DOI: 10.19650/j.cnki.cjsi.J2513731

CNN 结合 Transformer 的高光谱图像和 LiDAR 数据协同地物分类方法*

吴海滨,左云逸,王爱丽,吕浩然,王敏慧
(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)

摘 要:在高光谱图像与 LiDAR 数据协同分类的研究领域中,尽管 CNN 和 Transformer 在图像处理和数据分析中分别展现出对局部特征和全局依赖关系的敏锐洞察力,但二者的协同机制尚未充分挖掘,跨模态特征互补潜力未被有效释放。故提出了一种 CNN 结合 Transformer 的高光谱图像和 LiDAR 数据的多模态遥感数据协同地物分类方法。首先,该模型通过主成分分析对高光谱图像进行降维处理以去除光谱的冗余信息,继而利用 CNN 分层捕获局部纹理特征,同时借助 Transformer 自注意力机制构建全局光谱-空间表征。然后通过双向特征交互机制,将 Transformer 输出的全局上下文信息注入 CNN 特征通道,同时将 CNN 通道提取的局部细节反馈至 Transformer 支路,经特征耦合单元实现跨尺度特征对齐,强化模型对高光谱图像全局结构与局部细节的联合提取能力。对于 LiDAR 数据,采用动态卷积级联模块有效捕获高程信息和上下文关系,最终通过跨模态特征融合模块实现双源数据特征的深度交互与融合,在双模态语义互补中提升复杂地物的分类精度。在 Houston2013、Trento 和 Augsburg 这 3 个公开数据集上的实验表明,该方法总体分类精度分别达到 99.85%、99.68% 和 97.34%,平均准确率分别达到 99.87%、99.34% 和 90.60%,较 GLT、HCT 等主流方法的分类精度有所提高,充分证明所提方法进行多模态数据协同分类的优势和有效性。

关键词: 高光谱图像;LiDAR 数据;Transformer;卷积神经网络;多模态数据
中图分类号: TH761 **文献标识码**: A **国家标准学科分类代码**: 420.2040

Collaborative land classification method using CNN combined with Transformer for hyperspectral images and LiDAR data

Wu Haibin, Zuo Yunyi, Wang Aili, Lyu Haoran, Wang Minhui
(*Heilongjiang Province Key Laboratory of Laser Spectroscopy Technology and Application, Harbin University of Science and Technology, Harbin 150080, China*)

Abstract: In the field of collaborative classification between hyperspectral images and LiDAR data, although CNN and Transformer have shown keen insight into local features and global dependencies in image processing and data analysis, their collaborative mechanisms have not been fully explored, and the potential for cross-modal feature complementarity has not been effectively unleashed. Therefore, this article proposes a multimodal collaborative land-cover classification method for remote sensing data that combines CNN with Transformer for hyperspectral images and LiDAR data. Firstly, the model performs dimensionality reduction on hyperspectral images through principal component analysis to remove redundant spectral information. Then, it uses CNN layers to capture local texture features, and constructs a global spectral-spatial representation using the Transformer self-attention mechanism. Then, through a bidirectional feature interaction mechanism, the global contextual information from the Transformer is injected into the CNN feature channels, while the local details extracted by the CNN are fed back into the Transformer branch. Cross-scale feature alignment is achieved through the feature coupling unit, enhancing the joint extraction ability of the model for the global structure and local details of hyperspectral images. For LiDAR data, a dynamic convolution cascade module is used to effectively capture elevation information and contextual relationships. Finally, a cross-modal feature fusion module is used to achieve deep interaction and fusion of dual source data

收稿日期:2025-02-06 Received Date: 2025-02-06
*基金项目:黑龙江省重点研发计划项目(JD2023SJ19)资助

features, improving the classification accuracy of complex land features in the complementary semantics of dual modalities. Experiments on three publicly available datasets—Houston 2013, Trento, and Augsburg—showed that the overall classification accuracy of our proposed method reached 99.85%, 99.68%, and 97.34%, respectively, with average accuracies of 99.87%, 99.34%, and 90.60%. This improvement in classification accuracy compared to mainstream methods such as GLT and HCT fully demonstrates the advantages and effectiveness of our proposed method for multimodal data collaborative classification.

Keywords: hyperspectral image; LiDAR data; Transformer; convolutional neural network; multimodal data

0　引　　言

遥感高光谱图像(hyperspectral image, HSI)包含数百个光谱波段,能够比较准确地反映不同地物的光谱特征[1]。但是,高光谱图像在区分光谱相似的地物时存在同谱异物、同物异谱的现象[2-3]。在这种情况下,光探测与测距(light detection and ranging, LiDAR)数据记录的物体高程信息能够有效地补充高光谱信息的不足。因此,结合高光谱图像与 LiDAR 数据以增强地物分类的精确性,已成为遥感探测领域国内外研究的前沿课题[4]。

针对 HSI 特征提取模块,基于卷积神经网络(convolutional neural network, CNN)的方法通常采用局部像素与上下文信息之间的连通性来学习高光谱图像的特征,包括 2-D CNN[5-6] 和 3-D CNN[7-8]。Yang 等[9] 结合 1-D 和 2-D CNN 设计了双分支网络以提取光谱和空间特征。Chen 等[10] 开发了一种基于正则化深度特征提取的 3-D CNN 方法,能够有效地提取 HSI 中的联合空间-光谱信息。He 等[11] 提出残差网络,通过残差链路减少信息损失。Zhu 等[12] 提出了一种端到端的残差频谱空间注意网络(residual spectral-spatial attention network, RSSAN)通过设计光谱关注模块和空间注意模块,自适应选择有用波段和邻域像素,提升了 HSI 分类性能。Lu 等[13] 提出的基于耦合对抗学习的分类方法(coupled adversarial learning based classification, CALC)通过无监督的耦合对抗特征学习子网络和有监督的多层特征融合分类子网络有效提取和融合 HSI 和 LiDAR 数据中的高阶语义特征,并通过自适应概率融合策略进一步提高分类性能。Lin 等[14] 提出动态跨模态特征交互网络(dynamic cross-modal feature interaction network, DCMNet),通过构建含双线性空间注意力块、双线性通道注意力块和集成卷积块的 3 层路由空间,利用动态路由机制实现基于输入数据的自适应特征交互路径,以进行 HSI 和 LiDAR 数据的联合分类。2-D CNN 架构在提取 HSI 的空间特征方面表现优异,但是在捕捉长距离依赖关系方面存在局限性[15]。

近年来,Transformer 架构凭借其在长距离序列依赖方面的显著优势,已被成功应用于图像处理领域,并展现出优异的性能[16-17]。Dosovitskiy 等[18] 将 Transformer 网络应用于图像分类任务上的表现尤为突出。Qing 等[19] 通过将频谱注意机制与 Transformer 中的多头注意机制相结合,成功捕获序列频谱关系,提高了 HSI 分类性能。Hong 等[20] 采用 Transformer 结构挖掘和表示频谱特征的局部序列属性,设计跳变连接进一步提高分类效果。Sun 等[21] 提出了光谱-空间特征标记化 Transformer(spectral-spatial feature tokenization transformer, SSFTT)方法,通过跨层编码器对局部空间上下文信息进行建模,以表征相邻序列间的关系。Liang 等[22] 提出邻域对比标记任务(neighborhood contrastive tokenization task, NeiCoT),使用预测器最大化局部点与全局平均锚点之间的共有信息,增强了 Transformer 在 HSI 编码过程中的特征学习相关性。Mei 等[23] 提出群感知层次 Transformer(group-aware hierarchical transformer, GAHT),通过引入分组像素嵌入(grouped pixel embedding, GPE)模块和多头自注意机制(multi-head self-attention, MHSA)作用到局部空间光谱上下文中,提高了 HSI 分类性能。

针对 HSI 与 LiDAR 数据的联合分类研究,Feng 等[24] 提出的全局-局部 Transformer 网络(global-local Transformer network, GLT),通过连接 HSI 和 LiDAR 的特征来捕获局部-全局特征。Ding 等[25] 提出动态尺度分层融合网络(dynamic scale hierarchical fusion network, DSHFNet),通过计算尺度空间中的相似度,动态选择合适的尺度特征,并有效进行特征降维。Xu 等[26] 提出的多支路交互 Transformer 同时提取光谱、空间和高程信息。Ni 等[27] 提出选择性光谱-空间聚合 Transformer 网络(selective spectral-spatial aggregation Transformer, S2ATNet),通过卷积特征选择模块动态捕获不同土地覆盖的上下文特征,利用级联空间-光谱学习和交互融合块进行特征交互学习,再经最大平均分类头获得最终分类结果,以实现高光谱图像和 LiDAR 数据的土地覆盖分类。Zhao 等[28] 设计了分层 CNN 和 Transformer(hierarchical CNN and Transformer, HCT)模型,先通过双分支 CNN 网络分别提取 HSI 和 LiDAR 特征,再经特征标记化和 Transformer 编码器处理,利用交叉令牌注意力融合编码器融合信息,最后通过分类器得到分类结果。然而,这些方法主要关注于不同传感器数据特征的叠加而缺乏有效的特征间交互,或仅将特征交互限制在浅层,没有充分考虑局部与全局特征之间的依赖关系。因此,本文充分利用 CNN 和 Transformer 的优势,提出了 CNN 结合 Transformer 的 HSI

和 LiDAR 数据协同的地物分类方法，提升了多模态遥感数据的分类准确率。

1 CNN 结合 Transformer 的多模态数据分类

1.1 算法架构

CNN 结合 Transformer 的多模态数据协同地物分类方法的架构如图 1 所示，简称 MMCT（multi modal classification combined CNN and Transformer）。首先，采用主成分分析法（principal component analysis，PCA）对 HSI 进行降维，降低数据冗余度并获取关键信息。在特征提取阶段，采用 CNN 分层捕获 HSI 局部纹理特征，同时通过 Transformer 自注意力机制构建全局光谱-空间表征。通过双向特征交互机制，将 Transformer 输出的全局上下文信息注入 CNN 特征图，并将 CNN 提取的局部细节反馈至 Transformer 的补丁嵌入层，最终通过特征耦合单元实现跨尺度特征对齐，有效增强模型对遥感影像全局结构和局部细节的联合提取能力。在 LiDAR 数据支路，引入动态卷积级联模块提取高程信息，有效捕捉空间信息和上下文关系。最后，利用跨模态特征融合模块（cross modal feature fusion module）通过编码器-解码器的多级堆叠结构，实现多模态特征的分层交互与融合，该机制在强化模型对复杂地物模式建模能力的同时，有效提升了特征表达的判别性与分类精度。

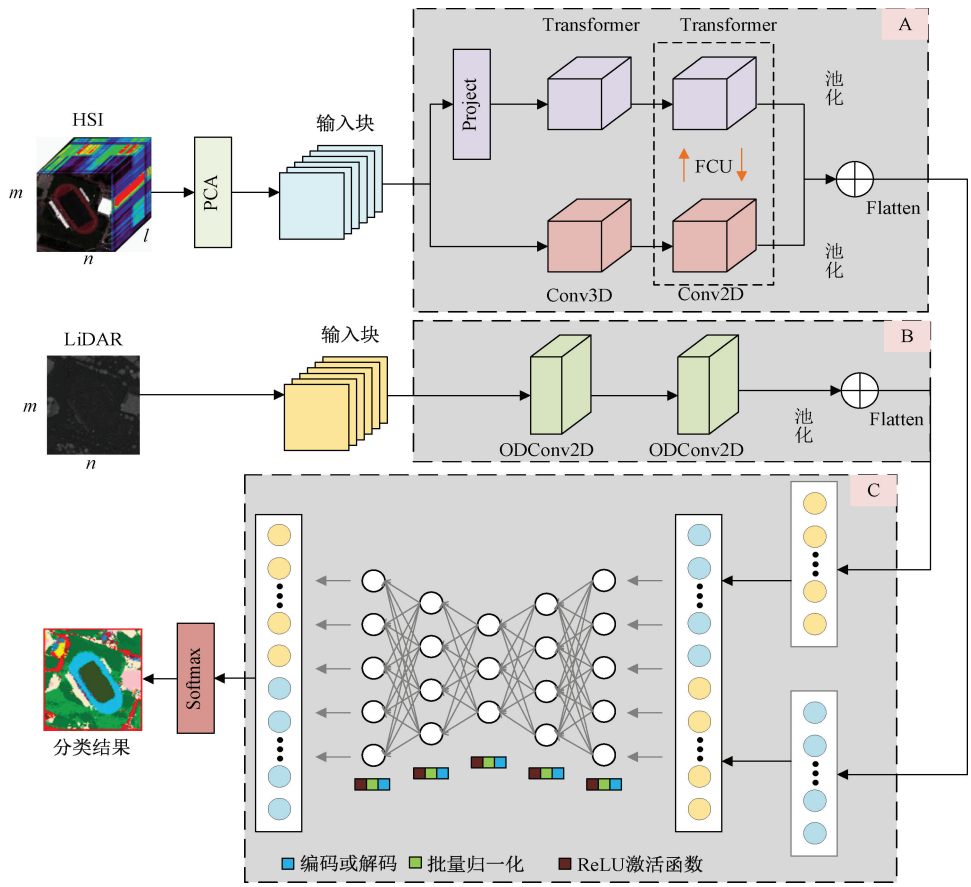


图 1 CNN 结合 Transformer 的多模态数据分类方法

Fig. 1 Multi modal data classification method diagram combined CNN with Transformer

1.2 HSI 空-谱特征提取

将 HSI 数据记为 $X_H \in \mathrm{R}^{m\times n\times l}$，覆盖同一区域的 LiDAR 数据记为 $X_L \in \mathrm{R}^{m\times n}$，其中 m 和 n 代表空间的长和宽，l 为 HSI 光谱维数。为了降低光谱维数，采用主成分分析法提取 X_H 的前 b 个主成分，即在保持空间维度不变的情况下将光谱维数从 l 降低到 b，X_H 转换为 $X_H^{PCA} \in \mathrm{R}^{m\times n\times b}$。

之后，对于每个像素及其周围的像素 patch 提取，分别得到一个立方体（$X_H^P \in \mathrm{R}^{s\times s\times b}$）和小 patch（$X_L^P \in \mathrm{R}^{s\times s}$），$s\times s$ 为 patch 大小。使用中心像素的标签来标记每一个 patch；对于边缘像素执行填充操作，填充的宽度为 $(s-1)/2$。

CNN 与 Transformer 特征耦合单元如图 2 所示，CNN 支路采用特征金字塔结构，特征图的分辨率随着网络深

度的增加而降低。通道中包括由 n 个瓶颈层构成的卷积块。瓶颈层包含 1 × 1 下投影卷积、3 × 3 空间卷积、1 × 1 上投影卷积以及瓶颈层输入和输出之间的残差连接。2 个卷积块中 n 的值分别为 1 和 2。

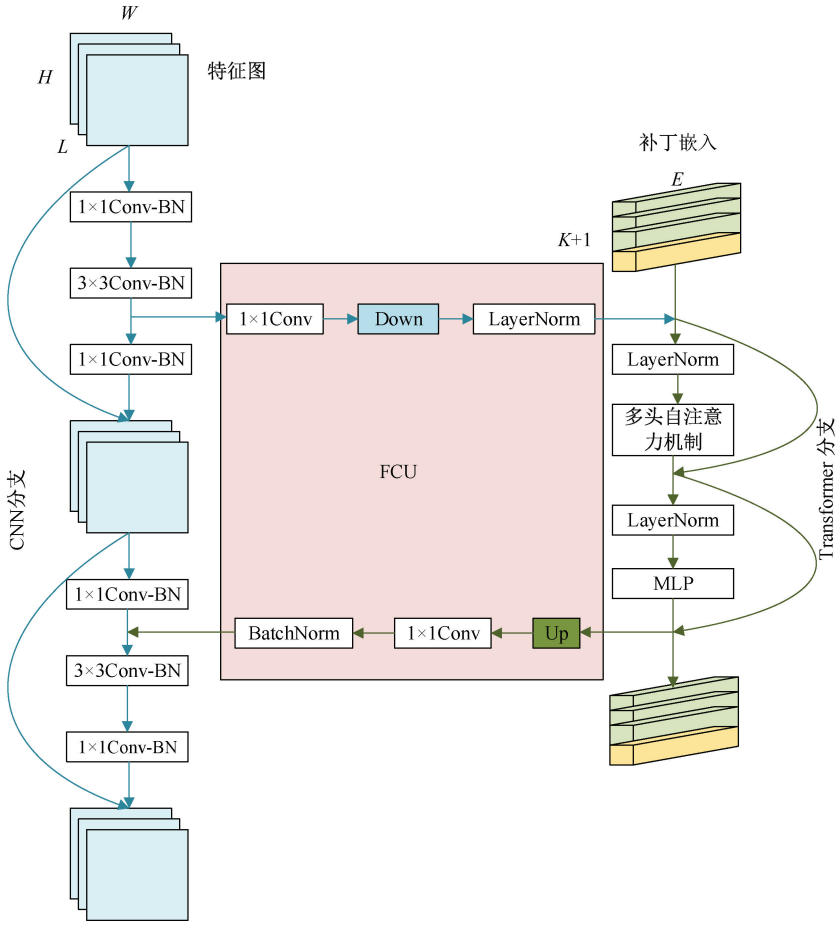


图 2　CNN 与 Transformer 特征耦合单元

Fig. 2　Feature coupling unit of CNN and Transformer

Transformer 支路由 2 个结构相似的 Transformer 块构成,每个 Transformer 块由一个多头自注意力机制模块和一个多层感知机(multilayer perceptron,MLP)块组成,其中层归一化(Layernorm)层被应用于自注意力机制模块和 MLP 模块的残差连接之前。在利用线性投影层将高 HSI 特征图压缩为无重叠的 14×14 补丁嵌入之后,通过一个步幅为 4 的 4×4 卷积核进行处理,随后对补丁嵌入标记执行分类。

通过特征交互实现局部特征与全局的连续耦合,以期消除 CNN 支路的特征图和 Transformer 支路的补丁嵌入之间的错位。CNN 的特征图的维数为 $C\times H\times W$,其中 C、H、W 分别为通道数、高度和宽度,补丁嵌入的形状为 $(K+1)\times E$,其中 K、1、E 分别表示补丁数、类标记和嵌入维数。在将局部特征输入到 Transformer 支路时,先采用 1×1 卷积使特征图对齐补丁嵌入的通道数,然后使用下采样模块完成空间维度的对齐,最后将特征图输入到 Transformer 支路中。全局信息从 Transformer 支路输送到 CNN 支路时,对补丁嵌入进行上采样以对齐空间尺度,然后通过 1×1 卷积将通道维数与 CNN 特征图的维数对齐,之后将全局信息添加到特征图中。在这个过程中,使用 LayerNorm 和 BatchNorm 模块对特征进行正则化。

1.3　LiDAR 数据特征提取

常规卷积层使用一个静态卷积核,该卷积核在所有输入样本中保持不变。而动态卷积层则通过多种卷积核的线性组合,采用注意力机制对这些卷积核进行动态加权,使模型能够自适应关注 LiDAR 数据中不同区域的关键高度变化。所以,本研究 LiDAR 数据特征提取模块由两个动态卷积块级联构成。

动态卷积由卷积核与注意力函数生成的权重系数构成。核空间具空间核大小、输入通道数、输出通道数及卷积核数 4 个维度。DyConv 等传统动态卷积的注意力函数仅为卷积核分配单一注意力权重,致使其空间、

输入通道和输出通道维度被忽视。基于此,ODConv 引入多维注意力机制,采用并行策略沿核空间四维度学习卷积核的互补性注意力。DyConv 和 ODConv 的结构图如图 3 所示。

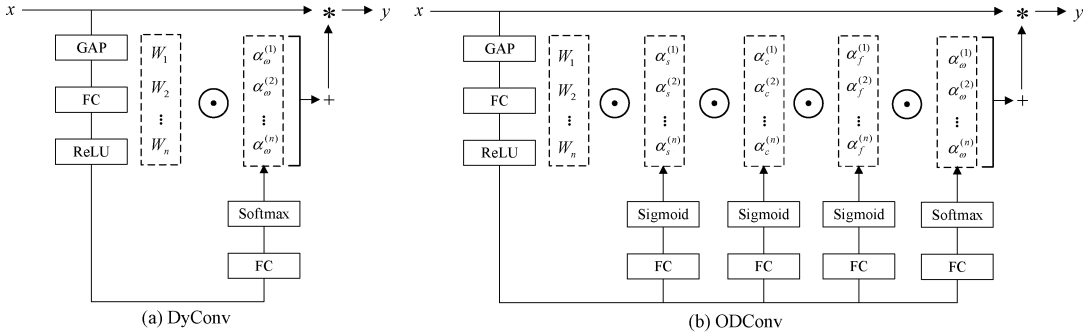


图 3 DyConv 和 ODConv 的结构

Fig. 3 Structure of DyConv and ODConv

ODConv 可以定义为:

$$y=(\alpha_{\omega}^{(1)}\odot\alpha_{f}^{(1)}\odot\alpha_{c}^{(1)}\odot\alpha_{s}^{(1)}\odot W_1+\cdots+\alpha_{\omega}^{(n)}\odot\alpha_{f}^{(n)}\odot\alpha_{c}^{(n)}\odot\alpha_{s}^{(n)}\odot W_n)*x \tag{1}$$

式中:$\alpha_{\omega}^{(i)}\in\mathrm{R}$ 为卷积核 W_i 的注意力权重系数;$\alpha_{s}^{(i)}\in\mathrm{R}^{k\times k}$、$\alpha_{c}^{(i)}\in\mathrm{R}^{c_{in}}$ 和 $\alpha_{f}^{(i)}\in\mathrm{R}^{c_{out}}$ 分别表示3个新引入的注意力项,分别沿卷积核 W_i 的核空间的空间维数、输入通道维数和输出通道维数计算;$\odot$ 表示沿核空间不同维数的乘法运算。

ODConv 为卷积核在空间、通道、滤波器和核 4 个维度分配差异化的注意力权重,4 种注意力相互补充,提升特征提取性能。其计算方法为:先对输入特征进行全局平均池化(global average pooling, GAP)处理,再通过全连接(full connection, FC)层和 4 个分支处理,分别生成对应的注意力权重。当使用 ODConv 处理 LiDAR 高程特征时,模型架构包括两层 ODConv,每层后跟批归一化和 ReLU 层,有助于优化训练过程与提升模型性能。

ODConv 中的 4 种注意力机制如图 4 所示,通过注意力机制显著提升了对 LiDAR 数据的特征提取能力。其核心在于为每个卷积核同时引入空间位置、通道维度、滤波器层级和整体核层级的动态权重分配(α_{ω},α_{s},α_{c},α_{f})。相比于传统的二维卷积,空间注意力使卷积核在不同位置具有差异化响应,适应地形局部特征。此外,通道注意力能够强化高程特征的关键信息。这种协同机制使 ODConv 能动态捕获 LiDAR 数据中的多维度上下文特征,通过堆叠 2 个 ODConv 块(配合批归一化+ReLU),在 LiDAR 高程信息处理中展现出强大的地形特征表征能力和收敛效率。

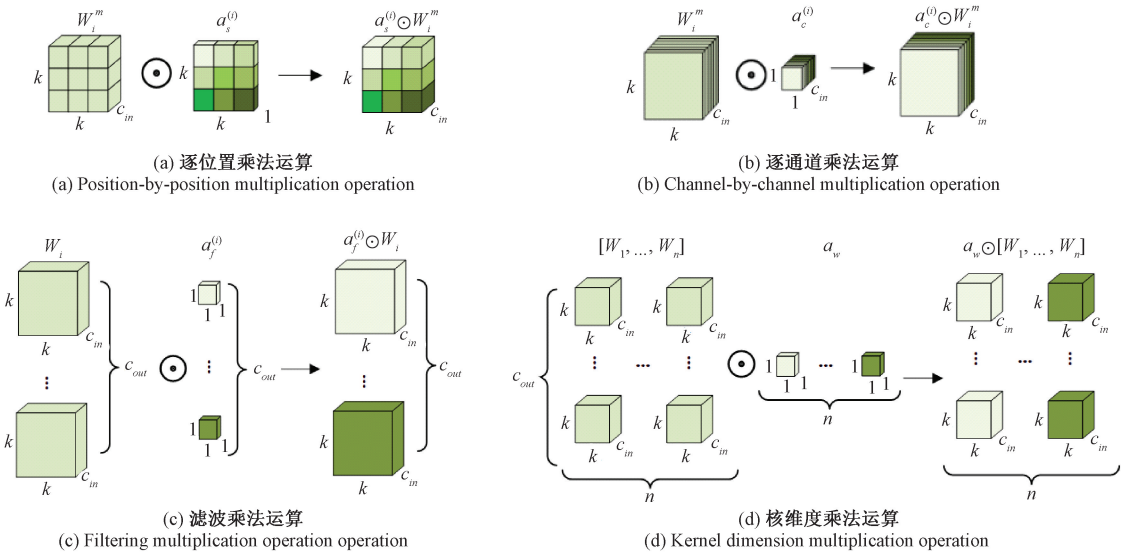


图 4 ODConv 中的 4 种注意力机制

Fig. 4 Four types of attention mechanism in ODConv

1.4　多模态数据跨通道特征融合

与传统方法(如简单拼接、对齐或浅层融合)相比,跨通道重建模块(cross-channel reconstruction,CCR)能够更充分地挖掘多模态数据的互补特性,实现模态间的双向信息交互,生成更紧凑且判别性更强的融合特征。跨通道编码器-解码器结构原理如图 5 所示。

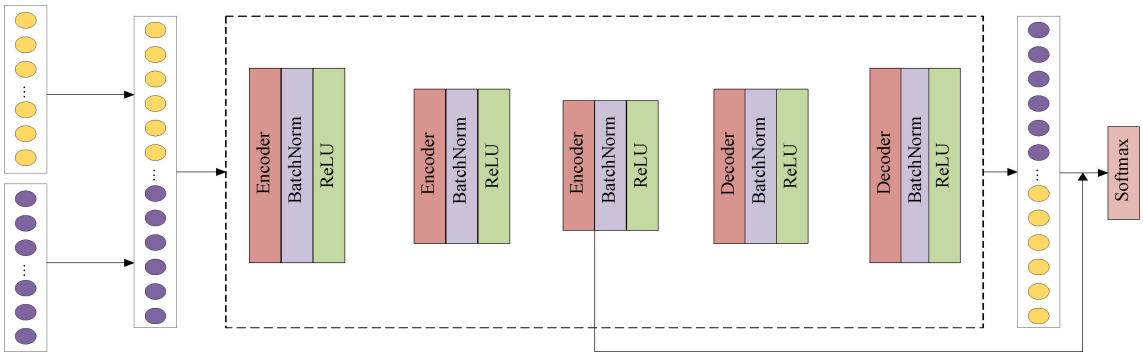


图 5　跨通道编码器-解码器结构原理

Fig. 5　Diagram of cross-channel encoder-decoder structure

首先将双支路提取的 HSI 特征与 LiDAR 特征拼接起来,之后采用编码器-解码器的结构进行特征级的交叉融合,过程为:

$$v_i^{(l)} = g_{en}(z_{s,i}^{(p)}, W_v^{(p)}, b_v^{(l)}) = f(W_v^{(l)}[z_{1,i}^{(p)}, z_{2,i}^{(p)}] + b_v^{(l)}), \quad l = p+1, \cdots, m \tag{2}$$

式中:$v_i^{(l)}$ 表示第 l 层的像素级特征融合;函数 g_{en} 定义为跨模态编码器-解码器结构中的编码器网络。解码器部分 g_{de} 表示为:

$$u_i^{(l)} = g_{de}(v_i^{(m)}, W_u^{(l)}, b_u^{(l)}) = f(W_u^{(l)} v_i^{(m)} + b_u^{(l)}), \quad l = m+1, \cdots, n \tag{3}$$

式中:$u_i^{(l)}$ 表示第 l 层的像素重构特征。跨模态编码器-解码器结构需要通过网络学习从输入特征 $\{z_{s,i}^{(p)}\}_{p=1}^{2}$ 到输出重构特征 $u_i^{(n)}$ 的映射。之后,通过优化总体损失函数更新网络参数,再将双支路交叉融合之后的特征输入 softmax 分类器进行分类,其中总体损失函数由两部分构成,分别为 PolyLoss 以及重构特征与跨通道特征之间的 L2 范数正则化损失。具体为:

$$L = L_{Poly-1} + \alpha L_{rec} \tag{4}$$

式中:α 是为了平衡上式中不同项的参数。在具体实验中确定为 1,从而产生相对稳定的性能。

PolyLoss 是通过在交叉熵损失 L_{CE}(cross-entropy loss)和 L_{FL}(focal loss)基础上进行优化得到的损失函数,两者的泰勒展开式如式(5)和(6)所示。

$$L_{CE} = -\log(P_t) = \sum_{j=1}^{\infty} \frac{1}{j(1-P_t)^j} = (1-P_t) + \frac{1}{2}(1-P_t)^2 + \cdots \tag{5}$$

$$L_{FL} = \&-(1-P_t)^{\gamma}\log(P_t) = \&\sum_{j=1}^{\infty} \frac{1}{j(1-P_t)^{j+\gamma}} = (1-P_t)^{1+\gamma} + \frac{1}{2(1-P_t)^{2+\gamma}} \tag{6}$$

式中:P_t 为模型对目标真值类的预测概率;Polyloss 作为新的损失函数的框架,有无穷多个多项式系数 α_j 需要调整。本研究采用 Poly-1 Loss,即只改动交叉熵损失的第 1 个多项式的系数:

$$L_{Poly-1} = (1+\epsilon_1)(1-P_t) + \frac{1}{2(1-P_t)^2} + \cdots = -\log(P_t) + \epsilon_1(1-P_t) \tag{7}$$

其中,通过 $(1+\epsilon_1)$ 来取代 1 作为第 1 个多项式的系数,$\epsilon_1 \in [1, \infty)$。$L_{rec}$ 可以表示为:

$$L_{rec} = \sum_{i}^{N} |[z_{2,i}^{(p)}, z_{1,i}^{(p)}] - u_i^{(n)}|_2^2 \tag{8}$$

式中:$u_i^{(n)}$ 为输出重构特征;L_{rec} 计算重构特征 $u_i^{(n)}$ 与跨通道特征 $[z_{2,i}^{(p)}, z_{1,i}^{(p)}]$ 之间经 L2 范数正则化的损失。

2　实验结果及分析

2.1　实验数据集描述

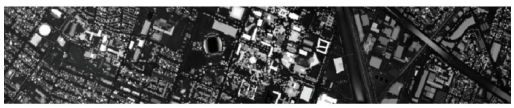
为了验证本研究提出的 HSI 和 LiDAR 数据协同分类模型的优势和有效性,采用 Houston2013、Trento 和 Augsburg 这 3 个公开的数据集进行实验。

1)Houston2013 数据集

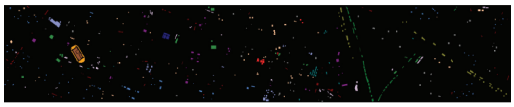
Houston2013 数据集由美国国家科学基金会资助的国家机载激光测绘中心于 2012 年收集,2013 年度的 IEEE GRSS 数据融合竞赛发布,覆盖休斯顿大学校园及周边城区。其中,HSI 数据包含 144 个光谱波段(0.38~1.05 μm)。对于同一区域,仅提供一个波段的 LiDAR 数据。HSI 和 LiDAR 数据均为 349 pixels×1 905 pixels,空间分辨率为 2.5 m,包含 15 个城市土地覆盖类别,HSI 的

伪彩色合成图像、LiDAR 数据的灰度图像和地面真值图如图 6 所示。

(a) HSI伪彩色合成图像
(a) Pseudo color composite image of HSI



(b) LiDAR的DSM灰度图像
(b) Grayscale DSM image of LiDAR



(c) 真值图
(c) Ground truth map

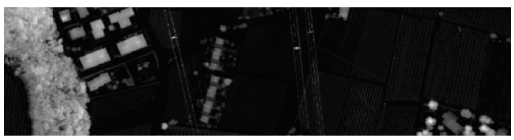
图 6 Houston2013 数据集伪彩色图、灰度图和真值图

Fig. 6 Pseudo-color, grayscale, and ground truth maps of the Houston2013 dataset

2) Trento 数据集

Trento 数据集是在意大利 Trento 南部的一个农村地区捕获的。HSI 数据由 AISA Eagle 系统的高光谱成像(AISA) Eagle 传感器获得,该传感器具有 63 个光谱波段,光谱分辨率为 0.42~0.99 μm;LiDAR 数据由 Optech ALTM 3100EA 传感器采集,具有 1 个光栅。该数据集的大小为 600 pixels×166 pixels,空间分辨率为 1 m,包含 6 种地物。HSI 的伪彩色合成图像、LiDAR 数据的灰度图像和地面真值图如图 7 所示。

(a) HSI伪彩色合成图像
(a) Pseudo color composite image of HSI



(b) LiDAR的DSM灰度图像
(b) Grayscale image of LiDAR DSM

(c) 真值图
(c) Ground truth map

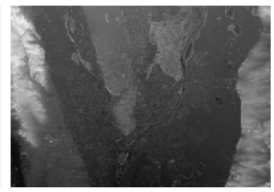
图 7 Trento 数据集伪彩色图、灰度图和真值图

Fig. 7 Pseudo-color, grayscale, and truth maps of the Trento dataset

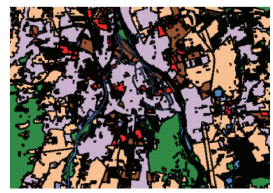
3) Augsburg 数据集

Augsburg 数据集是在德国奥格斯堡市上空捕获的。HSI 数据由 DAS-EOC HySpex 传感器获取,LiDAR 数据由 DLR-3K 系统收集,空间分辨率为 30 m。HSI 数据包含 180 个波段(0.4~2.5 μm),大小是 332 pixels×485 pixels,包含 7 种地物类别。HSI 的伪彩色合成图像、LiDAR 数据的灰度图像和地面真值图如图 8 所示。

(a) HSI伪彩色合成图像
(a) Pseudo color composite image of HSI



(b) LiDAR的DSM灰度图像
(b) Grayscale image of LiDAR DSM



(c) 真值图
(c) Ground truth map

图 8 Augsburg 数据集伪彩色图、灰度图和真值图

Fig. 8 Pseudo-color, grayscale, and truth maps of the Augsburg dataset

2.2 实验平台及超参数设置

实验基于配备 Intel(R) Xeon(R) 4208 CPU @ 2.10 GHz 处理器和 Nvidia GeForce RTX 2080Ti 显卡的高性能计算平台。采用 PyCharm 软件作为主要的开发环境,利用 PyTorch 深度学习框架进行模型的构建、训练和评估。为了降低实验误差并提高结果可靠性,从训练集中随机抽取有限样本用于模型训练,并将 epoch 设定为 100。为确保实验结果的稳定性和可信度,所有表格的实验结果均为 10 次独立实验的平均值。为了优化网络,选择 Adam 优化器作为初始优化器,mini-batch 大小为 64。同时,使用权重衰减率为 0.000 1 和动量为 0.9 的正则化策略来增强模型的泛化能力。

在评估方法性能方面,采用总体准确率(overall accuracy,OA)、平均准确率(average accuracy,AA)、Kappa (K)系数和每类精度作为评价指标。OA 是指模型在全部测试样本中正确预测的样本数量与总样本数之比。AA 则表示在每个类别中,正确预测的样本数量与该类别样本总数之比的平均值。Kappa 系数是一种用于评估分类准确性的统计指标,它衡量了遥感分类结果图与地面真实情况图之间的吻合程度。

2.3　实验对比及分析

1)不同分类方法的对比

为了验证所提出模型的有效性,将提出的网络与 8 种具有代表性的分类方法在 Houston2013、Trento 和 Augsburg 数据集上进行比较,包括 RSSAN[11]、SSFTT[20]、NeiCoT[21]、GAHT[22]、CALC[12]、DSHF[23]、GLT[24] 和 HCT[26],其中 RSSAN、SSFTT、NeiCoT 和 GAHT 用于 HSI 分类。不同方法在 Houston2013、Trento、Augsburg 数据集上的分类精度对比如表 1~3 所示。不同方法在 Houston 2013、Trento、Augsburg 数据集上的分类结果如图 9~11 所示。

表 1　不同方法在 Houston2013 数据集上的分类精度对比

Table 1　Comparison of classification accuracy of different methods on the Houston2013 dataset

类别号	HSI				HSI+ LiDAR				
	RSSAN	SSFTT	NeiCoT	GAHT	CALC	DSHF	GLT	HCT	MMCT
C01	98.14±1.85	99.42±0.71	99.70±0.33	99.84±0.28	98.66±1.10	97.78±0.91	99.26±0.70	99.43±0.53	99.77±0.36
C02	99.80±0.14	99.28±0.41	98.92±1.05	100.00	99.60±0.57	99.06±0.58	99.59±0.32	99.87±0.10	99.74±0.27
C03	98.34±1.32	99.31±0.87	99.31±0.69	98.99±0.91	99.78±0.35	99.96±0.08	99.92±0.10	99.96±0.08	99.92±0.10
C04	97.09±1.21	98.84±1.05	99.21±0.40	98.90±0.34	99.55±0.41	98.90±0.51	99.62±0.41	99.83±0.11	99.78±0.18
C05	98.58±1.07	96.43±2.48	99.96±0.08	99.88±0.26	99.93±0.09	99.79±0.18	100.00	99.82±0.04	100.00
C06	99.79±0.37	99.55±0.72	100.00	100.00	100.00	100.00	100.00	100.00	100.00
C07	99.53±0.22	99.59±0.55	98.32±1.18	99.91±0.11	99.18±0.92	98.43±1.33	99.08±0.81	98.66±1.14	99.82±0.31
C08	98.19±1.92	99.02±0.75	98.94±0.68	98.18±0.71	98.40±0.91	97.42±2.57	99.58±0.41	99.47±0.62	99.63±0.74
C09	98.67±1.95	99.51±0.34	98.90±0.95	99.97±0.04	98.38±0.85	97.41±1.98	98.07±1.44	97.24±1.68	99.98±0.05
C10	97.96±0.83	99.47±0.32	99.95±0.14	98.88±0.97	100.00	99.27±1.32	99.73±0.29	99.44±0.74	99.80±0.45
C11	99.37±0.89	99.62±0.29	99.85±0.19	99.06±0.95	99.94±0.10	98.35±1.37	99.79±0.17	99.72±0.29	99.90±0.19
C12	98.37±1.83	99.60±0.22	99.25±0.62	99.84±0.20	99.44±0.45	98.21±1.26	99.04±0.64	99.85±0.19	99.81±0.22
C13	99.74±0.35	99.78±0.42	99.07±0.88	98.70±0.84	98.91±2.01	98.74±0.91	99.93±0.14	99.70±0.61	99.95±0.10
C14	99.46±0.69	99.51±0.48	100.00	99.07±0.94	100.00	99.60±0.63	100.00	100.00	100.00
C15	99.03±0.71	99.65±0.24	100.00	99.82±0.09	99.98±0.06	100.00	100.00	100.00	100.00
OA/%	98.65±0.53	99.04±0.43	99.27±0.35	99.39±0.19	99.29±0.18	98.99±0.58	99.46±0.17	99.43±0.15	99.85±0.05
AA/%	99.06±0.39	99.12±0.32	99.35±0.24	99.44±0.15	99.45±0.20	99.21±0.45	99.57±0.13	99.55±0.13	99.87±0.06
K×100	98.30±0.58	98.63±0.40	99.12±0.39	99.22±0.33	99.23±0.20	98.91±0.63	99.41±0.19	99.38±0.16	99.83±0.06

表 2　不同方法在 Trento 数据集上的分类精度对比

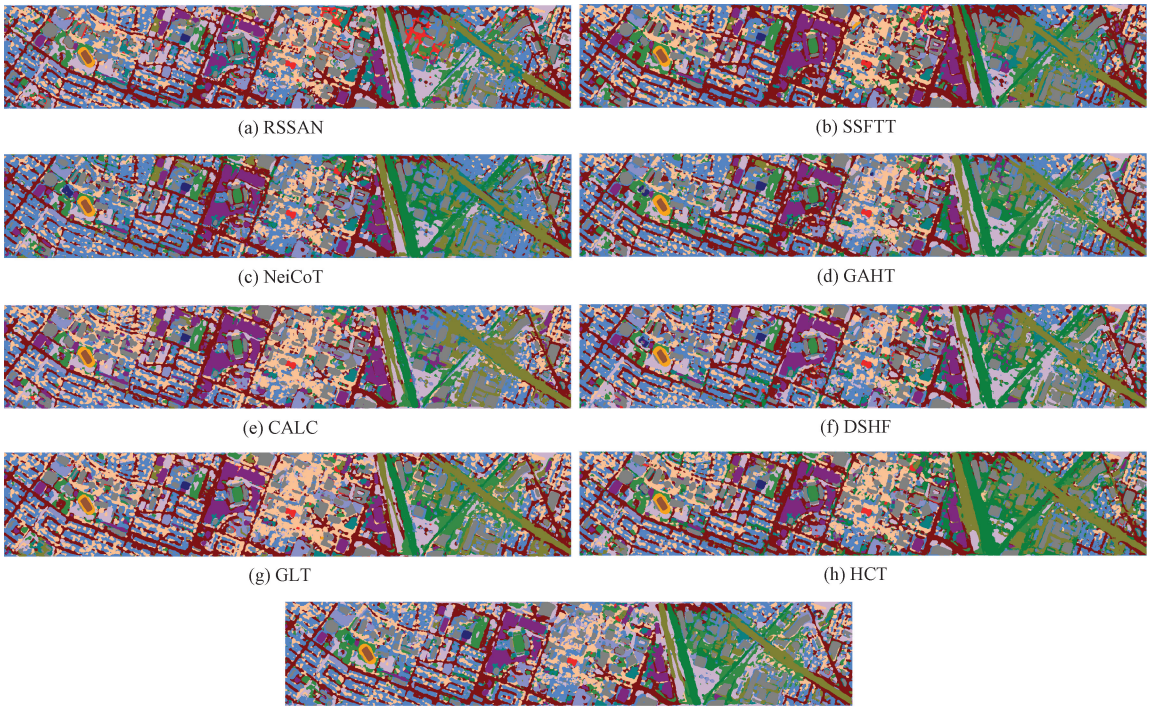
Table 2　Comparison of classification accuracy of different methods on the Trento dataset

类别号	HSI				HSI+ LiDAR				
	RSSAN	SSFTT	NeiCoT	GAHT	CALC	DSHF	GLT	HCT	MMCT
C01	95.12±1.55	96.55±1.66	96.24±1.50	97.25±1.16	97.14±2.56	99.39±0.40	98.79±2.46	99.15±0.49	99.97±0.03
C02	94.21±0.68	96.77±0.96	97.98±0.94	98.15±0.84	98.91±1.20	99.31±0.63	98.20±1.74	98.67±0.42	99.21±0.52
C03	98.82±0.32	99.22±0.12	99.11±0.15	99.45±0.22	97.17±1.55	99.23±1.09	98.96±0.86	100.00	98.17±0.90
C04	99.02±0.71	98.61±1.06	99.43±0.53	99.87±0.11	99.20±0.21	99.97±0.05	99.89±0.15	100.00	100.00
C05	99.93±0.07	99.98±0.02	99.99±0.01	99.94±0.04	98.90±0.99	99.29±0.56	99.93±0.10	99.96+0.08	100.00
C06	99.63±0.15	99.55±0.38	99.84±0.09	99.92±0.03	97.36±1.01	95.68±3.07	96.24±2.54	98.94±0.48	99.04±0.80
OA/%	98.17±0.47	98.91±0.14	99.00±0.10	99.26±0.19	98.54±0.76	99.14±0.30	99.21±0.42	99.55±0.10	99.68±0.03
AA/%	96.72±0.92	97.74±0.69	98.70±0.53	98.52±0.81	97.91±1.37	98.81±0.51	98.67±0.61	99.30±0.14	99.34±0.11
K×100	97.59±0.62	98.56±0.28	98.99±0.25	99.02±0.13	98.38±0.90	98.84±0.44	98.94±0.56	99.43±0.13	99.57±0.05

表 3 不同方法在 Augsburg 数据集上的分类精度对比

Table 3 Comparison of classification accuracy of different methods on the Augsburg dataset

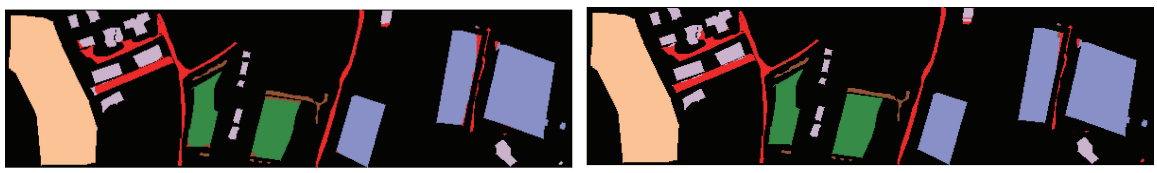
类别号	HSI				HSI+ LiDAR				
	RSSAN	SSFTT	NeiCoT	GAHT	CALC	DSHF	GLT	HCT	MMCT
C01	94.43±7.71	96.18±5.89	97.15±1.64	98.17±0.83	97.85±0.76	99.42±0.71	96.79±1.87	98.71±0.16	98.41±0.36
C02	93.02±4.67	93.38±2.61	98.16±0.58	99.04±0.49	93.73±2.41	98.31±1.87	98.71±0.99	98.65±0.33	99.13±0.28
C03	83.33±11.22	85.01±8.07	93.10±1.82	92.18±1.54	90.11±3.77	86.65±2.84	88.23±1.99	91.68±2.42	91.35±3.01
C04	96.03±5.61	96.46±2.16	96.90±1.34	97.06±1.12	98.25±1.03	96.43±2.48	97.78±1.60	98.83±0.33	97.43±1.04
C05	36.09±10.82	83.68±17.04	76.27±7.08	85.13±4.82	64.97±3.86	70.49±3.40	79.96±3.95	83.07±5.72	91.18±5.61
C06	53.89±18.83	50.56±20.24	65.97±9.79	75.55±2.23	46.60±2.25	52.38±3.59	66.95±3.54	70.49±5.27	82.78±4.33
C07	56.57±25.32	60.79±20.60	56.02±10.99	68.73±1.29	58.39±7.11	62.71±5.62	71.34±4.41	66.22±3.15	77.22±2.23
OA/%	91.02±2.59	93.11±2.22	95.85±0.49	96.31±0.55	94.30±1.43	94.68±1.23	95.44±0.69	97.04±0.14	97.34±0.17
AA/%	68.99±14.21	80.87±7.11	83.37±2.86	86.52±1.17	80.63±3.82	82.44±1.80	86.45±1.78	86.81±0.63	90.60±1.43
K×100	87.11±7.61	90.12±3.24	93.99±0.71	95.09±0.58	93.65±1.69	94.08±1.30	93.94±0.92	95.76±0.19	96.17±0.24



(a) RSSAN (b) SSFTT (c) NeiCoT (d) GAHT (e) CALC (f) DSHF (g) GLT (h) HCT (i) MMCT

图 9 不同方法对 Houston2013 数据集的分类结果

Fig. 9 Classification results of different methods on the Houston2013 dataset



(a) RSSAN (b) SSFTT

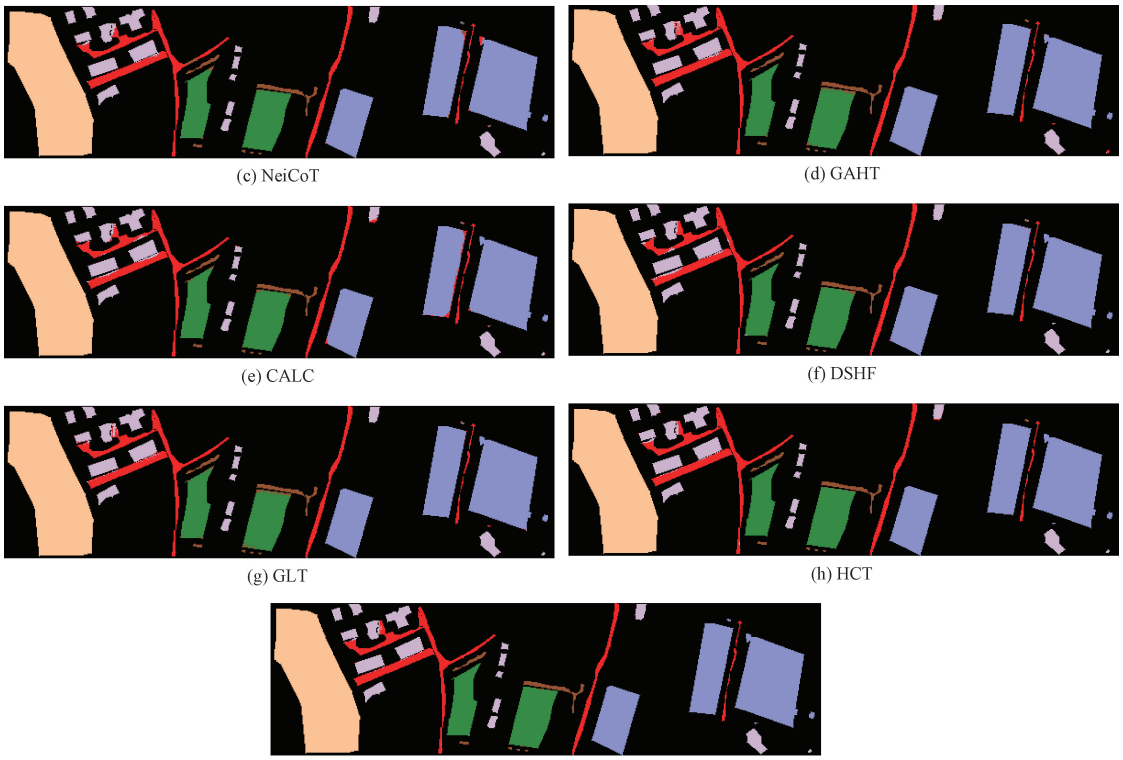


图 10　不同方法对 Trento 数据集的分类结果

Fig. 10　Classification results of different methods on the Trento dataset

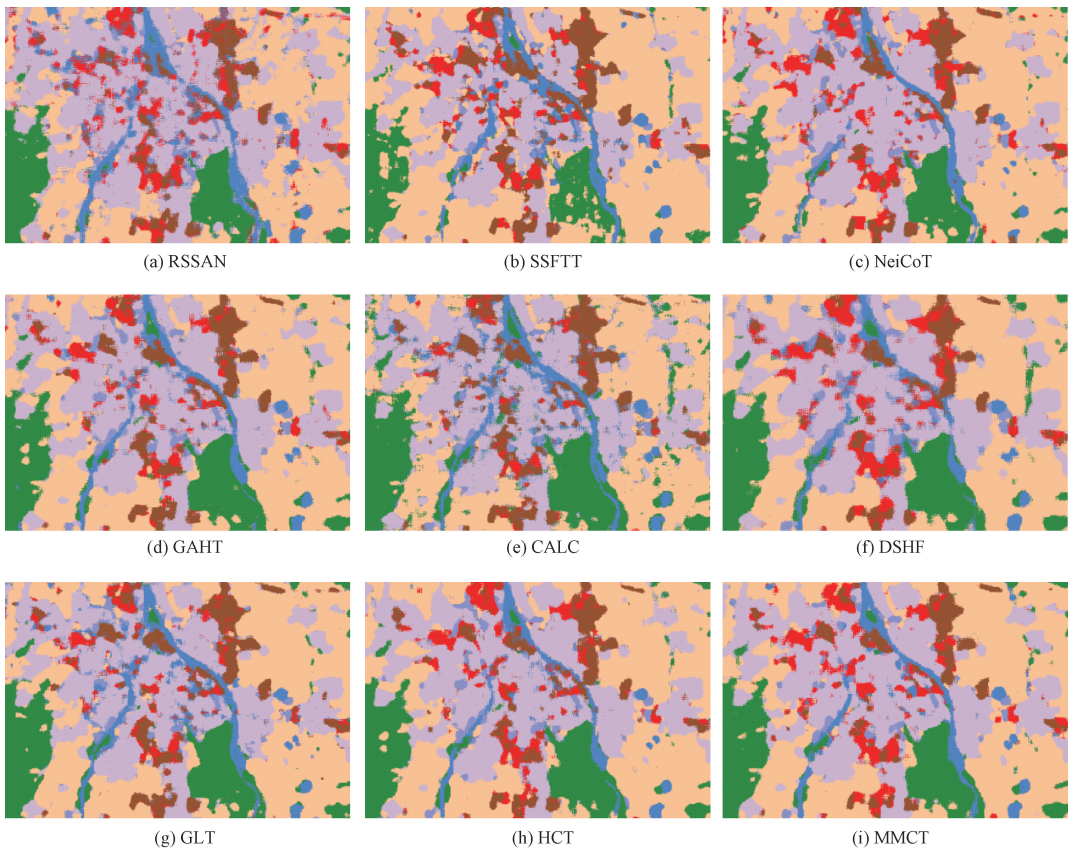


图 11　不同方法对 Augsburg 数据集的分类结果

Fig. 11　Classification results of different methods on the Augsburg dataset

由表 1 可知,双传感器协同分类模型的分类精度明显优于单传感器分类方法。与其他方法相比,本研究提出的方法在 3 个评价指标上都有明显的改善,尤其对 Residential、Commercial 和 Road 有显著提升。其中,Residential 的整体精度达到了 99.82%,Road 的整体精度为 99.98%。观察图 9 可知,分类精度的提升与椒盐噪声的减少呈正相关性,这表明本研究提出的方法在提升协同分类性能方面具有有效性。

如表 2 所示,在 Trento 数据集上,本文方法在 Apple Tree 和 Vineyard 的分类性能方面也有明显提升,分别达到了 99.97% 和 100%。图 10 展示了本研究方法在描绘 Building 和 Road 边缘方面的精确性,能够呈现出更为清晰和光滑的轮廓,表明本方法在细粒度特征表达和提取方面具有更强大的能力。

从表 3 可以看出,在 Augsburg 数据集上,本研究所提出的方法对 Allotment、Commercial Area 和 Water 类的分类结果相较于其他方法展现了显著的优越性,分类精度分别达到了 91.18%、82.78% 和 77.22%。通过对图 11 的观察,亦可证实本研究提出的方法在描绘 Allotment、Commercial Area 以及 Water 区域的边缘方面更为精确,呈现出了更为清晰的轮廓。

2)模型参数对分类性能的影响

为了确定最佳模型结构,本部分针对多个关键参数进行了系统实验,包括 PCA 降维后的的主成分数、输入数据 patch size 的大小、学习率以及特征融合策略中的编码器层数。通过全面评估不同参数组合下的实验结果,得到最优模型参数配置,为后续研究提供了有力的支持和参考。

(1) 主成分数

第 1 个参数是应用 PCA 时对高光谱图像选择的主成分数,旨在提取光谱的主要成分,以提升算法效率并降低噪声干扰。为确定主成分数的最佳选择,本研究采用控制变量法,patch size、学习率和深度特征融合策略的参数保持不变。主成分数对 OA、AA、Kappa 系数的影响如图 12 所示。在 3 个高光谱遥感影像数据集上,随着主成分数的增加,3 个评价指标都呈现出先上升后下降的趋势。综合考虑计算复杂度与分类性能,本研究将主成分数设置为 25。

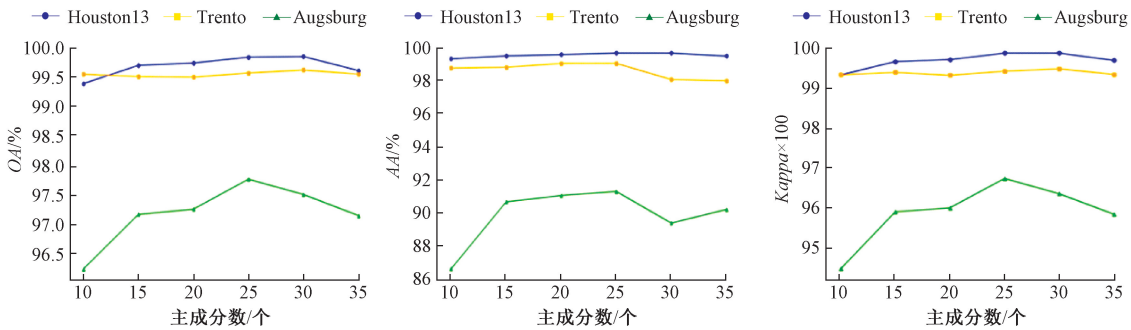


图 12 主成分数对 OA、AA 和 Kappa 系数的影响

Fig. 12 Influence of PCA's number on OA, AA and Kappa coefficient

(2) Patch size

与主成分数的分析类似,其他超参数保持不变。在候选集{7,9,11,13,15,17}中选择不同的 patch 大小进行效果评估。Patch size 对 OA、AA、Kappa 系数的影响如图 13 所示。结果表明,Houston2013 数据集的最佳 patch 大小为 13,而 Trento 数据集和 Augsburg 数据集的最佳 patch 大小均为 11。

(3)学习率

学习率是深度学习模型中一个重要的超参数,对目标函数收敛到局部最优值有重要影响。目标函数可以在适当的学习率下快速达到局部最小值。在实验中,学习

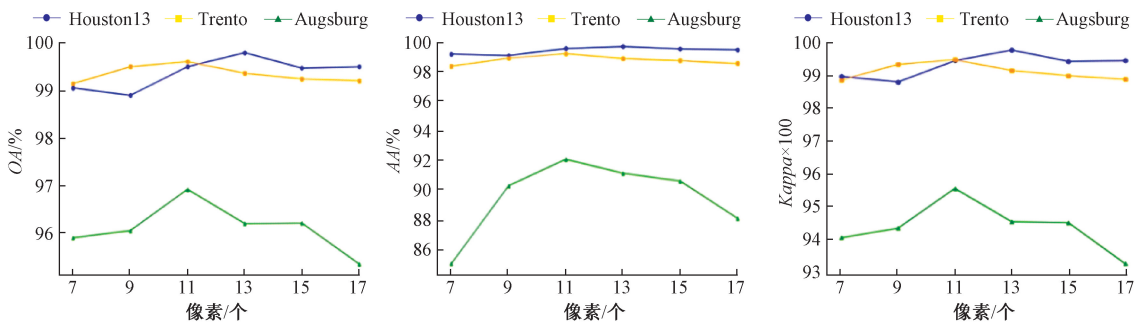


图 13 Patch size 对 OA、AA 和 Kappa 系数的影响

Fig. 13 Influence of patch size on OA, AA and Kappa coefficient

率从候选集{5×10^{-5},1×10^{-4},5×10^{-4},1×10^{-3},5×10^{-3}}中选取。学习率对 OA、AA、Kappa 系数的影响如图 14 所示。图 14 显示了通过设置不同的学习率,本方法在 3 个数据集上得到的评价指标。可以看出,对于 Houston2013 数据集的最佳学习率为 1×10^{-3},Trento 数据集和 Augsburg 数据集的最佳学习率为 5×10^{-3}。

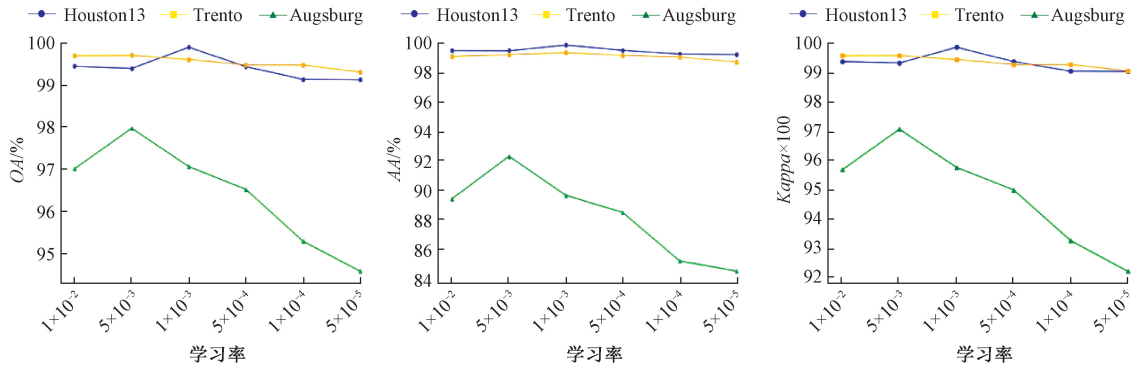


图 14　学习率对 OA、AA 和 Kappa 系数的影响

Fig. 14　Influence of learning rate on OA, AA and Kappa coefficient

(4)编码器层数

如图 6 所示,将跨通道重构机制的编码器和解码器视为一个统一的编码器整体,编码器的层数会影响特征融合的效果。为了验证编码器层数对分类性能的影响,将候选集设置为{1,2,3,4,5}。在 3 个数据集上对编码层数对 OA、AA、Kappa 系数的影响如图 15 所示。可以看出,随着编码器层数的增加,模型在 3 个数据集上的性能呈先上升后下降的趋势。当编码器层数为 3 时,所提出的网络可以达到最优的分类效果。这种现象说明更深的网络并不能带来更好的性能。

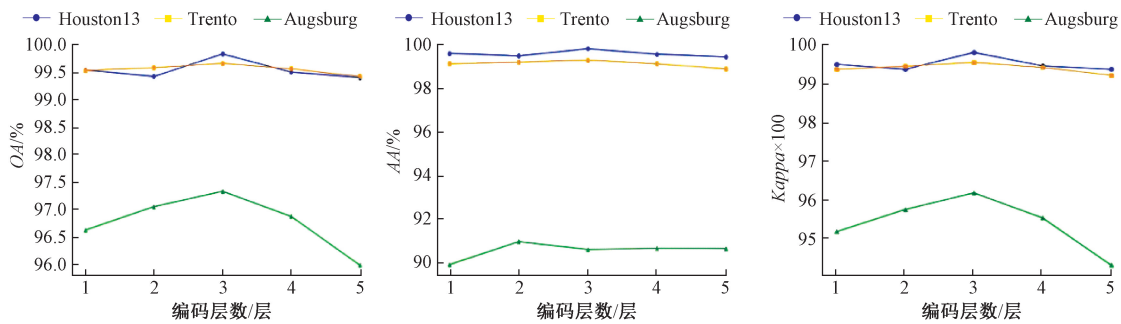


图 15　编码器层数对 OA、AA 和 Kappa 系数的影响

Fig. 15　Influence of the number of encoder layers on OA, AA, and Kappa coefficient

3)消融实验

(1) 各模态输入数据的消融分析

考虑到不同模态数据对模型分类性能的影响,使用 3 个数据集分别进行输入不同传感器数据进行分类,不同模态数据输入的消融分析如表 4 所示。通过对比 3 个数据集的 OA 值、AA 值和 Kappa 值,可以认为多模态数据融合方法相较于单模态方法会获得更好的分类效果,验证了多模态数据对提高分类性能有积极的作用,也验证了所设计的双分支网络可以充分利用不同模态数据的有效信息。

表 4　不同模态数据输入的消融分析

Table 4　Ablation analysis of different modal data inputs

数据集	Houston2013			Trento			Augsburg		
	OA/%	*AA*/%	*K* × 100	*OA*/%	*AA*/%	*K* × 100	*OA*/%	*AA*/%	*K* × 100
HSI	99.28±0.38	99.44±0.31	99.22±0.42	99.20±0.16	98.57±0.41	98.92±0.21	95.60±2.27	88.52±2.73	93.72±3.12
LiDAR	68.81±5.53	69.63±6.54	66.23±5.96	91.76±6.37	88.13±4.25	89.19±7.88	74.04±11.91	59.41±11.03	65.01±14.30
HSI+LiDAR	99.85±0.05	99.87±0.06	99.84±0.06	99.68±0.03	99.34±0.11	99.57±0.05	97.34±0.17	90.60±1.43	96.17±0.24

不同传感器数据类型下的 Houston2013、Trento、Augsburg 数据集 t-SNE 可视化如图 16~18 所示。在 3 个数据集上,仅采用 HSI 进行分类任务时,不同类别的数据点分布表现出显著的重叠。这说明单纯依赖 HSI 所提供的光谱空间信息进行分类有一定的局限性。当仅使用 LiDAR 数据进行分类时,数据点的分布较为分散,这表明仅依赖 LiDAR 数据所提供的高程信息进行分类的性能亦不尽人意,且其分类效果略逊于 HSI 图像。相比之下,利用 HSI 与 LiDAR 数据进行协同分类时,不同类别的数据点展现出更为显著的聚类效果,充分表明,本文提出的协同分类模型能够有效地整合两种数据源的互补信息,增强对不同地物类别的识别能力,进而实现对单一数据源分类性能的提升。

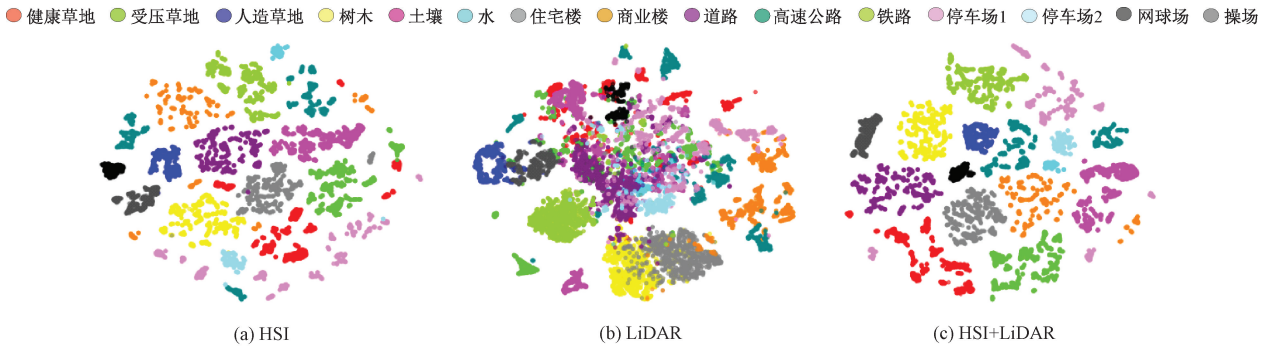


图 16 不同传感器数据类型下的 Houston2013 数据集 t-SNE 可视化

Fig. 16 t-SNE visualization of the Houston2013 dataset under different sensors

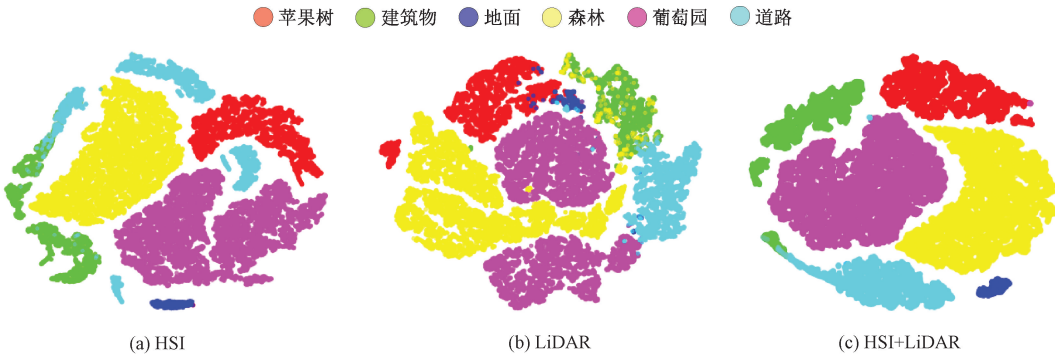


图 17 不同传感器数据类型下的 Trento 数据集 t-SNE 可视化

Fig. 17 t-SNE visualization of the Trento dataset under different sensors

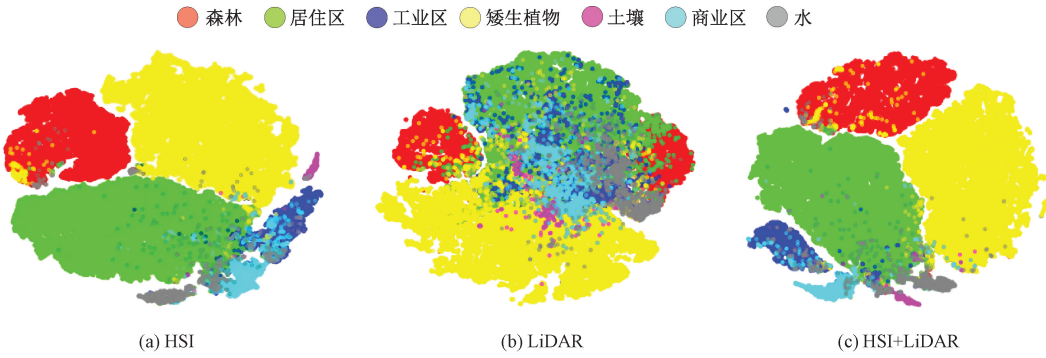


图 18 不同传感器数据类型下的 Augsburg 数据集 t-SNE 可视化

Fig. 18 t-SNE visualization of the Augsburg dataset under different sensors

(2) 特征提取与融合模块的消融分析

本研究通过系统的消融实验评估了网络各核心模块对协同分类性能的贡献。各个模块在 Houston2013 数据集上的消融如表 5 所示,通过逐一移除各核心模块构建了 3 种变体模型(变体 A、B 和 C),深入分析了各组件对分类性能的影响机制。

表 5　各个模块在 Houston2013 数据集上的消融实验

Table 5　Ablation experiments of various modules on the Houston2013 dataset

模块	A	B	C	*OA*/%	*AA*/%	*K* × 100
1	—	√	√	99.38±0.44	99.38±0.39	99.32±0.56
2	√	—	√	99.62±0.19	99.70±0.16	99.59±0.21
3	√	√	—	99.56±0.22	99.66±0.16	99.52±0.23
4	√	√	√	99.85±0.05	99.87±0.06	99.84±0.06

实验框架由 3 个核心组件构成:HSI 支路中 CNN 与 Transformer 协同的特征提取模块,LiDAR 支路的 ODConv 模块,以及实现跨模态特征融合的通道重组模块。实验数据显示,当移除 HSI 支路的特征提取模块(变体 A)时,模型整体分类准确率显著下降至 99.38%,成为性能表现最差的配置,这充分验证了复合特征提取架构对高光谱数据表征的重要性。在移除 LiDAR 支路的 ODConv 模块(变体 B)后,模型虽仍保持 99.62% 的总体精度,但与完整模型仍存在明显差距,表明动态卷积结构对空间特征提取具有不可替代的作用。

值得注意的是,即使将双模态特征直接拼接后进行分类(变体 C),模型仍能获得 99.56% 的总体精度,这说明各支路的特征提取模块已具备较强的表征能力。然而,该结果较完整模型仍存在 0.39 个百分点的性能差距,有力证实了通道重组模块在实现跨模态特征深度融合中的关键作用——其通过建立模态间的语义关联,有效提升了特征的判别性表达能力。

这些实验结论从 3 个方面验证了网络架构设计的有效性:首先,多尺度卷积与注意力机制的协同显著增强了高光谱特征的层次化表征;其次,动态卷积结构有效提升了 LiDAR 数据的空间特征提取能力;最后,基于通道重组的融合策略实现了跨模态特征的互补优化。研究结果表明,构建层次化的特征提取体系与设计自适应的特征融合机制,是提升多源遥感数据分类性能的关键路径,这为后续研究提供了重要的架构设计参考。通过系统的模块化分析,本研究不仅验证了各组件的作用机理,更为网络结构的优化改进提供了明确的方向。

3　结　　论

本研究提出一种融合卷积神经网络与 Transformer 的高光谱图像和 LiDAR 数据协同地物分类方法,该方法通过主成分分析对高光谱数据降维以去除冗余,利用 CNN 分层捕获局部纹理特征并结合 Transformer 自注意力机制构建全局光谱-空间表征,通过双向特征交互与特征耦合单元实现跨尺度特征对齐,同时采用动态卷积级联模块提取 LiDAR 高程信息,借助跨通道重建模块完成多模态特征的分层交互与融合;实验表明,该方法在 Houston2013、Trento 和 Augsburg 数据集上分别取得 99.85%、99.68% 和 97.34% 的总体精度,较 HCT 等主流模型分别提升 0.42%、0.13% 和 0.30%,有效验证了多模态特征融合策略对全局结构与局部细节的联合提取能力,未来将探索自监督学习框架以适应少标签场景下的地物分类需求。

参考文献

[1] 涂潮, 刘万军, 赵琳琳, 等. 有限训练样本下的多尺度空洞密集网络高光谱影像分类[J]. 仪器仪表学报, 2024, 45(4):206-216.
TU CH, LIU W J, ZHAO L L, et al. Multiscale dilated dense network for hyperspectral image classification with limited training samples[J]. Chinese Journal of Scientific Instrument, 2024, 45(4):206-216.

[2] 杨明莉, 范玉刚, 李宝芸. 基于 LDA 和 ELM 的高光谱图像降维与分类方法研究[J]. 电子测量与仪器学报, 2020, 34(5):190-196.
YANG M L, FAN Y G, LI B Y. Research on dimensionality reduction and classification of hyperspectral images based on LDA and ELM[J]. Journal of Electronic Measurement and Instrumentation, 2020, 34(5):190-196.

[3] 赵雪松, 付民, 刘雪峰. 基于深度特征提取残差网络的高光谱图像分类[J]. 电子测量技术, 2024, 47(18):120-129.
ZHAO X S, FU M, LIU X F. Hyperspectral image classification based on deep feature extraction residual network[J]. Electronic Measurement Technology, 2024, 47(18):120-129.

[4] ROY S K, DERIA A, HONG D F, et al. Multimodal fusion Transformer for remote sensing image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-20.

[5] ZHANG H T, YAO J, NI L, et al. Multimodal attention-aware convolutional neural networks for classification of hyperspectral and LiDAR data[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2022, 16: 3635-3644.

[6] ZHAO X D, TAO R, LI W, et al. Joint classification of hyperspectral and LiDAR data using hierarchical random walk and deep CNN architecture[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(10): 7355-7370.

[7] AHMAD M, KHAN A M, MAZZARA M, et al. A fast and compact 3-D CNN for hyperspectral image classification[J]. IEEE Geoscience and Remote Sensing Letters, 2020, 19: 1-5.

[8] ROY S K, KRISHNA G, DUBEY S R, et al. HybridSN: Exploring 3-D-2-D CNN feature hierarchy for hyperspectral image classification[J]. IEEE Geoscience and Remote Sensing Letters, 2019, 17(2): 277-281.

[9] YANG J X, ZHAO Y Q, CHAN J C W. Learning and transferring deep joint spectral-spatial features for hyperspectral classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(8): 4729-4742.

[10] CHEN Y SH, JIANG H L, LI CH Y, et al. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(10): 6232-6251.

[11] HE K M, ZHANG X Y, REN SH Q, et al. Deep residual learning for image recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016:770-778.

[12] ZHU M H, JIAO L CH, LIU F, et al. Residual spectral-spatial attention network for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59(1): 449-462.

[13] 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.

[14] LIN J Y, GAO F, QI L, et al. Dynamic cross-modal feature interaction network for hyperspectral and LiDAR data classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2025, 63:1-16.

[15] PAOLETTI M E, HAUT J M, FERNANDEZ-BELTRAN R, et al. Deep pyramidal residual networks for spectral-spatial hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 57(2): 740-754.

[16] XUE ZH X, TAN X, YU X CH, et al. Deep hierarchical vision transformer for hyperspectral and LiDAR data classification[J]. IEEE Transactions on Image Processing, 2022, 31: 3095-3110.

[17] 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, 2020, 19: 1-5.

[18] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: Transformers for image recognition at scale[J]. ArXiv preprint arXiv: 2010.11929, 2020.

[19] QING Y H, LIU W Y, FENG L Y, et al. Improved transformer net for hyperspectral image classification[J]. Remote Sensing, 2021, 13(11): 2216.

[20] HONG D F, HAN ZH, YAO J, et al. SpectralFormer: Rethinking hyperspectral image classification with transformers[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1-15.

[21] SUN L, ZHAO G R, ZHENG Y H, et al. Spectral-spatial feature tokenization transformer for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-14.

[22] LIANG M M, ZHANG X H, YU X CH, et al. An efficient Transformer with neighborhood contrastive tokenization for hyperspectral images classification[J]. International Journal of Applied Earth Observation and Geoinformation, 2024, 131: 103979.

[23] MEI SH H, SONG CH, MA M Y, et al. Hyperspectral image classification using group-aware hierarchical transformer[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-14.

[24] FENG Y N, SONG L Y, WANG L, et al. DSHFNet: Dynamic scale hierarchical fusion network based on multi-attention for hyperspectral image and LiDAR data classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61:1-14.

[25] DING K X, LU T, FU W, et al. Global-local transformer network for HSI and LiDAR data joint classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-13.

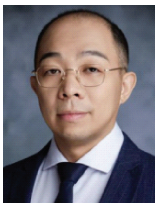
[26] XU H T, ZHENG T, LIU Y ZH, et al. A joint convolutional cross ViT network for hyperspectral and light detection and ranging fusion classification[J].

Remote Sensing, 2024, 16(3): 489.

[27] NI K, LI Z R, YUAN CH Y, et al. Selective spectral-spatial aggregation Transformer for hyperspectral and LiDAR classification[J]. IEEE Geoscience and Remote Sensing Letters, 2024, 22:1-5.

[28] ZHAO G R, YE Q L, SUN L, et al. Joint classification of hyperspectral and LiDAR data using a hierarchical CNN and Transformer [J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 61: 1-16.

作者简介



吴海滨,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 received his B. Sc. and M. Sc. degrees both 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 research interests include computer vision, virtual reality, and remote sensing image processing.

王爱丽(通信作者),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 (Corresponding author) received the 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 research interest includes remote sensing image processing.