

基于改进 HybridSN 网络的融合多尺度特征的多光谱—极化 SAR 影像分类

高永刚, 梁文戈, 刘雨婷, 詹俊杰, 李芳颖

福州大学 环境与安全工程学院, 福州 350108

摘要: 与单一遥感影像相比, 多光谱与合成孔径雷达 SAR (Synthetic Aperture Radar) 影像的融合分类可有效实现数据互补, 提升地物识别精度。HybridSN (Hybrid Spectral CNN) 是一种基于 3D 卷积与 2D 卷积的高光谱影像分类模型, 能有效融合光谱和空间特征以提升分类精度。然而, HybridSN 在处理多光谱与极化 SAR 等异构数据时, 易将 SAR 噪声误判为有效特征, 且难以协调 SAR 的结构信息与多光谱影像的高层语义特征, 导致特征融合效率受限。针对上述问题, 本文提出一种融合混合注意力模块 HAM (Hybrid Attention Module) 与循环层注意力机制 RLA (Recurrent Layer Attention) 的 HAR-HybridSN 模型 (Hybrid and Recurrent Layer Attention HybridSN)。该模型采用 HAM 模块通过通道—空间协同注意力增强关键特征响应并抑制 SAR 噪声干扰, 利用 RLA 模块优化网络深层与浅层特征的融合路径, 从而提升多尺度异质特征的整合能力; 以 GF-1 多光谱与 GF-3 极化 SAR 影像为实验数据, 开展与 HybridSN 等 6 种算法的对比实验验证。研究表明: HAR-HybridSN 的总体精度、平均精度和 Kappa 系数分别高于 93%、91% 和 91%, 显著优于对比模型; 在地物交错及边界模糊的区域, HAR-HybridSN 能有效抑制冗余信息的干扰, 准确识别关键语义特征; 相较于其他模型, 其总体精度提升约 1%, 在建筑、耕地等类别中表现出更优的空间一致性与分类连贯性。

关键词: 多光谱影像, 极化 SAR, 多尺度特征, HAR-HybridSN, 地物分类

中图分类号: TP701/P2

引用格式: 高永刚, 梁文戈, 刘雨婷, 詹俊杰, 李芳颖. 2026. 基于改进 HybridSN 网络的融合多尺度特征的多光谱—极化 SAR 影像分类. 遥感学报, 30(7): 2178–2193

Gao Y G, Liang W G, Liu Y T, Zhan J J and Li F Y. 2026. Improved HybridSN Network for multispectral-polarimetric SAR image classification with multi-scale feature. National Remote Sensing Bulletin, 30(7): 2178–2193 [DOI:10.11834/jrs.20265376]

1 引言

作为遥感影像解译中的核心技术, 影像分类旨在从复杂的遥感数据中提取并识别各类地物, 如建筑物、道路、水体、植被和裸土等, 其在农业管理 (张伟雄 等, 2024)、土地利用 (刘小华, 2024)、森林覆盖评估 (薛宇飞 等, 2024) 等领域中发挥着重要作用。多光谱影像由于蕴含丰富的光谱特征和纹理特征, 已被广泛应用于地物分类中, 但其成像质量易受云层等气象条件的影响, 从而导致影像质量下降甚至影像缺失, 使其分类精度受到限制。相较于多光谱影像, 合成孔径雷

达 SAR (Synthetic Aperture Radar, SAR) 具备全天时、全天候及强穿透能力的优势, 其成像质量不受天气因素的制约, 且数据获取较为稳定; 然而, SAR 数据极易受到干涉效应的影响, 致使其影像解译较难且成本较高。因此, 将 SAR 与多光谱影像融合应用成为现阶段分类研究的重要趋势 (明义森 等, 2023)。

融合多源遥感影像的分类研究始于 20 世纪 90 年代, 早期研究已证实 SAR 与多光谱协同分类可显著提升精度 (Solberg 等, 1996; Huang 等, 2007)。Xu 和 Gong (2007) 探索了 EO-1 高光谱与多光谱数据的协同潜力。Schmitt 和 Zhu (2016) 验

收稿日期: 2025-09-12; 预印本: 2026-05-14

基金项目: 福建省自然科学基金 (编号: 2023J01064)

第一作者简介: 高永刚, 研究方向为遥感图像处理与应用。E-mail: yggao@fzu.edu.cn

通信作者简介: 刘雨婷, 研究方向为 SAR 图像处理与应用。E-mail: 1942249903@qq.com

证了多光谱与SAR融合在土地覆盖制图中的有效性。Joshi等(2016)指出,特征级融合最能发挥多模态互补优势。赵诣和蒋弥(2019)利用多光谱影像引导SAR同质区选取,显著提升了分类性能。Montanaro等(2022)将半监督学习引入多模态框架,缓解了标注样本稀缺的问题。此外,融合策略可有效缓解不透水层提取中的光谱混淆现象(Gamba和Aldrichi, 2012; Zhang和Xu, 2018),凸显其在复杂城市环境中的潜力。

近年来,深度学习凭借强大的自动特征提取能力,推动了多源遥感影像分类的发展(Davari等, 2022; Sisay等, 2023)。然而,多光谱与SAR影像存在显著的非线性辐射差异与模态鸿沟,直接融合易引入噪声或导致特征冲突(Hu等, 2024)。为此,主流方法采用双分支网络分别提取异构特征,再通过融合模块实现信息整合(Xu等, 2024)。Zhang等(2019)采用线性拼接融合多模态特征,虽易于实现,但缺乏冗余抑制能力。Shu等(2022)引入一致性约束以保持结构对齐,却可能抑制互补信息传递。Sun等(2024)利用通道—空间注意力增强特征响应,但忽视跨层一致性,易丢失边缘细节。近期,Deng等(2024)提出FASONet (Feature Alignment-Based SAR and Optical Image Fusion Network),通过特征对齐模块和多模态挤压激励模块实现自适应融合。Zhang等(2025)提出M²Caps模型,结合胶囊机制与级联融合,在WHU-OPT-SAR和GF-2/GF-3数据集上显著提升精度,验证了多尺度建模的有效性。可见,当前多模态融合的核心挑战在于:如何在增强跨模态互补性的同时,有效抑制噪声与冗余,并保持多尺度特征的空间与语义一致性。

针对上述问题,本文提出HAR-HybridSN (Hybrid and Recurrent Layer Attention HybridSN),一种面向多光谱—极化SAR融合分类的改进模型。该模型通过引入混合注意力模块HAM (Hybrid Attention Module) (Li等, 2022)利用通道—空间协同注意力增强关键特征响应并抑制SAR噪声干扰,采用循环层注意力机制RLA (Recurrent Layer Attention) (Fang等, 2023)优化网络深层与浅层特征的融合路径,促进局部细节与全局语义的协同表达,从而提升多尺度异质特征的整合能力;并以GF-1多光谱与GF-3极化SAR影像为实验数据,开展与HybridSn等6种算法的对比实验验证,

以期为高精度土地覆盖分类提供新的技术路径。

2 研究方法

2.1 HAR-HybridSN模型结构

HybridSN (Hybrid Spectral CNN)是一种融合3D卷积与2D卷积的深度网络,其先通过3D卷积捕获空间—光谱联合特征,再通过2D卷积增强空间特征表达并降低计算复杂度,最终通过全连接层实现分类,具有结构简洁、计算高效的优势(Roy等, 2020)。然而,在多光谱与极化SAR异构数据融合场景下,该模型因缺乏显式的噪声抑制机制与特征选择机制,易将SAR相干斑噪声误判为有效特征;同时,其线性堆叠结构限制了跨层级语义交互,使得模型难以协调SAR的结构化散射特性与多光谱数据的高层语义信息,导致模态间融合不充分,影响分类鲁棒性。为解决这些问题,本文在保留HybridSN主干架构优势的基础上,引入混合注意力模块HAM和循环层注意力机制RLA,并据此提出改进模型HAR-HybridSN,以增强跨模态特征的判别能力与多层级特征的融合能力。HAR-HybridSN模型结构如图1所示。

HAM是一种用于深度卷积神经网络的轻量化注意力模块。由于HAM的通道注意力中没有使用多层感知机MLP (Multilayer Perceptron),在空间注意力中也没有冗余卷积,使其相较于卷积块注意力模块CBAM (Convolutional Block Attention Module) (Woo等, 2018)所需训练参数量更少、计算量更低,并在减少显存占用的同时能加速模型收敛。此外,在通道注意力中,HAM可以自适应调整平均池化特征与最大池化特征的权重比例,从而实现两种池化特征的动态融合。在SAR影像处理中,该机制既可以抑制通道内的随机相干斑噪声,又可以避免噪声峰值被误判为特征,从而筛选出噪声占比低的有效信号,减少后续空间处理的噪声输入。而CBAM直接拼接两种池化特征,通过MLP生成通道注意力,未考虑不同特征提取阶段中两种池化的重要性差异。与高效通道注意力ECA (Efficient Channel Attention) (Wang等, 2020)相比,HAM可以通过空间注意力聚焦于图像中的关键区域,并且其采用更大的卷积核(7×7)捕捉全局纹理信息,同时保留局部细粒度特征,有效维持纹理连续性。考虑到HAM的特性,本文将其嵌

入于 HybridSN 的 2D 卷积层之前,使模型融合通道注意力与空间注意力(图2),以实现关键特征的自适应增强与对噪声干扰的有效抑制,从而显

著提升模型在低信噪比、地物边界模糊等复杂场景下的鲁棒性与分类精度。

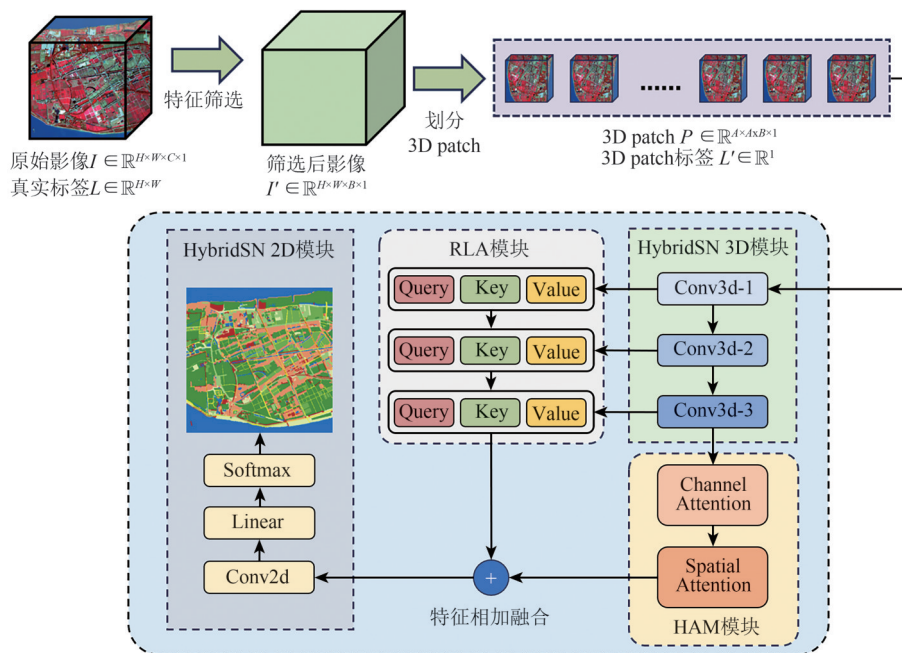


图1 HAR-HybridSN模型的结构

Fig. 1 The architecture of the HAR-HybridSN model

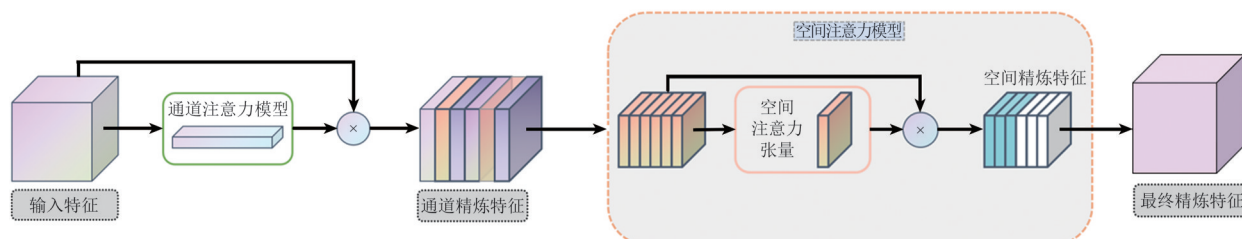


图2 本文所使用HAM的构架图

Fig. 2 The architecture of the HAM adopted in this paper

RLA 是一种用于增强深度神经网络层间交互的注意力机制,其具有跨层检索相关信息提升模型的代表能力。ResNet (He 等, 2016) 中的残差连接和 DenseNet (Huang 等, 2017) 中的密集连接缺乏对层间信息的筛选,而 RLA 通过注意力机制量化各前层对当前层的重要性,使其跨层交互更精准。RLA 打破了传统线性堆叠结构的限制,有效解决了深层网络中特征传递不充分的问题,因此本文使用 RLA (图3) 以实现浅层细节(如纹理、边缘)与深层语义(如类别结构)的高效融合,从而增强模型对局部细节与全局上下文的综合建模能力,提升对复杂地物类别的识别性能。

2.2 HAR-HybridSN 模型实现流程

假设原始影像数据为 $I \in \mathbb{R}^{H \times W \times C \times 1}$, 真实标签为 $L \in \mathbb{R}^{H \times W}$, 其中 H 与 W 分别代表输入影像的高度和宽度, C 代表输入特征的数量(即影像深度)。为了提高分类精度和计算速度,避免所谓的“维度诅咒”或高维数据冗余造成分类精度下降等问题,需对原始影像数据进行特征筛选。特征筛选时,首先利用 Filter 方法对特征进行初步筛选去除低相关性特征,然后使用 XGBoost 算法评估特征的重要性并优选判别性强的特征,最后采用 SuperPCA (Jiang 等, 2018) 进一步压缩维度并保留最优空间结构特征,实现高维特征的有效降维与判别力增

强, 获得筛选后的影像 $I' \in \mathbb{R}^{H \times W \times B \times 1}$ (B 代表特征筛选后的特征数量)。将 I' 进一步划分为一系列 3D patch $P \in \mathbb{R}^{A \times A \times B \times 1}$, 其中 A 代表 3D patch 的高度和宽度。每个 3D patch 的中心像元的标签确定了该 3D patch 整体的标签 $L' \in \mathbb{R}^1$ 。以 P 和 L' 为输入, 使用 HybridSN 模型中的 3D 卷积网络对 P 进行卷积, 提取特征 F_1 、 F_2 和 F_3 。卷积公式如下所示:

$$F_1 = \text{ReLU}(\text{BN}(\text{Conv3d}_1(P))) \quad (1)$$

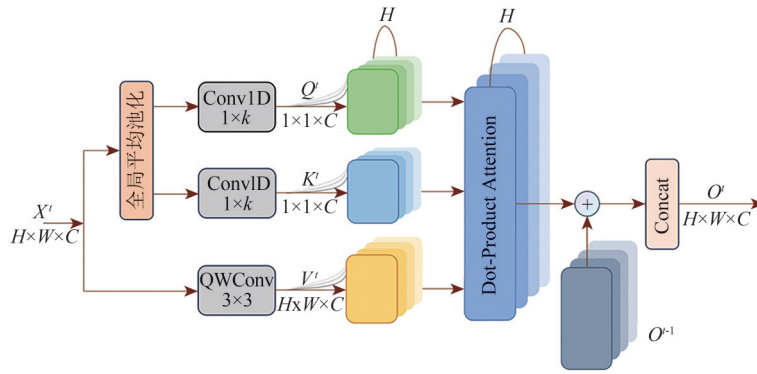


图3 本文所使用RLA的构架图

Fig. 3 The architecture of the RLA adopted in this paper

由于后续执行HAM和RLA都是在二维特征图上进行, 因此需要调整输出特征的维度。假设 F_3 的维度为 $\mathbb{R}^{M \times M \times D \times 32}$ 。式中: M 代表特征图的高度和宽度; D 代表特征图的深度; M 和 D 都由 3D 卷积核的尺寸决定; F_3 的通道数为 32, 与 3D 卷积层中第三层的输出通道数一致。将三维特征 $F_3 \in \mathbb{R}^{M \times M \times D \times 32}$ 的深度维度和通道维度进行合并, 以转换为二维特征 $F_3 \in \mathbb{R}^{M \times M \times 32D}$, 便于后续计算。

通道注意力机制CA (Channel Attention) 通过显式建模各通道对最终判别结果的贡献程度, 能够自适应地增强具有判别性的特征通道。CA以二维特征 F_3 作为输入, 具体计算步骤如下:

(1) 全局池化: 对 F_3 的空间维度进行全局池化, 得到通道级的全局信息, 全局池化可以通过平均池化 (Average Pooling) 和最大池化 (Max Pooling) 实现, 分别得到 2 个向量 F_{avg} 和 F_{max} 。公式如下:

$$F_{\text{avg}} = \frac{1}{M \times M} \sum_{i=1}^M \sum_{j=1}^M F_3(i, j) \quad (3)$$

$$F_{\text{max}} = \max_{i,j} F_3(i, j) \quad (4)$$

(2) 权重计算: 将 F_{avg} 和 F_{max} 输入到一个共享的全连接层 FC (Fully Connected Layer), 输出 2 个权重向量。后将这 2 个权重向量相加并通过 Sigmoid

$$F_i = \text{ReLU}(\text{BN}(\text{Conv3d}_i(F_{i-1}))), i = 2, 3 \quad (2)$$

式中, i 代表层数; F_i 为第 i 层的输出特征; $\text{BN}(\cdot)$ 表示批归一化 (Batch Normalization) 层; $\text{ReLU}(\cdot)$ 表示线性整流函数 (Rectified Linear Unit) 层; $\text{Conv3d}_i(\cdot)$ 表示第 i 层的 3D 卷积层, 第一、二、三层的输入通道数分别为 1、8、16, 输出通道数分别为 8、16、32, 卷积核大小分别为 $3 \times 3 \times 7$ 、 $3 \times 3 \times 5$ 、 $3 \times 3 \times 3$ 。

激活函数计算, 输出通道注意力权重 A_c , 其公式如下:

$$A_c = \text{Sigmoid}(\text{FC}(F_{\text{avg}}) + \text{FC}(F_{\text{max}})) \quad (5)$$

(3) 特征输出: 之后将 A_c 与 F_3 逐元素相乘, 输出通道精炼特征 F' , $\odot$ 表示哈达玛积, 计算公式如下:

$$F' = A_c \odot F_3 \quad (6)$$

空间注意力机制SA (Spatial Attention) 旨在捕捉特征图在空间维度上的显著性。其以通道精炼特征 F' 为输入, 计算步骤如下:

(1) 特征融合: 首先, 对 F' 在通道维度上进行平均池化和最大池化操作, 分别得到 2 个向量 F'_{avg} 和 F'_{max} , 其计算公式如下:

$$F'_{\text{avg}} = \frac{1}{32D} \sum_{k=1}^{32D} F'(:, :, k) \quad (7)$$

$$F'_{\text{max}} = \max_k F'(:, :, k) \quad (8)$$

(2) 卷积运算: 将 F'_{avg} 和 F'_{max} 沿通道维度进行拼接 (Concatenate), 然后输入进一个卷积层, 再通过 Sigmoid 激活函数计算, 输出空间注意力权重 A_s , 其计算公式如下:

$$A_s = \text{Sigmoid}(\text{Conv}(\text{Concat}[F'_{\text{avg}}, F'_{\text{max}}])) \quad (9)$$

(3) 特征输出: 将 A_s 与 F' 逐元素相乘, 输出

最终精炼特征 F_{HAM} , 计算公式如下:

$$F_{HAM} = A_s \odot F' \quad (10)$$

混合注意力模块将通道注意力和空间注意力结合起来, 通过依次引入CA和SA, 使模型同时考虑通道维度和空间维度的特征重要性。这种方式可以提升模型对不同图像区域的空间特征及各通道特征的表达能力, 尤其适用于处理具有复杂特征的图像分类任务。

RLA机制用于增强不同卷积层之间的信息交互。将3D卷积网络中各层输出的三维特征的深度维度和通道维度进行合并, 以转换为二维特征。以第*i*层的二维特征 F_i 为输入, 计算第*i*层注意力机制的查询 Q^i 、键 K^i 和值 V^i , 其公式如下所示:

$$Q^i = f_Q^i(F_i) \quad (11)$$

$$K^i = \text{Concat}[f_K^i(F_1), f_K^i(F_2), \dots, f_K^i(F_i)] \quad (12)$$

$$V^i = \text{Concat}[f_V^i(F_1), f_V^i(F_2), \dots, f_V^i(F_i)] \quad (13)$$

式中: $f_Q^i(\cdot)$ 是提取当前层 (第*i*层) 信息的函数; $f_K^i(\cdot)$ 和 $f_V^i(\cdot)$ 是从第一层到第*i*层的特征中提取信息的函数; $\text{Concat}[\cdot]$ 表示将多个向量拼接在一起。之后, 采用一种递归的方式计算第*i*层的层注意力 O_i , 公式如下所示:

$$O_i = \lambda_o \odot O_{i-1} + Q^i (K^i)^T V^i \quad (14)$$

式中, λ_o 是一个可训练的向量, 如果*i*为1, 则 O_{i-1} 等于0。之后将第三层的层注意力 O_3 与 F_3 相加, 输出最终的层注意力特征 F_{RLA} 。 F_{RLA} 与 F_{HAM} 相加便可输出最终特征 F_{final} , 计算公式如下:

$$F_{RLA} = O_3 + F_3 \quad (15)$$

$$F_{final} = F_{RLA} + F_{HAM} \quad (16)$$

最终, F_{final} 输入至 HybridSN 模型中的 2D 卷积网络, 该网络仅有一层, 卷积核尺寸为 3×3。输出结果经过批归一化和 ReLU 激活函数计算后, 再通过由线性层和 Softmax 层组成的分类器 $\Phi_{classifier}$ 进行分类处理, 得到 3D patch P 预测的分类结果 L_{pred} 。公式如下所示:

$$L_{pred} = \Phi_{classifier}(\text{ReLU}(\text{BN}(\text{Conv2d}(F_{final})))) \quad (17)$$

综上, HAR-HybridSN 在保留 HybridSN 光谱—空间联合建模优势的基础上, 融合混合注意力与循环层注意力机制, 在通道—空间特征提取和层级语义融合方面实现多维度提升。其具体流程如下:

模型: HAR-HybridSN 模型

输入: 原始影像 $I \in \mathbb{R}^{H \times W \times C \times 1}$, 真实标签 $L \in \mathbb{R}^{H \times W}$ 。

输出: 完成训练的模型。

步骤:

- (1) 初始化: 迭代次数 Epoch=100; 学习率 $r=0.001$; 批大小 Batch Size=128;
- (2) 对原始影像进行特征筛选得到 $I' \in \mathbb{R}^{H \times W \times B \times 1}$;
- (3) 在特征筛选后的影像上划分 3D patch, 得到 3D patch $P \in \mathbb{R}^{A \times A \times B \times 1}$ 和 3D patch 标签 $L' \in \mathbb{R}^1$;
- (4) 当训练次数小于 Epoch 时, 依次执行以下操作, 否则算法跳转至步骤 13
- (5) 对 P 执行 3 次 3D 卷积操作, 输出三维特征 F_1 、 F_2 和 F_3 , 将这些特征的维度调整至二维;
- (6) 计算二维特征 F_1 、 F_2 和 F_3 之间的层注意力, 得到特征 O_3 ;
- (7) O_3 与 F_3 进行相加得到最终的层注意力特征 F_{RLA} ;
- (8) 以二维特征 F_3 为输入, 执行通道注意力模块, 输出通道精炼特征 F' ;
- (9) 以通道精炼特征 F' 为输入, 执行空间注意力模块, 输出最终精炼特征 F_{HAM} ;
- (10) 将 F_{RLA} 和 F_{HAM} 进行相加融合输出最终特征 F_{final} ;
- (11) 以最终特征 F_{final} 为输入, 执行 2D 卷积, 并使用线性层和 Softmax 函数预测分类结果 L_{pred} ;
- (12) 使用交叉熵损失函数计算预测结果 L_{pred} 和标签 L' 的损失, 并使用优化器更新模型参数, 进行下一轮训练;
- (13) 训练完模型后, 保存模型权重参数, 方便后续进行测试。

2.3 评价指标

本文采用总体精度 OA (Overall Accuracy)、平均精度 AA (Average Accuracy)、Kappa 系数、生产者精度 PA (Producer's Accuracy)、用户精度 UA (User's Accuracy) 和 F1-Score 作为评价指标。具体计算公式如下:

$$OA = \frac{\sum_{i=1}^n P_{ii}}{N} \quad (18)$$

$$AA = \frac{1}{n} \sum_{i=1}^n \left(\frac{P_{ii}}{P_{+i}} \right) \quad (19)$$

$$\text{Kappa} = \frac{N \sum_{i=1}^n p_{ii} - \sum_{i=1}^n p_{i+} p_{+i}}{N^2 - \sum_{i=1}^n p_{i+} p_{+i}} \quad (20)$$

$$\text{PA} = \frac{\sum_{i=1}^n p_{ii}}{\sum_{i=1}^n p_{+i}} \quad (21)$$

$$\text{UA} = \frac{\sum_{i=1}^n p_{ii}}{\sum_{i=1}^n p_{i+}} \quad (22)$$

$$\text{F1-Score} = 2 \times \frac{\text{UA} \times \text{PA}}{\text{UA} + \text{PA}} \quad (23)$$

式中, N 为总样本数; n 为分类任务中包含 n 个类别; p_{ii} 为实际类别为 i , 且正确分类为 i 的样本数

量; p_{+i} 为实际类别为 i 的样本总数; p_{i+} 为分类类别为 i 的样本总数。

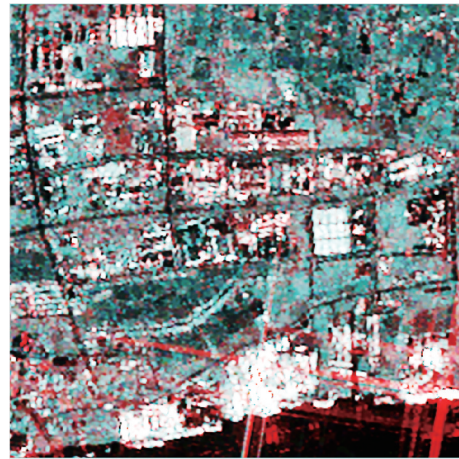
3 实验过程与结果分析

3.1 研究数据

本文以空间分辨率为 8 m 的 GF-1 PMS 多光谱影像 (2023 年 10 月 18 日) 和 GF-3 QPSI SAR 影像 (2023 年 10 月 13 日) 为数据源 (图 4)。研究区覆盖典型城乡交错带, 地物类型复杂, 涵盖农业、林业、城市与水体等多种覆盖类型。为确保分类体系的系统性与地理可解释性, 依据实地调查与高分辨率参考影像, 将研究区划分为 8 类地物: 农作耕地、休耕裸地、有林地、灌木林地、交通运输用地、建筑用地、水域以及其他。



(a) 高分一号多光谱影像
(a) GF-1 multispectral image



(b) 高分三号 SAR 影像
(b) GF-3 SAR

图 4 研究区高分一号多光谱影像和高分三号 SAR 影像

Fig. 4 Multispectral image of GF-1 and GF-3 SAR image of study area

3.2 特征集构建

为充分挖掘多源遥感数据的判别潜力, 本文构建了融合光谱、纹理、极化与指数特征的多维度特征集。具体而言, 针对 GF-1 影像, 本文提取 NDVI 指数 (Huete 等, 2002) 和 NDWI 指数 (McFeeters, 1996)。针对 GF-3 极化 SAR 影像, 本文提取多种极化分解特征: Pauli 分解 (Lee 和 Pottier, 2009)、Freeman 三份量分解 (Freeman 和 Durden, 1998)、Yamaguchi 分解 (Yamaguchi 等, 2005)、H/A/Alpha 分解 (Cloude 和 Pottier, 1996)、An 分解 (An 等, 2010)、MCSM (Multiple-Component

Scattering Model) 分解 (Zhang 等, 2008)、6SD (Six-component Scattering matrix power Decomposition) 分解 (Singh 和 Yamaguchi, 2018) 和 Bhattacharya 分解 (Bhattacharya 等, 2015)。此外, 为增强地物的空间结构表征能力, 本文采用灰度共生矩阵 GLCM (Gray-Level Co-occurrence Matrix) (Guan 等, 2015) 从多角度提取纹理特征, 包括对比度、同质性、能量、相关性 with 粗糙度等。最终, 本文结合原始波段与上述衍生特征, 构建了一个包含 120 个特征的多模态特征数据集, 各类特征信息详见表 1。

表1 用于分类的特征信息

Table 1 Feature information for classification

特征类别	特征数目	特征名称
原始波段	4	GF-1:R,G,B,NIR
后向散射系数	4	GF-3:VH,VV,HV,HH
指数特征	2	GF-1:NDVI,NDWI
纹理特征	32	GF-1:R,G,B,NIR,PAN的纹理特征
	32	GF-3:VH,VV,HV,HH的纹理特征
极化分解特征	9	GF-3:T3矩阵
	3	GF-3:Pauli分解特征
	3	GF-3:Freeman分解特征
	4	GF-3:Yamaguchi分解特征
	7	GF-3:H/A/Alpha全极化分解特征
	4	GF-3:An分解特征
	6	GF-3:MCSM分解特征
	6	GF-3:6SD分解特征
	4	GF-3:Bhattacharya分解特征

3.3 实验参数设置

为保证对比实验的公平性，针对HAR-HybridSN和其他对比模型，本文使用了相同的随机种子，所有模型的初始权重采用了随机初始化，并训练相同的Epoch（100轮）。在所有实验中，原始数据的H、W和C分别设置为500、500和120。特征筛选后数据的特征数量B为16。利用滑动窗口方式进行3D patch划分，其步长设置为1；同时，为了保证边缘像素能被提取，对特征筛选后数据进行填充（padding）处理，其3D patch的高度和宽度均为15。实验采用统一的数据集划分策略，其中训练集与测试集像元数量比例为7：3，且未使用任何数据增强手段。为实现整幅影像全覆盖预测，同时避免空间数据泄漏，本文设计并采用基于空间隔离的交叉验证方案。具体而言，针对500×500的研究区影像，保持宽度不变，依次选取高度1—150、151—300、301—450的子区域（150×500）作为测试集，每次均以除当前测试区外的其余区域作为训练集，例如选取高度151—300的子区域作为测试集，那么高度范围1—150以及301—500的子区域则作为训练集，二者像元数量比例为7：3。仅在训练集内提取3D patch用于模型训练，再以训练完成的模型对当前测试集进行逐像元预测。单次预测结束后重置模型参数，避免数据泄漏，并继续选取下一高度区间的子区域开展模型训练

与预测。需要说明的是，高度451—500的剩余边缘区域（50×500）统一纳入最后一次测试。待全部子区域均完成测试与预测后，将各子区域结果按空间位置拼接，最终得到整幅影像的完整预测结果。HAR-HybridSN运行在配备32 GB RAM、Intel Core i5-12600KF处理器和NVIDIA RTX 4060 Ti 16GB的计算机上。代码基于Python 3.9，采用PyTorch框架实现。HAR-HybridSN训练时使用的批大小（Batch Size）为128，优化器为Adam（学习率0.001，权重衰减 1×10^{-6} ）。

3.4 结果与分析

为验证HAR-HybridSN在地物分类中的性能优势，本文选取PolSARFormer（Jamali等，2023）、DeepCNN（Deep Convolutional Neural Network）（Jamali等，2022）、HybridSN、CRPM-Net（Xiao等，2020）、MP-ResNet（Multipath Residual Network）（Ding等，2022）和MCPT（Mixed Convolutional Parallel Transformer）（Wang等，2023）6种深度学习模型作为对比方法，从定性和定量两个层面进行分析。

3.4.1 分类结果的定性分析

图5展示了各模型的分类结果。总体而言，所有深度学习模型均能较好地识别主要地物类别，分类结果与参考影像吻合度较高，整体表现良好，反映出较强的分类能力。为更细致地评估模型在地物边界与局部结构提取方面的性能差异，本文对关键区域进行局部放大分析，放大结果详见图6。

从图6可以看出，7种模型的分类结果中均存在不同程度的椒盐噪声现象。其中，HAR-HybridSN的分类结果最为平滑，图像质量更优；CRPM-Net分类结果中的椒盐噪声最为显著，对影像解译造成明显影响；DeepCNN、PolSARFormer、HybridSN及MCPT的椒盐噪声程度介于上述两者之间，表现中等。

在水体识别方面，MP-ResNet与CRPM-Net的分类表现最差，存在较明显的漏分与错分问题；HAR-HybridSN的水体识别表现最优，其分类结果与参考影像的一致性最高；其余算法的水体识别精度则处于中等水平。

在“其他”地物的识别上，MP-ResNet的表

现最差，不仅存在部分“其他”地物被错分为水体的问题，还出现了建筑用地被误判为“其他”地物的情况；PolSARFormer与DeepCNN同样存在

明显的错分与漏分问题，分类表现较差；HAR-HybridSN的分类边界清晰、错分最少，整体与参考影像的一致性最高，表现最优。

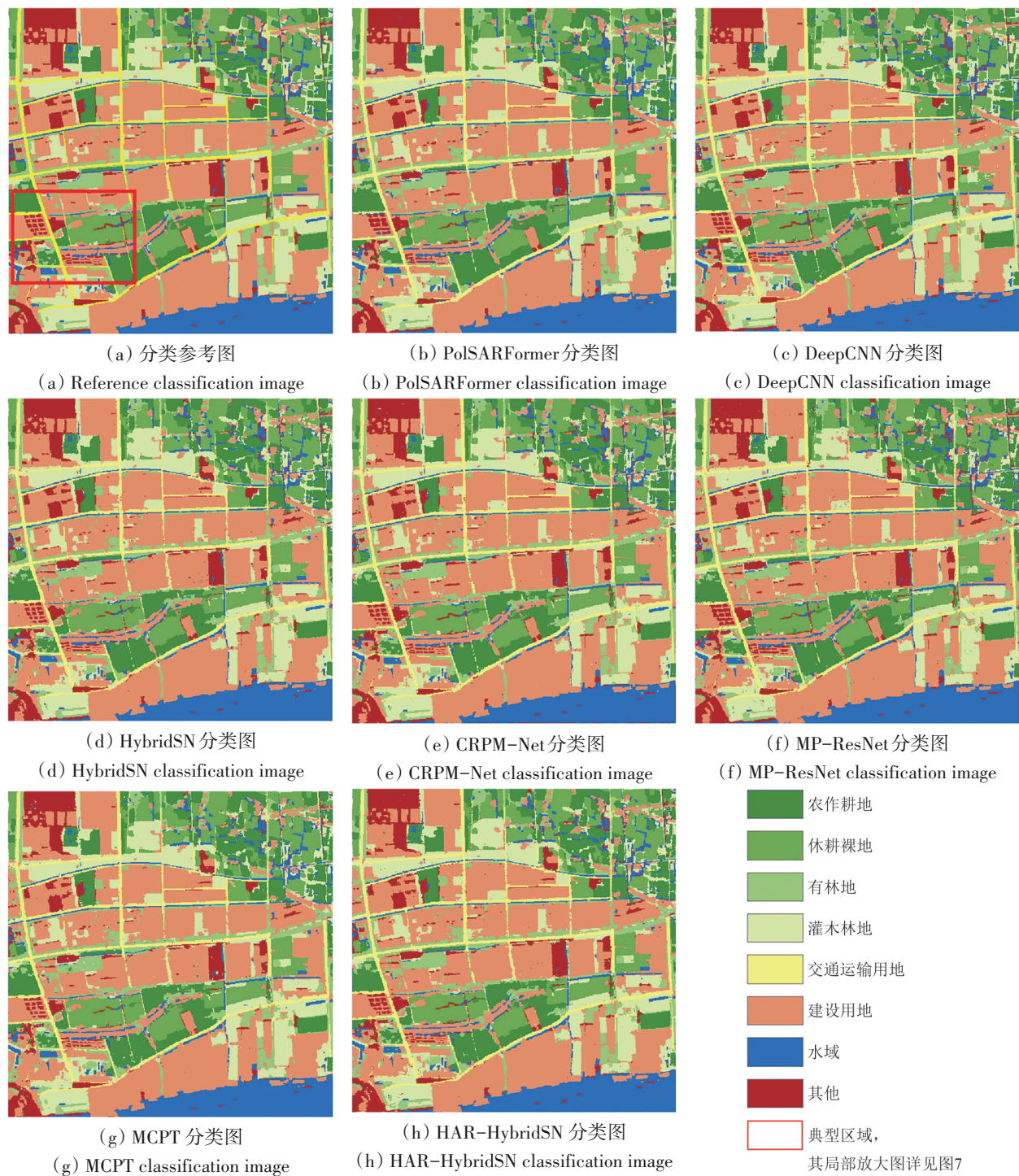


图5 各模型的分类结果图
Fig. 5 Classification results for each model

对于农作耕地的识别，HAR-HybridSN表现最优，分类边界清晰且精准；HybridSN表现次之，也能较好区分农作耕地与其他类别；PolSARFormer与MP-ResNet存在部分有林地被误分为耕地的情况，

而DeepCNN则存在农作耕地被误分为有林地的问题；此外，MCPT、MP-ResNet及HybridSN在该区域的分类结果中存在较多椒盐噪声，并且地物边界比较模糊，表现中等。

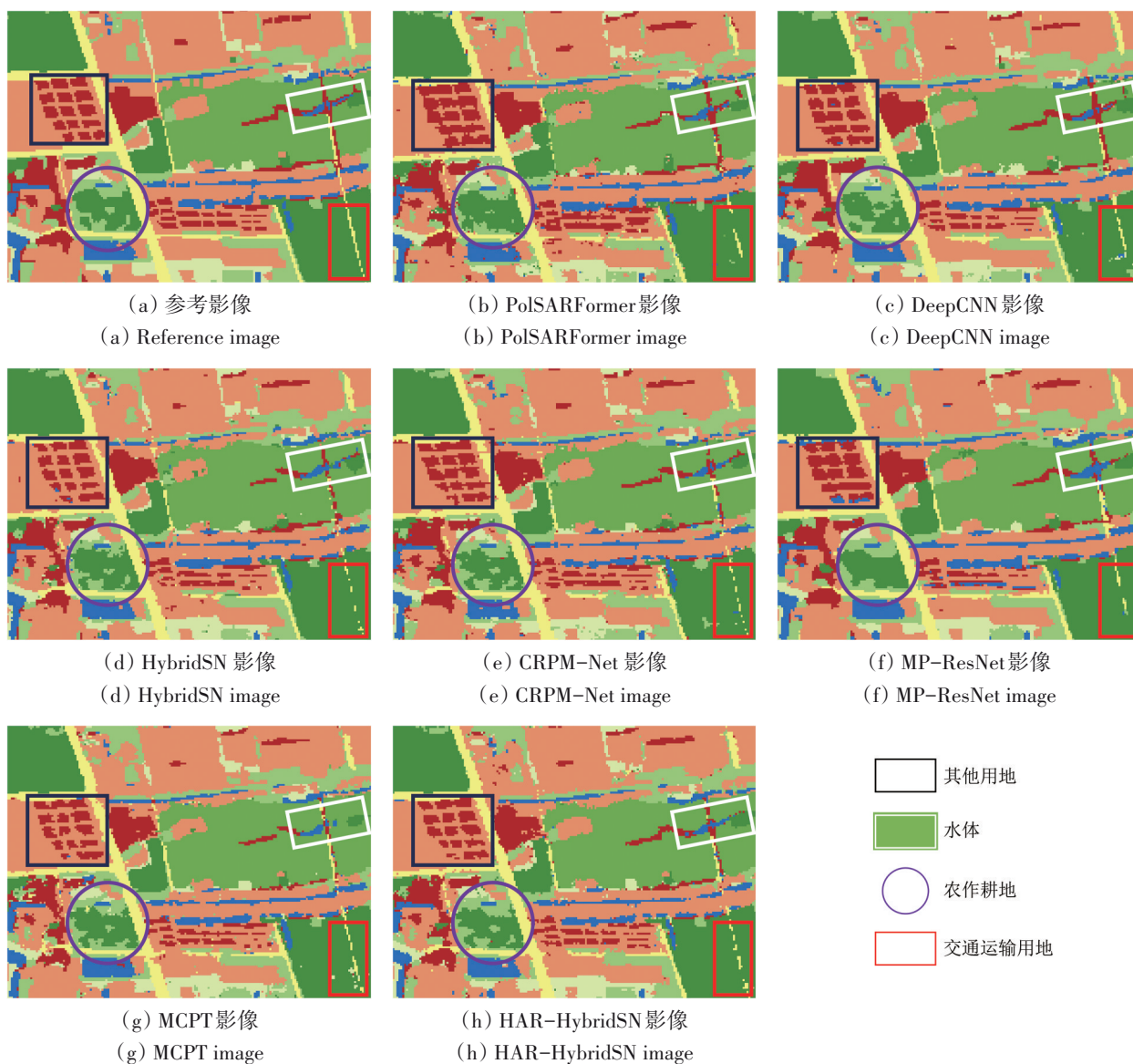


图6 各种模型的分类结果的典型地物区域局部放大图

Fig. 6 Local enlarged views of the classification results of various models

在交通运输用地的识别方面, HAR-HybridSN 的识别结果最为精准, 地物形态与参考影像的一致性最高; MCPT 虽能基本识别出交通运输用地, 但存在较多误分问题, 部分农作耕地被误判为交通运输用地; 而 PolSARFormer 主要存在漏分问题; 其余算法不仅存在错分与漏分问题, 还伴随地物边界模糊及椒盐噪声现象, 整体表现较差。

在有林地与灌木林地的区分上, CRPM-Net 的表现最差, 其误分与漏分问题较为突出; 其余算法的分类表现则处于中等水平。7 种算法在休耕裸地分类中整体表现良好, 其分类结果与参考影像的差异较小; 仅在个别区域存在轻微的错分或漏分现象, 整体而言, 各算法间的差异不显著。

综上, HAR-HybridSN 在各类地物识别中均表现出更优的稳定性与判别能力, 得益于混合注意力模块 HAM 与循环层注意力机制 RLA, 模型能有效增强关键特征、抑制噪声干扰, 实现多源特征的高效融合。相较而言, PolSARFormer 与 HybridSN 整体精度一般; MP-ResNet 与 CRPM-Net 在边界清晰度与类别区分度上表现欠佳; MCPT 与 DeepCNN 处于中等水平。HAR-HybridSN 展现出更均衡的泛化性能与应用潜力。

3.4.2 分类结果的定量评价

分析表 2 可知, HAR-HybridSN 在多数地物类别的生产者精度与用户精度上表现最优, 其中在

有林地（0.9386）、交通运输用地（0.9514）和建筑用地（0.9772）的生产者精度均为第一，表明其在复杂地物识别中具有良好的判别能力与分类稳健性。相比之下，PolSARFormer、DeepCNN、HybridSN、CRPM-Net与MP-ResNet在有林地和其他类别上精度偏低，尽管在建筑用地、休耕裸地等简单类别中表现尚可，但整体性能仍有不足。

从错分与漏分误差看，HAR-HybridSN在多数类别中误差最低，分类可靠性最高。其在建筑用地与休耕裸地的错分误差分别为0.0457和0.0239，显著优于其他对比模型；而HybridSN、CRPM-Net与MP-ResNet在复杂地物上误差较高，且伴随地物边界模糊与类别混淆问题，反映出其对高维冗余特征的建模能力有限。

表 2 各模型的分类结果精度评价

Table 2 Accuracy assessment for classification results of various models

分类模型	类别	生产者精度	用户精度	漏分误差	错分误差	F1-Score
PolSARFormer	农作耕地	0.9365	0.9166	0.0635	0.0834	0.9264
	休耕裸地	0.9408	0.9337	0.0592	0.0663	0.9372
	有林地	0.8503	0.8528	0.1497	0.1472	0.8516
	灌木林地	0.8922	0.9008	0.1078	0.0992	0.8965
	交通运输用地	0.8610	0.8879	0.1390	0.1121	0.8743
	建筑用地	0.9517	0.9429	0.0483	0.0571	0.9473
	水域	0.9115	0.9179	0.0885	0.0821	0.9147
	其他	0.8397	0.8729	0.1603	0.1271	0.8560
DeepCNN	农作耕地	0.9223	0.9341	0.0777	0.0659	0.9282
	休耕裸地	0.9586	0.9127	0.0414	0.0873	0.9351
	有林地	0.8620	0.8334	0.1380	0.1666	0.8475
	灌木林地	0.8733	0.9159	0.1267	0.0841	0.8941
	交通运输用地	0.8720	0.8626	0.1280	0.1374	0.8673
	建筑用地	0.9459	0.9473	0.0541	0.0527	0.9466
	水域	0.9024	0.9303	0.0976	0.0697	0.9161
	其他	0.8706	0.8415	0.1294	0.1585	0.8558
HybridSN	农作耕地	0.9417	0.9212	0.0583	0.0788	0.9313
	休耕裸地	0.9466	0.9461	0.0534	0.0539	0.9463
	有林地	0.8580	0.8612	0.1420	0.1388	0.8596
	灌木林地	0.9018	0.9102	0.0982	0.0898	0.9060
	交通运输用地	0.8798	0.9078	0.1202	0.0922	0.8936
	建筑用地	0.9518	0.9486	0.0482	0.0514	0.9502
	水域	0.9370	0.9175	0.0630	0.0825	0.9271
	其他	0.8594	0.8897	0.1406	0.1103	0.8743
CRPM-Net	农作耕地	0.8956	0.9291	0.1044	0.0709	0.9120
	休耕裸地	0.9440	0.8897	0.0560	0.1103	0.9160
	有林地	0.8168	0.8187	0.1832	0.1813	0.8178
	灌木林地	0.8464	0.8672	0.1536	0.1328	0.8566
	交通运输用地	0.8416	0.8536	0.1584	0.1464	0.8476
	建筑用地	0.9451	0.9300	0.0549	0.0700	0.9357
	水域	0.8760	0.9151	0.1240	0.0849	0.8951
	其他	0.8361	0.7642	0.1639	0.2358	0.7985

						续表
分类模型	类别	生产者精度	用户精度	漏分误差	错分误差	F1-Score
MP-ResNet	农作耕地	0.9144	0.9318	0.0856	0.0682	0.9230
	休耕裸地	0.9444	0.9315	0.0556	0.0685	0.9379
	有林地	0.8245	0.8456	0.1755	0.1544	0.8349
	灌木林地	0.8718	0.9137	0.1282	0.0863	0.8922
	交通运输用地	0.8864	0.8671	0.1136	0.1329	0.8766
	建筑用地	0.9605	0.9278	0.0395	0.0722	0.9439
	水域	0.9111	0.9249	0.0889	0.0751	0.9180
	其他	0.8245	0.8760	0.1755	0.1240	0.8495
MCPT	农作耕地	0.9082	0.9479	0.0918	0.0521	0.9276
	休耕裸地	0.9349	0.9575	0.0651	0.0425	0.9461
	有林地	0.9290	0.8059	0.0710	0.1941	0.8630
	灌木林地	0.9279	0.9093	0.0721	0.0907	0.9185
	交通运输用地	0.8698	0.9341	0.1302	0.0659	0.9008
	建筑用地	0.9587	0.9547	0.0413	0.0453	0.9567
	水域	0.9507	0.9066	0.0493	0.0934	0.9288
	其他	0.8199	0.9063	0.1801	0.0937	0.8609
HAR-HybridSN	农作耕地	0.9150	0.9430	0.0850	0.0570	0.9288
	休耕裸地	0.9319	0.9761	0.0681	0.0239	0.9535
	有林地	0.9386	0.7978	0.0614	0.2022	0.8625
	灌木林地	0.8859	0.9323	0.1141	0.0677	0.9085
	交通运输用地	0.9514	0.9074	0.0486	0.0926	0.9289
	建筑用地	0.9772	0.9543	0.0228	0.0457	0.9656
	水域	0.9491	0.9314	0.0509	0.0686	0.9402
	其他	0.8096	0.9563	0.1904	0.0437	0.8768

在 F1-Score 指标上, HAR-HybridSN 在大多数类别中均超过 0.90, 建筑用地与休耕裸地表现尤为突出; MCPT 在建筑用地与水域表现较好, 但在有林地和“其他”地物中性能下降明显; PolSARFormer 与 DeepCNN 虽然在简单地物上表现均衡, 但是其复杂类别识别能力较弱; HybridSN、CRPM-Net 与 MP-ResNet 整体 F1-Score 偏低, 尤其在复杂地物分类任务中表现不佳。

为进一步评估各模型的整体分类性能, 表 3 列出了不同模型的总体精度评价结果。HAR-HybridSN 的总体精度 OA、平均精度 AA 和 Kappa 系数分别为 0.9316、0.9176 和 0.9154, 表现最好, 表明其在复杂地物分类任务中具有优异的判别能力与分类一致性。HybridSN 与 MCPT 的表现稍差, 这两个模型的 OA 分别为 0.9238 和 0.9249, PolSARFormer 与 DeepCNN 的表现一般, 说明其分

类能力尚可。CRPM-Net 与 MP-ResNet 的整体性能较差, 其中 CRPM-Net 的 OA 为 0.8906, Kappa 系数为 0.8649, 表明其分类能力不足。

表 3 各模型的整体分类结果精度评价

Table 3 Accuracy evaluation of the overall classification results of each model

分类模型	OA	AA	Kappa 系数	运行时间
PolSARFormer	0.9159	0.8979	0.8960	60 min 42 s
DeepCNN	0.9143	0.9008	0.8943	68 min 15 s
HybridSN	0.9238	0.9095	0.9058	17 min 18 s
CRPM-Net	0.8906	0.8775	0.8649	22 min 24 s
MP-ResNet	0.9126	0.8922	0.8917	25 min 26 s
MCPT	0.9249	0.9011	0.9074	23 min 17 s
HAR-HybridSN	0.9316	0.9176	0.9154	19 min 39 s

在运行效率方面, PolSARFormer 与 DeepCNN 耗时最长, 分别为 60 min 42 s 和 68 min 15 s; 而

HybridSN 与 HAR-HybridSN 的运行效率更高,耗时分别为 17 min 18 s 和 19 min 39 s, 其在保持高精度的同时具备良好的计算效率。CRPM-Net、MP-ResNet 与 MCPT 运行时间为 22—25 min, 处于中等水平, 可以在精度与效率间实现一定平衡。

综上, HAR-HybridSN 在分类精度上显著优于其他模型, 且计算效率优于多数对比模型, 展现出优越的综合性能; HybridSN 虽精度略低, 但运行最快, 适合对效率敏感的应用场景。

3.5 HAR-HybridSN 算法的消融实验

3.5.1 训练集比例对分类精度的影响

为评估 HAR-HybridSN 在小样本条件下的分类性能与鲁棒性, 本文设置不同训练样本比例(50%、60%、70%、80%、90%), 其余样本作为测试集, 并且固定其他超参数, 基于 OA、AA 和 Kappa 系数开展消融实验, 结果如表 4 所示。

从表 4 中可知, 随着训练集比例从 50% 增至 90%, HAR-HybridSN 的 OA、AA 与 Kappa 系数整体呈上升趋势, 表明更多样本有助于模型充分学习地物特征, 提升分类性能。当训练集比例达到 70% 后, 精度增长趋于平缓, 提升幅度显著减小, 而训练耗时则明显增加。值得注意的是, 在 50% 训练集比例下, OA、AA 与 Kappa 较 70% 条件下的差异均<0.01, 表明模型在少量标注数据下仍能有

效挖掘判别特征, 具备良好的小样本学习能力与泛化性能。当训练集比例进一步提升至 90% 时, 精度仅小幅提高, 但训练时间显著增长, 计算成本增加, 性能增益有限。这表明, 尽管模型结构增强有助于精度提升, 但需权衡其带来的资源开销。因此, 在实际应用中应综合考虑精度需求与运算效率, 合理设置训练比例, 实现性能与成本之间的最优平衡, 对于在保证精度的同时提升运行效率具有重要意义。

表 4 不同训练集比例的分类精度
Table 4 Classification accuracy with different training set ratios

训练集比例/%	OA	AA	Kappa 系数	训练时间
50	0.9247	0.9072	0.9004	9 min 34 s
60	0.9294	0.9107	0.9059	10 min 21 s
70	0.9316	0.9176	0.9154	12 min 15 s
80	0.9329	0.9211	0.9196	13 min 49 s
90	0.9348	0.9234	0.9223	16 min 56 s

3.5.2 各模块对分类精度的影响

为研究 HAM 和 RLA 对 HAR-HybridSN 模型性能的影响, 本文设置了 4 组实验 (HybridSN、HybridSN+HAM、HybridSN+RLA、HAR-HybridSN), 训练设置均保持一致, 以 OA、AA 和 Kappa 系数为精度指标进行评估, 结果如表 5 所示。

表 5 不同模块组合下的分类精度
Table 5 Classification accuracy under different module combinations

模块组合	参数量/ 10^6	FLOPs/ 10^9	OA	AA	Kappa 系数
HybridSN	0.9309	1.8069	0.9238	0.9095	0.9058
HybridSN+HAM	0.9331	1.8097	0.9307	0.9098	0.9090
HybridSN+RLA	1.1133	4.0637	0.9310	0.9158	0.9108
HybridSN+HAM+RLA (HAR-HybridSN)	1.1155	4.0666	0.9316	0.9176	0.9154

从表 5 中可知, HybridSN+HAM、HybridSN+RLA 的精度 (OA、AA、Kappa 系数) 均高于原始的 HybridSN, 且融合 HAM 和 RLA 的 HAR-HybridSN 精度最优, 表明二者在功能上的互补性。从功能定位上来看, HAM 的注意力机制局限于当前层, 无法利用前层特征, 导致特征表示仅依赖单层信息, 难以较好地融合多光谱和 SAR 影像的多尺度特征, 而 RLA 可以弥补 HAM 的单层局限, 通过递归跨层交互主动整合不同层级的特征。尽管 RLA 实现了跨层特征交互, 但在层内特征优化方面存在不足, 其

没有专门建模通道间的依赖关系, 也没有对空间局部区域进行精细化筛选, 导致层内关键特征无法有效表达, 而 HAM 通过通道注意力和空间注意力可以弥补这一点。这种互补关系使得 HAR-HybridSN 的精度比添加单一模块的模型精度更高。

从模型复杂度的视角分析, HybridSN 作为基础架构, 参数量仅为 0.9309×10^6 , FLOPs 为 1.8069×10^9 , 是所有组合中参数量和计算量最小的。由于 HAM 是轻量化模块, 因此在 HybridSN 的基础上单独嵌入 HAM 模块后, 参数量和 FLOPs 仅略微增

加，而精度也可以得到小幅度提升。由于RLA模块需要计算3D卷积层中每层的层注意力，因此HybridSN+RLA组合的参数数量和FLOPs有明显增加，而这种组合的精度也优于HybridSN+HAM和HybridSN。尽管HAR-HybridSN的精度最优，但是其模型复杂度也最高。因此需要权衡精度和模型复杂度来选择合适的模型。

3.5.3 特征数量对分类精度的影响

为评估HAR-HybridSN在不同特征数量下的分类性能，本文采用原始8个特征（GF-1的4个波段与GF-3的4个后向散射系数）、降维后12个特征、降维后16个特征和降维后20个特征开展对比试验，结果如表6所示。

表6 不同特征数量下的分类精度
Table 6 Classification accuracy under different feature numbers

特征数量	FLOPs/G	OA	AA	Kappa系数
原始8个特征	1.7150	0.9241	0.9049	0.9059
降维后12个特征	2.0361	0.9276	0.9105	0.9097
降维后16个特征	4.0666	0.9316	0.9176	0.9154
降维后20个特征	13.1846	0.9213	0.8927	0.9012

从表6中可知，采用降维后的16个特征对HAR-HybridSN进行训练，可得到最优的分类精度。使用特征筛选后的影像进行训练，其分类精度较原始波段影像提升约1%，由此说明HAR-HybridSN同样能从原始数据中自动学习到有效的高级特征。此外，FLOPs随特征数量的增加而上升，但降维后20个特征的分类精度不及降维后16个特征与原始8个特征的精度，表明保留的特征数量过多反而会降低分类精度，还会造成计算资源的浪费。在实际应用中，如果可以实现更高的分类精度，进行一定的特征工程是值得考虑的；在追求便捷高效的应用场景中，以原始影像为输入也能得到较好的分类精度。

4 结 论

为实现融合多光谱影像与SAR影像的高精度地物分类，本文在HybridSN模型的基础上提出了HAR-HybridSN模型，并对比验证了该算法与PolSARFormer、DeepCNN、HybridSN、CRPM-Net、MP-ResNet和MCPT6种模型在研究区地物分类性能上的差异，主要结论如下：

（1）HAR-HybridSN通过引入混合注意力模块和循环层注意力机制，增强了不同尺度与层级特征之间的语义融合能力，有效突出关键信息区域，抑制噪声，提升模型判别力。此外该模型实现了跨层级的特征互补，增强上下文理解。在SAR与多光谱影像融合的复杂应用场景中，HAR-HybridSN能够更好地适配异构数据特性，增强对多源地物信息的整合能力，提高分类结果的一致性与稳定性。

（2）HAR-HybridSN的总体精度较其他模型提升约1.0%。从运行时间来看，HAR-HybridSN在保持较高分类精度的同时，具备良好的计算效率，整体性能优于其他模型。此外，在小样本条件下，HAR-HybridSN也展现出良好的分类能力和泛化性能。

尽管HAR-HybridSN在地物分类中取得一定进展，但也存在着不足。首先是HAM和RLA的融合方式比较简单，仅仅是将HAM输出的特征和RLA输出的特征相加融合，并且没有对这两个特征进行权重调整。这种融合方式导致两个模块仅能通过影响HybridSN 3D模块来实现彼此交互，二者在梯度传播中的联系不明显。此外，RLA中的注意力机制使用的是单头注意力机制，其特征表示能力会弱于多头注意力机制。在未来的研究中，可以RLA输出的特征作为HAM的输入，以强化这两个模块的交互能力，加强彼此在梯度传播中的联系。此外，还将采用多头注意力机制来替换RLA中的单头注意力机制。

参考文献(References)

An W T, Cui Y, Yang J and Zhang H J. 2010. Fast alternatives to H/α for polarimetric SAR. IEEE Geoscience and Remote Sensing Letters, 7(2): 343-347 [DOI: 10.1109/LGRS.2009.2035135]
Bhattacharya A, Muhuri A, De S, Manickam S and Frery A C. 2015. Modifying the Yamaguchi four-component decomposition scattering powers using a stochastic distance. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8(7): 3497-3506 [DOI: 10.1109/JSTARS.2015.2420683]
Cloude S R and Pottier E. 1996. A review of target decomposition theorems in radar polarimetry. IEEE Transactions on Geoscience and Remote Sensing, 34(2): 498-518 [DOI: 10.1109/36.485127]
Davari N, Akbarizadeh G and Mashhour E. 2022. Corona detection and power equipment classification based on GoogleNet-AlexNet: an accurate and intelligent defect detection model based on deep

- learning for power distribution lines. *IEEE Transactions on Power Delivery*, 37(4): 2766-2774 [DOI: 10.1109/TPWRD.2021.3116489]
- Deng F, Huang M Y, Bao W, Ji N and Xiang X S. 2024. FASNet: a feature alignment-based SAR and optical image fusion network for land use classification//6th Chinese Conference on Pattern Recognition and Computer Vision. Xiamen: Springer: 298-310 [DOI: 10.1007/978-981-99-8549-4_25]
- Ding L, Zheng K, Lin D, Chen Y X, Liu B, Li J S and Bruzzone L. 2022. MP-ResNet: multipath residual network for the semantic segmentation of high-resolution PolSAR images. *IEEE Geoscience and Remote Sensing Letters*, 19: 4014205 [DOI: 10.1109/LGRS.2021.3079925]
- Fang Y W, Cai Y X, Chen J T, Zhao J Y, Tian G J and Li G D. 2023. Cross-layer retrospective retrieving via layer attention. *arXiv: 2302.03985* [DOI: 10.48550/arXiv.2302.03985]
- Freeman A and Durden S L. 1998. A three-component scattering model for polarimetric SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 36(3): 963-973 [DOI: 10.1109/36.673687]
- Gamba P and Aldrichi M. 2012. SAR data classification of urban areas by means of segmentation techniques and ancillary optical data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(4): 1140-1148 [DOI: 10.1109/JSTARS.2012.2195774]
- Guan D D, Tang T, Zhao L J and Lu J. 2015. A feature combining spatial and structural information for SAR image classification//2015 IEEE International Geoscience and Remote Sensing Symposium. Milan: IEEE: 4396-4399 [DOI: 10.1109/IGARSS.2015.7326801]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- Hu Z H, Qian Y R, Xiao Z Q, Yang G Q, Jiang H and Sun X. 2024. SABNet: self-attention bilateral network for land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17: 8559-8569 [DOI: 10.1109/JSTARS.2024.3382096]
- Huang G, Liu Z, van der Maaten L and Weinberger K Q. 2017. Dense-ly connected convolutional networks//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE: 2261-2269 [DOI: 10.1109/CVPR.2017.243]
- Huang H, Legarsky J and Othman M. 2007. Land-cover classification using Radarsat and Landsat imagery for St. Louis, Missouri. *Photogrammetric Engineering and Remote Sensing*, 73(1): 37-43 [DOI: 10.14358/PERS.73.1.37]
- Jamali A, Mahdianpari M, Mohammadimanesh F, Bhattacharya A and Homayouni S. 2022. PolSAR image classification based on deep convolutional neural networks using wavelet transformation. *IEEE Geoscience and Remote Sensing Letters*, 19: 4510105 [DOI: 10.1109/LGRS.2022.3185118]
- Jamali A, Roy S K, Bhattacharya A and Ghamisi P. 2023. Local window attention transformer for polarimetric SAR image classification. *IEEE Geoscience and Remote Sensing Letters*, 20: 4004205 [DOI: 10.1109/LGRS.2023.3239263]
- Jiang J J, Ma J Y, Chen C, Wang Z Y, Cai Z H and Wang L Z. 2018. SuperPCA: a superpixelwise PCA approach for unsupervised feature extraction of hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 56(8): 4581-4593 [DOI: 10.1109/TGRS.2018.2828029]
- Joshi N, Baumann M, Ehammer A, Fensholt R, Grogan K, Hostert P, Jepsen MR, Kuemmerle T, Meyfroidt P, Mitchard ETA, Reiche J, Ryan CM and Waske B. 2016. A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring. *Remote Sensing*, 8(1): 70 [DOI: 10.3390/rs8010070]
- Lee J S and Pottier E. 2009. *Polarimetric Radar Imaging*. New York: CRC Press
- Li G Q, Fang Q, Zha L L, Gao X and Zheng N G. 2022. HAM: hybrid attention module in deep convolutional neural networks for image classification. *Pattern Recognition*, 129: 108785 [DOI: 10.1016/j.patcog.2022.108785]
- Liu X H. 2024. Multi-scale attention fusion network for remote sensing image scene classification. *Science of Surveying and Mapping*, 49(12): 28-42 (刘小华. 2024. 基于多尺度注意力融合网络的遥感影像场景分类. *测绘科学*, 49(12): 28-42) [DOI: 10.16251/j.cnki.1009-2307.2024.12.004]
- McFeeters S K. 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7): 1425-1432 [DOI: 10.1080/01431169608948714]
- Ming Y S, Liu Q H, Bai H and Huang C. 2023. Classification and change detection of vegetation in the Ruogai Wetland using optical and SAR remote sensing data. *National Remote Sensing Bulletin*, 27(6): 1414-1425 (明义森, 刘启航, 柏荷, 黄昌. 2023. 利用光学和SAR遥感数据的若尔盖湿地植被分类与变化监测. *遥感学报*, 27(6): 1414-1425) [DOI: 10.11834/jrs.20221767]
- Montanaro A, Valsesia D, Fracastoro G and Magli E. 2022. Semi-supervised learning for joint SAR and multispectral land cover classification. *IEEE Geoscience and Remote Sensing Letters*, 19: 2506305 [DOI: 10.1109/LGRS.2022.3195259]
- Huete A, Didan K, Miura T, et al. 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices[J]. *Remote Sensing of Environment*, 83: 195-213 [DOI: 10.1016/S0034-4257(02)00096-2]
- Roy S K, Krishna G, Dubey S R and Chaudhuri B B. 2020. HybridSN: exploring 3-D-2-D CNN feature hierarchy for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 17(2): 277-281 [DOI: 10.1109/LGRS.2019.2918719]
- Schmitt M and Zhu X X. 2016. Data fusion and remote sensing: an ever-growing relationship. *IEEE Geoscience and Remote Sensing Magazine*, 4(4): 6-23 [DOI: 10.1109/MGRS.2016.2561021]
- Shu X B, Yang J W, Yan R and Song Y. 2022. Expansion-squeeze-exci-

- tation fusion network for elderly activity recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(8): 5281-5292 [DOI: 10.1109/TCSVT.2022.3142771]
- Singh G and Yamaguchi Y. 2018. Model-based six-component scattering matrix power decomposition. *IEEE Transactions on Geoscience and Remote Sensing*, 56(10): 5687-5704 [DOI: 10.1109/TGRS.2018.2824322]
- Sisay G, Gessesse B, Fürst C, Kassie M and Kebede B. 2023. Modeling of land use/land cover dynamics using artificial neural network and cellular automata Markov chain algorithms in Goang watershed, Ethiopia. *Heliyon*, 9(9): e20088 [DOI: 10.1016/j.heliyon.2023.e20088]
- Solberg A H, Jain A K and Tast T. 1996. A Markov random field model for classification of multisource satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 34(1): 100-113 [DOI: 10.1109/36.481897]
- Sun C W, Zhang Q, Zhuang C Y and Zhang M Q. 2024. BMFNet: bifurcated multi-modal fusion network for RGB-D salient object detection. *Image and Vision Computing*, 147: 105048 [DOI: 10.1016/j.imavis.2024.105048]
- Wang Q L, Wu B G, Zhu P F, Li P H, Zuo W M and Hu Q H. 2020. ECA-Net: efficient channel attention for deep convolutional neural networks//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE: 11531-11539 [DOI: 10.1109/CVPR42600.2020.01155]
- Wang W K, Wang J L, Lu B B, Liu B Y, Zhang Y K and Wang C Y. 2023. MCPT: mixed convolutional parallel transformer for polarimetric SAR image classification. *Remote Sensing*, 15(11): 2936 [DOI: 10.3390/rs15112936]
- Woo S, Park J, Lee J Y and Kweon I S. 2018. CBAM: convolutional block attention module//15th European Conference on Computer Vision. Munich: Springer: 3-19 [DOI: 10.1007/978-3-030-01234-2_1]
- Xiao D L, Liu C, Wang Q, Wang C and Zhang X. 2020. PolSAR image classification based on dilated convolution and pixel-refining parallel mapping network in the complex domain. *arXiv*: 1909.10783 [DOI: 10.48550/arXiv.1909.10783]
- Xu B and Gong P. 2007. Land-use/land-cover classification with multi-spectral and hyperspectral EO-1 data. *Photogrammetric Engineering and Remote Sensing*, 73(8): 955-965 [DOI: 10.14358/PERS.73.8.955]
- Xu G Z, Jiang X, Zhou Y, Li S T, Liu X Z and Lin P W. 2024. Robust land cover classification with multimodal knowledge distillation. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 4500516 [DOI: 10.1109/TGRS.2023.3344448]
- Xue Y F, Zhang S H, Bai N N, Yuan F, Liu J, Chen Y, Huang X H, Xiong L L and Fu Y C. 2024. Classification of mangrove species using coupled Sentinel-1/2 multi-source time-series remote sensing image features. *Journal of Geo-Information Science*, 26(11): 2626-2642 (薛宇飞, 张声哈, 白娜娜, 原峰, 刘杰, 陈烨, 黄晓慧, 熊兰兰, 付迎春. 2024. 耦合 Sentinel-1/2 多源时序遥感影像特征的红树林物种分类方法. *地球信息科学学报*, 26(11): 2626-2642) [DOI: 10.12082/dqxxkx.2024.240334]
- Yamaguchi Y, Moriyama T, Ishido M and Yamada H. 2005. Four-component scattering model for polarimetric SAR image decomposition. *IEEE Transactions on Geoscience and Remote Sensing*, 43(8): 1699-1706 [DOI: 10.1109/TGRS.2005.852084]
- Zhang H D, Yu A Z, Gao K L, Lu X B, Cao X F, Guo W Y and Lian W Q. 2025. M²Caps: learning multi-modal capsules of optical and SAR images for land cover classification. *International Journal of Digital Earth*, 18(1): 2447347 [DOI: 10.1080/17538947.2024.2447347]
- Zhang H S, Wan L M, Wang T, Lin Y Y, Lin H and Zheng Z Z. 2019. Impervious surface estimation from optical and polarimetric SAR data using small-patched deep convolutional networks: a comparative study. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(7): 2374-2387 [DOI: 10.1109/JSTARS.2019.2915277]
- Zhang H S and Xu R. 2018. Exploring the optimal integration levels between SAR and optical data for better urban land cover mapping in the Pearl River Delta. *International Journal of Applied Earth Observation and Geoinformation*, 64: 87-95 [DOI: 10.1016/j.jag.2017.08.013]
- Zhang L M, Zou B, Cai H J and Zhang Y. 2008. Multiple-component scattering model for polarimetric SAR image decomposition. *IEEE Geoscience and Remote Sensing Letters*, 5(4): 603-607 [DOI: 10.1109/LGRS.2008.2000795]
- Zhang W X, Tang P, Meng Y, Zhao L J, Zhao Z T and Zhang Z. 2024. Crop type classification of remote sensing image time series based on multi-scale spatial-temporal global attention model. *National Remote Sensing Bulletin*, 28(11): 2865-2877 (张伟雄, 唐婷, 孟瑜, 赵理君, 赵智韬, 张正. 2024. 基于多尺度时空全局注意力的遥感影像时间序列农作物分类. *遥感学报*, 28(11): 2865-2877) [DOI: 10.11834/jrs.20243557]
- Zhao Y and Jiang M. 2019. Integration of SAR polarimetric parameters and multi-spectral data for object-based land cover classification. *Acta Geodaetica et Cartographica Sinica*, 48(5): 609-617 (赵诣, 蒋弥. 2019. 极化 SAR 参数优化与光学波谱相结合的面对象土地覆盖分类. *测绘学报*, 48(5): 609-617) [DOI: 10.11947/j.AGCS.2019.20170746]

Improved HybridSN Network for multispectral–ppolarimetric SAR image classification with multi–scale feature

GAO Yonggang, LIANG Wenge, LIU Yuting, ZHAN Junjie, LI Fangying

College of Environment and Safety Engineering, Fuzhou University, Fuzhou 350108, China.

Abstract: Multispectral images are widely applied in land-use classification, but weather conditions have a considerable impact on their quality. Synthetic Aperture Radar (SAR), with its all-weather, all-time capabilities and strong penetration, holds significant application value in land-use classification. However, traditional SAR-based classification methods mainly rely on backscattering features, which restricts improvements in classification accuracy and stability. Compared with land-use classification methods that use single remote sensing images, the fusion of multispectral and SAR images can achieve data complementarity and improve classification accuracy. Hybrid Spectral CNN (HybridSN) is a hyperspectral image classification model based on a hybrid architecture of 3-D and 2-D convolutional neural networks. It effectively fuses spectral and spatial features to improve classification accuracy. However, the HybridSN model tends to misclassify noise as valid features when processing heterogeneous data such as fused SAR and multispectral images, struggling to coordinate the structural features of SAR images with high-level semantic information in multispectral images. To address these issues, we proposed the Hybrid and Recurrent Layer Attention HybridSN (HAR-HybridSN), which introduces a Hybrid Attention Module (HAM) and a Recurrent Layer Attention (RLA) based on HybridSN. The HAM in the proposed model integrates channel and spatial attention. The channel attention module generates channel attention maps through global average and max pooling to highlight discriminative features while suppressing redundant information. After channel aggregation, the spatial attention module then produces spatial attention maps that guide the model to focus on key regions within the image. These two modules work synergistically to enhance the model's robustness and classification accuracy in complex scenarios such as low signal-to-noise ratio and blurred boundaries. RLA introduces semantic interaction channels between different convolutional layers, efficiently fusing shallow and deep features. Through bidirectional information transmission, low-level details and high-level semantics are co-modeled, which enhances the comprehensive recognition capability of the model for textures, boundaries, and semantic categories. Using GF-3 SAR images and GF-1 multispectral images as data sources, we conducted qualitative and quantitative evaluations of the classification results of the HAR-HybridSN model in comparison with six models, including HybridSN. Results show that the overall accuracy, average accuracy, and kappa coefficient of HAR-HybridSN are higher than 93%, 91%, and 91%, respectively. In areas with mixed ground objects and blurred boundaries, it can accurately identify key semantic features and effectively suppress interference from redundant information, leading to enhanced discriminative ability for complex ground objects. Specifically, in complex ground objects, the overall accuracy of the HAR-HybridSN model is about 1% higher than that of other models, and it maintains better consistency and continuity in categories such as buildings and cultivated land. In terms of running time, HAR-HybridSN maintains high classification accuracy while exhibiting favorable computational efficiency, with its overall performance surpassing that of other models. Furthermore, under small-sample conditions, it demonstrates robust classification capability and generalization performance.

Key words: multispectral images, polarimetric SAR, multi-scale features, HAR-HybridSN, land use classification

Supported by Natural Science Foundation of Fujian Province (No. 2023J01064)