

DOI: 10.19650/j.cnki.cjsi.J2614912

频域特征融合结合超连接机制的多源 传感器遥感数据协同分类*

王爱丽¹, 戴诗语¹, 陈寅生¹, 于亮², 吴海滨¹

(1. 哈尔滨理工大学黑龙江省激光光谱技术及应用重点实验室 哈尔滨 150080;

2. 哈尔滨工业大学超精密光电仪器工程研究所 哈尔滨 150001)

摘要:针对高光谱图像(HSI)与激光雷达(LiDAR)、合成孔径雷达(SAR)等多源传感器遥感数据在协同分类中因成像机理与测量特性差异导致的特征提取不充分及协同利用受限问题,本文提出一种结合频域特征融合与超连接机制的跨传感器频域感知融合网络(CFF-Net)。该方法从频率域视角刻画多源观测信息的互补关系,实现异构传感器特征的协同建模与有效融合。在特征提取阶段,构建多维特征感知编码器,通过并行建模高频局部细节信息与低频全局结构信息,实现对不同传感器观测特性的互补表征;在特征协同融合阶段,引入跨传感器超融合模块,将超连接机制与频率自注意力相结合,实现多源传感器特征在频域上的自适应对齐与协同增强,从而提升融合特征的一致性与判别能力。在 MUUFL Gulfport (MU)、Houston 2013 (HU) 和 Augsburg (AU) 3 个公开多源传感器遥感数据集上的实验结果表明,CFF-Net 的总体分类精度(OA)分别达到 92.23%、93.51% 和 91.54%,平均精度(AA)分别为 92.05%、93.77% 和 64.81%,均优于所有对比方法。模型参数量仅为 0.30 M,浮点运算量为 60.55 M,在保持轻量级结构的同时实现了分类性能与计算效率的合理平衡,验证了所提方法在多源传感器遥感数据协同分类中的有效性与适用性。

关键词: 高光谱图像;激光雷达;合成孔径雷达图像;超连接机制;Transformer

中图分类号: TP391.4 TH74 **文献标识码:** A **国家标准学科分类代码:** 410.7510

Multi-sensor remote sensing data collaborative classification combining frequency-domain feature fusion with hyper-connection mechanism

Wang Aili¹, Dai Shiyu¹, Chen Yinsheng¹, Yu Liang², Wu Haibin¹

(1. Heilongjiang Province Key Laboratory of Laser Spectroscopy Technology and Application, Harbin University of Science and Technology, Harbin 150080, China; 2. Institute of Ultra-Precision Optoelectronic Instrument Engineering, Harbin Institute of Technology, Harbin 150001, China)

Abstract: In the joint classification of multi-sensor remote sensing data, such as hyperspectral images (HSI), light detection and ranging (LiDAR), and synthetic aperture radar (SAR), differences in imaging mechanisms and measurement characteristics often lead to insufficient feature extraction and limited collaborative utilization. To address these issues, this paper proposes a cross-sensor frequency-aware fusion network (CFF-Net) that integrates frequency-domain feature fusion with a hyper-connection mechanism. From a frequency-domain perspective, the proposed method characterizes the complementary relationships among multi-source observations, enabling the collaborative modeling and effective fusion of heterogeneous sensor features. In the feature extraction stage, a multi-dimensional feature perception encoder is constructed to jointly model high-frequency local detail information and low-frequency global structural information in parallel, thereby achieving complementary representations of different sensor observation characteristics. In the feature fusion stage, a cross-sensor hyper-fusion module is introduced, combining the hyper-connection mechanism with frequency-based self-attention to realize adaptive alignment and collaborative enhancement of multi-sensor features in the frequency domain, thereby improving the consistency and discriminability of the fused representations. Extensive experiments are conducted on three public multi-sensor remote sensing datasets: MUUFL Gulfport (MU), Houston 2013 (HU), and Augsburg (AU). The results demonstrate that CFF-Net achieves overall

收稿日期:2026-01-30 Received Date: 2026-01-30

* 基金项目:黑龙江省重点研发计划资助项目(JD2023SJ19)资助

accuracy (OA) of 92.23%, 93.51%, and 91.54%, alongside average accuracy (AA) of 92.05%, 93.77%, and 64.81%, respectively, significantly outperforming all compared methods. With only 0.30 M parameters and 60.55 M FLOPs, the proposed method achieves a favorable balance between classification performance and computational efficiency, validating its effectiveness and applicability in multi-sensor remote sensing data collaborative classification.

Keywords: hyperspectral image; LiDAR; synthetic aperture radar image; hyper-connection mechanism; Transformer

0 引 言

地物分类作为遥感领域的基础性与核心任务,是实现土地利用/覆盖变化监测、生态环境评估、城市规划及资源管理等应用的重要环节。单一传感器往往难以兼顾光谱、空间与结构特征,因此多传感器数据协同分类成为提升分类精度的重要途径^[1]。

高光谱图像(hyperspectral image, HSI)以数百个连续波段记录地物的光谱信息,具备极高的光谱分辨率与“图谱合一”的优势,能够精确刻画地物的物质组成与内部结构,为类别区分提供坚实基础^[2]。然而,受限于成像机理与空间分辨率,HSI对地物的三维形态和几何结构表征能力有限。为此,激光雷达(light detection and ranging, LiDAR)通过生成数字表面模型(digital surface model, DSM),可精确反映地表目标的高度与空间结构,有效区分光谱相似但高度差异显著的地物。同时,合成孔径雷达(synthetic aperture radar, SAR)作为主动成像传感器,能在各种气象与光照条件下获取地表散射信息,对地物的粗糙度、介电常数及形态特征高度敏感,凭借全天候、全天时的观测能力,为地物识别提供了重要补充。

尽管上述多源数据具有天然互补性,但如何从中高效提取并融合异构特征仍是待解决的关键问题。现有方法的主要局限体现在两方面:特征提取难以同时兼顾局部细节与全局上下文;跨传感器协同融合策略对异构传感器间关系建模不足,导致协同效应受限。

近年来,随着深度学习端到端特征学习框架的发展,卷积神经网络(convolutional neural networks, CNN)^[3]与Transformer模型^[4]等方法逐渐被引入多源传感器遥感影像分析领域^[5-7]。在此基础上,国内外学者围绕遥感多源传感器数据的协同分类开展了大量研究。早期工作如双分支卷积神经网络(the two-branch convolution neural network, Two-Branch)^[8]、深度编码器-解码器(deep encoder-decoder network architecture, EndNet)^[9]及多模态深度学习(multimodal deep learning, MDL-Middle)^[10]主要采用基于CNN的双分支结构,对不同模态特征进行并行建模并实现初步融合。随后,扩展视觉Transformer(extended vision Transformer, ExViT)^[11]与深层对称融合Transformer(deep symmetric fusion Transformer, DSymFuser)^[12]通过引入交叉注意力机制和对称分支设

计,进一步增强了跨传感器特征之间的交互建模能力;多尺度跨域融合网络(multiscale cross-domain fusion network, MCFNet)^[13]则从空间尺度角度出发,对HSI与LiDAR数据的多尺度互补特性进行了探索^[14-16]。

现有方法大多集中于空间域特征建模,而对异构传感器间频率域关联关系的研究相对不足。相关研究表明^[17],CNN更擅长提取纹理和边缘等高频局部细节信息,而Transformer依赖自注意力机制,更适合建模空间布局 and 上下文语义等低频全局结构信息^[18-19]。

这一差异为高低频信息的联合建模提供了理论基础。近年来,部分研究将CNN-Transformer混合架构引入多模态特征提取过程,在频率域实现高低频信息协同编码,并通过频率域自适应对齐机制增强多传感器特征融合效果^[20]。这些方法通过学习复杂的非线性光谱与空间特征表示,为多传感器遥感影像的特征建模与分类提供了新的实现路径。

此外,针对深层CNN-Transformer混合网络在训练过程中可能面临的特征传递效率与稳定性问题,本文引入超连接机制(hyper-connection, HC)^[21],对传统残差连接进行拓展。该机制通过构建可学习的跨层连接关系,增强多尺度、多路径特征传递与跨层语义融合能力,有助于提升网络训练稳定性和特征复用效率。

基于此,本文提出一种结合频域特征融合与超连接机制的跨传感器频域感知融合网络(cross-sensor frequency-aware fusion network, CFF-Net)。该方法通过联合建模高频局部细节与低频全局结构信息,并结合超连接机制实现跨层稳定特征传递,从而提升异构多传感器数据的协同表征能力,实现更精确的地物分类。

1 CFF-Net 模型

1.1 CFF-Net 网络架构

如图1所示,CFF-Net采用“异构特征提取—频域对齐增强—协同分类”的整体框架,从频率域视角建模多源观测的互补关系,实现异构传感器特征的协同表征与有效融合。

1) 多维特征感知编码器

多维特征感知编码器(multidimensional feature perception encoder, MFPE)作为网络底层特征编码模块,采用双分支结构对多传感器数据进行特征提取,主要包

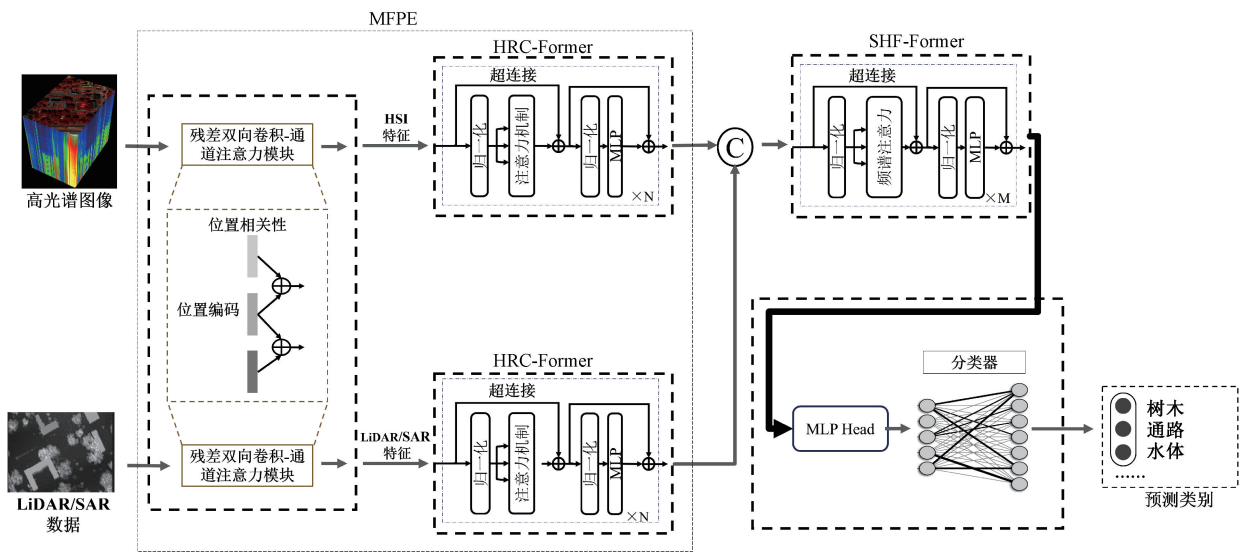


图1 多源传感器遥感数据协同分类模型 CFF-Net 网络架构

Fig. 1 Network architecture of the proposed CFF-Net for the collaborative classification of multi-sensor remote sensing data

括高频分支和低频分支。高频分支采用残差双向卷积-通道注意力模块 (residual dual-way squeeze-and-excitation, ResDW-SE), 利用卷积算子的局部建模能力提取光谱变化、空间纹理及边缘细节等高频信息; 低频分支引入超连接上下文 Transformer (hyper-connected context transformer, HRC-Former), 通过自注意力机制建模长距离依赖关系, 提取空间分布及上下文语义等低频全局结构信息。

通过高低频双分支协同建模, 实现对光谱、空间及散射等多维特征的分层表达, 为后续跨传感器融合提供稳定且具有判别性的特征表示。

2) 跨传感器超融合 Transformer

在特征融合阶段, 引入跨传感器超融合 Transformer (super-heterogeneous fusion Transformer, SHF-Former) 模块。该模块结合超连接机制与频率自注意力机制, 对不同传感器特征的频谱分布差异进行自适应对齐与协同增强, 促进多源特征在统一表示空间中的深层交互, 提升融合特征的一致性与判别能力。

3) 分类头

网络末端采用全连接层将融合特征映射至目标类别空间, 完成多传感器遥感地物分类。

1.2 多维特征感知编码器

针对异构遥感数据在特征表达上的差异性, 尤其是单一特征提取方式难以同时兼顾局部细节与全局结构的问题, 本文在特征提取阶段引入 MFPE。其中, ResDW-SE 模块主要用于高频信息建模, 侧重刻画局部空间结构、光谱变化及散射纹理等细粒度特征; HRC-Former 模块则面向低频全局信息建模, 借助超连接机制捕获跨区域及跨传感器的长距离依赖关系, 以表征宏观

空间分布与上下文语义。通过上述双分支协同设计, 网络能够在频率域上对高、低频特征进行区分建模, 并在传感器层面实现特征对齐与信息互补, 为后续跨传感器特征融合提供较为稳定且具判别性的特征表示。

1) ResDW-SE 模块

如图 1 所示, MFPE 中的各传感器数据, 其中 HSI 为 $X_H \in \mathbf{R}^{h \times w \times d}$, 其中 h 和 w 分别表示空间维度的高度和宽度, d 表示光谱波段的数量; LiDAR 或其他数据表示为 $X_0 \in \mathbf{R}^{h \times w \times b}$, b 表示 LiDAR 或其他数据的通道数。首先各传感器数据经过 ResDW-SE 模块进行初步特征编码。该模块旨在强化高频特征的提取能力, 在高维、多源传感器遥感数据处理中兼顾局部细节捕获与通道判别增强, 从而提升特征编码能力。为此, ResDW-SE 集成了残差连接、深度可分离卷积 (depthwise separable convolution, DWConv)、压缩与激励 (squeeze-and-excitation, SE) 注意力机制以及全局关系网络 (global relation network, GRN), 实现了轻量化的高频特征建模与局部精细特征表征能力, 模块结构示意图如图 2 所示, 图 2 中 BN (batch normalization) 为批量归一化函数, GELU 为激活函数。

ResDW-SE 模块对输入多源传感器遥感数据 X_H 或 X_0 进行处理。首先, ResDW-SE 模块采用一个 3×3 残差连接的 DWConv, 残差连接保证信息流的连续性, 缓解深层网络训练中的梯度消失问题, 在有效提取局部特征的同时, 显著减少参数量与计算开销。该设计对大规模、高维遥感数据尤为重要, 可在保证特征表达能力的同时提升模型效率并降低资源消耗。公式如式 (1) 所示。

$$u_c = \text{DWConv}_{3 \times 3}(X) + X \quad (1)$$

随后, 采用 SE 注意力机制对频域特征进行通道级自

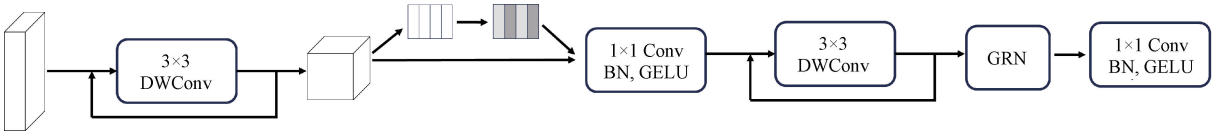


图2 ResDW-SE 模块示意图

Fig. 2 Network architecture diagram of the multidimensional ResDW-SE Block

适应校准,该机制通过全局平均池化聚合空间信息,再经全连接层与 Sigmoid 激活生成通道权重:

$$F_{sq}(u_c) = \frac{1}{H' \times W'} \sum_{i=1}^{W'} \sum_{j=1}^{H'} u_c(i, j) \quad (2)$$

$$s = \text{Sigmoid}(W_2 \times \text{RELU}(W_1 \times F_{sq}(u_c))) \quad (3)$$

其中, W_1 和 W_2 为两个全连接层。

SE 作为一种动态通道加权机制,可显著增强信息量丰富的通道,同时抑制不重要的通道,从而提升特征的判别能力和表示有效性,以精确捕捉关键地物语义信息:

$$F_{SE} = u_c \cdot s \quad (4)$$

接着,应用带有 BN 和 GELU 激活函数的 1×1 卷积,用于融合 SE 模块学习到的通道信息并调整通道维度,以适应后续处理。

$$F_1 = \text{GELU}(\text{BN}(\text{Conv}_{1 \times 1}(F_{SE}))) \quad (5)$$

最后,特征 F_1 经过一个带有残差连接的 3×3 的 DWConv,随后应用带有 BN 和 GELU 激活的 1×1 卷积,用于融合信息并保持梯度稳定。该设计不仅保留了原始信息流,还缓解了局部感受野的限制,从而更全面地表征遥感场景中地物类别的空间布局与语义上下文。

$$F_2 = \text{DWConv}_{3 \times 3}(F_1) + F_1 \quad (6)$$

$$F_{\text{ResDW-SE}} = \text{GELU}(\text{BN}(\text{Conv}_{1 \times 1}(\text{GRN}(F_2)))) \quad (7)$$

ResDW-SE 模块在高频特征建模过程中侧重于局部结构、纹理及光谱细节的刻画,有助于增强特征对细粒度差异的响应能力。同时,通过引入注意力与通道自适应机制,该模块在一定程度上缓解了不同传感器特征在表示空间中的分布差异,为后续全局依赖建模及跨传感器特征融合提供了较为稳定的输入特征。

2) 超连接上下文 Transformer

如图 1 所示,经 ResDW-SE 提取的各传感器特征输入至对应的 HRC-Former 模块,实现由局部高频细节向全局低频结构的层次化特征建模。在标准 Transformer 架构中,残差连接采用固定权重形式,缺乏对层间依赖关系的自适应调节能力,信息传递方式相对静态,限制了深层特征间的交互效率,使模型在建模跨区域低频结构特征时存在局限。

针对上述问题,HRC-Former 在自注意力模块中引入超连接(Hyper-Connection)机制,以增强深层网络中的信息流动能力并提升特征表示的灵活性,其结构示意图 3 所示。

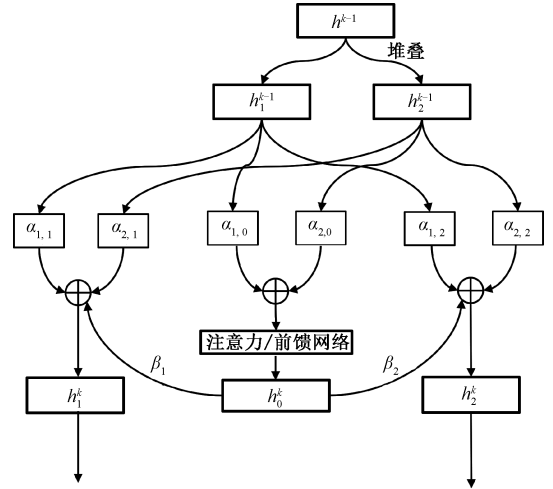


图3 HC-Transformer 示意图

Fig. 3 Network of HC-Transformer

该机制通过显式建模宽度连接和深度连接,对传统残差连接进行了有效泛化,实现了跨层次、跨传感器特征的动态交互与充分表达。设第 $k-1$ 层的隐状态向量为 $h^{k-1} \in \mathbf{R}^D$, D 为特征维度,其首先被扩展为多个子路径状态: $\{h_1^{k-1}, h_2^{k-1}, \dots, h_n^{k-1}\}$, n 为扩展比例,这些子路径可通过不同的非线性变换进行处理(例如注意力机制或前馈网络),形成中间表示。

具体而言,首先扩展后的信息在不同分支之间通过可学习的权重 $\alpha_{i,j}$ 进行横向融合,实现跨路径的信息交互,以图 3 中($n=2$)为例,得到横向融合后的分支为:

$$\tilde{h}_0^{k-1} = \alpha_{1,0} h_1^{k-1} + \alpha_{2,0} h_2^{k-1} \quad (8)$$

$$\tilde{h}_1^{k-1} = \alpha_{1,1} h_1^{k-1} + \alpha_{2,1} h_2^{k-1} \quad (9)$$

$$\tilde{h}_2^{k-1} = \alpha_{1,2} h_1^{k-1} + \alpha_{2,2} h_2^{k-1} \quad (10)$$

其中, $\alpha_{i,j}$ 控制从第 j 子路径到第 i 子路径的连接强度。其次,对主分支 $\tilde{h}_0^{k-1}$ 经过独立的变换函数 $\mathcal{T}$ (可以为 Transformer 中的自注意力层或前馈网络)得到激活表示:

$$h_0^k = \mathcal{T}(\tilde{h}_0^{k-1}) \quad (11)$$

最终,各子分支的输出通过深度连接系数 β_1, β_2 加权聚合,生成下一层的输入:

$$h_1^k = \beta_1 h_0^k \quad (12)$$

$$h_2^k = \beta_2 h_0^k$$

其中, β_1 和 β_2 可为标量参数,也可由轻量级网络动态预测,从而根据输入内容自适应调整拓扑结构。

该设计通过增强自注意力模块中层间信息的动态流动能力,在结构层面缓解了固定残差连接对特征传递路径的约束,使模型能够更充分地利用跨层特征信息,从而提升对低频全局依赖关系及跨区域上下文结构的建模能力。

1.3 跨传感器超融合 Transformer

针对多传感器融合中传感器异质性显著、跨传感器数据特征交互受限的问题,本文设计了 SHF-Former 模块,旨在统一表示空间内实现异构传感器特征的深层协同建模。

在结构上,SHF-Former 以 MFPE 输出的多源特征作为输入,通过联合表示学习方式对不同传感器特征进行融合建模。SHF-Former 以 MFPE 输出的多源特征为输入,在 HRC-Former 的基础上拓展超连接机制,并与频谱注意力相结合,从结构层与频域层两个维度实现跨传感器特征的自适应协同融合。

如图 4 所示,遥感多源传感器数据经过特征拼接进入超连接机制,对主分支 $\tilde{h}_0^{k-1}$ 进行二维离散傅里叶变换(fast Fourier transform, FFT) 可得到其频域表达 $Z(u, v)$,其中 u 和 v 是频率域的变量:

$$Z(u, v) = \sum_x \sum_y \tilde{h}_0^{k-1} e^{-2\pi j(ux+vy)}$$

(13)

随后,定义频域滤波器 $W_c(u, v)$,与 $Z(u, v)$ 进行逐点乘法,得到加权后的频域函数 $Z'(u, v)$:

$$Z'(u, v) = Z(u, v) W_c(u, v)$$

(14)

最后,对 $Z'(u, v)$ 进行反傅里叶变换(inverse fast Fourier transform, IFFT),以取得空间域的输出函数 h_0^k :

$$\begin{cases} \mathcal{F}^{-1}[Z'(u, v)] = \sum_u \sum_v Z'(u, v) e^{2\pi j(ux+vy)} \\ h_0^k = \mathcal{F}^{-1}[Z'(u, v)] \end{cases}$$

(15)

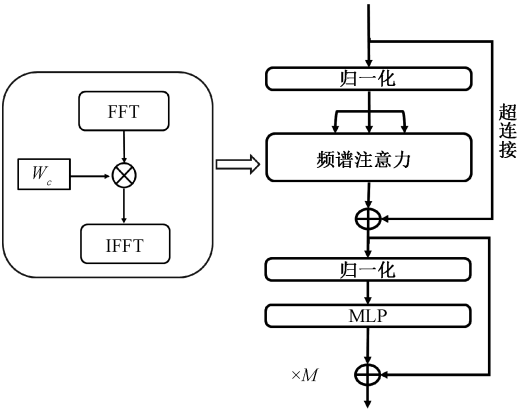


图 4 频谱注意力示意图

Fig. 4 Network of spectrum attention

上述设计使网络在融合阶段能够对不同频谱成分进

行自适应调节,在缓解异构传感器间表示不一致的同时,使融合特征兼具全局结构一致性与类别判别性。

2 实验结果与分析

2.1 实验数据集描述

实验选用 MUUFL Gulfport (MU)^[22]、Houston 2013 (HU)^[23] 和 Augsburg (AU)^[24] 3 个公开多传感器数据集,详细信息见表 1。

表 1 多源传感器数据集描述
Table 1 Description of the multi-model datasets.

数据集	MU	HU	AU
传感器类型	HSI, DSM	HSI, DSM	HSI, SAR
通道	64, 2	144, 1	180, 4
数据尺寸	325×220	349×1 905	332×485
训练样本数量	1 100	2 832	761
测试样本数量	52 587	12 197	77 533
类别	11	15	7

MU 数据集由 ITRES CASI-1500 高光谱成像仪与 Optech ALTM 激光雷达仪搭载同一航空平台同步获取,观测区域为美国密西西比州。HSI 光谱范围为 375 ~ 1 050 nm,经去除低信噪比波段后保留 64 个有效波段。数据经辐射定标、大气校正及几何配准预处理,空间规格为 325×220 像元,地面采样距离 1.0 m,涵盖 11 类典型城市地物,可视化结果见图 5 及表 2。

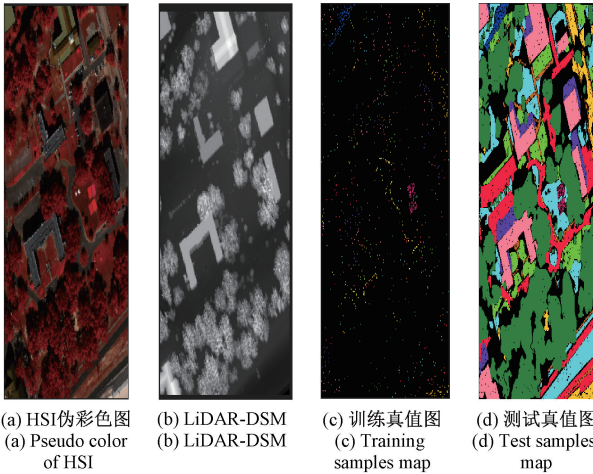


图 5 MU 数据集的伪彩色图和真值图

Fig. 5 Pseudo color map and ground-truth map of MU

HU 数据集由 NCALM 于 2013 年对休斯顿大学校园及周边区域执行航测任务获取,集成 HSI(ITRES CASI-

表 2 MU 数据集地物类别详情

Table 2 Land cover class details in MU dataset			
类别编码	类别名称	训练数量	测试数量
C1	树木 (trees)	100	23 146
C2	草地为主区域 (mostly grass)	100	4 170
C3	混合地表 (mixed ground surface)	100	6 782
C4	泥土和沙地 (dirt and sand)	100	1 726
C5	道路 (road)	100	6 587
C6	水体 (water)	100	366
C7	建筑物阴影 (buildings shadow)	100	2 133
C8	建筑物 (buildings)	100	6 140
C9	人行道 (sidewalk)	100	1 285
C10	黄色路缘 (yellow curb)	100	83
C11	布料板 (cloth panels)	100	169
合计		1 100	52 587

1500,380~1 050 nm,144 波段)与 LiDAR (DSM)两种传感器数据。数据经几何校正与重采样,空间规格为 349×1 905 像元,地面采样距离 2.5 m,涵盖 15 类城市地物,可视化结果见图 6 及表 3。

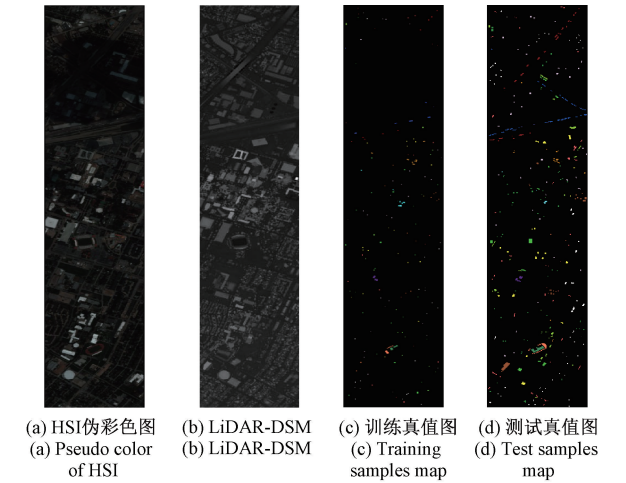


图 6 HU 数据集的伪彩色图和真值图

Fig. 6 Pseudo color map and ground-truth map of HU

AU 数据集观测区域位于德国奥格斯堡,集成 HSI (HySpex VNIR-1600,400~2 500 nm,180 波段)与 SAR (Sentinel-1,提取 VV/VH 极化强度及极化协方差矩阵实部与虚部共 4 个特征)两种传感器数据。数据经精密配准与重采样,空间规格为 332×485 像元,地面采样距离 30.0 m,涵盖 7 类典型地物,可视化结果见图 7 及表 4。

2.2 实验平台及参数

实验基于 Ubuntu 20.04 系统,硬件配置为 NVIDIA Quadro RTX 6000 GPU 与 Intel Xeon E5-2640 v4 CPU,使

表 3 HU 数据集地物类别详情

Table 3 Land cover class details in HU dataset.			
类别编码	类别名称	训练数量	测试数量
C1	健康草地 (healthy grass)	198	1 053
C2	胁迫草地 (stressed grass)	190	1 064
C3	人造草地 (synthetic grass)	192	505
C4	树木 (trees)	188	1 056
C5	土壤 (soil)	186	1 056
C6	水体 (water)	182	143
C7	住宅区 (residential)	196	1 072
C8	商业区 (commercial)	191	1 053
C9	道路 (road)	193	1 059
C10	公路 (highway)	191	1 036
C11	铁路 (railway)	181	1 054
C12	停车场 1 (parking Lot1)	192	1 041
C13	停车场 2 (parking Lot2)	184	285
C14	网球场 (tennis court)	181	247
C15	跑道 (running track)	187	473
合计		2 832	12 197

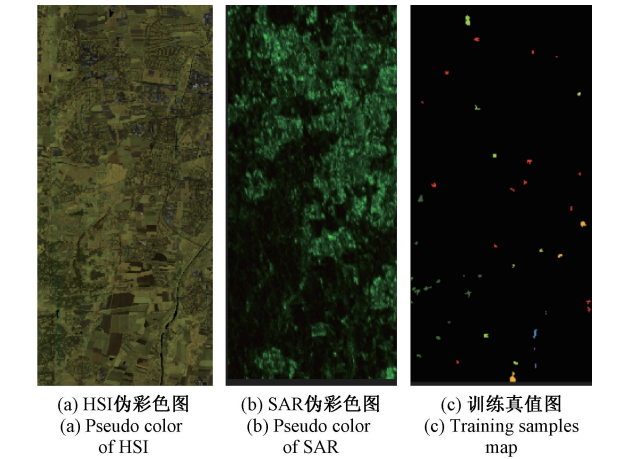


图 7 AU 数据集的伪彩色图和真值图

Fig. 7 Pseudo color map and ground-truth map of AU

表 4 AU 数据集地物类别详情

Table 4 Land cover class details in AU dataset			
类别编码	类别名称	训练数量	测试数量
C1	树木 (trees)	146	13 361
C2	住宅区 (residential area)	264	30 065
C3	工业区 (industrial area)	21	3 830
C4	低矮植被 (low plants)	248	26 609
C5	小块耕地 (allotment)	52	523
C6	商业区 (commercial area)	7	1 638
C7	水体 (water)	23	1 507
合计		761	77 533

用 Python 3.10 与 PyTorch 2.5.1 框架。训练参数设置如下:批大小(batch size)为 64,训练轮次 500,训练/验证集比例 8:2,优化器为 AdamW,学习率调度策略为 cosine 退火,初始学习率 5×10^{-4} ,权重衰减系数 0.1。

2.3 实验对比及分析

1) 不同分类方法的对比

为验证所提模型在多源传感器遥感数据协同分类任务中的有效性,本文选取多种具有代表性的先进遥感分类方法作为对比基准。Two-Branch^[8]和 EndNet^[9]采用双分支与编解码结构以实现多源传感器特征融合;MDL-Middle^[10]通过中间层融合增强传感器协同;多注意力层次融合网络(multi-attentive hierarchical fusion net, MAHiDFNet)^[25]引入多重注意力与密集连接机制以强化特征交互;特征融合与提取框架(feature fusion and extraction framework, FusAtNet)^[26]利用自注意力与交叉注意力联合建模空间-光谱特征;基于耦合对抗学习的分类方法(coupled adversarial learning based classification, CALC)^[27]借助对抗学习优化传感器对齐;可分离动态分组卷积残差网络(separable dynamic grouping convolution ResNet50, SepG-ResNet50)^[28]在单流卷积框架下提升特征表达能力;ExViT^[11]与 DSymFuser^[12]则基于 Transformer

结构实现跨传感器特征的深层融合;MCFNet^[13]通过构建多尺度跨域融合算子,进一步探索了 HSI 与 LiDAR 数据在不同空间分辨率下的互补性;

本文采用总体分类精度(overall accuracy, OA)、平均精度(average accuracy, AA)和 Kappa 系数(K)作为实验评价指标。其中,OA 表示模型在所有测试样本中正确分类的样本数占总样本数的比例,用于衡量总体分类性能;AA 则为各类别分类精度的平均值,反映模型在不同类别间的平衡性;Kappa 系数用于评估分类结果与地面真实标签之间的一致性,能够更加客观地反映模型的实际分类能力。

表 5~7 列出了 3 个数据集上各方法的分类精度。CFF-Net 在 MU、HU 和 AU 数据集上的 OA 分别为 92.23%、93.51%和 91.54%,均优于所有对比方法。在 MU 数据集上,CFF-Net 在树木(trees: 94.79%)、阴影(buildings shadow: 96.11%)等光谱相似类别上优势明显, cloth panels 达 100.00%;在 HU 数据集上, railway (99.53%)、tennis court (100.00%)等类别精度显著提升;在 AU 数据集上, low plants 精度达 93.36%,整体 OA 较次优方法 MCFNet 提升 1.44 个百分点,验证了模型在 HSI 与 SAR 融合场景下的泛化能力。

Table 5 Comparison of classification accuracy of different methods on MU dataset											(%)
类别	Two-Branch	EndNet	MDL-Middle	MAHiDFNet	FusAtNet	CALC	SepG-ResNet50	ExVit	DSymFuser	MCFNet	CFF-Net
C1	90.29	83.55	87.31	89.87	80.36	91.97	86.78	92.64	91.24	90.75	94.79
C2	75.68	79.38	76.38	63.19	73.57	81.77	78.47	77.48	79.57	66.76	81.15
C3	69.71	76.28	68.33	75.85	68.24	77.56	71.20	82.07	82.23	91.24	85.37
C4	93.97	87.08	78.74	96.18	70.74	95.19	89.98	93.74	89.17	96.29	95.54
C5	91.79	89.59	83.76	88.52	80.95	89.19	76.38	90.28	85.60	85.97	92.90
C6	99.73	95.90	88.52	85.25	81.15	100.00	99.73	99.73	100.00	99.45	99.73
C7	91.84	88.28	92.12	92.72	89.40	95.17	91.70	91.47	91.80	94.33	96.11
C8	94.79	92.07	89.69	95.44	87.92	96.91	87.31	97.48	96.29	97.35	96.91
C9	72.30	76.96	77.04	75.80	77.98	69.81	69.88	73.54	77.12	75.49	78.52
C10	96.39	95.18	86.75	91.57	78.31	95.18	90.36	93.98	91.57	95.18	91.57
C11	97.63	97.63	99.41	99.41	99.41	100.00	99.41	99.41	100.00	99.41	100.00
OA	87.03	84.33	83.54	86.45	79.24	89.31	82.90	89.94	88.78	89.13	92.23
AA	88.56	87.45	84.37	86.80	80.73	90.25	85.56	90.17	89.54	90.20	92.05
K×100	83.10	79.85	78.84	82.37	73.81	86.05	77.94	86.82	85.37	85.80	89.76

为直观评估 CFF-Net 的分类性能,本文在 MU、HU 及 AU 3 个基准数据集上进行了可视化对比实验,结果分别如图 8~10 所示。总体而言,CFF-Net 所生成的分类图在

区域轮廓上更为清晰、连续,与真实地物的空间分布吻合度更高;而对比方法在地物边缘处普遍存在锯齿状伪影或边界断裂现象,空间一致性明显不足。具体来看,在

表 6 HU 数据集分类不同方法的分类精度对比											
Table 6 Comparison of classification accuracy of different methods on HU dataset											(%)
类别	Two-Branch	EndNet	MDL-Middle	MAHiDF-Net	FusAt-Net	CALC	SepG-ResNet50	ExVit	DSym-Fuser	MCFNet	CFF-Net
C1	82.90	81.58	83.10	82.91	80.72	86.51	72.36	81.67	80.06	82.91	83.10
C2	84.31	83.65	85.06	84.68	97.46	84.59	77.35	100.00	78.67	100.00	100.00
C3	96.44	100.00	99.60	100.00	90.69	90.50	34.85	99.01	99.41	93.66	97.43
C4	96.59	93.09	91.57	93.37	99.72	91.86	86.84	98.96	95.55	98.67	98.39
C5	99.62	99.91	98.86	99.43	97.92	100.00	91.38	99.91	98.67	99.72	99.91
C6	82.52	95.10	100.00	99.30	93.71	99.30	95.10	100.00	95.80	95.80	88.11
C7	85.54	82.65	97.64	83.58	91.98	88.71	81.62	93.28	87.22	93.84	90.67
C8	76.64	81.29	88.13	81.96	85.19	83.19	61.73	89.27	85.85	95.25	87.75
C9	87.35	88.29	85.93	83.76	85.93	91.60	86.31	89.42	93.29	89.24	90.46
C10	60.71	89.00	74.42	66.41	69.50	65.44	46.26	71.33	67.95	86.10	87.84
C11	90.61	83.78	84.54	74.57	85.48	95.92	69.35	92.88	79.13	96.68	99.53
C12	90.78	90.39	95.39	88.38	89.15	90.78	86.94	89.63	91.07	89.72	92.60
C13	86.67	82.46	87.37	88.42	77.19	91.93	78.25	89.47	84.21	85.96	91.23
C14	92.31	100.00	95.14	100.00	84.21	94.74	87.04	99.60	99.19	100.00	100.00
C15	99.79	98.10	100.00	100.00	87.53	100.00	18.82	97.89	100.00	98.73	99.58
OA	86.68	88.52	89.55	85.87	88.14	88.84	72.67	91.57	86.37	93.47	93.51
AA	87.52	89.95	91.05	88.55	87.76	90.34	71.63	92.82	88.40	93.75	93.77
K×100	85.56	87.59	87.59	84.76	87.12	87.92	70.40	90.85	85.29	92.90	92.96

表 7 AU 数据集分类不同方法的分类精度对比											
Table 7 Comparison of classification accuracy of different methods on AU dataset											(%)
类别	Two-Branch	EndNet	MDL-Middle	MAHiDF-Net	FusAt-Net	CALC	SepG-ResNet50	ExVit	DSym-Fuser	MCFNet	CFF-Net
C1	92.21	94.21	91.72	95.27	90.59	93.81	92.83	93.64	90.58	95.22	93.77
C2	95.41	86.08	95.19	96.84	91.07	95.69	92.90	97.71	96.32	98.82	98.22
C3	27.75	27.29	62.61	62.22	56.96	90.99	60.89	65.82	72.45	49.84	79.11
C4	85.27	92.75	87.76	86.27	85.08	94.79	73.03	84.88	89.65	91.39	93.36
C5	38.24	22.94	50.86	52.33	12.59	30.40	34.99	46.85	62.91	52.77	31.55
C6	3.42	23.91	24.24	28.45	14.11	2.63	4.00	24.54	17.22	21.06	8.61
C7	17.98	55.78	29.04	13.87	14.64	21.96	16.59	24.49	13.27	38.35	49.04
OA	84.20	84.59	87.74	87.94	80.45	90.99	80.50	87.72	88.35	90.10	91.54
AA	51.47	60.47	63.06	63.61	54.20	61.47	53.51	62.56	63.20	63.92	64.81
K×100	76.92	77.78	82.69	82.83	72.91	87.07	80.50	82.49	83.36	85.77	87.81

MU 数据集上,CFF-Net 对 trees 等类别的边界刻画更为精确,而对比方法在 buildings shadow 等类别边缘出现明显的区域混叠;在 HU 数据集上,CFF-Net 对 railway 等线状的识别结果更精准;在 AU 数据集上,CFF-Net 对 residential area 等大面积类别的边界定位更为准确。上述结果表明,CFF-Net 在高频细节建模、多源异质数据协同融合方面均具有显著优势,实现更加精细且连贯的地物识别。

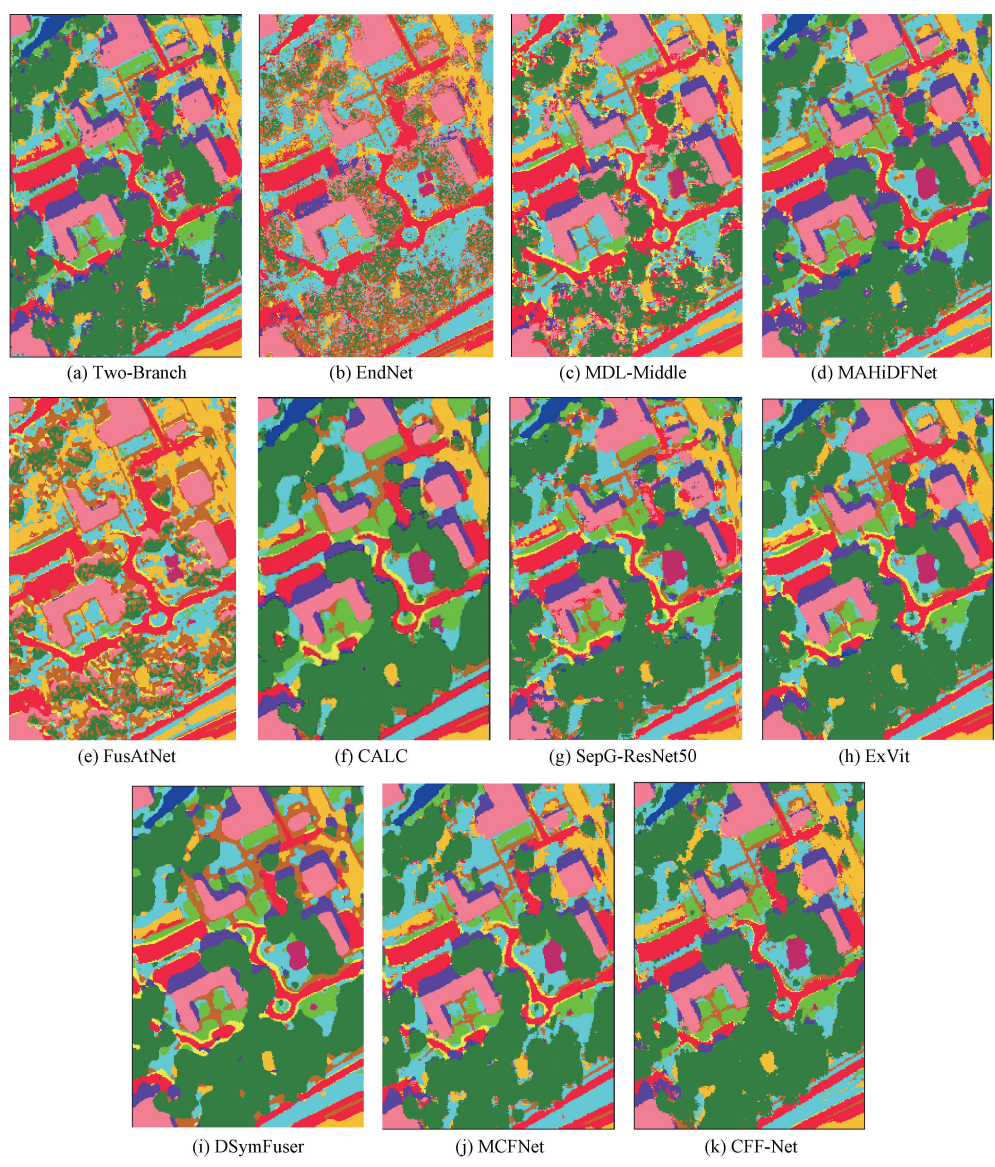


图 8 MU 数据集的分类结果
Fig. 8 Classification results of MU dataset

2) 模型参数对分类性能的影响

针对多源传感器遥感的地物分类任务,本文深入探究了图像块大小选择对分类效果的影响、超连接结构在不同模块中的作用。

(1) 图像块大小选择对分类性能的影响

本文以每个像素为中心,构建大小为 $n \times n$ 的图像块作为模型输入。图像块的大小对分类性能具有显著影响。为评估其作用,本文在不同块尺寸下进行了模型性能测试。图 11(a)~(c)展示了在 MU、HU 和 AU 数据集上,当图像块大小分别为 7×7 、 9×9 、 11×11 、 13×13 、 15×15 和 17×17 时的分类结果。

从图 11 可以看出,在 MU 数据集上,随着图像块尺寸由 7×7 逐步增大至 15×15 ,模型分类精度整体呈提升

趋势,并在 15×15 时取得最高 OA;当块尺寸进一步增大至 17×17 时,性能出现一定下降。HU 数据集的变化趋势表现出一定波动性,其中 9×9 与 11×11 尺寸间存在局部变化,但总体上 15×15 时仍取得最优分类效果,而 17×17 时下降较为明显。对于 AU 数据集,OA 在 13×13 时达到峰值,随后随着图像块尺寸继续增大而明显下降,尤其在 17×17 时下降更为显著。这主要是由于大尺寸图像块引入了更多与中心像素类别不一致的外围像素,增加了噪声干扰。因此,综合 MU 与 HU 两个主要数据集的实验结果及模型整体稳定性,本文最终选择 15×15 作为最优图像块尺寸。

(2) 超连接结构在不同模块中的作用

针对 Transformer 模块中 HC 结构在不同位置的影

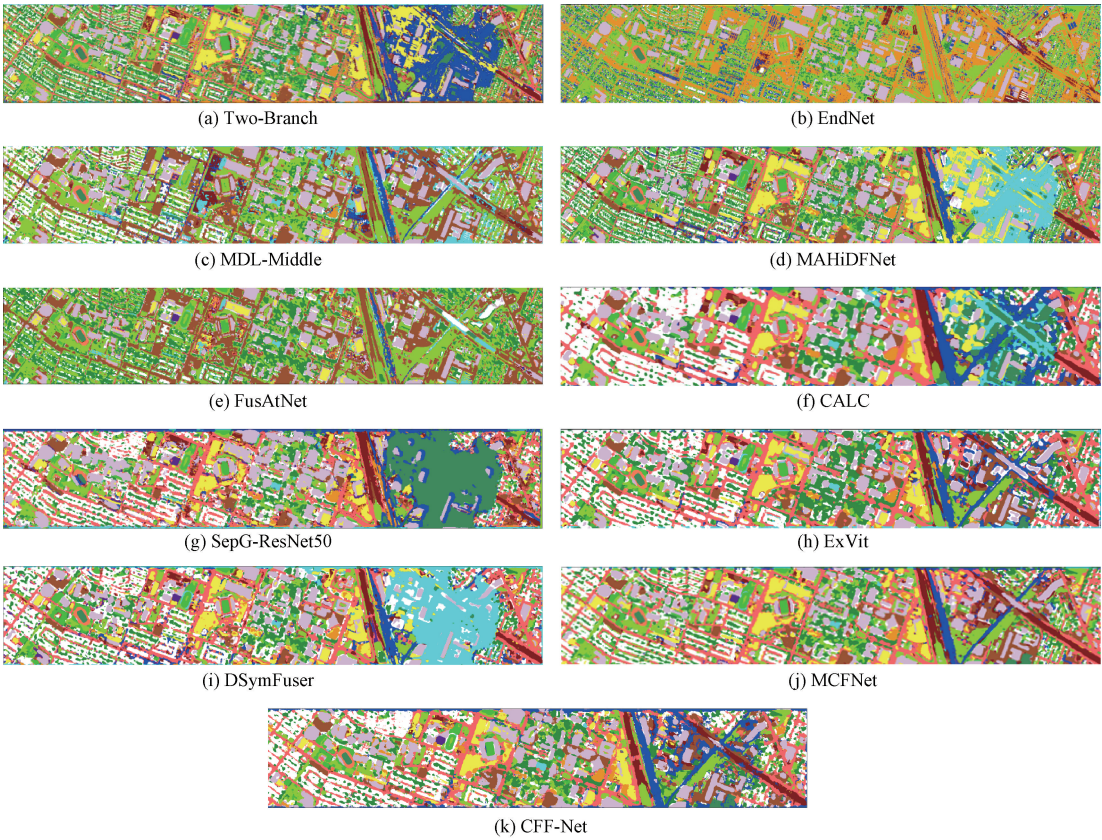


图 9 HU 数据集的分类结果
Fig. 9 Classification results of HU dataset

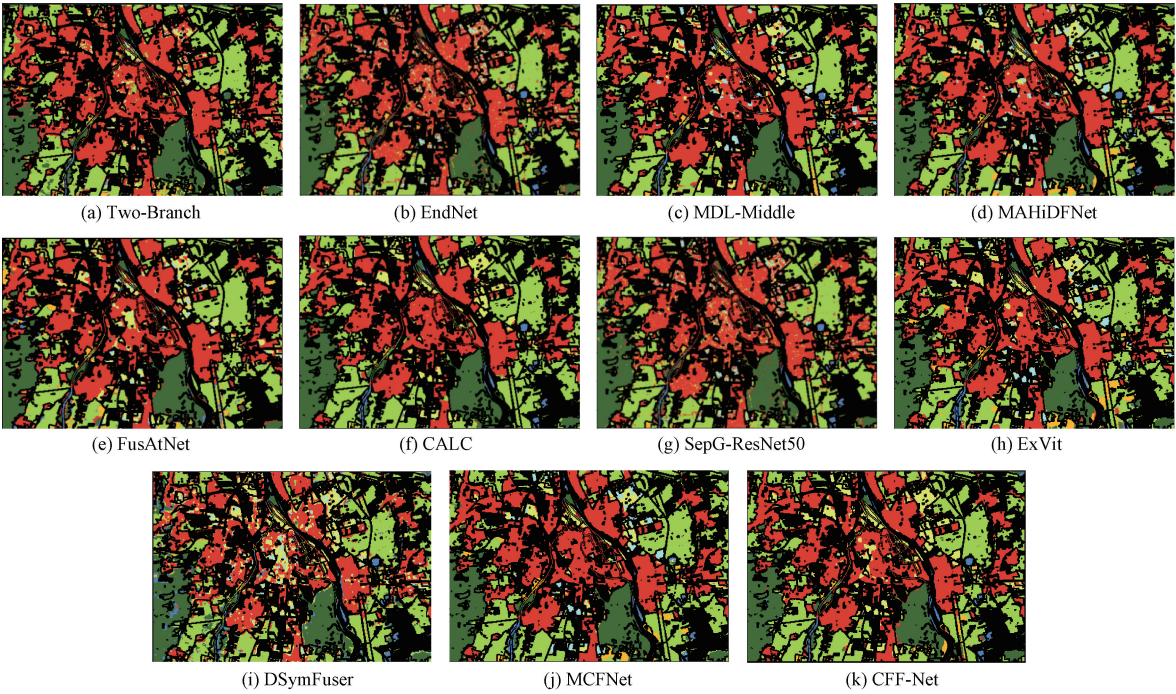


图 10 AU 数据集的分类结果
Fig. 10 Classification results of AU dataset

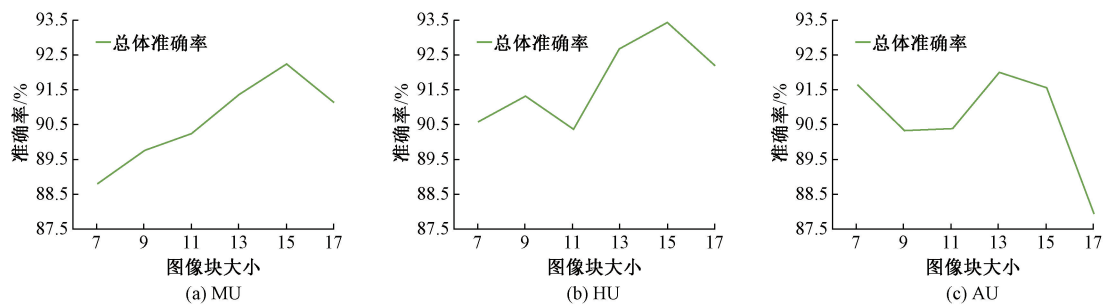


图 11 3 个数据集上图像块大小选择的影响

Fig. 11 Effects of image patch size on three datasets

响,本文分别在 Attention 子模块与 FFN 中引入 HC 结构进行对比实验,结果如表 8 所示,Attention-HC 配置(即在 Attention 模块中引入 HC 结构)在 MU 和 HU 数据集上分别取得 92.23% 和 93.51% 的最高 OA,因此本文将其作为最终最优模型配置;虽然 AU 数据集上 FFN-HC 略优,但综合整体稳定性与泛化能力,最终选择 Attention-HC 作为统一配置方案。

为进一步验证 HC 结构在 Transformer 模块内部的最优配置,本文通过 t-SNE 可视化分析对实验结果进行了深入探讨。如图 12 所示,Attention-HC 配置下的特征分

表 8 超连接结构在不同模块中的作用

HC		OA/%		
Attention	FFN	MU	HU	AU
×	×	90.98	92.98	90.25
×	√	91.23	93.06	91.83
√	×	92.23	93.51	91.54
√	√	91.63	93.18	91.25

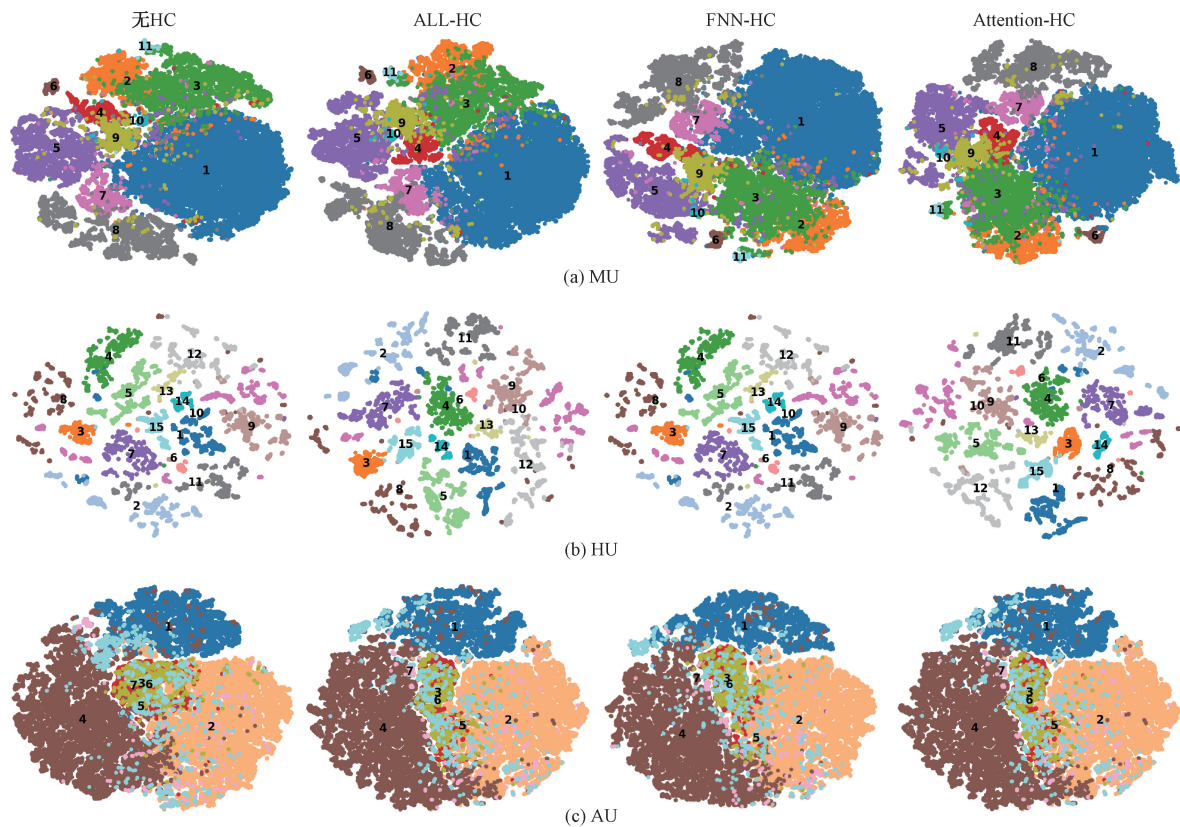


图 12 3 个数据集上 t-SNE 可视化结果

Fig. 12 Visualization Results of t-SNE on three datasets

布相较于无 HC 及其他配置具有更好的类间分离性和较高的类内紧致性,部分易混淆类别的边界更加清晰,说明超连接机制有助于增强特征判别能力,提高模型在多传感器异构特征表征中的稳定性与鲁棒性。

2.4 计算复杂性分析

本文采用浮点运算量(FLOPs)和参数量(Param)评估模型计算复杂度,结果如表 9 所示。

表 9 不同分类模型的 FLOPs 和参数数量 (M=10⁶)
Table 9 FLOPs and parameters of different classification models (M=10⁶)

模型	Param/M	FLOPs/M
Two-Branch	5.60	120.81
EndNet	0.09	0.09
MDL-Middle	0.10	5.28
MAHiDFNet	77.00	155.00
FusAtNet	36.90	3 460.31
CALC	0.30	28.79
SepG-ResNet50	14.70	48.28
ExVit	0.20	104.56
DSymFuser	0.80	78.54
MCFNet	0.30	39.32
CFF-Net	0.30	60.55

由表 9 可知,MAHiDFNet (77.00 M) 和 FusAtNet (36.90 M) 参数量最大,FusAtNet 的 FLOPs 亦高达 3 460.31 M,计算开销显著高于其他方法。CFF-Net 的参数量为 0.30 M, FLOPs 为 60.55 M,与参数量相当的 CALC (28.79 M) 和 MCFNet (39.32 M) 相比,FLOPs 略高,但在 3 个数据集上取得了更优的分类精度,表明 CFF-Net 在模型紧凑性与分类性能之间实现了合理平衡,具备较好的工程实用性。

3 结 论

针对多传感器遥感数据中因成像机理、光谱响应及散射特性差异导致的特征表达不一致问题,本文提出了一种结合频域特征融合与超连接机制的跨传感器频域感知融合网络(CFF-Net)。通过高低频联合建模与跨传感器频率自注意力融合,CFF-Net 在 MU、HU 和 AU 3 个数据集上分别取得 92.23%、93.51% 和 91.54% 的总体分类精度,在保持 0.30 M 参数量的轻量级结构下实现了性能与效率的平衡。当前方法仍存在一定局限:频域建模的计算路径相对固定,对训练样本极少的类别分类精度有

待提升。未来工作也将尝试将所提框架推广至更多传感器组合的协同分类场景。

参考文献

[1] 李盛阳,刘志文,刘康,等. 航天高光谱遥感应用研究进展(特邀)[J]. 红外与激光工程,2019,48(3):9-23.
LI SH Y, LIU ZH W, LIU K, et al. Advances in application of space hyperspectral remote sensing (invited) [J]. Infrared and Laser Engineering, 2019, 48(3): 9-23.

[2] 刘城洋,赵琳,于亮,等. 对比学习结合 DenseNet 的高光谱图像开放集分类[J]. 光学精密工程,2025, 33(23):3737-3753.
LIU CH Y, ZHAO L, YU L, et al. Contrastive learning combined with DenseNet for open-set classification of hyperspectral images [J]. Optics and Precision Engineering, 2025, 33(23):3737-3753.

[3] ZHANG M M, LI W, DU Q, et al. Feature extraction for classification of hyperspectral and LiDAR data using patch-to-patch CNN [J]. IEEE Transactions on Cybernetics, 2020, 50(1):100-111.

[4] LYU P Y, WU W J, ZHONG Y F, et al. Review of vision transformer models for remote sensing image scene classification [C]. IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium. Kuala Lumpur, 2022: 2231-2234.

[5] LI J X, HONG D F, GAO L R, et al. Deep learning in multimodal remote sensing data fusion: A comprehensive review [J]. International Journal of Applied Earth Observation and Geoinformation, 2022, 112:102926.

[6] 吴海滨,左云逸,王爱丽,等. CNN 结合 Transformer 的高光谱图像和 LiDAR 数据协同地物分类方法[J]. 仪器仪表学报,2025,46(8):286-301.
WU H B, ZUO Y Y, WANG AI L, et al. Collaborative classification method of hyperspectral image and LiDAR data based on CNN combined with Transformer [J]. Chinese Journal of Scientific Instrument, 2025, 46(8): 286-301.

[7] 吴海滨,戴诗语,王爱丽,等. CNN-Transformer 结合对比学习的高光谱与 LiDAR 数据协同分类[J]. 光学精密工程, 2024, 32(7): 1087-1100.
WU H B, DAI SH Y, WANG AI L, et al. Collaborative classification of hyperspectral and LiDAR data based on CNN-Transformer [J]. Optics and Precision Engineering,

- 2024, 32(7):1087-1100.
- [8] XU X D, LI W, RAN Q, et al. Multisource remote sensing data classification based on convolutional neural network [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 56(2):937-949.
- [9] HONG D F, GAO L R, HANG R L, et al. Deep encoder-decoder networks for classification of hyperspectral and LiDAR data [J]. *IEEE Geoscience and Remote Sensing Letters*, 2022, 19: 5500205.
- [10] HONG D F, GAO L R, YOKOYA N, et al. More diverse means better: Multimodal deep learning meets remote-sensing imagery classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 59(5): 4340-4354.
- [11] YAO J, ZHANG B, LI CH Y, et al. Extended vision transformer for land use and land cover classification: A multimodal deep learning framework [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 5514415.
- [12] CHANG H H, BI H X, LI F, et al. Deep symmetric fusion transformer for multimodal remote sensing data classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 5644115.
- [13] SONG Q Y, MO F, DING K X, et al. MCFNet: multiscale cross-domain fusion network for HSI and LiDAR data joint classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 4703912.
- [14] 左云逸, 吴海滨, 王爱丽. 双域掩码与多尺度局部重建的 HSI 和 LiDAR 数据协同分类方法[J]. *红外与激光工程*, 2025, 54(10): 294-306.
- ZUO Y Y, WU H B, WANG AI L. Collaborative classification method for HSI and LiDAR data based on dual-domain masking and multi-scale local reconstruction[J]. *Infrared and Laser Engineering*, 2025, 54(10):294-306.
- [15] XU K, WANG B J, ZHU ZH, et al. A contrastive learning enhanced adaptive multimodal fusion network for hyperspectral and LiDAR data classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 63: 4700319.
- [16] ZHAO F, FAN J D, ZHANG J J, et al. Hybrid feature extraction and Mamba fusion for multi-modal remote sensing image classification [J]. *IEEE Geoscience and Remote Sensing Letters*, 2026, 23: 5501605.
- [17] PARK N, KIM S. How do vision transformers work? [J]. *ArXiv preprint arXiv:2202.06709*, 2022.
- [18] CHEN L W, GU L, ZHENG D ZH, et al. Frequency-adaptive dilated convolution for semantic segmentation[C]. *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024: 3414-3425.
- [19] BAI J W, YUAN L, XIA SH T, et al. Improving vision transformers by revisiting high-frequency components [C]. *European Conference on Computer Vision*, 2022:1-18.
- [20] TATSUNAMI Y, TAKI M. FFT-based dynamic token mixer for vision [C]. *AAAI Conference on Artificial Intelligence*, 2024: 15328-15336.
- [21] ZHU D F, HUANG H ZH, HUANG Z H, et al. Hyperconnections [C]. *International Conference on Learning Representations*, 2025.
- [22] GADER P, ZARE A, CLOSE R, et al. MUUFL Gulfport hyperspectral and LiDAR airborne data set [R]. Gainesville: University of Florida, 2013.
- [23] DEBES C, MERENTITIS A, HEREMANS R, et al. Hyperspectral and LiDAR data fusion: Outcome of the 2013 GRSS data fusion contest [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2014, 7(6):2405-2418.
- [24] HONG D F, HU J L, YAO J, et al. Multimodal remote sensing benchmark datasets for land cover classification with a shared and specific feature learning model [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2021, 178: 68-80.
- [25] WANG X H, FENG Y N, SONG R X, et al. Multi-attentive hierarchical dense fusion net for fusion classification of hyperspectral and LiDAR data [J]. *Information Fusion*, 2022, 82: 1-18.
- [26] MOHLA S, PANDE S, BANERJEE B, et al. FusAtNet: Dual attention based spectrospatial multimodal fusion network for hyperspectral and LiDAR classification [C]. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020:92-93.
- [27] LU T, DING K X, FU W, et al. Coupled adversarial learning for fusion classification of hyperspectral and LiDAR data [J]. *Information Fusion*, 2023, 93: 118-131.
- [28] YANG Y, ZHU D Y, QU T T, et al. Single-stream CNN with learnable architecture for multisource remote sensing

data[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5409218.

作者简介



王爱丽, 2002 年于哈尔滨工业大学获得学士学位, 2004 年于哈尔滨工业大学获得硕士学位, 2008 年于哈尔滨工业大学获得博士学位, 现为哈尔滨理工大学教授, 主要研究方向为遥感图像处理。

E-mail: aili925@hrbust.edu.cn

Wang Aili received her B. Sc. , M. Sc. and Ph. D. degrees all from Harbin Institute of Technology in 2000, 2004 and 2008, respectively. She is currently a professor at Harbin University of Science and Technology. Her main research interest includes remote sensing image processing.



戴诗语, 2013 年于哈尔滨理工大学获得学士学位, 2016 年于哈尔滨理工大学获得硕士学位, 2022 年于哈尔滨理工大学仪器科学与技术专业攻读博士学位, 主要研究方向为机器视觉、深度学习。

E-mail: 2210604079@stu.hrbust.edu.cn

Dai Shiyu received her B. Sc. degree from Harbin University of Science and Technology in 2013, her M. Sc. degree from Harbin University of Science and Technology in 2016. She is currently a Ph. D. candidate in Instrument Science and Technology at Harbin University of Science and Technology from 2022. Her main

research interests include machine vision and deep learning.



于亮, 2013 年于哈尔滨工业大学获得学士学位, 2020 年于哈尔滨工业大学获得博士学位, 现为哈尔滨工业大学副教授。主要研究方向为精密仪器测量。

E-mail: liangyu@hit.edu.cn

Yu Liang received his B. Sc. and Ph. D. degree from the Harbin Institute of Technology in 2013 and 2020, respectively. He is currently an associate professor at Harbin Institute of Technology. His main research interests include precision instrument measurement.



吴海滨 (通信作者), 2000 年于哈尔滨工业大学获得学士学位, 2002 年于哈尔滨工业大学获得硕士学位, 2008 年于哈尔滨理工大学获得博士学位, 现为哈尔滨理工大学教授, 主要研究方向为计算机视觉、虚拟现实、遥感图像处理。

E-mail: woo@hrbust.edu.cn

Wu Haibin (Corresponding author) received his B. Sc. and M. Sc. degrees from Harbin Institute of Technology in 2000 and 2002, respectively, and his Ph. D. degree from Harbin University of Science and Technology in 2008. He is currently a professor at Harbin University of Science and Technology. His main research interests include computer vision, virtual reality, and remote sensing image processing.