

DOI: 10.20189/j.cnki.CN/61-1527/E.202603005

自适应多跳图卷积与 Transformer 融合的高光谱图像分类方法

周正阳, 李文浩, 王 念, 朱良玉, 崔智高*
(火箭军工程大学, 西安 710025)

摘 要: 针对高光谱图像分类中传统图卷积网络 (graph convolutional network, GCN) 感受野受限、深层堆叠易导致过平滑以及空谱交互存在信息冗余的问题, 提出一种自适应多跳图卷积与 Transformer 融合的高光谱图像分类方法。该方法设计了自适应多跳图注意力层, 通过可学习的注意力向量对 0 至 K 跳邻域特征进行自适应加权融合, 同时构建了增强型 GCN-Transformer 融合接口, 引入通道注意力机制和一维邻域卷积模块, 从而有效压缩了节点嵌入的通道冗余并增强了光谱判别性。在 Indian Pines、Salinas 和 Pavia University 数据集上的实验结果表明, 所提方法的分类精度分别达到 96.05%、96.92% 和 96.28%, 相比经典 GTFN 方法分别提升了 2.05、0.11 和 1.14 个百分点, 表明多跳空间拓扑建模与增强光谱序列融合能有效克服高光谱“维度灾难”, 显著提升模型对不规则边界的刻画精度; 当引入空间图建模时, Transformer 模型的总体精度提升了 27.91 个百分点, 且当跳数 $K=3$ 时模型的自适应加权分类性能达到最优, 表明自适应多跳图注意力机制能在较低算力开销下拓宽空间视野并利用权重衰减规律有效降低过平滑风险; 在通道注意力模块中引入特征接口后, 模型的综合性能达到最优, 表明增强型接口能有效压缩交互阶段的信息冗余, 大幅增强模型对高度相似地物光谱的判别能力。

关键词: 高光谱图像分类; 图卷积网络; Transformer; 自适应多跳聚合; 空谱融合
中图分类号: TP751.1 **文献标志码:** A **开放科学标识码(OSID):**
文章编号: 2097-2741(2026)03-0069-11



听语音
与作者互动
聊科研

高光谱图像是一种将二维几何空间信息与一维光谱辐射信息有机结合起来特殊遥感图像, 其在每个像素点中记录了数十个乃至数百个连续窄波段的辐射信息, 蕴含丰富的地物光谱特征, 目前已被广泛应用于精准农业、矿物勘探、城市规划 and 军事监测等领域^[1-2]。然而, 受维度灾难^[3]、有限标注样本以及复杂空谱结构等因素影响, 使得高光谱图像分类一直是遥感领域内的研究热点与技术难点。

传统高光谱图像分类方法, 如支持向量机 (support vector machine, SVM)^[4]、随机森林^[5]等, 多依赖于准确的人工提取特征, 对空谱关联的建

模能力有限。随着深度学习技术的不断发展, 卷积神经网络 (convolutional neural network, CNN)、残差网络、动态卷积网络^[6]等深度学习网络为高光谱图像分类问题提供了新的解决方案。其中, CNN 凭借局部感受野和参数共享机制, 在空间特征提取方面表现出色^[1,7], 特别是其引入三维卷积模块, 进一步将光谱维度纳入到卷积核中, 实现了空谱联合特征提取^[8]。但受限于 CNN 模型的局部卷积结构, 导致模型难以有效捕获长程光谱依赖关系。近年来, 基于自注意力机制的 Transformer^[9]在自然语言处理和计算机视觉领域取得突破性进展^[10]。例如: Dosovitskiy 等^[11]提出的

收稿日期: 2026-04-03 修回日期: 2026-05-08 录用日期: 2026-05-29

第一作者: 周正阳 (2003—), 男, 硕士研究生, 主要从事高光谱图像处理研究。

*通信作者: 崔智高 (1984—), 男, 教授, 主要从事模式识别与智能控制技术研究。

引用格式: 周正阳, 李文浩, 王念, 等. 自适应多跳图卷积与 Transformer 融合的高光谱图像分类方法[J]. 火箭军工程大学学报, 2026, 40 (3): 69-79. ZHOU Zhengyang, LI Wenhao, WANG Nian, et al. Adaptive multi-hop graph convolution and Transformer fusion network for hyperspectral image classification[J]. Journal of Rocket Force University of Engineering, 2026, 40 (3): 69-79.

视觉 Transformer (vision Transformer, ViT) 将图像划分为块序列, 通过多头自注意力建模全局依赖, 开创了 Transformer 直接应用于图像处理的先河, 但该方法对空间邻域细节的捕捉能力相对较弱, 且受训练样本规模和算力资源影响较大; 此外, Hong 等^[12]提出 SpectralFormer 网络架构, 将波段划分为光谱 Token 序列并进行跨层融合, 在光谱维度分类性能优异, 但其空间建模能力薄弱^[13], 无法有效描述地物的不规则边界与空间拓扑结构。

针对 CNN 和 Transformer 各自在空间与光谱特征建模方面存在的局限性, 研究者开始探索图神经网络 (graph convolutional network, GCN) 与 Transformer 的融合方法^[14]。其中, 最具代表性的工作有光谱-空间图 Transformer 网络, 图 Transformer 与卷积融合网络等^[15-18], 从而验证了 GCN-Transformer 融合范式在高光谱图像分类中的有效性。GCN 具有对非欧几里得数据建模的天然优势^[19], 能通过聚合邻域节点信息实现空间上下文建模, 在高光谱分类问题中表现出色^[20-25]。Yang 等^[26]将 GCN 的空间建模能力与 Transformer 的光谱表达能力以串联方式深度融合, 提出了一种基于 GCN 和 Transformer 的融合网络 (GCN and Transformer fusion network, GTFN), 该模型能够利用 GCN 灵活的邻域聚合能力来捕获不规则空间上下文, 同时借助 Transformer 建模光谱维度的长程依赖, 从而实现空谱特征的互补融合。然而, GTFN 存在以下两方面问题: 一是 GCN 模块采用标准单跳谱域图卷积^[19], 每层仅聚合一阶邻域信息, 虽然可通过堆叠多层 GCN 来间接扩大感受野, 但深层堆叠会引发过平滑问题^[24,27], 从而导致节点表示趋同, 且无法根据地物空间分布特征自适应选择最优聚合范围^[28]; 二是 GCN 与 Transformer 之间的特征接口仅为简单线性投影, 未对节点嵌入进行通道维度的增强处理, 存在信息冗余和判别力不足的问题。

针对标准图卷积的过平滑问题, 多跳图卷积方法^[29]通过同时聚合多阶邻域信息, 能够在单层内实现多尺度特征捕获, 从而避免深层堆叠所带来的过平滑问题, 但现有方法大多采用固定权重的多跳聚合策略, 未能自适应调整各跳特征的贡献度。为解决上述问题, 本文在 GTFN 框架基础上提出了一种自适应多跳图卷积与 Transformer 融合的高光谱图像分类网络。该方法设计了自适应多跳图注意力层, 将标准单跳 GCN 升级为多跳自适

应聚合结构, 通过可学习的注意力向量对各跳特征进行自适应加权融合, 从而在单层内实现多尺度空间上下文捕获。此外, 该方法还构建了增强型 GCN-Transformer 融合接口, 包含通道注意力模块和一维邻域卷积模块, 从而协同提升了节点嵌入的光谱判别性。

1 算法描述

1.1 整体框架与训练过程

本文方法的整体框架如图 1 所示, 主要由数据预处理与图构建模块、自适应多跳图注意力模块、增强型 GCN-Transformer 融合接口模块以及 SpectralFormer 分类模块构成。具体流程为: 1) 给定高光谱图像 $I \in \mathbb{R}^{H \times W \times B}$ (其中, H 为图像高度, W 为图像宽度, B 为光谱波段维度), 经 PCA 降维和归一化预处理后, 为每个像素提取空间邻域块并基于高斯核构建邻接矩阵; 2) 将空间邻域块送入自适应多跳 GCN 模块, 通过 4 层自适应多跳图卷积层递归聚合 0~ K 跳邻域特征, 并经注意力加权融合与通道注意力、邻域卷积增强后输出形状为 $(B, 64, N)$ 的图感知特征; 3) 将图感知特征直接作为 SpectralFormer 的输入, 经过 Token 线性嵌入、位置编码与 Classification Token 拼接后, 由 5 层 Transformer 编码器 (含跨层自适应融合 (cross-layer adaptive fusion, CAF)) 建模全局光谱依赖, 最后经多层感知机 (multi-layer perceptron, MLP) 头输出分类结果。整个流程中自适应多跳 GCN 模块与 Transformer 端到端联合训练, 梯度从分类损失反向传播至 2 个网络, 分别由独立的 Adam 优化器更新参数。

1.2 数据预处理与图构建

本文采用跟随分块 (Follow Patch) 策略, 以充分利用分类像素的局部空间信息并降低 GCN 的计算复杂度。假设分类像素为 $P(x_{\text{center}}, y_{\text{center}})$, 以该像素为中心划定大小为 $p \times p$ 的 patch, patch 的位置通过标定其左上角像素点 $C_1(x_1, y_1)$ 和右下角像素点 $C_2(x_2, y_2)$ 的坐标来确定。像素点 C_1 横、纵坐标与中心点像素坐标的关系为

$$x_1 = \begin{cases} x_{\text{center}} - \frac{p-1}{2} & x_{\text{center}} - \frac{p-1}{2} < x_{\text{center}} < W - \frac{p-1}{2} \\ 0 & x_{\text{center}} \leq \frac{p-1}{2} \\ W - p + 1 & x_{\text{center}} \geq W - \frac{p-1}{2} \end{cases} \quad (1)$$

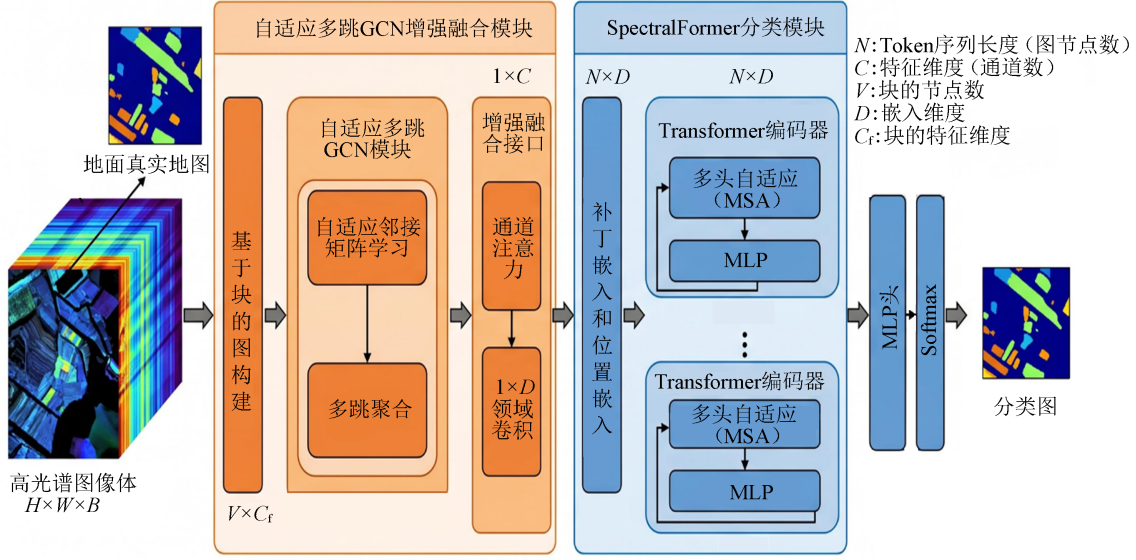


图 1 本文方法的整体框架

Fig. 1 Overall framework of the proposed method

$$y_2 = \begin{cases} y_{\text{center}} - \frac{p-1}{2} & \frac{p-1}{2} < y_{\text{center}} < H - \frac{p-1}{2} \\ 0 & y_{\text{center}} \leq \frac{p-1}{2} \\ H - p - 1 & y_{\text{center}} \geq H - \frac{p-1}{2} \end{cases} \quad (2)$$

式中: p 为 patch 块的边长, 通常为奇数。

右下角像素点 C_2 的坐标计算方式与 C_1 类似。和 ViT 中简单的 patch 划分不同, Follow Patch 扩展了图像边缘点的空间信息, 使分类像素始终位于 patch 中心, 避免了边缘化问题, 同时以局部空间信息聚合的形式去除了全局冗余信息, 使分类和计算更加高效。

对于 Follow Patch 中的 $N = p \times p$ 个像素节点, 将每个像素经 PCA 降维后的特征向量作为节点特征。节点 i 与节点 j 之间的边权定义为高斯核相似度, 表达式为

$$A_{ij} = \exp\left(-\frac{|f_i - f_j|^2}{\sigma^2}\right) \quad (3)$$

式中: A_{ij} 为节点 i 与节点 j 之间的邻接边权重; f_i 与 f_j 分别为节点 i 和节点 j 经 PCA 降维后的特征向量; $|\cdot|^2$ 表示计算 2 个特征向量间的欧氏距离平方; σ 为控制径向基函数作用范围的高斯核带宽参数。

采用对称归一化得到

$$\hat{A} = D^{-1/2} A D^{-1/2} \quad (4)$$

式中: $\hat{A}$ 为对称归一化后的邻接矩阵; A 为由式 (3) 计算所得的原始高斯邻接矩阵; D 为该空间

局部图的度矩阵, 其对角线元素定义为

$$D_{ii} = \sum_j A_{ij} \quad (5)$$

上述数据预处理步骤与 GTFN 一致, 以保证实验对比的公平性。在图构建的具体规模上, 对于每个大小为 $p \times p$ 的局部空间图, 其节点总数为 p^2 。在建立边连接时, 本文采用局部空间邻域窗口 (设窗口大小为 $l \times l$) 来稀疏化邻接矩阵。对于图中的任一中心节点, 最多只与其周围的 $l^2 - 1$ 个邻居节点建立有向连接。在本文的默认实验设置中 ($p = 9, l = 3$), 单个局部图的节点数量仅为 81, 最大节点度数为 8, 网络总边数严格限制在 648 条以内, 形成了高度稀疏的图拓扑结构。

此种跟随分块的图构建策略在理论上显著降低了图卷积的计算复杂度。若直接在整幅高光谱图像 (空间尺寸为 $H \times W$) 上构建全连接的全局图, 单层传统 GCN 的矩阵乘法计算复杂度将达到 $O(H^2 W^2 C)$ 。其中, $O(\cdot)$ 为计算复杂度函数。面对大量的像素点 (如 Indian Pines 数据集中图像空间尺寸为 145 像素 $\times$ 145 像素), 全局图的内存开销与计算成本极为高昂。而本文采用 Follow Patch 策略, 将大图拆解为局部子图, 针对单个样本的图计算复杂度被严格控制在 $O(p^4 C)$ 。由于局部块尺寸远小于整图尺寸, 该策略能够在保留目标像素关键空间上下文的同时, 从算法理论层面上将计算复杂度显著降低, 为后续的多跳图聚合提供了算力可行性。

1.3 自适应多跳图注意力层

GTFN 采用标准单跳谱域图卷积^[26], 该方法

的聚合公式为

$$\mathbf{h}^{l+1} = \varphi(\hat{\mathbf{A}}\mathbf{h}^l\mathbf{w}^l) \quad (6)$$

式中： $\mathbf{h}^{l+1}$ 与 $\mathbf{h}^l$ 分别为图卷积网络第 $l+1$ 层与第 l 层的节点特征矩阵； $\mathbf{w}^l$ 为第 l 层的可学习权重参数矩阵； $\varphi(\cdot)$ 为非线性激活函数。

GTFN每层仅利用一阶邻域信息，通过堆叠4层GCN来间接扩大感受野，并在最后一层引入光谱注意力机制来增强模型对光谱域特征的提取能力。然而，上述标准单跳GCN存在以下2个关键局限：一是固定单跳聚合对所有空间尺度一视同

仁，无法根据地物的空间分布特征自适应选择最优聚合范围；二是深层堆叠虽可间接扩大感受野，但会引发过平滑问题^[24,27]，导致节点表示趋同、判别力下降。

本文提出的自适应多跳图注意力层（adaptive multi-hop graph attention layer, AMGAL），在单层内可同时计算多跳邻域特征，并通过可学习的注意力向量进行自适应加权融合。

AMGAL的结构如图2所示，其具体操作流程包含以下3个关键步骤。

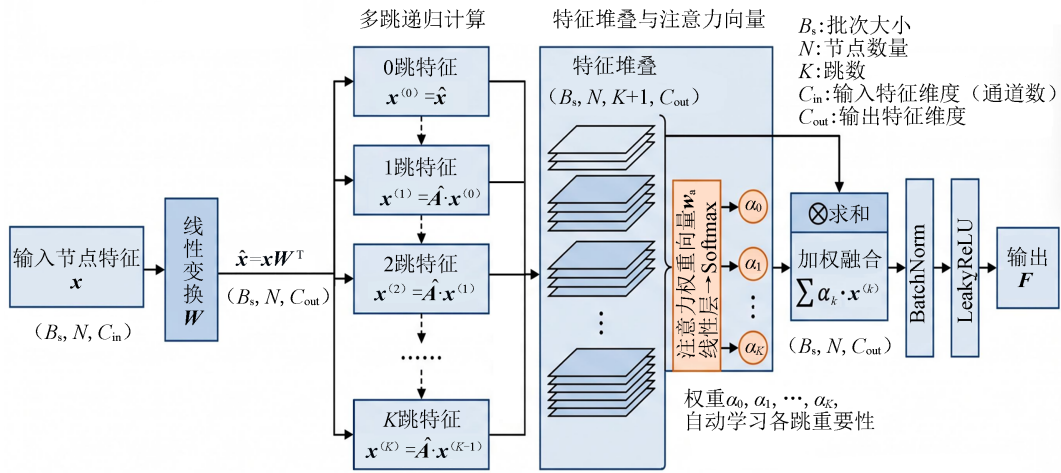


图 2 自适应多跳图注意力层结构

Fig. 2 Architecture of the adaptive multi-hop graph attention layer

1) 多跳特征递归计算。对于第 l 层输入，首先进行线性变换，然后递归计算各跳特征，表达式为

$$\mathbf{x}^{(k)} = \hat{\mathbf{A}} \cdot \mathbf{x}^{(k-1)} \quad k = 1, 2, \dots, K \quad (7)$$

式中： $\mathbf{x}^{(0)}$ 为经过初始线性变换后的节点特征； $\mathbf{x}^{(k)}$ 为第 k 跳聚合后的邻域特征矩阵，其编码以每个节点为中心、拓扑半径为 k 的邻域空间信息^[25]； K 为设定的最大多跳数。

2) 自适应注意力加权融合。将各跳特征沿新维度堆叠，通过可学习注意力向量 $\mathbf{w}_a$ 计算自适应权重，表达式为

$$\mathbf{F} = \sum_{k=0}^K \alpha^{(k)} \cdot \mathbf{x}^{(k)} \quad (8)$$

$$\alpha = \text{Softmax}(\mathbf{w}_a)$$

式中： $\mathbf{w}_a$ 为自适应学习的注意力权重向量； $\text{Softmax}(\cdot)$ 为归一化指数函数； $\alpha^{(k)}$ 为经过 Softmax 归一化后计算得出的第 k 跳特征的全局共享标量权重； $\mathbf{F}$ 为经过各阶邻域自适应加权融合后输出的增强节点特征矩阵。

3) 输出经 BatchNorm 和 LeakyReLU 激活。本文采用4层自适应多跳图注意力层，维度变化为 $B \rightarrow 128 \rightarrow 128 \rightarrow 128 \rightarrow 64$ 。

在计算复杂度方面，尽管 AMGAL 在单层内执行了 K 跳特征聚合，但由于 Follow Patch 策略已将节点数极度压缩，其计算代价完全在可控范围内。与 GTFN 通过堆叠多层物理 GCN 网络来扩大感受野相比，AMGAL 凭借极其微小的参数增量（仅 $K+1$ 个注意力权重参数）和较低的计算复杂度，实现了更为高效的多尺度感受野扩展。

AMGAL 的关键优势在于：1) 单层即可覆盖 0 至 K 跳邻域，无需深层堆叠；2) 注意力权重可自动降低高跳数特征的贡献，缓解过平滑；3) 相比堆叠多层 GCN，AMGAL 的参数增量极小。

1.4 增强型 GCN-Transformer 融合接口

在 GTFN 模型中，GCN 与 Transformer 之间的特征接口仅依赖于简单的线性投影。这种设计存在 2 个明显的局限性：一是 GCN 输出的节点嵌入特征中，并非所有通道都对分类贡献相等，直接输入会导致冗余通道干扰后续的光谱序列建模；二是简单的线性投影未能充分挖掘相邻光谱通道间潜在的局部相关性，导致特征的判别力不足。

针对上述问题，本文构建了一种增强型 GCN-Transformer 融合接口，以自适应多跳图注意力模

块输出的节点嵌入矩阵 F 作为输入, 通过依次串联通道注意力模块与邻域卷积模块, 协同提升节点

嵌入的光谱判别性。增强型 GCN-Transformer 融合接口结构如图 3 所示, 其具体流程如下。

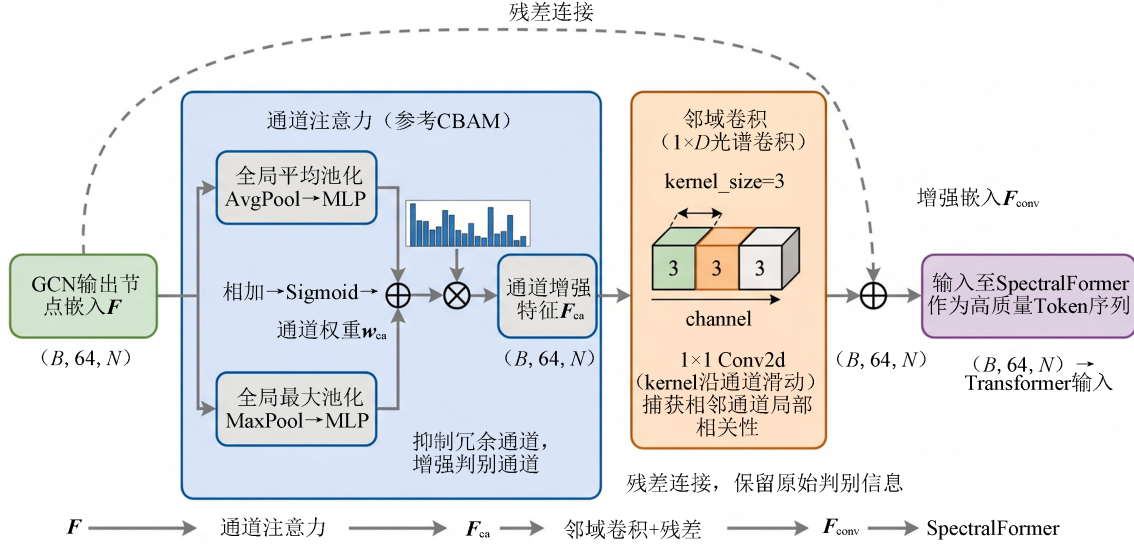


图 3 增强型 GCN-Transformer 融合接口结构

Fig. 3 Architecture of the enhanced GCN-Transformer fusion interface

1) 引入通道注意力模块。借鉴卷积注意力模块 (convolutional block attention module, CBAM) 的通道注意力机制^[30], 对 64 维节点嵌入进行通道维度的自适应增强^[28], 表达式为

$$w_{ca} = \psi(\text{MLP}(\text{GAP}(F))) + \text{MLP}(\text{GMP}(F)), F_{ca} = F \odot w_{ca} \quad (9)$$

式中: w_{ca} 为计算生成的通道注意力权重向量; F 为自适应多跳图注意力模块输出的节点嵌入矩阵; $\text{GAP}(\cdot)$ 与 $\text{GMP}(\cdot)$ 分别代表沿空间维度执行的全局平均池化与全局最大池化操作; $\text{MLP}(\cdot)$ 为共享权重的多层感知机网络; $\psi(\cdot)$ 为非线性激活函数 (具体为 Sigmoid 激活函数, 用于将权重映射至 0 到 1 之间); F_{ca} 为经过通道注意力机制加权增强后的输出特征; $\odot$ 表示矩阵的逐元素相乘。该模块通过全局统计, 自适应增强对分类贡献大的通道、抑制冗余通道。

2) 在通道维度增强之后, 引入一维邻域卷积模块以捕获相邻通道间的局部依赖关系。经邻域卷积模块处理后的最终输出特征 F_{conv} 为

$$F_{conv} = \text{Conv1D}(F_{ca}) + F_{ca} \quad (10)$$

式中: $\text{Conv1D}(\cdot)$ 表示沿光谱通道维度执行的一维卷积操作。 F_{ca} 同时作为残差连接项被引入, 以确保原始通道的判定信息不丢失。

3) 辅以 BN 层和 ReLU 函数正则化后, 融合接口输出的增强节点嵌入 $Z \in \mathbb{R}^{B \times 64 \times N}$ 将作为高质量的 Token 序列输入至 SpectralFormer 模块中。这

种增强型接口设计不仅有效压缩了信息冗余, 更显著提升了空谱联合特征的代表能力。

1.5 SpectralFormer 模块与分类输出

采用 SpectralFormer^[12] 作为 Transformer 的骨干网络, 包括线性投影层、可学习位置编码、Classification token 和多层 Transformer 编码器。CAF 机制在第 1 层和第 3 层执行跨层融合。

设 Transformer 堆叠的层数为 5, 多头注意力机制的头数为 4, 特征向量的维度为 64, 前馈神经网络隐藏层维度为 8。目标损失函数采用交叉熵损失函数, 对自适应多跳图注意力模块和 SpectralFormer 分别设置独立的 Adam 优化器, 学习率均为 5×10^{-4} , 使用 StepLR 调度 (step_size = epochs/10, $\gamma = 0.9$) 实现端到端联合训练。

2 实验验证

2.1 数据集与实验设置

本文在 3 个标准高光谱基准数据集上进行实验。其中, Indian Pines 数据集由 AVIRIS 传感器采集, 每张图像的空间尺寸为 145 像素 $\times$ 145 像素, 涵盖 200 个波段、16 个地物类别和 10 249 个标注样本; Salinas 数据集由 AVIRIS 传感器采集, 每张图像的空间尺寸为 512 像素 $\times$ 217 像素, 包含 204 个波段、16 个地物类别和 54 129 个标注样本; Pavia University 数据集由 ROSIS 传感器采集, 每

张图像的空间尺寸为 610 像素 $\times$ 340 像素, 包含 103 个波段, 9 个地物类别和 42 776 个标注样本。

为验证本文方法的有效性, 选择多个代表性方法进行对比实验, 对比方法包括 SVM-RBF^[4]、3D-CNN^[8]、DFFN^[31]、RSSAN^[32]、DBDA^[33]、M-DGCN^[22]、WFCG^[34]、AMGCFN^[35]、SpectralFormer^[12]、SSFTT^[36]、GAHT^[37] 以及 GTFN^[26]。所有实验在 NVIDIA GPU 上运行, 采用 PyTorch 框架, 随机种子固定为 0。PCA 降维至 70 维, patch=9, band_patches=1, n_gcn=21, batch_size=64, epochs=200, weight_decay=0。评价指标采用总体精度 (overall accuracy, OA)、平均精度 (average accuracy, AA) 和 Kappa 系数。其中, OA 衡量模型在整个数据集上的宏观分类正确率, OA

越高, 表示模型整体预测正确的样本比例越高; AA 衡量模型对每一个单独类别的平均分类能力, AA 越高, 表示模型对每一个地物类别的识别能力越均衡; Kappa 系数用于不平衡数据集的分类一致性检验, 该系数越高, 表示模型的预测结果与真实标签的一致性越高, 分类结果更可靠。

2.2 定性分析

为直观验证本文方法的空谱联合特征建模能力与实际分类性能, 本节基于 Indian Pines 数据集, 选取 SpectralFormer、GTFN 作为对比方法, 以数据集官方地物真值为标准, 从区域一致性、边界刻画精度、噪声抑制能力、易混淆类别区分度 4 个维度开展可视化定性分析。各方法在 Indian Pines 数据集的可视化分类结果如图 4 所示。

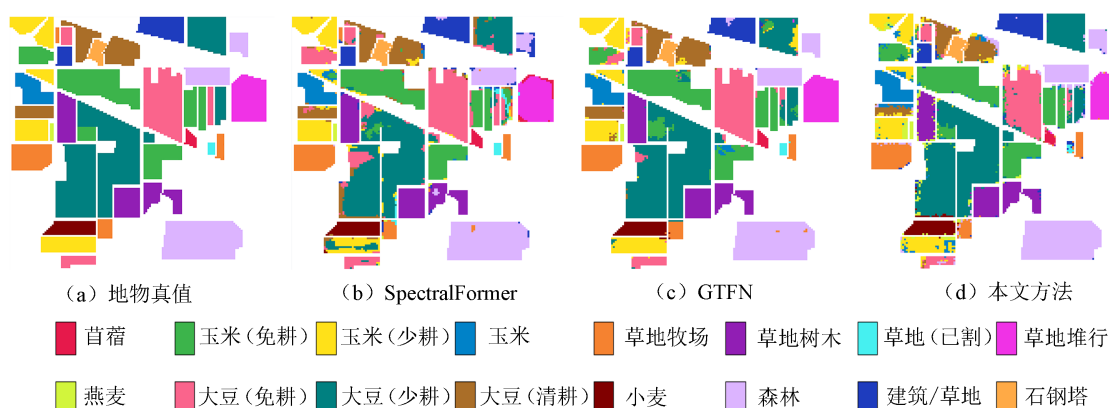


图 4 不同方法在 Indian Pines 数据集的分类结果

Fig. 4 Classification results of different methods on the Indian Pines dataset

SpectralFormer 作为纯 Transformer 架构的代表性算法, 仅聚焦光谱维度的长程依赖进行建模, 缺乏对空间拓扑结构与不规则地物边界的有效表征能力。可视化结果显示, 该方法的分类图中存在大量椒盐噪声与孤立错分像素, 同质区域内部类别混淆严重, 地物边界与真值偏差显著; 对燕麦、小麦等小尺寸地物, 以及免耕/少耕玉米、免耕/少耕大豆等光谱特征高度相似的易混淆类别, 区分能力极差, 分类结果与真实地物分布匹配度低。

GTFN 通过单跳 GCN 与 Transformer 的串联融合, 实现了对空谱特征的联合建模, 较 SpectralFormer 方法显著减少了分类噪声, 地物区域完整性得到明显提升, 从而验证了空间拓扑建模对高光谱分类的增益作用。但该方法仍存在两处核心局限: 一是标准单跳 GCN 的感受野固定, 无法自适应捕获多尺度空间上下文信息, 在地物交界区域存在明显的边界像素错分, 小目标与背景的过渡区域类别混淆问题突出; 二是 GCN 与 Trans-

former 间仅采用简单线性投影作为特征接口, 未能充分挖掘节点嵌入的通道间相关性, 同质区域内部仍存在零散错分像素, 边界刻画精度与区域一致性仍有明显提升空间。

本文方法的分类结果与地物真值具有高度一致性, 分类图中椒盐噪声较其他方法更少, 各类地物区域完整, 不规则边界与真值贴合度高, 充分验证了本文改进模型的有效性。在空间维度, 本文设计的自适应多跳图注意力层, 通过可学习注意力向量实现 0~K 跳邻域特征的自适应加权融合, 在单层内完成多尺度空间上下文建模, 既有效扩展了感受野, 又通过权重自适应调节缓解了过平滑问题。无论是大面积农田、林地等同质区域, 还是钢结构塔、干草堆等小尺寸地物, 本文方法均能实现精准的空间结构表征, 解决了对比方法边界错分、小目标识别精度低的问题。在光谱维度, 本文构建的增强型 GCN-Transformer 融合接口, 通过通道注意力机制自适应增强有效特

征通道、抑制冗余通道，同时借助一维邻域卷积挖掘相邻光谱通道的局部相关性，大幅提升了节点嵌入的光谱判别性，对光谱特征高度相似的易混淆地物类别实现了精准区分，进一步提升了分

类结果的边界精度与区域一致性。

2.3 定量比较

表 1 给出了不同高光谱图像分类方法在 3 个数据集上的总体分类性能对比。

表 1 不同数据集上的总体分类性能对比
Table 1 Comparison of overall classification performance on different datasets

方法	Indian Pines			Salinas			Pavia University		
	OA/%	AA/%	Kappa	OA/%	AA/%	Kappa	OA/%	AA/%	Kappa
SVM-RBF	66.96	62.30	0.628	87.31	92.01	0.859	79.74	78.32	0.740
3D-CNN	65.83	79.12	0.614	87.59	92.96	0.862	77.87	83.90	0.718
DFFN	83.61	91.96	0.815	91.35	95.22	0.904	78.00	84.41	0.719
RSSAN	73.78	80.94	0.705	89.66	94.79	0.885	82.61	83.33	0.772
DBDA	87.06	93.13	0.853	92.18	96.14	0.913	92.23	91.98	0.898
MDGCN	86.77	93.97	0.850	92.20	94.63	0.913	92.10	94.38	0.897
WFCG	87.68	86.12	0.860	92.68	96.83	0.919	92.98	91.77	0.908
AMGCFN	91.25	95.21	0.900	94.87	97.88	0.943	93.12	93.55	0.910
SpectralFormer	65.24	76.07	0.611	88.80	93.09	0.876	76.03	79.91	0.694
SSFTT	82.96	91.55	0.807	90.67	95.34	0.896	91.08	92.34	0.883
GAHT	85.96	92.96	0.841	94.21	97.37	0.936	92.81	94.57	0.906
GTFN	94.00	96.33	0.932	96.81	98.30	0.965	95.14	94.58	0.936
本文方法	96.05	95.67	0.955	96.92	98.12	0.966	96.28	94.09	0.951

注：粗体表示每列最优值。

由表 1 可知，本文模型在 India Pines、Salinas 以及 Pavia University 基准数据集上的 OA 值分别为 96.05%、96.92% 和 96.28%，Kappa 系数分别为 0.955、0.966 和 0.951，整体分类性能均优于其他对比方法。其中，本文模型的 OA 值较次优方法 GTFN 分别提升了 2.05、0.11 和 1.14 个百分点，表明多跳空间拓扑建模与增强光谱序列融合的协同效应，能有效克服高光谱“维度灾难”，显著提升复杂地物的分类精度；同时，本文模型在 3 个数据集上的 Kappa 系数较次优方法 GTFN 分别提升了 0.023、0.001 和 0.015，表明本文方法的预测结果与真实标签具备更高的一致性，分类结果更具可靠性。

尽管本文方法在 3 个数据集上的 OA 和 Kappa 系数均显著高于其他几种对比模型，但其在全部数据集上的 AA 值却略低于 GTFN 方法。这主要是因为本文引入的自适应多跳图注意力机制在大幅扩展感受野、提升大面积同质区域分类精度的同时，多阶广域邻域信息的聚合不可避免地极少数且样本极度不平衡的小类别（如 Indian Pines

中的燕麦类）产生了一定程度的平滑干扰，导致这些少数类别的分类召回率出现微弱下降，进而对宏观平均精度造成轻微影响。综合来看，虽然本文方法较次优方法 GTFN 的 AA 指标稍有下降，但其 OA 和 Kappa 性能指标均有明显提升，仍能证明本文模型在整体分类性能和面对复杂地物场景时具备更强的优越性与鲁棒性。

2.4 消融实验

为验证各模块的有效性，在 Indian Pines 数据集上设计消融实验，设置变体①为 Transformer (SpectralFormer)，变体②为单跳 GCN+Transformer，变体③为多跳 GCN（等权）+简单投影+Transformer，变体④为多跳 GCN（注意力）+简单投影+Transformer，变体⑤为本文方法，结果如表 2 所示。

由表 2 可得，各模块均对最终性能有正向贡献。①→②（OA+27.91%）：引入空间图建模带来显著增益，验证了空间拓扑信息对高光谱分类的重要性。②→③（OA+1.15%）：将单跳 GCN 升级为等权多跳融合，扩展了单层感受野，证明多尺

度邻域信息的引入具有正向作用。③→④ (OA+0.95%)：在多跳基础上引入注意力加权，进一步提升性能，说明自适应加权优于等权融合。④→⑤ (OA+0.80%)：加入通道注意力模块增强信息通道响应，模型 OA 进一步提升至 96.05%，综合性能相对最优。

表 2 消融实验结果 (Indian Pines)
Table 2 Results of the ablation experiment (Indian Pines)

变体配置	OA/%	AA/%	Kappa
①Transformer	65.24	76.07	0.611
②单跳 GCN+Transformer	93.15	92.50	0.923
③多跳 GCN(等权)+简单投影+Transformer	94.30	93.85	0.936
④多跳 GCN(注意力)+简单投影+Transformer	95.25	94.90	0.946
⑤本文方法	96.05	95.67	0.955

为探究最大跳数 K 对分类性能的影响，在 Indian Pines 数据集上进行了对比实验，结果如表 3 所示。

表 3 不同最大跳数 K 的性能对比 (Indian Pines)
Table 3 Performance comparison under different maximum hop counts K (Indian Pines)

K 值	OA/%	Kappa
1	92.14	0.926
2	95.26	0.934
3	96.05	0.955
4	95.88	0.953
5	95.52	0.949

由表 3 可得，当 $K=1$ 时，自适应多跳图注意力层退化为标准单跳 GCN，OA 值为 92.14%，此时与 GTFN 中的 GCN 模块等价。随着 K 值增大，OA 逐步提升， $K=2$ 时模型 OA 为 95.26%， $K=3$ 时达到最大值 96.05%。然而，继续增大 K 至 4 和 5 时，OA 分别小幅回落至 95.88% 和 95.52%。这一趋势表明，适度扩展感受野能够有效捕获多尺度空间上下文信息，但过大的跳数会引入过多平滑信号，即便注意力机制对高跳特征施加了抑制，仍难以完全消除其负面影响。对 $K=3$ 时的注意力权重进行观察发现，各跳权重呈单调递减分布，即零跳（自身特征）权重最大（0.254 184），三跳特征权重最小（0.248 246），这与“近邻像素对分类贡献更大”的先验认知一致。

2.5 架构分析总结

为进一步阐明本文方法的性能增益来源，本节从空间拓扑演化与特征交互机制 2 个维度，深层解构其与基线模型 GTFN 的架构差异，两者架构如表 4 所示。

1) 空间拓扑建模的范式转换。GTFN 依赖于标准单跳谱域图卷积的层级堆叠来被动扩张空间感受野。这种“深层解”在算力开销随层数线性增长的同时，极易诱发节点表征趋同，导致过平滑表征，从而削弱模型对地物不规则边界的判别精度。相比之下，本文提出的自适应多跳图注意力机制，将“网络深度”转化为“拓扑广度”。通过单层内递归计算 $0 \sim K$ 跳邻域特征，并在非欧几何空间引入可学习的注意力向量，实现了多尺度上下文的自适应加权融合。该机制利用权重的动态衰减规律，在机理上中和了高跳数所引发的信息泛化风险，在不增加网络深度的前提下实现了复杂边界的高精度刻画。

表 4 本文方法与 GTFN 的架构对比
Table 4 Architectural comparison between the proposed method and GTFN

维度	GTFN	本文方法
GCN 类型	标准单跳 GCN	自适应多跳注意力 GCN
单层感受野	1 跳	0 至 K 跳 ($K=3$)
多尺度策略	堆叠多层 GCN	单层内多跳融合
过平滑缓解	未显式处理	注意力加权自动抑制
特征接口	简单线性投影	通道注意力+邻域卷积
通道增强	无	CBAM 式通道注意力

2) 特征接口的深度提纯。在跨模块特征交互阶段，GTFN 仅采用全连接线性投影作为映射接口，忽略了节点嵌入的通道间相关性，导致空谱交互阶段存在显著的信息冗余。本文方法突破了传统的简单线性级联范式，构建了串联通道注意力模块与一维邻域卷积模块的增强型融合接口。前者通过全局统计量挖掘非欧节点嵌入的通道响应差异，自适应抑制冗余维度，后者沿光谱序列维度捕获局域依赖关系，并辅以残差连接引入原特征锚点。这种“深度提纯与局域重组”的协同效应，极大增强了模型对高度相似同质光谱地物的微弱谱线差异判别力。

3) 从参数开销与泛化鲁棒性来看，本文方法仅引入了极少的标量权重与轻量卷积参数，却在三大标准基准数据集上均斩获了显著的分类增益。

3 结论

为解决高光谱图像分类中空间感受野受限、深层网络易出现过度平滑以及空谱特征交互冗余等问题, 提出一种自适应多跳图卷积与 Transformer 融合的高光谱图像分类方法。该方法设计自适应多跳图注意力层, 并构建了集成通道注意力与一维邻域卷积的增强型特征融合接口, 以实现更深层次的空谱联合特征提取。通过在基准测试集上的对比实验, 得出如下结论。

1) 针对图卷积感受野受限与深层堆叠所引发的过平滑问题, 自适应多跳图注意力机制对多阶拓扑邻域进行了动态赋权。消融实验结果显示, 当最大跳数 $K=3$ 时, 模型在 Indian Pines 数据集上的总体精度达到最优的 96.05%, 相较于退化为标准单跳 GCN 时提升了 3.91 个百分点。同时, 网络自发学习到的注意力权重呈现出零跳自身特征权重最大、三跳特征权重最小的单调递减分布, 表明该机制能够在不增加网络深度的前提下, 利用权重的自动衰减规律中和过平滑风险, 进而提取多尺度空间上下文特征。

2) 针对 GCN 与 Transformer 特征交互阶段的信息冗余现象, 所构建的增强型复合特征融合接口能够对光谱序列进行提纯与局域特征重组。实验结果显示, 在多跳 GCN 的基础上, 将传统的简单线性投影替换为引入了通道注意力与邻域卷积的增强型接口后, 模型的总体精度从 95.25% 进一步提升至 96.05%, 验证了增强型接口通过自适应筛选通道并捕获局部相关性, 增强了模型对高度相似光谱地物的判别能力。

3) 多跳空间拓扑建模与增强光谱序列融合的协同架构, 为复杂地物场景分类提供了有效支持。在 Indian Pines、Salinas 和 Pavia University 基准数据集上, 本文方法的总体分类精度分别达到了 96.05%、96.92% 和 96.28%, 相较于次优的 GTFN 融合架构方法分别提升了 2.05、0.11 和 1.14 个百分点; 对应的 Kappa 系数也分别提升了 0.023、0.001 和 0.015, 证明了本文方法在处理高光谱图像复杂空谱耦合结构时具有较好的分类一致性与鲁棒性。

参考文献

- [1] Hu Wei, Huang Yangyu, Wei Li, et al. Deep convolutional neural networks for hyperspectral image classification[J]. *Journal of Sensors*, 2015(1): 258619.
- [2] Li Shutao, Song Weiwei, Fang Leyuan, et al. Deep learning for hyperspectral image classification: an overview[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(9): 6690-6709.
- [3] Hughes G. On the mean accuracy of statistical pattern recognizers[J]. *IEEE Transactions on Information Theory*, 1968, 14(1): 55-63.
- [4] Melgani F, Bruzzone L. Classification of hyperspectral remote sensing images with support vector machines[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2004, 42(8): 1778-1790.
- [5] Amini S, Homayouni S, Safari A, et al. Object-based classification of hyperspectral data using random forest algorithm[J]. *Geo-spatial Information Science*, 2018, 21(2): 127-138.
- [6] 迟骋, 李向阳, 张志利, 等. 基于动态卷积的混合模型性能评估方法[J]. *火箭军工程大学学报*, 2024, 38(5): 53-58, 80.
Chi Cheng, Li Xiangyang, Zhang Zhili, et al. Performance evaluation method of hybrid model based on dynamic convolution[J]. *Journal of Rocket Force University of Engineering*, 2024, 38(5): 53-58, 80.
- [7] 李邦杰, 武健, 张大巧, 等. 基于卷积神经网络的景象匹配区适配性筛选方法[J]. *火箭军工程大学学报*, 2024, 38(3): 45-50, 59.
Li Bangjie, Wu Jian, Zhang Daqiao, et al. A CNN-based scene matching adaptability screening method[J]. *Journal of Rocket Force University of Engineering*, 2024, 38(3): 45-50, 59.
- [8] Ben Hamida A, Benoit A, Lambert P, et al. 3-D deep learning approach for remote sensing image classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(8): 4420-4434.
- [9] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//*Advances in Neural Information Processing Systems*. Long Beach: Curran Associates, 2017: 5998-6008.
- [10] 赵晶钰, 李敏, 陈谢发, 等. 融合注意力机制与全局信息的 SAR 舰船目标检测算法[J]. *火箭军工程大学学报*, 2024, 38(4): 40-46.
Zhao Jingyu, Li Min, Chen Xiefa, et al. SAR ship target detection algorithm combining attention mechanism and global information[J]. *Journal of Rocket Force University of Engineering*, 2024, 38(4): 40-46.
- [11] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: transformers for image recognition at scale[EB/OL]. (2021-06-03) [2026-03-27]. <https://doi.org/10.48550/arXiv.2010>.

- 11929.
- [12] Hong Danfeng, Han Zhu, Yao Jing, et al. SpectralFormer: rethinking hyperspectral image classification with transformers[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5518615.
 - [13] Yang Xiaofei, Cao Weijia, Lu Yao, et al. Hyperspectral image transformer classification networks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5528715.
 - [14] Ahmad M, Shabbir S, Roy S K, et al. Hyperspectral image classification: traditional to deep models: a survey for future prospects[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2022, 15: 968-999.
 - [15] Shi Haotian, Luo Zihang, Ma Yiyang, et al. SS-GTN: spectral spatial graph transformer network for hyperspectral image classification[J]. Remote Sensing, 2026, 18(2): 199.
 - [16] Zhao Xiaofeng, Ma Junyi, Wang Lei, et al. GTCFN: a graph-based transformer and convolution fusion network for hyperspectral image classification [J]. IEEE Transactions on Geoscience and Remote Sensing, 2025, 63: 5530819.
 - [17] Chen Yuanjiang, Yang Bin, Bian Taonian, et al. Dual-scale graph transformer with spectral-spatial convolution for hyperspectral image classification[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2026, 19: 12430-12444.
 - [18] Zhang Zilin, Yu Haoyang, Jiang Zhaohui, et al. CentralGraphFormer: center-guided spectral spatial graph transformer for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2026, 64: 5507918.
 - [19] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks[EB/OL]. (2017-02-22)[2026-03-27]. <https://doi.org/10.48550/arXiv.1609.02907>.
 - [20] Qin Anyong, Shang Zhaowei, Tian Jinyu, et al. Spectral spatial graph convolutional networks for semisupervised hyperspectral image classification[J]. IEEE Geoscience and Remote Sensing Letters, 2019, 16(2): 241-245.
 - [21] Hong Danfeng, Gao Lianru, Yao Jing, et al. Graph convolutional networks for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(7): 5966-5978.
 - [22] Wan Sheng, Gong Chen, Zhong Ping, et al. Multiscale dynamic graph convolutional network for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(5): 3162-3177.
 - [23] 赵晓枫, 段永屹, 马鉴懿, 等. 多尺度卷积与超像素级图卷积融合的高光谱图像分类网络[J]. 火箭军工程大学学报, 2025, 39(5): 94-106.
 - Zhao Xiaofeng, Duan Yongyi, Ma Junyi, et al. Multiscale-convolutional and superpixel-level graph convolutional fusion network for hyperspectral image classification[J]. Journal of Rocket Force University of Engineering, 2025, 39(5): 94-106.
 - [24] Ding Yao, Zhao Xiaofeng, Zhang Zhili, et al. Semi-supervised locality preserving dense graph neural network with ARMA filters and context-aware learning for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-12.
 - [25] Jing Haoyu, Wu Sensen, Zhang Laifu, et al. Aggregative and contrastive dual-view graph attention network for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 5530817.
 - [26] Yang Aitao, Li Min, Ding Yao, et al. GTFN: GCN and transformer fusion network with spatial-spectral features for hyperspectral image classification [J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 6600115.
 - [27] Li Qimai, Han Zhichao, Wu Xiaoming. Deeper insights into graph convolutional networks for semi-supervised learning[C]//Proceedings of the AAAI Conference on Artificial Intelligence. New Orleans: AAAI Press, 2018, 32(1): 3538-3545.
 - [28] Luo Xingwen, Tu Bing, Liu Bo, et al. Hyperspectral image classification via dynamic adaptive graph convolutional network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2026, 64: 5507818.
 - [29] Zhang Hao, Ku Junhua, Zhao Jie. Multi-scale graph wavelet convolutional network for hyperspectral image classification[J]. Frontiers in Remote Sensing, 2025, 6: 1637820.
 - [30] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of the 15th European Conference on Computer Vision. Cham: Springer, 2018: 3-19.
 - [31] Song Weiwei, Li Shutao, Fang Leyuan, et al. Hyperspectral image classification with deep feature fusion network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(6): 3173-3184.
 - [32] Zhu Minghao, Jiao Licheng, Liu Fang, et al. Residual spectral-spatial attention network for hyper-

- spectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(1): 449–462.
- [33] Li Rui, Zheng Shunyi, Duan Chenxi, et al. Classification of hyperspectral image based on double-branch dual-attention mechanism network[J]. Remote Sensing, 2020, 12(3): 582.
- [34] Dong Yanni, Liu Quanwei, Du Bo, et al. Weighted feature fusion of convolutional neural network and graph attention network for hyperspectral image classification[J]. IEEE Transactions on Image Processing, 2022, 31: 1559–1572.
- [35] Zhou Hao, Luo Fulin, Zhuang Huiping, et al. Attention multihop graph and multiscale convolutional fusion network for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 5508614.
- [36] Sun Le, Zhao Guangrui, Zheng Yuhui, et al. Spectral-spatial feature tokenization transformer for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5522214.
- [37] Mei Shaohui, Song Chao, Ma Mingyang, et al. Hyperspectral image classification using group-aware hierarchical transformer[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5539014.

Adaptive Multi-Hop Graph Convolution and Transformer Fusion Network for Hyperspectral Image Classification

ZHOU Zhengyang, LI Wenhao, WANG Nian,
ZHU Liangyu, CUI Zhigao*

(Rocket Force University of Engineering, Xi'an 710025, China)

ABSTRACT: To address the issues of limited receptive fields in conventional graph convolutional networks (GCNs), over-smoothing caused by deep network stacking, and information redundancy in spatial-spectral interactions for hyperspectral image classification, an adaptive multi-hop graph convolution and Transformer fusion network was proposed. An adaptive multi-hop graph attention layer was designed to achieve the adaptive weighted fusion of 0- to K -hop neighborhood features via learnable attention vectors. Concurrently, an enhanced GCN-Transformer fusion interface was constructed by integrating a channel attention mechanism and a one-dimensional neighborhood convolution module, thereby effectively compressing the channel redundancy of the node embeddings and elevating the spectral discriminability. The results on the Indian Pines, Salinas, and Pavia University datasets demonstrate that the proposed method achieves classification accuracies of 96.05%, 96.92%, and 96.28%, outperforming the classic GTFN method by 2.05, 0.11, and 1.14 percentage points, respectively. These results confirm that the synergistic effect of multi-hop spatial topology modeling and enhanced spectral sequence fusion can effectively overcome the “curse of dimensionality” in hyperspectral and significantly improve the delineation accuracy of irregular boundaries. The introduction of spatial graph modeling boosts the overall accuracy by 27.91 percentage points, with adaptive weighting achieving optimal performance at a hop count of $K=3$, indicating that the adaptive multi-hop graph attention mechanism can broaden the spatial perspectives with lower computational overhead and reduce over-smoothing risks by utilizing learned weight decay regularities. By incorporating a feature interface into the channel attention module, the model yields the best comprehensive performance, demonstrating that the enhanced interface effectively compresses information redundancy during the interaction phase, which strengthens the discriminative capability for highly similar spectral land covers.

KEYWORDS: hyperspectral image classification; graph convolutional network; Transformer; adaptive multi-hop aggregation; spectral-spatial fusion

(编辑: 刘建民)