损失:

$$\mathcal{L}_{\text{vgg-per}} = \frac{1}{C_i W_i H_i} \left\| \mathbf{f}_i^{\text{vgg}}(\mathbf{x}_{\text{ref}}^{\text{rgb}}) - \mathbf{f}_i^{\text{vgg}}(\mathbf{x}^{\text{rgb}}) \right\|_2 \quad (3)$$

其中, $\mathcal{L}_{\text{vgg-per}}$ 是 VGG 感知损失, $\mathbf{f}_i^{\text{vgg}}(\cdot)$ 表示 VGG-19 的第 i 层的特征图, (C_i, W_i, H_i) 表示该层特征图的尺寸。 $\mathbf{x}_{\text{ref}}^{\text{rgb}}$ 和 $\mathbf{x}^{\text{rgb}}$ 分别是参考 HR-HIS 图像 $\mathbf{x}_{\text{ref}}$ 和预测 HR-HIS 图像 $\mathbf{x}$ 的合成 RGB 图像。

(3)转移感知损失函数

转移感知损失函数主要为了约束预测的高光谱图像 HSI,使其具有与从 HyperTransformer 转移的纹理特征 T 相似的纹理特征,这样可以使得纹理特征传输过程更有效。

$$\mathcal{L}_{\text{t-per}} = \frac{1}{C_s W_s H_s} \left\| \mathbf{f}_{\text{FE-HSI}}(\mathbf{x})^s - T^s \right\|_2 \quad (4)$$

其中, $\mathcal{L}_{t-per}$ 是转移感知损失, s 表示空间尺度, 可以取值为 1、2 或 4, 分别对应于不同的分辨率级别。 $s = 1$ 表示 LR-HSI 的原始分辨率 (即 $\times 1 \uparrow$)。 $s = 2$ 表示 LR-HSI 的两倍上采样分辨率 (即 $\times 2 \uparrow$)。 $s = 4$ 表示 LR-HSI 的四倍上采样分辨率 (即 $\times 4 \uparrow$, 目标高分辨率)。其中 T^s 表示第 s 个空间尺度的转移特征图, $f_{FE-HSI}(x)^s$ 是高光谱图像特征提取器在第 s 个空间尺度的特征图, (C_s, W_s, H_s) 分别表示该尺度特征图的尺寸, C_s 是通道数, W_s 是宽度, H_s 是高度。

(4) 重建损失

为了确保生成的高分辨率多光谱图像与输入的低分辨率多光谱图像在结构上保持高度一致, 在损失函数中引入重建损失函数。该损失函数通过以下公式计算:

$$\mathcal{L}_{nrec} = \mathcal{L}_{rec-spatial} + \mathcal{L}_{rec-spectral} \quad (5)$$

其中, $\mathcal{L}_{rec-spatial}$ 和 $\mathcal{L}_{rec-spectral}$ 代表的是重建空间损失与重建光谱损失, 作用是衡量生成图像与输入图像在空间结构和光谱特性上的一致性, 从而确保生成的图像在细节和整体结构上与输入图像保持一致。其计算的过程如下:

$$\mathcal{L}_{rec-spatial} = \sum_{n=1}^N \|I^{(n)} - AP(\widehat{Y_P^{(n)}})\|_1 \quad (6)$$

$$\mathcal{L}_{rec-spectral} = \sum_{n=1}^N \|P^{(n)} - \downarrow(\widehat{Y_P^{(n)}})\|_1 \quad (7)$$

$I^{(n)}$ 表示输入的低分辨率多光谱图像的第 n 个样本。 $\widehat{Y_P^{(n)}}$ 表示生成的高分辨率多光谱图像的第 n 个样本。 $AP(\cdot)$ 表示空间处理操作, 通常用于将生成的高分辨率图像转换为与输入低分辨率图像一致的空间结构。 $\|\cdot\|_1$ 表示 L1 范数, 用于计算两个图像之间的像素级绝对误差。 $P^{(n)}$ 表示输入的低分辨率多光谱图像的第 n 个样本的光谱特征。 $\downarrow(\cdot)$ 表示降采样操作, 用于将生成的高分辨率图像的分辨率降低到与输入低分辨率图像一致。

(5) 图像质量损失

为了提高锐化图像的质量, 在损失函数中引入图像质量损失, 其计算方式如下所示:

$$L_{nrf} = 1 - QNR \quad (8)$$

L_{nrf} 表示图像质量损失, 用于衡量生成图像与目标图像在整体质量上的差异, QNR 表示图像质量综合指标, 它的取值范围在 0 到 1 之间, 值越大代表图像质量越好。 QNR 的计算如下:

$$QNR = (1 - D_\lambda)(1 - D_s) \quad (9)$$

其中 D_λ 表示光谱指标, 评估生成图像与目标图像在光谱特性上的相似度。 D_s 表示空间指标, 衡量生成图像与目标图像在空间结构上的一致性。 D_λ 的计算公式如下:

$$D_\lambda = \sqrt{\frac{2}{K(K-1)} \sum_{i=1}^K \sum_{j=1}^K |Q(\widehat{Y_{P_i}}, \widehat{Y_{P_j}}) - Q(P_i, P_j)|} \quad (10)$$

K 表示输入的低分辨率多光谱 (LRMS) 图像的通道数量。 P_i 表示输入的低分辨率多光谱图像的第 i 个通道。 $\widehat{Y_{P_i}}$ 表示生成的高分辨率多光谱 (HRMS) 图像的第 i 个通道。 $Q(x, y)$ 表示 QI 指标, 用于评估两个通道之间的相似性, 计算公式如下:

$$Q(x, y) = \frac{4\sigma_{xy}\bar{x}\bar{y}}{(\sigma_x^2 + \sigma_y^2)(\bar{x}^2 + \bar{y}^2)} \quad (11)$$

σ_{xy} 代表的是 x 和 y 的协方差, σ_x^2 和 σ_y^2 分别代表的是 x 和 y 的方差, $\bar{x}$ 和 $\bar{y}$ 分别代表的是 x 和 y 的均值, 空间指标 D_s 的计算公式如下:

$$D_s = \sqrt{\frac{1}{K} \sum_{i=1}^K |Q(\widehat{Y_{P_i}}, I) - Q(P_i, \downarrow I)|} \quad (12)$$

I 表示生成器输入的 PAN 图像, $\downarrow I$ 是对 I 进行 4 倍下采样后的图像, $Q(\widehat{Y}_p, I)$ 表示评估生成的 HRMS 图像的第 i 个通道与 PAN 图像之间的相似性, $Q(P_i, \downarrow I)$ 表示衡量输入的 LRMS 图像的第 i 个通道与下采样后的 PAN 图像之间的相似性。

4. 实验部分

4.1. 模型整体应用实验

在遥感图像锐化任务中, 图像代理注意力机制通过引入相对位置偏移和对图像共性特征的学习, 该机制能够增强模型对局部位置信息的捕捉能力, 更好地理解 and 处理遥感图像的结构信息。为了证明基于图像代理注意力的 HyperTransformer 模型的有效性, 将基于图像代理注意力的 HyperTransformer 模型与基于自注意力、轴注意力和代理注意力的 HyperTransformer 模型进行了比较, 损失函数包括 L_1 损失、VGG 感知损失、转移感知损失、重建损失、图像质量损失, 实验中设置为 $\lambda_1 = 0.5$, $\lambda_{vgg-per} = 0.1$, $\lambda_{t-per} = 0.05$, $\lambda_{nrec} = 0.5$, $\lambda_{nrf} = 0.05$, 图像代理向量数量默认为 8, 卷积核大小为 3×3 , 实验结果如表 1 所示, 可以看出与分别基于自注意力、轴注意力和代理注意力的 HyperTransformer 模型相比, 图像代理注意力模型在参数量 (Params) 和浮点运算 (FLOPs) 次数上明显降低。具体来说, 图像代理注意力模型的参数量为 68.5M, 浮点运算量为 14189M。与自注意力模型相比, 参数量减少了 21.0%, 浮点运算量分别减少了 22.6%。与轴注意力模型相比, 参数量减少了 13.4%, 浮点运算量减少了 15.7%。与代理注意力模型相比, 参数量减少了 12.0%, 浮点运算量减少了 11.2%。

在降低计算复杂度的同时, 图像代理注意力模型在影像锐化质量评价指标上表现优异。交叉相关系数 (CC) 为 0.993, 与原模型相当; 光谱角匹配 (SAM) 为 3.82, 也与原模型相当。此外, 相对全局合成误差 (ERGAS) 指标优化明显, 图像代理注意力模型的 ERGAS 值为 0.48, 优于其他模型。

这些定量比较结果证明了基于图像代理注意力的 HyperTransformer 模型相对于分别基于自注意力、轴注意力和代理注意力的 HyperTransformer 模型的优越性。图像代理注意力模型不仅在计算复杂度上显著降低, 而且在影像锐化质量上保持了较高的水平, 甚至在某些指标上有所提升。这表明图像代理注意力机制能够更有效的平衡模型性能和计算复杂度。

Table 1. Performance Comparison of HyperTransformer Models Based on Graph Proxy Attention with Other Attention Mechanisms

表 1. 基于图像代理注意力的 HyperTransformer 模型与其他注意力性能比较

	Params	FLOPs	CC	SAM	RMSE	ERGAS	PSNR
自注意力	108M	19649M	0.992	3.74	0.0080	1.97	43.98
轴注意力	98.5M	17964M	0.991	3.84	0.0091	0.49	42.89
代理注意力	97.3M	17135M	0.990	3.83	0.0090	0.50	42.98
图像代理注意力	85.5M	15168M	0.993	3.82	0.0084	0.48	43.87

图 4 展示了高光谱图像的合成 RGB 图像以及重建高光谱图像与参考高光谱图像之间的平均绝对误差 MAE 图像。尽管合成 RGB 图像之间的差异很小, 但从 MAE 图像中观察到每种方法之间的差异性。如 MAE 图像所示, 本研究提出的基于图像代理注意力机制的 HyperTransformer 模型相比于分别基于自注意力机制、轴注意力机制和代理注意力机制的 HyperTransformer 模型实现了更优效果, 这些视觉结果进一步证明了基于图像代理注意力机制的 HyperTransformer 模型的卓越能力。

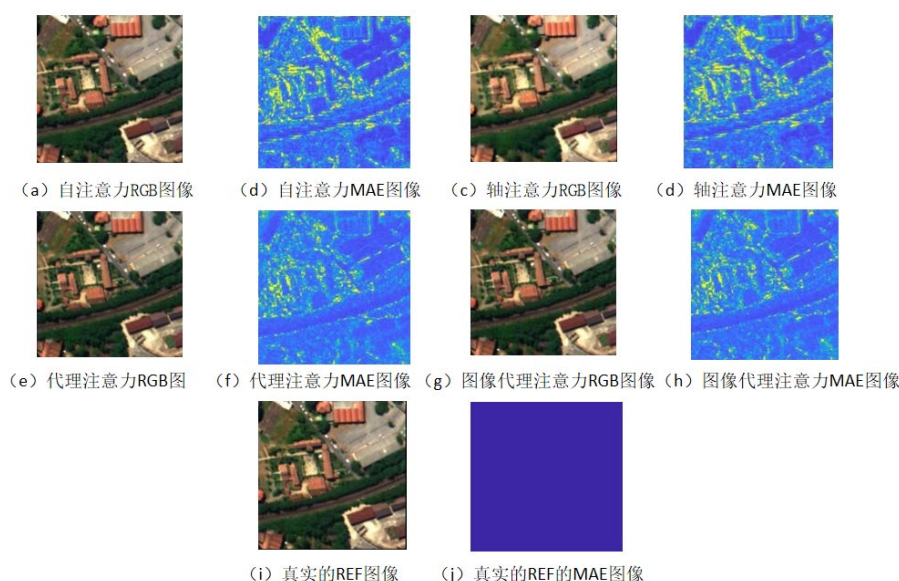


Figure 4. Schematic Diagram of RGB and MAE Images

图 4. RGB 与 MAE 图像示意图

图 5 展示了在 Pavia Center 数据集的建筑区域, 在不同注意力机制下的遥感影像锐化实验结果。具体来说, 图中对比了自注意力机制、轴注意力机制、代理注意力机制和图像代理注意力机制四种机制在生成高分辨率高光谱图像 HR-HSI 时的表现。实验结果表明, 图像代理注意力机制在降低计算复杂度的同时, 能够保持或提高遥感影像锐化的质量。这些机制在参数量、计算复杂度和锐化质量指标上均表现出色, 特别是在相对全局合成误差 (ERGAS) 方面有显著提升, 这些实验可视化结果证明了分别基于图像代理注意力机制的 HyperTransformer 模型在高光谱图像 Pansharpening 任务中的有效性和适用性。

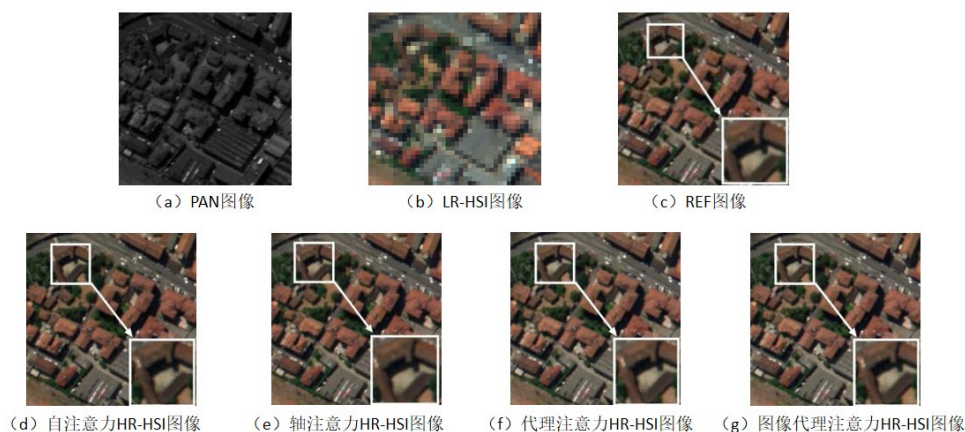


Figure 5. Visualization Results of Building Images in the Dataset

图 5. 数据集中建筑图片可视化结果

4.2. 图像代理向量数量的消融实验

本实验旨在探究不同代理向量数量对 HyperTransformer 模型性能的影响, 以确定最优的 n 值, 在保证模型性能的同时最大化计算效率。实验基于 Pavia Center 遥感数据集, 采用图像代理注意力机制, 系统地调整代理向量数量 $n \in \{4, 8, 16, 32\}$, 默认代理向量数量 $n=8$, 卷积核大小 3×3 等核心参数不变, 以

评估模型在图像锐化任务中的表现,评估指标包括计算效率指标(Params、FLOPs)和图像质量指标(CC、SAM、RMSE、ERGAS、PSNR)。

实验结果如表 2 所示,代理向量数量对模型性能具有显著影响,当 $n=8$ 时,模型在特征表达能力和计算效率之间达到了最佳平衡。 $n=4$ 的设置虽然能够将 FLOPs 降至 13.94G 并获得最小的计算复杂度,但其图像质量指标明显下降,CC 值仅为 0.985, SAM 升至 4.05, RMSE 达到 0.0092, ERGAS 为 0.55, PSNR 仅为 42.5,表明过少的图像代理向量严重限制了模型的特征提取能力。相比之下, $n=8$ 的配置在仅增加 8.8% 计算成本的情况下(FLOPs 从 13.94G 增至 15.17G),将 CC 提升至 0.993, SAM 降至 3.82, RMSE 降至 0.0084, ERGAS 降至 0.48, PSNR 提升至 43.87,实现了显著的性能改善。当进一步增加代理向量数量至 16 和 32 时,虽然 CC 值有 0.001-0.002 的微弱提升($n=16$ 时 $CC=0.994$, $n=32$ 时 $CC=0.995$),但 FLOPs 分别增加了 25.8% ($n=16$ 时 17.54G) 和 42.7% ($n=32$ 时 19.89G),这种性能提升与计算成本的增加不成正比。

可视化结果如图 6 所示, $n=4$ 时会出现部分光谱失真,特别是在建筑区域表现出部分颜色偏差,而 $n=8$ 、 $n=16$ 和 $n=32$ 的结果在视觉上难以区分。综合实验结果可以得出,当代理向量数量过少时,模型的特征表达能力受限,难以捕捉到图像中的关键特征,导致图像锐化质量下降。当代理向量数量超过 8 时,虽然特征表达能力稍有提升,但提升幅度非常有限,同时计算量显著增加,导致计算资源的浪费。代理向量数量存在饱和阈值,在遥感图像锐化任务中图像代理向量 $n=8$ 是最优选择,既能保证足够的特征表达能力,又不会引入过多计算开销。

Table 2. Performance Comparison Results of Graph Proxy Attention HyperTransformer Models with Different Numbers of Image Proxy Vectors

表 2. 使用不同图像代理向量数量的图像代理注意力 HyperTransformer 模型性能比较结果

Method	图像代理向量数量 n	Params	FLOPs	CC	SAM	RMSE	ERGAS	PSNR
HyperTransformer	$n=4$	77.9M	13938M	0.985	4.05	0.0092	0.55	42.51
	$n=8$	85.5M	15168M	0.993	3.82	0.0084	0.48	43.87
	$n=16$	98.6M	17541M	0.994	3.80	0.0082	0.47	43.91
	$n=32$	112.3M	19893M	0.995	3.79	0.0081	0.45	44.98

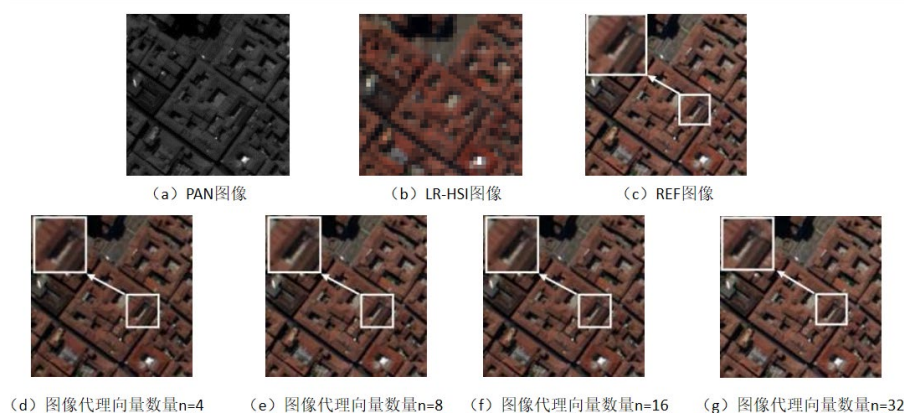


Figure 6. Visualization Results of HyperTransformer Models with Different Numbers of Image Proxy Vectors

图 6. 使用不同图像代理向量数量的 HyperTransformer 模型可视化结果

5. 结束语

本文针对遥感影像锐化任务中传统 Pansharpening 方法存在的计算复杂度高、模型复杂等问题,将本文提出的图像代理注意力机制应用在 HyperTransformer 模型中。通过在公开数据集 Pavia Center 上的实验,验证了所提出模型在计算效率和影像锐化性能方面的优越性。实验结果表明,基于图像代理注意力的 HyperTransformer 模型在参数量和浮点运算次数上显著减少,在交叉相关系数、光谱角度映射、均方根误差、峰值信噪比等评价指标上取得了不错的效果,证明了所提出的基于图像代理注意力模型的有效性。

参考文献

- [1] Wiley V, Lucas T. Computer vision and image processing: a paper review[J]. International journal of artificial intelligence research, 2018, 2(1): 29-36.
- [2] Cai W, Wei Z. Remote sensing image classification based on a cross-attention mechanism and graph convolution[J]. IEEE Geoscience and Remote Sensing Letters, 2020, 19: 1-5.
- [3] Dou W, Chen Y, Li X, et al. A general framework for component substitution image fusion: An implementation using the fast image fusion method[J]. Computers & Geosciences, 2007, 33(2): 219-228.
- [4] Vivone G, Restaino R, Licciardi G, et al. Multiresolution analysis and component substitution techniques for hyperspectral pansharpening[C]//2014 IEEE Geoscience and Remote Sensing Symposium. IEEE, 2014: 2649-2652.
- [5] Fasbender D, Radoux J, Bogaert P. Bayesian Data Fusion for Adaptable Image Pansharpening[J]. IEEE Transactions on Geoscience & Remote Sensing, 2008, 46(6): 1847-1857.
- [6] 刘蕙绮,陈应霞.融合跨阶段局部网络与注意力的遥感图像锐化方法[J].半导体光电, 2024, 45(5): 837-846.
- [7] 郑剑,刘子龙,于祥春.基于多尺度无参数注意力机制增强网络的遥感图像超分辨率重建[J].激光与光电子学进展, 2024, 61(24): 279-285.
- [8] Bandara W G C, Patel V M. Hypertransformer: A textural and spectral feature fusion transformer for pansharpening[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 1767-1777.
- [9] Plaza A, Plaza J. Recent advances in techniques for hyperspectral image processing[J]. Remote Sensing of Environment, 2009, 113(1): S110-S122.
- [10] Hoffer A, Swain P H, Davis S M. Remote sensing: The quantitative approach[J]. Biological and Physical considerations in applying Computer-Aided Analysis Techniques to Remote Sensor Data, 1978: 227-289.
- [11] Kruse F A, Lefkoff A B, Boardman J W, et al. The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data[J]. Remote sensing of environment, 1993, 44(2-3): 145-163.
- [12] Fonseca L M G, Namikawa L M, Castejon E F. Digital image processing in remote sensing[C]//2009 tutorials of the XXII Brazilian symposium on computer graphics and image processing. IEEE, 2009: 59-71.
- [13] Wald L. Quality of high resolution synthesised images: Is there a simple criterion?[C]//Third conference" Fusion of Earth data: merging point measurements, raster maps and remotely sensed images". SEE/URISCA, 2000: 99-103.
- [14] Pitas I, Venetsanopoulos A N. Order statistics in digital image processing[J]. Proceedings of the IEEE, 1992, 80(12): 1893-1921.