

基于交叉注意力融合多光谱与SAR数据的森林分类

谢一帆, 贾紫晗, 张晓丽

1. 北京林业大学 林木资源高效生产全国重点实验室, 北京 100083;
2. 北京林业大学 精准林业北京市重点实验室, 北京 100083;
3. 北京林业大学 森林培育与保护教育部重点实验室, 北京 100083

摘要: 遥感技术凭借其广覆盖、高时效和多维信息获取能力, 已成为林业调查的重要工具。多光谱遥感影像具备较高的空间与光谱分辨率, 能够有效捕捉不同地物之间的光谱差异, 合成孔径雷达数据则提供了稳定的地表结构信息与纹理特征, 可作为光谱特征的重要补充。然而, 主被动遥感数据在模态结构与信息表达上的差异, 常导致融合效果有限, 影响分类精度。针对这一问题, 本文以云南省普洱市部分区域为研究区, 选取3类森林树种(组)(思茅松、桉树、栎类)与一种经济林地类型(茶园)及其他3类地物作为分类对象, 构建了一个面向小样本条件下的融合主被动遥感数据的森林分类深度学习网络。该方法融合哨兵二号(Sentinel-2)多光谱影像与哨兵一号(Sentinel-1) SAR影像数据, 提出一种以多光谱影像为主导的交叉注意力融合网络, 并引入可学习Dropout比率的正则化门控机制, 实现自注意力与交叉注意力特征的动态融合调节。在融合过程中, 模型通过自注意力机制提取多光谱的显著特征, 结合交叉注意力引导其对SAR关键区域的响应, 在特征融合阶段动态控制各模态信息的贡献度。实验结果表明, 所提方法在多个融合策略与融合层级的对比中取得最优性能, 总体分类精度达到95.24%, 桉树、茶园、栎类和思茅松的分类精度分别为96.78%、94.07%、91.73%、92.90%。本研究验证了交叉注意力机制在主被动遥感信息协同建模中的有效性, 为复杂环境下的多源遥感森林分类提供了可行思路与技术支持。

关键词: Sentinel-2, Sentinel-1, 交叉注意力, 特征融合, 森林分类, 原型网络, Concrete Dropout

中图分类号: S757/P237/TP79/P2

引用格式: 谢一帆, 贾紫晗, 张晓丽. 2025. 基于交叉注意力融合多光谱与SAR数据的森林分类. 遥感学报, 29(10): 2989-3005

Xie Y F, Jia Z H and Zhang X L. 2025. Forest classification via cross-attention fusion of multispectral and SAR remote sensing data. National Remote Sensing Bulletin, 29(10): 2989-3005 [DOI: 10.11834/jrs.20255159]

1 引言

森林生态系统在维护全球生态平衡、应对气候变化和促进生物多样性保护中具有重要作用(Mori等, 2017)。在森林资源调查与监测任务中, 森林树种分布信息为估算森林生物量与碳储量(Tian等, 2022)、监测森林病虫害(Pimentel等, 2020)、检测外来入侵物种(Arasumani等, 2021)等提供了数据基底支持。因此, 如何在大范围区域内快速、准确获取森林类型和树种分布信息, 成为当前林业遥感研究的重点方向。

传统的森林资源清查方式获取主要依赖地面

调查, 在大尺度区域的应用中存在效率低、成本高等问题(Smith, 2002)。遥感技术凭借其大范围、周期性和高时效性的特点, 为森林识别提供了高效替代方案(White等, 2016)。近年来, 多光谱遥感MSI(Multispectral Image)在林业遥感中的应用日益广泛, 尤其是欧洲航天局推出的Sentinel-2卫星, 凭借获取便捷、光谱信息丰富、重访周期短等优势, 被广泛用于森林类型与树种(或树种组)分类(Phiri等, 2020; Persson等, 2018; Immitzer等, 2019; 夏清等, 2023)。Sentinel-2在植被红边与近红外波段上对植被类型、健康状况及树种差异的刻画能力显著(Hościło和Lewandowska,

收稿日期: 2025-05-20; 预印本: 2025-07-03

基金项目: 天空地一体化森林资源监测技术示范(编号: 2023YFD2201700); 近地—地基主被动多模态遥感联合的单木参数高效提取研究(编号: 32171779)

第一作者简介: 谢一帆, 研究方向为林业遥感与信息技术。E-mail: 18801495791@163.com

通信作者简介: 张晓丽, 研究方向为林业遥感与信息技术。E-mail: zhangxl@bjfu.edu.cn

2019; Grabska 等, 2019; 张琍 等, 2023), 为开展精细化森林分类奠定了坚实的数据基础。然而, 单一依赖光学数据进行树种识别仍面临挑战, 部分树种在光谱响应上存在高度相似性, 特别是在季节变化不显著或冠层结构差异不明显的森林中, 易造成分类混淆。合成孔径雷达 SAR (Synthetic Aperture Radar) 影像作为主动遥感的代表, 能够敏锐地捕捉不同树种的表面和结构特征, 并通过多种极化方式的应用, 有效区分树种之间的细微差异, 如树冠形态、枝叶结构以及树干特征等, 其独特的探测机制在树种间光谱特征相似的情况下, 能够提供另一角度的结构特征 (Dubois 等, 2020), 避免了光学分类时“同谱异物”现象。在森林分类中, SAR 影像通过其纹理、散射机制等结构性特征可有效弥补光学数据在信息维度上的不足。因此, SAR 影像作为光学数据的有益补充, 为提升分类的全面性与鲁棒性提供了数据支持。在此背景下, 主被动遥感数据的融合成为当前遥感森林分类的重要发展方向 (Adrian 等, 2021; Liu 等, 2018; 李晓雅 等, 2023)。通过融合多光谱影像与 SAR 影像信息, 可实现对树种的多维度综合表征, 增强分类模型对地物异质性的判别能力 (Bjerreskov 等, 2021)。然而, 在实际应用中, 由于两类数据在物理属性、尺度特征及分布形式上的显著差异, 如何在保证特征完整性的前提下进行有效融合, 仍面临挑战。因此, 如何高效融合光学与 SAR 数据的空谱特征, 丰富森林观测信息, 挖掘不同类型树种差异性, 是提升区域尺度森林分类效果的关键问题。

根据数据融合的层次与方式, 遥感数据融合方法可分为像素级融合、特征级融合和决策级融合 (Gao 等, 2020), 分别对应深度学习模型的数据输入、特征提取和分类决策等 3 个阶段。其中, 像素级融合是指在数据预处理阶段对不同源的遥感影像进行融合, 直接将融合后的数据作为模型的输入, 常用的融合方法包括多通道拼接、主成分分析融合 (Shahdoosti 和 Ghassemian, 2012)、小波变换融合 (Li 等, 2017)、Gram-Schmidt 变换融合 (Chen 等, 2017) 等。特征级融合指在深度学习模型中对来自不同数据源的中间特征进行融合, 相比像素级融合, 特征级融合直接融合经过神经网络编码器提取的高层语义特征, 主要的特征级

融合方法包括特征拼接、加权融合 (Tang 等, 2023)、注意力机制 (Dai 等, 2021) 和融合网络结构 (Wang 等, 2020) 等。决策级融合应用于分类决策阶段, 指在多数数据源完成特征提取和初步分类之后, 对其输出的分类结果进行整合以生成最终的分类结果, 包括多数投票法、加权投票法、Dempster-Shafer 证据理论 (Shafer, 1992) 融合方法、集成学习法 (Ghassemian, 2016) 等。在这 3 种层级的融合中, 像素级融合可以保留大部分的信息, 但计算量较大, 且对信息噪声较为敏感 (Chen 等, 2020)。决策级融合在信息整合过程中更为灵活, 但该方法依赖于每个数据源的最终分类结果, 可能忽略数据间的深层协同信息, 限制了其对复杂地物的识别能力 (Hu 等, 2021)。相比之下, 特征融合可以减少数据冗余, 且支持在深度学习模型中进行端到端的可推理融合, 是当前多源遥感数据融合中的研究热点之一 (Ghassemian, 2016)。

在特征级融合方法中, Two-Branch CNN (Xu 等, 2018) 是一种经典的特征融合方法, 即通过双分支特征提取网络分别提取高光谱影像和激光雷达的空谱特征, 在特征融合阶段将提取的深度特征进行拼接融合实现分类。尽管该方法实现了模态特征的整合, 但其简单的特征拼接策略导致特征维度膨胀, 增加计算负担, 并且缺乏对特征重要性的自适应建模能力, 限制了其性能表现。近年来, 注意力机制 (Bahdanau 等, 2016) 因其对特征之间依赖关系建模的能力被广泛引入融合任务中。注意力机制可以对特征赋予不同的权重, 从而引导网络更关注具有判别力的模态或区域, 实现更高效准确的特征融合。例如, AMM-FuseNet (Ma 等, 2022) 结合通道注意力机制和密集连接的空洞空间金字塔池化模块融合多源数据, 实现了有效的信息融合。ADMOHM (Zheng 等, 2024) 模型提出了一个由 CBAM 注意力增强的 2D-CNN 网络用于融合多源遥感数据。然而, 上述的方法本质上是对堆叠后的原始波段进行特征融合, 忽略了不同模态下数据特征的差异, 其机制仍然属于“单模态注意力”, 难以捕捉不同模态之间的协同关系, 无法实现真正意义上的跨模态互补。

交叉注意力 (Vaswani 等, 2017) 机制能够在不同模态之间构建“查询—键—值”关系, 通过

一个模态的信息引导另一个模态的特征选择，实现模态间的显式对齐与动态融合，弥补了单模态注意力对模态间关系建模能力的不足。这为实现多源数据之间更高效的协同建模提供了思路。因此，本文针对现有主被动遥感融合方法存在的融合方式粗糙、信息冗余、模态协同能力不足等问题，提出一种以多光谱影像为主导的交叉注意力融合网络，高效融合特征。该网络通过自注意力提取MSI内部显著特征，交叉注意力引导其对SAR特征的响应区域，提升模态间语义对齐能力。同时，设计可学习Dropout比率的正则化门控机制，实现融合特征间动态调节，控制融合强度与信息冗余，从而实现更高精度的森林分类。此外，考虑到现实场景中高质量、均衡的树种样本数据获取成本高、数量有限的问题（Boudiaf等，2020），本文特别关注小样本条件下的模型性能表现，探索所提方法在样本受限条件下的泛化能力与稳定性。

本文选取具有代表性的云南省普洱市部分区域作为研究区，基于Sentinel-1与Sentinel-2影像，结合人工目视解译与已有森林调查数据构建分类样本集，开展森林分类研究。通过与主流特征融合方法的对比验证，系统评估所提方法在分类精度、模型稳定性与适应性等方面的表现。研究结果有望为主被动遥感数据融合与森林分类提供有效的技术路径与方法参考。

2 研究区与数据

2.1 研究区

研究区位于中国云南省普洱市思茅区与宁洱县的交界地带（22.81°N—22.96°N，100.86°E—101.17°E），总面积约为531 km²，如图1（a）所示。研究区地处亚热带季风气候区，年平均气温19.5℃，年降雨量在1100—2780 mm。区域地貌以山地为主，海拔范围为1142—1827 m，地势起伏较大，坡度介于5°—47°。研究区内森林资源丰富，森林覆盖以天然林为主，少量人工林分散其中。森林组成和结构具有典型的亚热带森林特征，树种多样，分布破碎不规则，研究区内复杂的森林结构给森林分类带来了挑战。本文根据研究区内的主要树种组成将地物划分8类，包括思茅松、

桉树、栎类、茶园、建设用地、耕地、水体和其他林地。其中，栎类包括麻栎、华南石栎等栎属树种的天然混交林，思茅松类别包括思茅松纯林与以思茅松为优势树种的混交林，桉树为纯林，其他林地包含采伐迹地在内的非优势树种区域。

2.2 数据

2.2.1 多光谱数据

本研究所使用的多光谱数据为Sentinel-2影像。Sentinel-2卫星的多光谱影像包含13个波段，光谱范围为400—2400 nm。使用的卫星数据在GEE（Google Earth Engine）平台（<https://developers.google.com/earth-engine/datasets> [2025-07-03]）处理并下载，选取Sentinel-2 MSI L2A级别的地表反射率SR（Surface Reflectance）数据产品作为数据源。为获取高质量的无云影像，在GEE平台筛选2023年3月到10月覆盖研究区的卫星影像，并指定筛选云量小于20%，为消除云的遮掩，利用QA60云概率波段进行去云处理，并按日合成去云后的影像，裁剪研究区范围内的影像并镶嵌，获得多幅去云后的候选影像序列。经过目视判别对比，选择高质量的无云影像，最终选取了2023年3月25日的影像作为研究区的多光谱数据源（图1（b））。由于卷云波段B10不参与大气校正，因此本研究的卫星数据只包含12个波段，所有波段的空间分辨率均重采样为10 m。

2.2.2 SAR数据

本研究所使用的SAR数据为Sentinel-1影像。与Sentinel-2影像相同，对Sentinel-1图像的处理与下载均在GEE平台完成。本研究选取的雷达数据为Sentinel-1 IW GRD（Interferometric Wide swath-Ground Range Detected）产品，由于SAR成像不受云层遮挡影响，无需进行云掩膜或多时相筛选，可直接选取与Sentinel-2光学影像同期获取的产品。本文选用了VV极化与VH极化两个波段，将影像裁剪至研究区域范围。为提高图像质量降低雷达图像中斑点噪声，将SAR影像在SNAP（Sentinel Application Platform）软件中进行了Refine Lee滤波处理。最终的SAR图像如图1（c）所示。

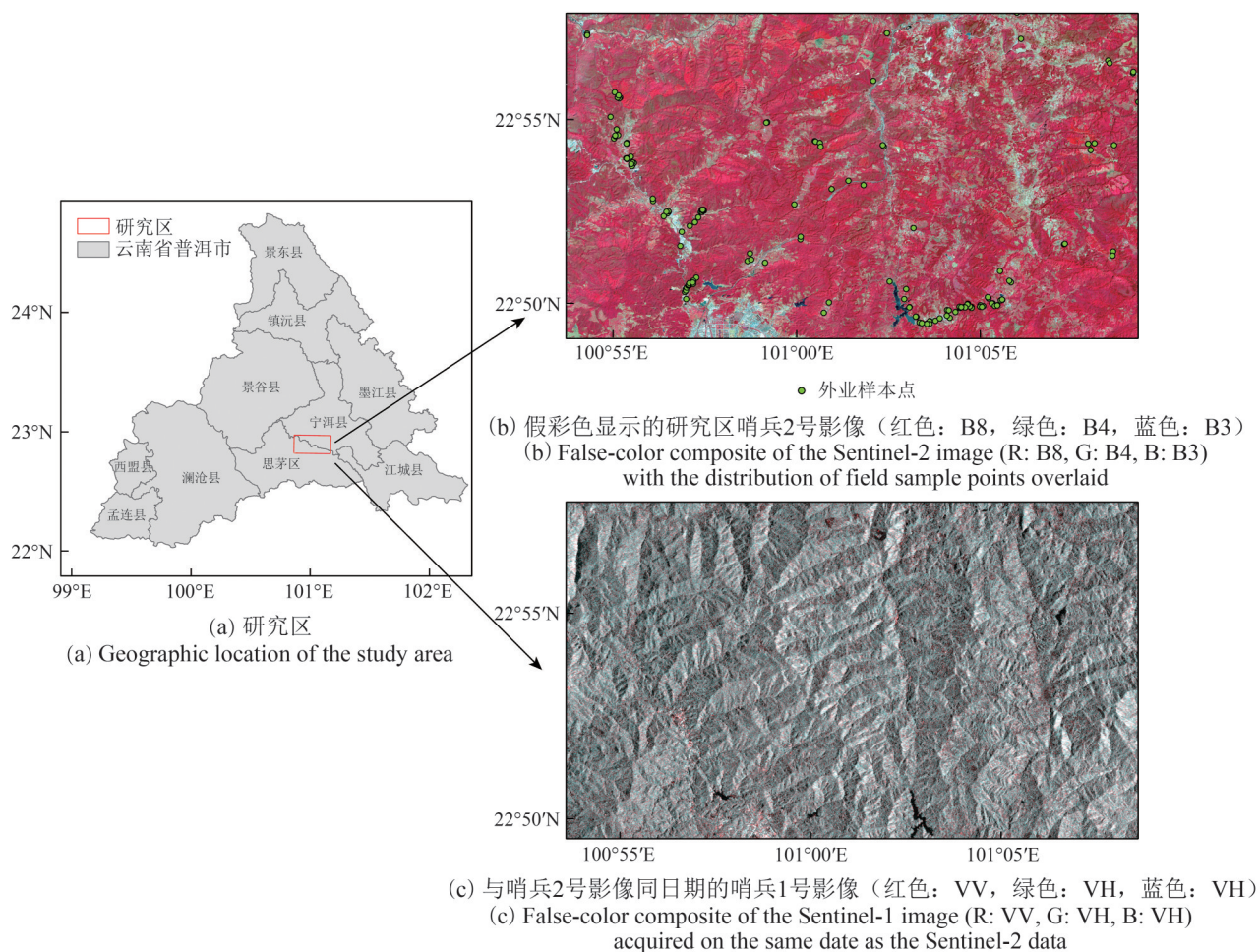


图1 研究区与遥感数据

Fig. 1 The study area and corresponding remote sensing data

2.2.3 地面调查数据与其他辅助数据

地面调查数据于2023年3月在普洱市展开收集。结合普洱市研究区内的地类调查资料，在国道、省道及县道等公路附近，在靠近树干的位置使用奥维互动地图软件（<https://www.ovital.com> [2025-07-03]）记录定位信息，按照数字序列编号，并记录树种的详细信息，包括树种名称、树种类别和树种组成等，并拍摄树木照片及周围环境照片，树种记录分布均匀且广泛，累计采集853个点（图1）。同时获取普洱市思茅区万掌山林场2020年森林资源调查小班数据以及宁洱县国有林场2016年森林资源调查小班数据，包括每个森林小班的优势树种、权属关系等基础信息。在地面调查无法覆盖的区域，小班数据为地物类型判断提供了可靠支撑。通过分析各类小班地物在遥感影像中的光谱响应特征和空间纹理特性，进一步归纳总结各类样本的判读特征，有效辅助高质量样本的目视解译与绘制。

2.3 样本制作与数据集划分

基于地面调查数据与小班数据，并结合 Google Earth Pro 高分辨率影像，归纳总结各类地物在 Sentinel-2 多光谱影像上的典型判读特征。随后，采用目视解译的方式在影像上识别并标注8类主要地物类别，构建对应的地面样本集（图2）。在完成地类判读后，进一步以解译样本为基础制作模型训练数据集。在样本构建过程，利用开源遥感处理工具GDAL以每个标注像元为中心，同时从多光谱影像和SAR影像裁剪13×13像素的窗口，生成标准化图像块（patch）。每个标注点仅对应一块patch，因而patch总数与标注点数量完全一致。模型的训练、验证与测试均以这些完整patch为基本输入单元。由于关注小样本场景下的森林分类，本文采用明显倾斜的比例分割数据集（Xu等，2024），按照1%和99%的比例划分为训练集与测试集。分类体系和样本数量如表1所示。

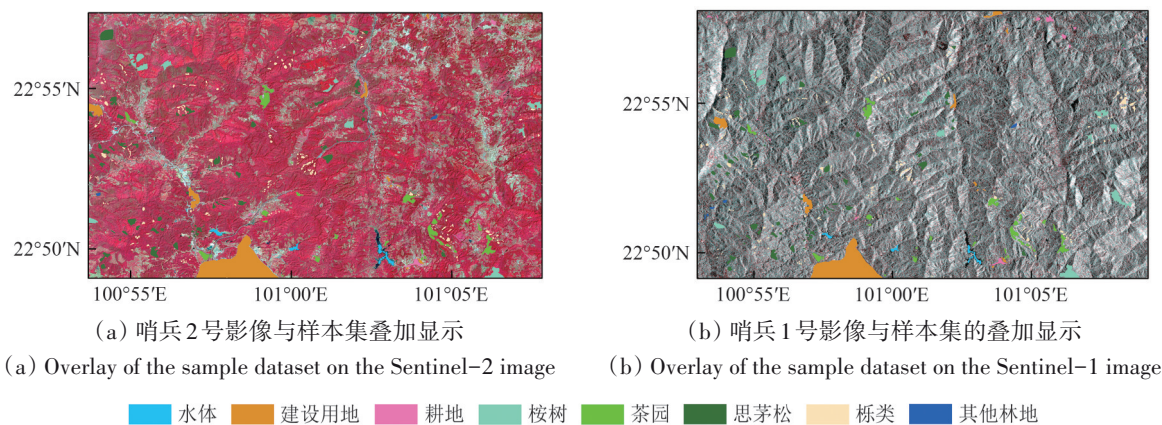


图2 研究区样本数据集

Fig. 2 Sample dataset for the study area

表 1 研究区分类系统与样本数据

Table 1 The classification system and sample data of the study area

地物类别		样本数量/像素
林地	桉树	26525
	茶园	23416
	栎类	24537
	思茅松	36779
	其他林地	6484
非林地	水体	4796
	建设用地	69452
	耕地	5858
—	总计	197847

3 研究方法

3.1 总体框架

本文所提出的 SAR 与 MSI 融合分类方法的总体框架如图 3 所示。该方法旨在充分挖掘光学影像与合成孔径雷达影像在光谱、空间与结构信息上

的互补性，通过注意力机制与门控融合策略实现主被动数据特征的有效整合。由于本文在小样本条件下分类，将改进的原型网络 SEPNet (Squeeze-and-Excitation-Prototypical Network) 作为基本分类框架。在特征融合框架中，MSI 与 SAR 数据首先分别输入至共享的特征提取网络 SEPNet Encoder 提取其深层语义特征。随后，各模态特征分别进入两个不同的注意力路径。MSI 特征通过自注意力模块建模其自身空间和光谱维度内的上下文依赖，强化其内部关键区域的表达；同时，MSI 特征还作为 Query 输入至交叉注意力模块中，与 SAR 特征作为 Key 和 Value 匹配，主动从 SAR 中获取互补的结构与纹理信息，提升跨模态信息感知能力。随后将 SEPNet 提取的 MSI 特征、MSI 自注意力特征、MSI 和 SAR 的交叉注意力特征进行整合，并引入残差式门控融合模块，以自适应地控制各通路特征在最终融合表示中的贡献比例。最后，融合后的特征被展平并输入至 SEPNet 原型分类模块中，得到最终的地物分类结果图。

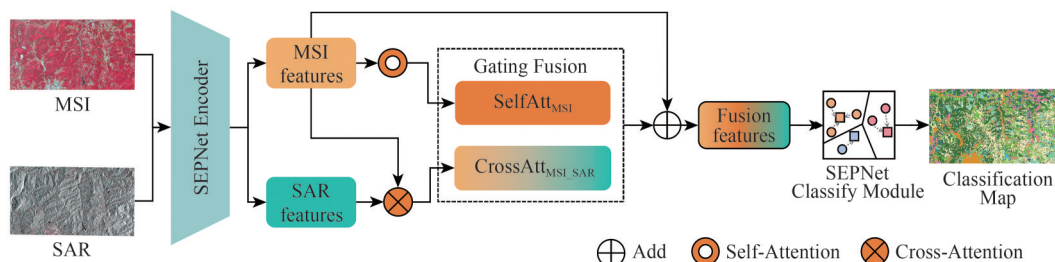


图3 基于 SAR 与多光谱特征融合森林分类的总体框架

Fig. 3 Overview of framework for forest classification based on SAR and multispectral feature fusion

3.2 SEPNet 分类网络

本文采用经典的小样本学习网络原型网络

PNet (Prototypical Networks) (Snell 等, 2017), 作为分类网络的基本结构, 并引入“激励—挤压”

SE (Squeeze-and-Excitation) (Hu 等, 2018) 通道注意力模块搭建 SEPNet 网络 (图 4)。该网络包括数据输入、特征提取和分类 3 个部分。在数据输入部分, 样本被划分为支持集和查询集, 利用支持集计算类别原型, 利用查询集优化原型。在特征提取部分, 采用四层卷积模块构成主干特征提取网络, 每层卷积块包括一个卷积层 (Conv2d, 输出维数为 64, 卷积核为 3×3)、一个非线性激活函数 (ReLU)、一个批归一化层 (Batch Normalization)、一个最大池化层 (Max_Pool2D, 池化核为 2×2)、

一个 Dropout 层 (丢弃比率为 0.2)。在每个卷积块后嵌入 SE 通道注意力机制模块, 经过四层 “卷积-SE” 模块的级联处理后, 输出被展平的高维特征表示。在分类部分, 通过计算支持集中每类样本的特征均值以构建类别原型, 并计算查询样本与各类原型之间的欧氏距离。随后使用 Softmax 函数将距离转换为概率分布, 从而估计查询样本的类别归属概率, 最终计算分类损失并评估模型精度, 完成地物类别的预测任务, 最终输出分类结果图。

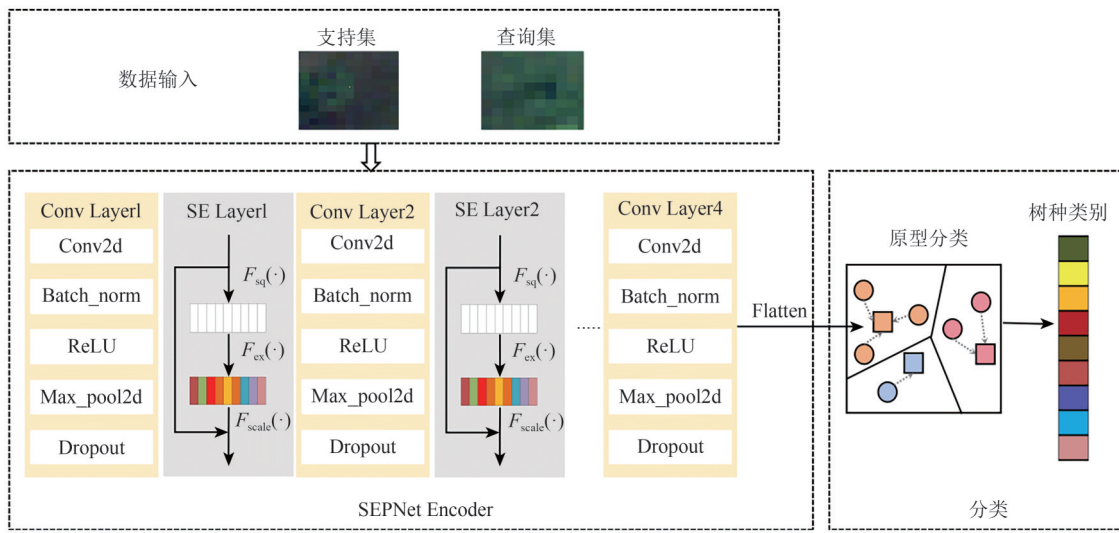


图 4 SEPNet 网络结构

Fig. 4 SEPNet network structure

3.3 自注意力模块

为充分挖掘 MSI 模态内部的信息冗余与局部依赖特性, 本文引入了基于 Transformer 架构的自注意力机制 (Self-Attention) (Vaswani 等, 2017) 以建模 MSI 特征图中不同空间位置之间的依赖关系, 如图 5 所示。

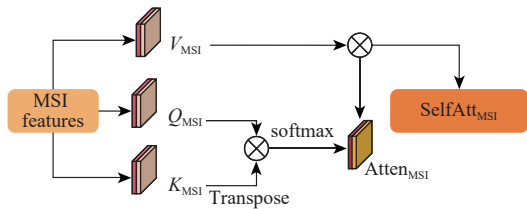


图 5 自注意力模块

Fig. 5 Self-attention module

首先对经过 SEPNet 编码器计算的 MSI 特征 $X_{\text{MSI}} \in \mathbf{R}^{B \times C \times H \times W}$ 进行展平, 转为序列形式

$X_{\text{MSI}} \in \mathbf{R}^{B \times L \times C}$, 其中 $L = H \times W$; 然后分别通过 3 组线性变换 (1×1 卷积) 获取 MSI 的 Query (Q , 式 (1))、Key (K , 式 (2))、Value (V , 式 (3)) 3 个向量序列。

$$Q = X_{\text{MSI}} W_Q \quad (1)$$

$$K = X_{\text{MSI}} W_K \quad (2)$$

$$V = X_{\text{MSI}} W_V \quad (3)$$

式中, $W_Q, W_K, W_V \in \mathbf{R}^{C \times d}$ 为可学习的投影矩阵, C 为通道数, d 为注意力维度。接下来计算注意力权重矩阵 $\text{Attention}(Q, K, V)$ (式 (4))。

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) \cdot V \quad (4)$$

该操作表示每个位置从其他位置聚合自身最相关的信息, 生成增强后的特征表示为 F_{selfAtt} (式 (5)):

$$F_{\text{selfAtt}} = \text{Attention}(X_{\text{MSI}} W_Q, X_{\text{MSI}} W_K, X_{\text{MSI}} W_V) \quad (5)$$

3.4 交叉注意力模块

为了进一步提升模态间互补信息的表达能力, 本文引入了交叉注意力机制, 用于引导 MSI 特征从 SAR 中提取相关信息, 增强其语义表达能力, 如图 6 所示。

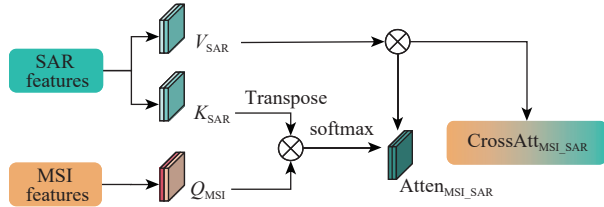


图 6 交叉注意力模块

Fig. 6 Cross-attention module

与自注意力不同, 交叉注意力模块使用 MSI 作为 Query, SAR 作为 Key 和 Value, 实现以 MSI 为导向提取 SAR 中的关键信息。 $X_{SAR} \in R^{B \times L \times C}$ 为经过编码器展平后的 SAR 特征; Query (Q , 式 (6)), Key (K , 式 (7)) 和 Value (V , 式 (8)) 分别为 MSI 的 Query, SAR 的 Key 和 Value。

$$Q = X_{MSI} W_Q \quad (6)$$

$$K = X_{SAR} W_K \quad (7)$$

$$V = X_{SAR} W_V \quad (8)$$

与自注意力的计算过程类似, 交叉注意力计算结果表示为

$$F_{crossAtt} = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) \cdot V \quad (9)$$

最终生成的交叉注意力表征将携带 SAR 中与 MSI 特征最相关的信息, 为后续的融合操作提供跨模态补充支持。

3.5 特征融合模块

特征融合模块结构如图 7 所示。该模块借鉴了 Transformer 架构中的残差连接 (Residual Connection) 机制, 将 MSI 特征与所提取的注意力特征进行逐元素相加, 从而保留原始特征信息并增强表示能力。随后对融合后的特征进行归一化处理 (LayerNorm), 以确保数值稳定性并促进模型的收敛。

为了有效融合两种注意力特征, 本研究提出了一种可学习 Dropout 比率的正则化门控融合方法, 以控制不同数据源特征的互补性。该方法借鉴 Transformer 中残差连接的 Dropout 机制, 用于增

强模型的泛化能力。在自注意力和交叉注意力特征上分别进行 Dropout 后进行融合, 通过不同的 Dropout 比率随机的丢弃神经元, 以间接调节两类特征的贡献。传统 Dropout 中, 丢弃神经元是基于伯努利分布的离散采样过程, 具有不可微性, 且依赖于手动调节。为实现可微分的随机采样过程, 引入 Concrete Dropout (Gal 等, 2017) 机制, 通过连续可导的 Gumbel-Softmax 分布对原始伯努利采样进行近似, 使得 Dropout 丢弃比率能够参与梯度更新, Dropout 丢弃比率 $z \in (0, 1)$ 通过式 (10) 进行采样。

$$z = \sigma \left(\frac{1}{t} (\log p - \log(1-p) + \log u - \log(1-u)) \right), \quad u \sim \text{Uniform}(0, 1) \quad (10)$$

式中, t 为温度系数, 设定为 0.1, 用于控制采样的平滑程度, $p \in (0.1, 0.3)$ 为初始化的 Dropout 丢弃比率 (在该区间内随机生成), 后续可被建模为可学习参数; u 为服从 $(0, 1)$ 均匀分布的随机变量, 用于生成 Gumbel 噪声, 从而模拟原本无法微分的伯努利采样过程; $\sigma(\cdot)$ 表示 Sigmoid 函数。基于 Concrete Dropout, 构建了特征融合模块, 最终的融合特征表示形式为

$$F_{fused} = \text{ConcreteDropout}_{z_1}(F_{selfAtt}) + \text{ConcreteDropout}_{z_2}(F_{crossAtt}) \quad (11)$$

式中, F_{fused} 表示融合特征, $F_{selfAtt}$ 表示 MSI 的自注意力特征, $F_{crossAtt}$ 表示 MSI 与 SAR 的交叉注意力特征, $z_1, z_2 \in (0, 1)$ 为可学习的 dropout 丢弃比率参数。

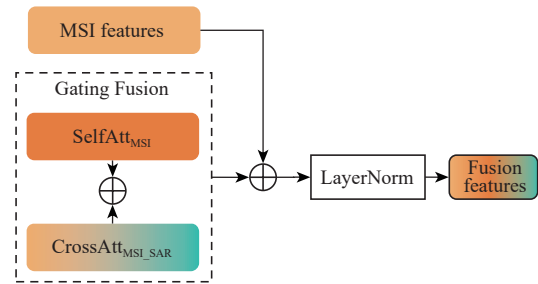


图 7 特征融合模块

Fig. 7 Feature fusion module

4 结果与讨论

4.1 实验设置

本研究基于 Pytorch 深度学习框架, 在 NVIDIA

RTX 3060 GPU 上进行训练和测试，图形尺寸设置为 13×13 像素。在训练过程中，采用 Adam 优化器，并引入 L2 正则化缓解过拟合现象，权重衰减系数设置为 0.001。采用 StepLR 学习率调度策略，初始学习率为 0.001，每 2000 次训练学习率减半（Chen 等，2021）。在元学习任务中，每轮进行 100 次情景采样。本研究地物共 8 类，按照每类 5 个支持集样本和 15 个查询集样本，构建 8way-5shot 的情景采样任务。训练过程中引入 Early Stopping 策略，根据测试集性能动态终止训练。在自注意力和交叉注意力特征计算中，注意力头数设置为 4，dropout 比率设为 0.1（Vaswani 等，2017）。模型分类性能以测试集的分类精度进行评估，评价指标包括总体精度 OA（Overall Accuracy）、Kappa 系数、准确度（Precision），和平均精度 AA

（Average Accuracy）（Chen 等，2021），并以总体精度作为最优模型的评估指标。

实验方案共分为 3 大类，即门控融合实验、不同融合策略对比实验和不同数据源对比实验。其中，门控融合实验主要聚焦于评估不同门控融合策略的效果，本文设计了 4 种门控对比方法，分为基于加权的特征融合和基于 Dropout 的正则化特征融合两类，旨在验证所提门控融合策略的有效性。不同融合策略对比实验从融合层级与融合方式两种对比角度出发，比较不同融合策略下的模型性能，以充分验证所提出方法相比其他融合方法的优势。不同数据源对比实验中，通过对比单一数据源与融合数据下的分类性能，验证融合数据对模型分类精度的提升效果。详细实验方案如表 2 所示。

表 2 实验方案
Table 2 Experimental settings

实验名称	对比角度	实验名称与代号	实验描述
门控融合实验	基于加权的特征融合	无门控:NoGate	对自注意力特征和交叉注意力特征进行逐元素简单相加
		基于可学习加权的门控融合:WeightGate	为自注意力和交叉注意力特征引入可学习的融合权重 w ，两种特征的权重相加为 1。 w 通过 Sigmoid 激活函数约束，初始化为 0.5
	基于 Dropout 的正则化特征融合	固定 Dropout 比率的正则化融合:FixedDrop	采用 Dropout 操作分别作用于自注意力和交叉注意力特征，手动设置不同的 Dropout 丢弃比率
		可学习 Dropout 比率的正则化融合:LearnDrop	本文提出的门控融合方法，如 3.5 节所述
不同融合方法对比实验	融合层级	像素级融合:InputFusion	在输入阶段将 SAR 波段作为额外通道与 MSI 拼接，形成多通道输入，实现像素级融合
	融合方式	特征直接堆叠融合:DirectStackFusion	对 SEPNet 提取得到的 MSI 与 SAR 特征在通道维度直接堆叠特征，实现特征融合
		特征降维堆叠融合:ReducedStackFusion	在特征直接堆叠的基础上，使用一个 1×1 卷积对堆叠后的特征进行通道降维处理，将其通道数压缩为原始维度的一半
不同数据源对比实验	数据源对比	仅只用 MSI 的分类:MSI-Only	仅使用原始的 MSI 图像作为输入，采用 SEPNet 分类
		仅使用 SAR 的分类:SAR-Only	仅使用原始的 SAR 图像作为输入，采用 SEPNet 分类

4.2 门控融合实验分类结果

在本研究中对多种门控融合方法进行了系统比较，以评估不同融合策略在分类任务中的表现，并验证本文所提门控策略的有效性。在 FixedDrop 方法中，本文通过手动调参对两种注意力的 Dropout 丢弃比率进行优化，范围从 0.1—0.9，步长为 0.1，最终选择 $p_1=p_2=0.1$ 作为该方法的最优实验设置。实验结果如表 3 所示。结果表明，LearnDrop 方法在 OA、AA 和 Kappa 系数 3 个评估指

标上均表现优异，分类精度高于其他方法，其 OA 达到了最优的 95.24%。相较于手动调参的 FixedDrop 方法，LearnDrop 在不仅分类精度上更为优越且减少了人工调控成本。这表明可学习的 Dropout 比率为模型带来了更高的适应性和精度。与基于加权的门控方法相比，基于 Dropout 正则化策略在特征融合中表现出更优性能，NoGate 和 WeightGate 的各项指标略低于 LearnDrop 和 FixedDrop，表明 Dropout 方法在特征融合中具有更强的泛化能力。

在加权方法中, NoGate 表现出相对较好的性能, 略高于 WeightGate 方法。这一现象可能与 WeightGate 方法的加权重有关, 其将权重限制在 (0, 1) 之间, 这可能影响了模型的表达能力和分类性能。而 NoGate 方法没有使用显式的加权机制, 因此其在某些场景下能够保持较好的分类精度。综上所述, LearnDrop 方法在所有评估指标上均表现最为优异, 证明了可学习的 Dropout 比率在特征融合中的有效性。相比之下, 基于 Dropout 的方法普遍优于基于加权的门控方法, LearnDrop 和 FixedDrop 展现了较强的适应性和性能优势。

表3 不同门控方法下的分类结果

分类结果	NoGate	WeightGate	FixedDrop	LearnDrop
OA/%	94.88	94.75	94.92	95.24
AA/%	93.85	93.64	94.14	94.48
Kappa×100	93.54	93.38	93.59	93.99

注: 黑色加粗数字代表模型分类最优结果。

4.3 不同融合方法下的对比实验结果

在本实验中对多种融合策略进行了对比分析, 以系统评估不同融合方式在遥感图像分类任务中的性能表现, 同时验证本研究所提方法的有效性和优越性。实验结果如表4所示。在融合层级的对比中, 像素级融合 (InputFusion) 方法实现简洁, 但仍不及本文提出的特征级融合方法。这表明仅在输入阶段进行波段拼接未能显式建模 SAR 与 MSI 间的互补关系, 从而限制了模型对 SAR 结构信息的有效利用。在特征融合方式上, 特征直接堆叠融合 (DirectStackFusion) 方法表现最差, 其 OA 仅为 90.48%。该方法将提取后的 MSI 和 SAR 特征简单堆叠, 导致通道维度翻倍, 特征冗余严重,

缺乏有效的信息筛选与对齐机制, 最终严重影响了分类效果。相比之下, 特征降维堆叠融合 (ReducedStackFusion) 方法在其基础上加入了 1×1 卷积进行通道降维, 有效缓解了冗余问题, 达到与 InputFusion 方法相当的分类性能。这一结果说明, 在特征堆叠后引入降维操作可在一定程度上提升模型性能。但由于其仍未涉及注意力机制等显式建模数据源间互补性的操作, SAR 提供的结构信息被多光谱遮掩, 模型性能存在上限。

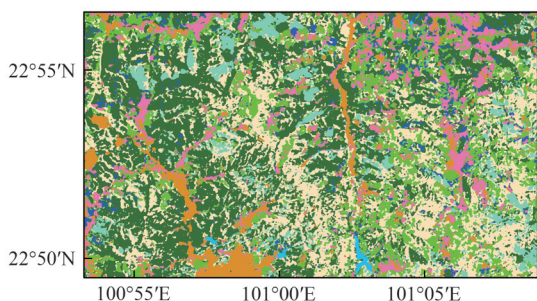
表4 不同融合方法下的分类结果

Table 4 Classification results by different fusion methods

融合方法		OA/%	AA/%	Kappa×100
本文方法		95.24	94.48	93.99
融合层级	InputFusion	94.82	93.50	93.47
特征融合 方式	DirectStackFusion	90.48	90.13	87.99
	ReducedStackFusion	94.82	93.81	93.47

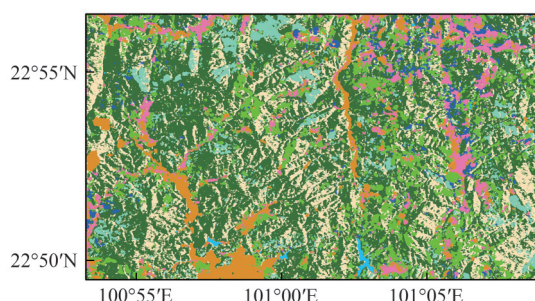
注: 黑色加粗代表模型最有结果。

为直观对比不同融合策略在森林分类中的表现, 本文对研究区影像进行了逐像素预测, 并生成对应的分类图像 (图8), 同时选取两个典型区域 (场景A和场景B) 进行放大分析 (图9)。在场景A中, 本文方法在茶园区域展现出更好的斑块完整性和边界光滑性。相比特征直接堆叠方法 (DirectStackFusion), 本文方法在栎类与思茅松两类容易混淆的植被类型间呈现出更强的类别分离能力。在场景B中, 特征直接堆叠方法 (DirectStackFusion) 方法在分析区域中出现了严重锯齿, 本文方法在保持类别判别准确性的同时, 提升了图像在空间尺度上的一致性, 分类斑块边界更规整, 内部连贯性更强, 进一步体现了融合策略在空间结构建模方面的优势。



(a) 本文方法

(a) Our method



(b) DirectStackFusion

(b) DirectStackFusion

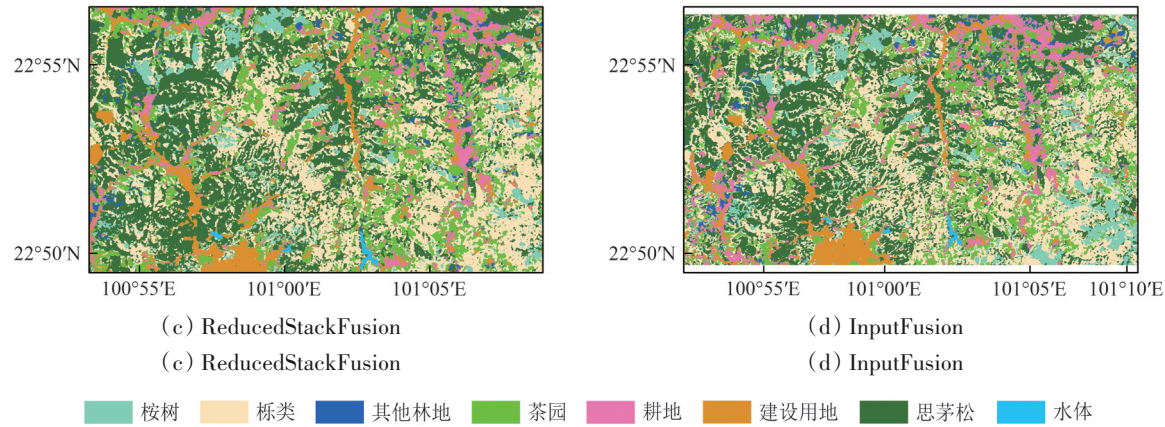


图8 不同融合方法的研究区分类结果图

Fig. 8 Classification results of the study area using different fusion methods

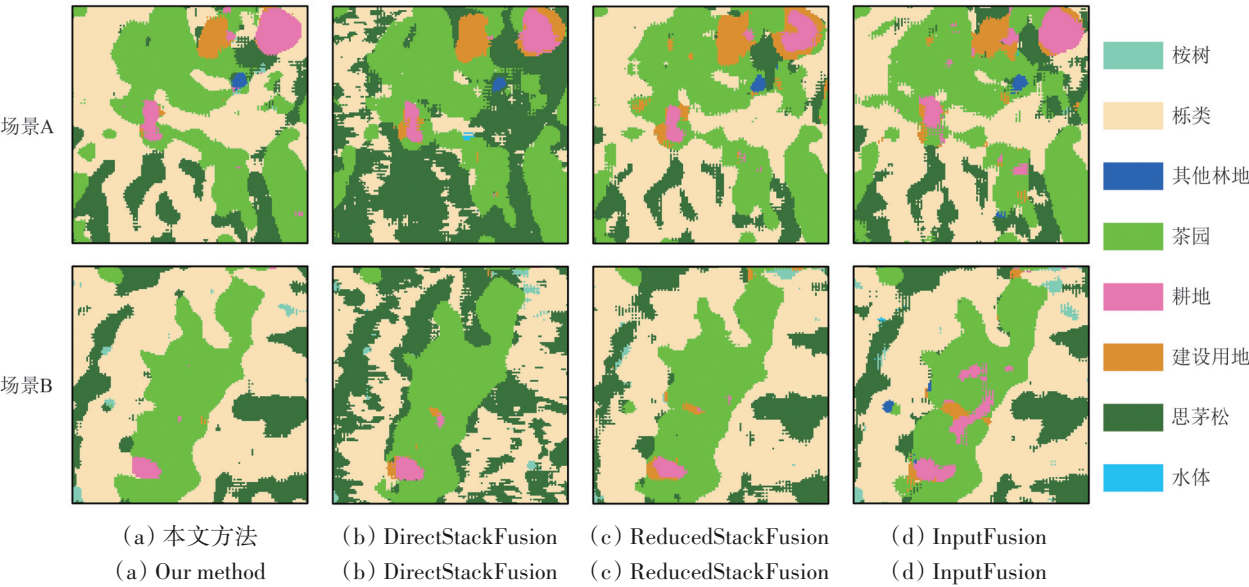


图9 不同融合方法典型区域分类结果图

Fig. 9 Classification results of representative regions based on different fusion methods

综上所述，以 MSI 为主导并结合自注意力与交叉注意力机制进行特征级融合的方法在本任务中表现最为优异。这种融合策略不仅充分发挥了 MSI 的光谱信息优势，还有效引入了 SAR 图像的结构补充，从而在分类精度上实现了提升。相较之下，输入层拼接或特征层堆叠虽然操作简便，但由于未能对数据源间复杂交互关系进行有效建模，最终在性能上仍存在明显差距，验证了注意力引导的深度融合策略在主被动数据遥感分类中的重要价值。

4.4 不同数据源下的对比实验结果

本文进一步对比了单源数据与主被动融合数据的分类效果，以验证融合后的数据对遥感图像分类性能的提升。对比方法包括仅使用 MSI 图像

(MSI-Only)、仅使用 SAR 图像 (SAR-Only)，以及将两者进行融合 (融合方法)，即本文提出的融合方法。实验结果如表 5 所示。从总体表现来看，本文的融合方法在所有评估指标上均优于单源数据方法，表明主被动遥感数据融合策略在本研究中可以提升分类性能，在复杂地物类型的判别上具有更强的表达能力和判别力。在单源分类结果中，MSI-Only 方法 OA 为 94.39%，远优于 SAR-Only 方法的 67.00%，显示出 MSI 在提供光谱判别信息方面的优势。SAR-Only 方法的性能整体较低，仅在边界清晰、特征突出的类别中 (水体和建设用地)，取得了较高的分类精度，这表明单独使用 SAR 图像在本场景下难以有效地区分多种地物类别。SAR 图像缺乏光谱信息，主要提供结构、

纹理特征, 而树种类别和其他类别均在光谱上差异显著, 使得其区分这些类别时表现有限。但 SAR 图像在某些特定类别中展现优势, 在思茅松上的表现相对较好, 分类准确度达到 63.95%, 明显优于其他森林类别的表现。这是由于 SAR 的后向散射特性与地表粗糙度密切相关, 思茅松属于针叶林, 在空间结构和冠层形态上与其他地类存在较大差异, 从而在 SAR 图像中具有较高的可分性。SAR 在该类地物上的优势进一步体现在了融合后的分类结果上。与 MSI-Only 方法相比, 融合方法在思茅松上取得了高达 92.90% 的分类准确度, 相比 MSI-Only 提高了 3.53%, 表明本方法能够有效挖掘 SAR 所提供的空间结构与纹理特征, 并与 MSI 的光谱特征形成互补, 从而增强模型的判别能力。

为直观展示不同数据源在森林分类任务中的表现, 本文对研究区遥感影像进行逐像素分类, 并生成对应的分类结果图。图 10 展示了基于 SAR、MSI 及融合数据源的分类结果, 同时选取了两个典型区域 (场景 C 和场景 D) 进行局部放大分析 (图 11), 以深入比较各方法在地物边界保持与类别区分方面的表现差异。从整体分类图像来看, SAR 数据的分类结果存在显著噪声, 像元类别混杂, 区域边界模糊, 分类精度较低。相比之下,

MSI 和融合数据源在空间一致性和类别区分方面表现更为优越, 图像结构清晰, 分类边界相对规整。在场景 C 中, 3 种数据源对特征显著的水体类别均表现出良好的识别能力, 但在其他地类上, SAR 分类结果仍呈现大范围混叠, 难以准确反映地物边界。而 MSI 与融合特征均能较好地地区分林地类型, 显示出较强的特征表征能力。在场景 D 中融合数据在区分栎类与思茅松两个地类时, 分类斑块更完整、边界更连续, 体现出融合结构在复杂类别间建模能力上的优势。

表 5 不同数据源下的分类结果

Table 5 Classification results by different data sources

分类精度	MSI-Only	SAR-Only	融合方法
OA/%	94.39	67.00	95.24
AA/%	93.39	61.17	94.48
Kappa×100	92.92	59.27	93.99
水体/%	99.53	96.39	100
建设用地/%	98.29	95.21	98.02
耕地/%	78.66	35.72	83.71
桉树/%	97.98	55.99	96.78
茶园/%	94.99	54.39	94.07
栎类/%	89.89	52.58	91.73
思茅松/%	89.31	63.95	92.90
其他林地/%	98.47	35.10	98.61

注: 黑色加粗代表最优结果。

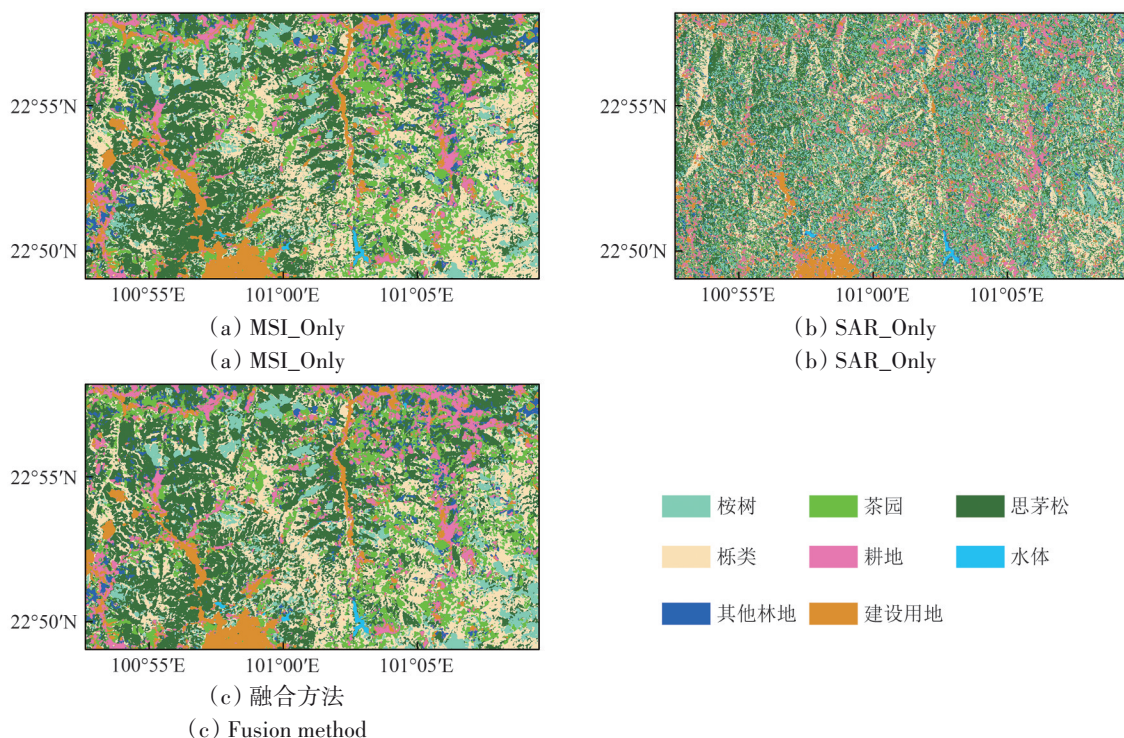


图 10 不同数据源研究区分类结果图

Fig. 10 Overall classification results of the study area using different data sources

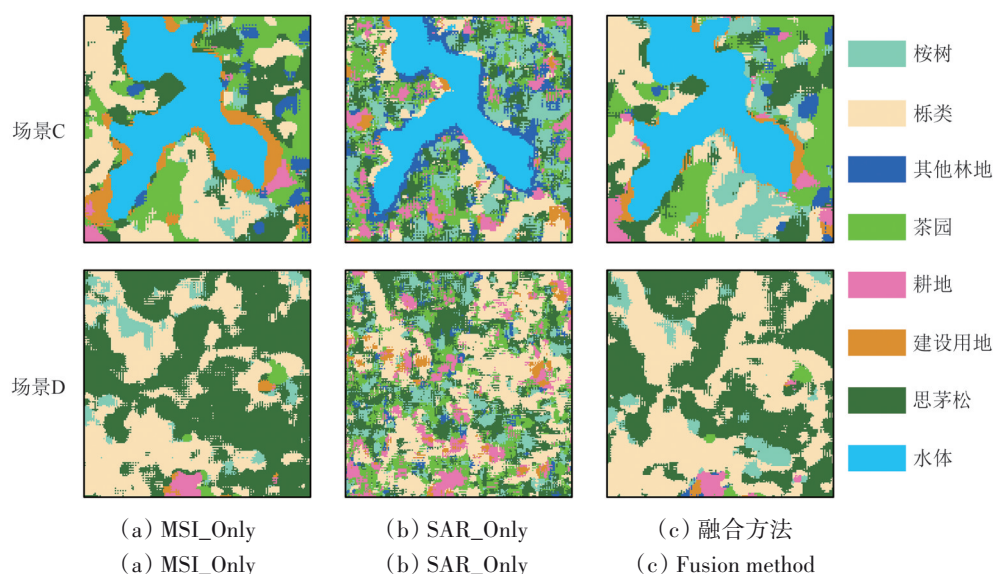


图 11 不同数据源典型区域分类结果图

Fig. 11 Classification results of representative regions based on different data sources

该实验结果表明,主被动数据融合方法在遥感图像分类任务中具有一定优势。通过将MSI的光谱信息与SAR的结构与纹理信息进行互补整合,融合模型有效提升了分类精度和稳定性,验证了主被动遥感数据互补性在遥感森林分类中的重要价值。

4.5 特征可视化分析

为直观揭示研究区内主要地物类型在光谱与极化特征空间的原始表达差异,本文绘制了Sentinel-2多光谱影像的平均光谱曲线(图12(a)),以及Sentinel-1 C波段VV与VH极化强度的散点分布图(图12(b))。图12(a)显示,不同地物类别在红边和近红外波段具有明显的响应差异。水体的反射率始终最低,呈现出典型的“暗而平”曲线特征;桉树、栎类等阔叶林在红边区域反射率快速上升,近红外波段达到峰值,表现出鲜明的植被红边特性;思茅松的红边跃升幅度和近红外峰值上略低于阔叶树种;茶园的近红外峰值介于阔叶林与针叶林之间。然而,从图12中亦可观察到多个类别之间存在光谱重叠现象。桉树与栎类在可见光至近红外的大部分波段中曲线相近;思茅松与栎类、桉树在SWIR波段的反射率几乎一致;这种类别间的特征混杂表明,仅依赖原始光谱数据进行分类,尤其在多类森林树种之间,难以实现有效的类别区分。图12(b)展示了各类别在VV-VH极化空间内的散点分布图。水体主要集中于左下象限,其VV和VH的值均较低,明显区别于其他地物,具有良好的可分性;建设用地散布较广,

部分样本出现在VV和VH的正值区间;其余地类在VV-VH空间中分布高度重叠,难以直接区分。总体来看, Sentinel-2多光谱影像在红边与近红外波段对森林植被类型的区分能力表现突出,但多个树种在光谱特征上仍存在重叠现象,给精细分类带来挑战。 Sentinel-1 SAR影像在原始极化特征层面对水体等结构差异显著的目标表现出良好的可分性,但在不同森林类别间,其判别能力相对有限。为提升分类精度,仍需借助高层次特征提取与融合机制,以强化类别判别能力并提升复杂森林场景下的识别效果。

为增强模型的可解释性,更直观展示特征融合方法在提取高判别性地类特征方面的有效性,本文采用t-SNE(t-Distributed Stochastic Neighbor Embedding)对MSI、SAR及其融合数据的原始特征与深度特征进行二维降维可视化(图13)。通过低维空间的映射,观察样本在特征空间中的聚类态势与分布规律,进而评价特征的可分性。原始特征由遥感图像的原始波段信息构成, MSI和SAR的深度特征为SEPN_{et}编码器提取的高级语义特征, MSI和SAR融合后深度特征为采用本文方法进行特征融合后的融合特征表示。结果显示, MSI、SAR以及波段叠加后的MSI和SAR,其样本的原始特征在二维空间中分布高度混杂,类间界限模糊,聚类结构不明显。 MSI的深度特征(图13(d))呈现出较好的聚类结构,类间间隔明显,类内点分布紧凑; SAR的深度特征(图13(e))在视觉上仍

表现出一定的混叠趋势,但相比原始特征略有改善;相比单一数据源,MSI与SAR融合后的深度特征(图13(f))在可视化中呈现出更优的聚类结构,类别分界明确,显示出更强的特征分离能力。尤其在桉树、思茅松和栎类3个主要树种类别

中,融合后的深度特征相比MSI深度特征其类间差异进一步增强。该结果表明,本文提出的特征融合网络能够有效整合主被动遥感信息,增强类别间特征差异性,提升样本在特征空间中的可分性。

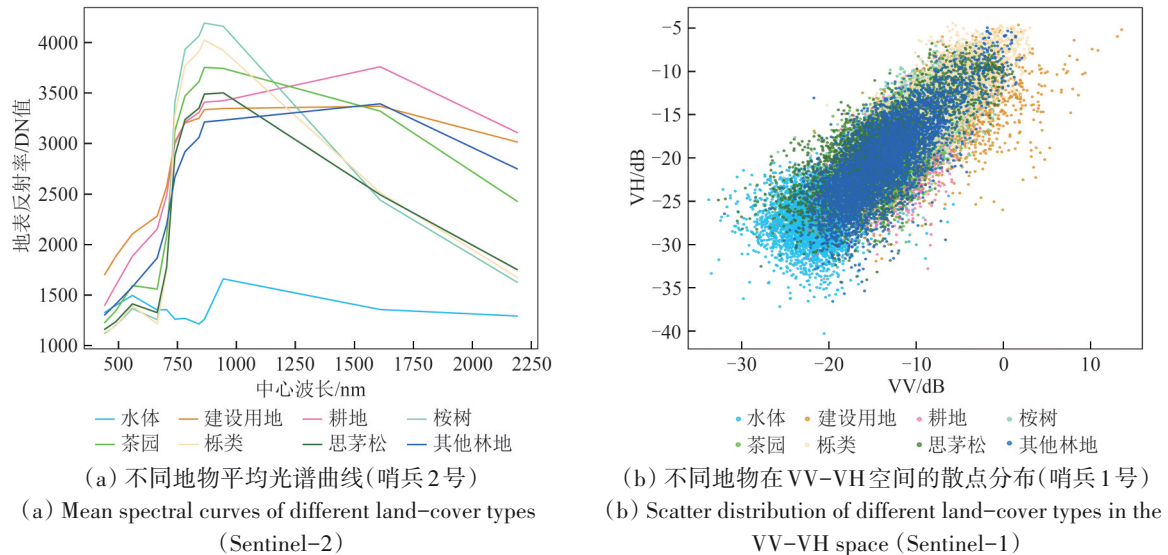


图12 不同地物在多光谱与 SAR 数据上的原始特征可视化

Fig. 12 Visualization of the multispectral and SAR features for the various classes

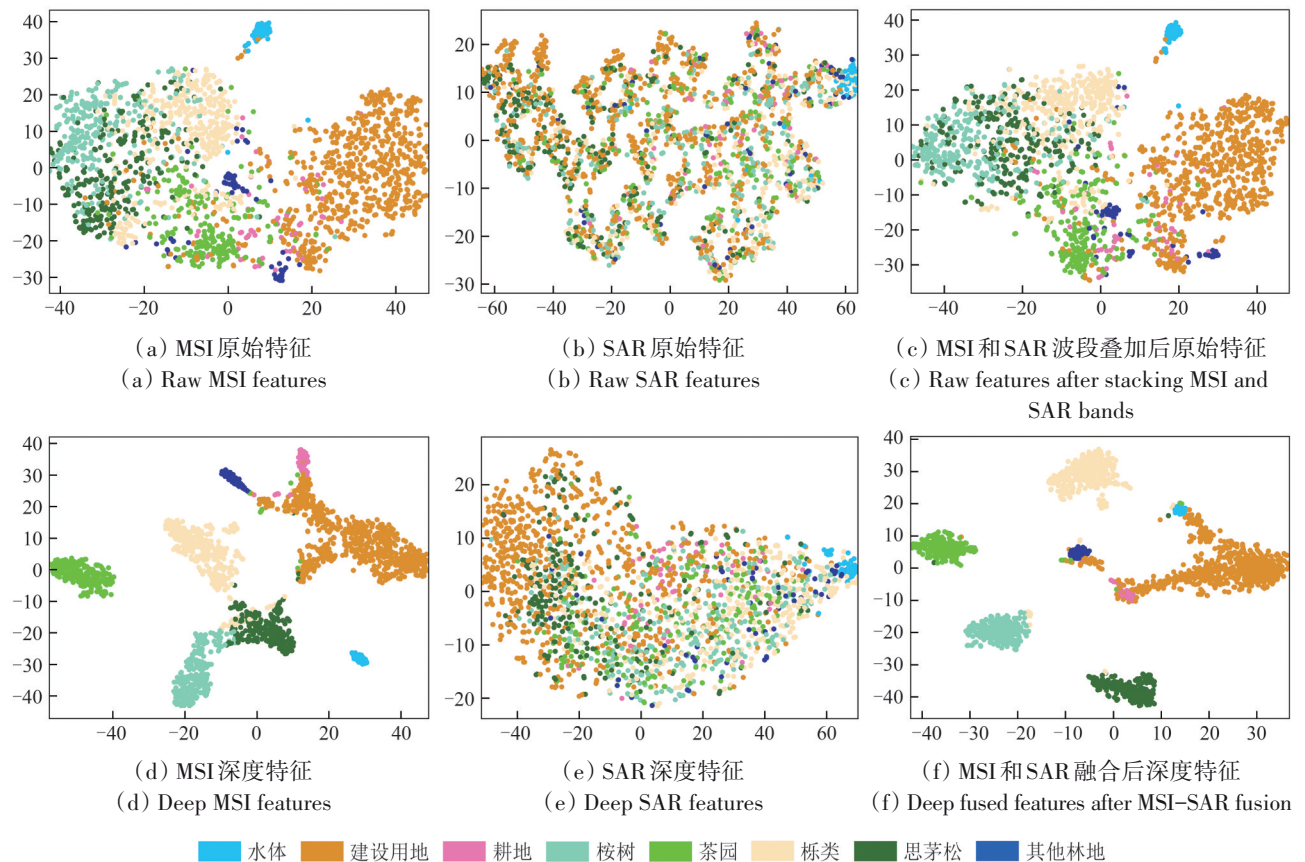


图13 t-SNE特征可视化

Fig. 13 t-SNE feature visualization

4.6 模型泛化能力评价

尽管模型在小样本条件下取得较高分类精度，其泛化能力与可迁移性对于跨区域部署、复杂场景适应及样本分布变化等实际应用仍至关重要。为此，本节引入空间非重叠区域的新增测试样本，对模型在未知区域的预测鲁棒性进行验证以评估其泛化性能。新增样本采自遥感影像中原未标注

的左上与右侧区域（图 14），与原样本区域（图 2）空间上完全不重叠。经目视解译，共新增 8 类地物样本，具体数量如表 6 所示。将新增样本并入原测试集，构建扩展测试集，并利用原训练集训练的最优模型对其进行分类精度评估。通过比较模型在原测试集与扩展测试集上的表现，量化其在空间独立区域的泛化能力。

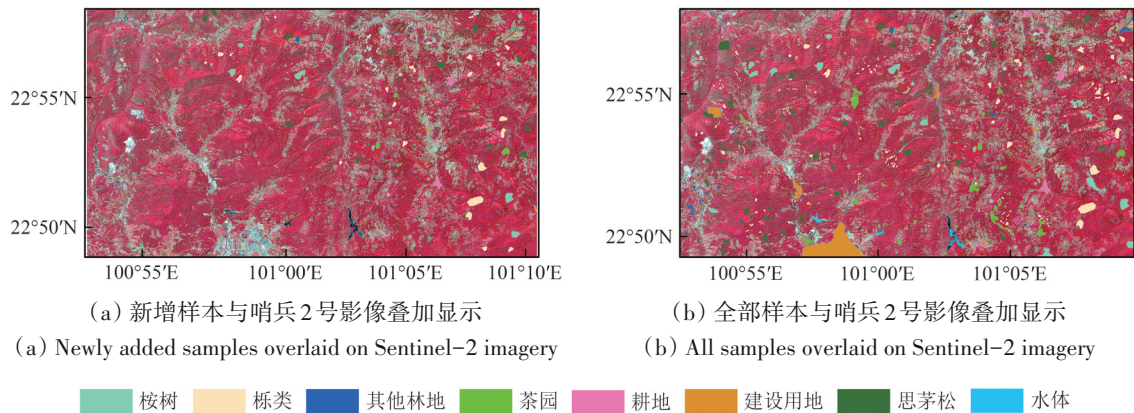


图 14 研究区新增样本分布
Fig. 14 Spatial distribution of newly added samples within the study area

表 6 研究区内新增样本分类系统与样本数量
Table 6 Classification system and sample counts of newly added samples in the study area

地物类别		新增样本数量/像素
林地	桉树	13943
	茶园	7070
	栎类	26744
	思茅松	16665
	其他林地	1458
非林地	水体	269
	建设用地	1430
	耕地	6596
—	总计	74175

表 7 展示了模型在新增测试样本前后分类性能的对比结果。尽管扩展测试集中引入了大量空间非重叠区域的样本，模型总体精度仍达 93.09%，相比原测试集的 95.24% 仅略有下降，整体分类性能保持在较高水平。多数地物类别的分类精度表现稳定，其中耕地与栎类在新增样本后精度略有提升。尽管部分林地类别（如思茅松）在扩展测试集上的精度有所下降，但整体仍处于可接受区间，未出现显著性能退化。该结果充分表明，在训练样本数量保持不变、测试区域扩展的条件下，

所提出的模型仍能在空间非重叠的未知区域中实现高精度分类，展现出其优异的泛化能力与空间迁移性。

表 7 研究区新增样本后模型分类结果
Table 7 Model classification results for the study area after incorporating the new samples

分类结果	样本新增后	样本新增前
OA/%	93.09	95.24
AA/%	93.81	94.48
Kappa×100	91.57	93.99
水体/%	99.01	100.00
建设用地/%	97.95	98.02
耕地/%	88.54	83.71
桉树/%	95.32	96.78
茶园/%	93.26	94.07
栎类/%	93.42	91.73
思茅松/%	85.58	92.90
其他林地/%	97.37	98.61

5 结 论

为有效实现主被动特征融合，本章提出了一种以 MSI 为主导的交叉注意力融合网络。该网络在融合模块中引入了自注意力与交叉注意力机制，分别用于挖掘 MSI 模态内部的显著特征与其对 SAR

模态的关注区域。在此基础上, 本文进一步提出了一种基于可学习 Dropout 比率的正则化门控融合方法, 用于动态调节自注意力特征与交叉注意力特征在最终融合表示中的贡献比例, 从而增强融合特征的判别性与表达能力。通过大量实验, 全面验证了所提出融合框架的有效性与泛化性。主要结论如下:

(1) 在以 MSI 为主导的交叉注意力融合网络框架下, 实现了研究区内 8 类地物的分类任务, 模型总体精度达到 95.24%。其中, 桉树、茶园、栎类和思茅松等主要森林类别的分类精度分别为 96.78%、94.07%、91.73% 和 92.90%, 表明所提方法具备实现高精度森林分类的能力。

(2) 对比了 4 种门控融合方法, 包括无门控、固定 Dropout 门控、可学习加权门控以及可学习 Dropout 门控融合。实验结果表明, 基于可学习 Dropout 比率的正则化门控融合方法在整体分类精度、平均精度与 Kappa 系数等多项指标上均取得了最优结果, 充分验证了该策略在调节注意力特征融合强度方面的有效性。

(3) 分别从特征融合方法、融合层级两个维度对当前融合模块的有效性进行了全面分析。结果表明, 在以 MSI 为主导的数据驱动下, 结合自注意力机制与交叉注意力机制进行特征级融合的策略能够充分捕捉模态内与模态间的关键信息, 显著优于其他融合方式, 说明该设计在结构合理性与任务适应性方面具有较强的优势。

(4) 分别对比了单一数据源与主被动融合下模型的性能。实验结果表明, 引入 SAR 数据作为补充模态在某些类别上带来了明显性能提升, 特别是在思茅松上, SAR 的补充信息显著提升了模型的判别能力, 进一步验证了融合模块的有效性。

然而, 研究仍存在以下局限性。首先, 当前所涉及的树种类型有限, 尚未涵盖更为复杂或细粒度的林种类别, 未来可通过引入更多样本及多样化地物类型, 扩展模型对复杂森林系统的适应能力; 其次, 研究区空间范围较小, 地形与生态结构相对单一, 将在后续工作中选取具有更丰富植被的区域, 以验证模型在树种分类任务上的鲁棒性和有效性; 最后, 模型仅利用单时相数据, 尚未充分挖掘地物物候特征在森林分类中的作用, 后续可引入多时相遥感数据, 融合时间维度特征,

从而提升分类性能。

参考文献 (References)

- Adrian J, Sagan V and Maimaitijiang M. 2021. Sentinel SAR-optical fusion for crop type mapping using deep learning and Google Earth Engine. *ISPRS Journal of Photogrammetry and Remote Sensing*, 175: 215-235 [DOI: 10.1016/j.isprsjprs.2021.02.018]
- Arasumani M, Bunyan M and Robin V V. 2021. Opportunities and challenges in using remote sensing for invasive tree species management, and in the identification of restoration sites in tropical montane grasslands. *Journal of Environmental Management*, 280: 111759 [DOI: 10.1016/j.jenvman.2020.111759]
- Bahdanau D, Cho K and Bengio Y. 2016. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv: 1409.0473* [DOI: 10.48550/arXiv.1409.0473]
- Bjerreskov K S, Nord-Larsen T and Fensholt R. 2021. Classification of nemoral forests with fusion of multi-temporal sentinel-1 and 2 data. *Remote Sensing*, 13(5): 950 [DOI: 10.3390/rs13050950]
- Boudiaf M, Masud Z I, Rony J, Dolz J, Piantanida P and Ben Ayed I. 2020. Information maximization for few-shot learning. *arXiv: 2008.11297* [DOI: 10.48550/arXiv.2008.11297]
- Chen B, Huang B and Xu B. 2017. Multi-source remotely sensed data fusion for improving land cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 124: 27-39 [DOI: 10.1016/j.isprsjprs.2016.12.008]
- Chen C, He X Y, Guo B Y, Zhao X and Chu Y L. 2020. A pixel-level fusion method for multi-source optical remote sensing image combining the principal component analysis and curvelet transform. *Earth Science Informatics*, 13(4): 1005-1013 [DOI: 10.1007/s12145-020-00472-7]
- Chen L, Tian X M, Chai G Q, Zhang X L and Chen E. 2021. A new CBAM-P-net model for few-shot forest species classification using airborne hyperspectral images. *Remote Sensing*, 13(7): 1269 [DOI: 10.3390/rs13071269]
- Dai Y M, Gieseke F, Oehmcke S, Wu Y Q and Barnard K. 2021. Attentional feature fusion//*Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision*. Waikoloa: IEEE: 3559-3568 [DOI: 10.1109/WACV48630.2021.00360]
- Dubois C, Mueller M M, Pathe C, Jagdhuber T, Cremer F, Thiel C and Schmullius C. 2020. Characterization of land cover seasonality in sentinel-1 time series data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-3-2020: 97-104 [DOI:10.5194/isprs-annals-V-3-2020-97-2020]
- Gal Y, Hron J and Kendall A. 2017. Concrete dropout//*Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: Curran Associates Inc.: 3584-3593
- Gao J, Li P, Chen Z K and Zhang J N. 2020. A survey on deep learning for multimodal data fusion. *Neural Computation*, 32(5): 829-864 [DOI: 10.1162/neco_a_01273]
- Ghassemian H. 2016. A review of remote sensing image fusion methods. *Information Fusion*, 32: 75-89 [DOI: 10.1016/j.inffus.2016.03.003]

- Grabska E, Hostert P, Pflugmacher D and Ostapowicz K. 2019. Forest stand species mapping using the Sentinel-2 time series. *Remote Sensing*, 11(10): 1197 [DOI: 10.3390/rs11101197]
- Hościło A and Lewandowska A. 2019. Mapping forest type and tree species on a regional scale using multi-temporal Sentinel-2 data. *Remote Sensing*, 11(8): 929 [DOI: 10.3390/rs11080929]
- Hu B X, Li Q and Hall G B. 2021. A decision-level fusion approach to tree species classification from multi-source remotely sensed data. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, 1: 100002 [DOI: 10.1016/j.ophoto.2021.100002]
- Hu J, Shen L and Sun G. 2018. Squeeze-and-excitation networks//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE: 7132-7141 [DOI: 10.1109/CVPR.2018.00745]
- Immitzer M, Neuwirth M, Böck S, Brenner H, Vuolo F and Atzberger C. 2019. Optimal input features for tree species classification in Central Europe based on multi-temporal Sentinel-2 data. *Remote Sensing*, 11(22): 2599 [DOI: 10.3390/rs11222599]
- Li S T, Kang X D, Fang L Y, Hu J W and Yin H T. 2017. Pixel-level image fusion: a survey of the state of the art. *Information Fusion*, 33: 100-112 [DOI: 10.1016/j.inffus.2016.05.004]
- Li X Y, Tian X, Duan T, Cao X M, Yang K J, Lu Q and Wang F. 2023. Estimation of fractional woody and herbaceous vegetation cover in Temperate Sparse Forest Grassland using fusion of UAV and Satellite imagery. *National Remote Sensing Bulletin*, 27(9): 2139-2152. (李晓雅, 田昕, 段涛, 曹晓明, 杨凯捷, 卢琦, 王锋. 2023. 融合无人机和卫星影像的温带疏林草原木本和草本植物覆盖度遥感估算. *遥感学报*, 27(9): 2139-2152) [DOI: 10.11834/jrs.20210605]
- Liu Y N, Gong W S, Hu X Y and Gong J Y. 2018. Forest type identification with random forest using Sentinel-1A, Sentinel-2A, multi-temporal Landsat-8 and DEM data. *Remote Sensing*, 10(6): 946 [DOI: 10.3390/rs10060946]
- Ma W L, Karakuş O and Rosin P L. 2022. AMM-FuseNet: attention-based multi-modal image fusion network for land cover mapping. *Remote Sensing*, 14(18): 4458 [DOI: 10.3390/rs14184458]
- Mori A S, Lertzman K P and Gustafsson L. 2017. Biodiversity and ecosystem services in forest ecosystems: a research agenda for applied forest ecology. *Journal of Applied Ecology*, 54(1): 12-27 [DOI: 10.1111/1365-2664.12669]
- Persson M, Lindberg E and Reese H. 2018. Tree species classification with multi-temporal Sentinel-2 data. *Remote Sensing*, 10(11): 1794 [DOI: 10.3390/rs10111794]
- Phiri D, Simwanda M, Salekin S, Nyirenda V R, Murayama Y and Ranagalage M. 2020. Sentinel-2 data for land cover/use mapping: a review. *Remote Sensing*, 12(14): 2291 [DOI: 10.3390/rs12142291]
- Pimentel C S, McKenney J, Firmino P N, Calvão T and Ayres M P. 2020. Sublethal infection of different pine species by the pine-wood nematode. *Plant Pathology*, 69(8): 1565-1573 [DOI: 10.1111/ppa.13241]
- Shafer G. 1992. Dempster-shafer theory. *Encyclopedia of Artificial Intelligence*, 1: 330-331
- Shahdoosti H R and Ghasseman H. 2012. Spatial PCA as a new method for image fusion//The 16th CSI International Symposium on Artificial Intelligence and Signal Processing. Shiraz: IEEE: 90-94 [DOI: 10.1109/AISP.2012.6313724]
- Smith W B. 2002. Forest inventory and analysis: a national inventory and monitoring program. *Environmental Pollution*, 116: S233-S242 [DOI: 10.1016/S0269-7491(01)00255-X]
- Snell J, Swersky K and Zemel R. 2017. Prototypical networks for few-shot learning//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc.: 4080-4090
- Tang C, Zheng X, Zhang W, Liu X W, Zhu X Z and Zhu E. 2023. Unsupervised feature selection via multiple graph fusion and feature weight learning. *Science China Information Sciences*, 66(5): 152101 [DOI: 10.1007/s11432-022-3579-1]
- Tian L, Tao Y, Fu W X, Li T, Ren F and Li M Y. 2022. Dynamic simulation of land use/cover change and assessment of forest ecosystem carbon storage under climate change scenarios in Guangdong Province, China. *Remote Sensing*, 14(10): 2330 [DOI: 10.3390/rs14102330]
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Ł and Polosukhin I. 2017. Attention is all you need//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc.: 6000-6010
- Wang Z J, Chen B, Lu R Y, Zhang H, Liu H W and Varshney P K. 2020. FusionNet: an unsupervised convolutional variational network for hyperspectral and multispectral image fusion. *IEEE Transactions on Image Processing*, 29: 7565-7577 [DOI: 10.1109/TIP.2020.3004261]
- White J C, Coops N C, Wulder M A, Vastaranta M, Hilker T and Tompalski P. 2016. Remote sensing technologies for enhancing forest inventories: a review. *Canadian Journal of Remote Sensing*, 42 (5): 619-641 [DOI: 10.1080/07038992.2016.1207484]
- Xia Q, Li J H, Dai S, Zhang H and Xing X M. 2023. Mapping high-resolution mangrove forests in China using GF-2 imagery under the tide. *National Remote Sensing Bulletin*, 27(6): 1320-1333. (夏清, 李建华, 代硕, 张涵, 邢学敏. 2023. 顾及潮汐影响的中国红树林高分二号遥感制图. *遥感学报*, 27(6): 1320-1333) [DOI: 10.11834/jrs.20221848]
- Xu X D, Li W, Ran Q, Du Q, Gao L R and Zhang B. 2018. Multi-source remote sensing data classification based on convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 56(2): 937-949 [DOI: 10.1109/TGRS.2017.2756851]
- Xu Y J, Zhou J and Zhang Z. 2024. A new Bayesian semi-supervised active learning framework for large-scale crop mapping using Sentinel-2 imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 209: 17-34 [DOI: 10.1016/j.isprsjprs.2024.01.023]
- Zhang L, Luo W T, Zhang H H, Yin X W and Li B. 2023. Classification scheme for mapping wetland herbaceous plant communities using time series Sentinel-1 and Sentinel-2 data. *National Remote Sensing Bulletin*, 27(6): 1362-1375 (张琨, 罗文庭, 张皓寰, 殷秀琬, 李斌. 2023. 时序 Sentinel-1 和 Sentinel-2 数据支持下的鄱阳

湖湿地草本植物群落制图分类. 遥感学报, 27(6):1362-1375)
[DOI: 10.11834/jrs.20222079]

Zheng Y Z, Dong W, Yang Z P, Lu Y H, Zhang X, Dong Y N and Sun
F Q. 2024. A new attention-based deep metric model for crop type

mapping in complex agricultural landscapes using multisource re-
mote sensing data. International Journal of Applied Earth Obser-
vation and Geoinformation, 134: 104204 [DOI: 10.1016/j. jag.
2024.104204]

Forest classification via cross-attention fusion of multispectral and SAR remote sensing data

XIE Yifan, JIA Zihan, ZHANG Xiaoli

1. State Key Laboratory of Efficient Production of Forest Resources, Beijing Forestry University, Beijing 100083, China;
2. Beijing Key Laboratory of Precision Forestry, College of Forestry, Beijing Forestry University, Beijing 100083, China;
3. Key Laboratory of Forest Cultivation and Protection, Ministry of Education, Beijing Forestry University, Beijing 100083, China

Abstract: Remote sensing, with its broad coverage, high timeliness, and ability to acquire multidimensional information, has become a vital tool for forestry surveys. Multispectral remote sensing imagery offers high spatial and spectral resolution, effectively capturing the spectral differences between various ground features. In parallel, Synthetic Aperture Radar (SAR) data provide stable surface structural information and textural features, serving as a crucial complement to spectral data. However, inherent discrepancies in modal structure and information representation between active (SAR) and passive (multispectral) remote sensing data often limit fusion effectiveness, which can ultimately influence classification accuracy. To address this challenge, our study focuses on a specific region in Pu'er City, Yunnan Province. We selected three key forest tree species groups—*Pinus kesiya*, *Eucalyptus*, and *Quercus* spp.—along with one type of economic forest, tea plantation, and three other ground object types as classification targets. Under small-sample conditions, we developed a deep learning network for forest classification that fuses active and passive remote sensing data. Our method, which integrates Sentinel-2 multispectral imagery with Sentinel-1 SAR imagery, proposes a multispectrum-dominated cross-attention fusion network. A key feature is the introduction of a regularized gating mechanism with a learnable dropout ratio, enabling dynamic regulation of self-attention and cross-attention feature fusion. Within this fusion process, the model uses a self-attention mechanism to extract prominent features from multispectral data. Meanwhile, cross-attention guides its response to critical regions in SAR data, dynamically controlling each modality's information contribution at the feature fusion stage. Our experimental results demonstrate that the proposed method achieves optimal performance when compared with diverse fusion strategies across various fusion levels. It yielded an overall classification accuracy of 95.24%. Specifically, the classification precision for *Eucalyptus*, tea plantation, *Quercus* spp., and *P. kesiya* were 96.78%, 94.07%, 91.73%, and 92.90%, respectively. This study successfully validates the effectiveness of the cross-attention mechanism in collaboratively modeling active and passive remote sensing information, offering a viable approach and technical support for multisource remote sensing-based forest classification in complex environments.

Key words: Sentinel-2, Sentinel-1, cross-attention, feature fusion, forest classification, prototypical network, Concrete Dropout

Supported by National Key Research and Development Program of China (No. 2023YFD2201700); National Natural Science Foundation of China (No. 32171779)