

# 面向不确定性融合的多模态遥感数据分类

何 欣, 陈雨时, 谷延锋\*, 刘天竹

(哈尔滨工业大学电子与信息工程学院, 黑龙江哈尔滨 150001)

**摘 要:** 分类是多模态遥感数据解译的关键技术和热点问题. 近年来, 深度学习方法在多模态遥感数据像素级分类中取得了长足的进展. 然而, 多模态遥感数据包含的不同模态经过特征提取后, 预测结果存在不一致的问题, 即预测结果不确定性, 影响多模态遥感数据分类方法的分类精度. 为降低预测结果不确定性, 本文提出了一种面向不确定性融合的多模态遥感数据分类框架. 该框架联合提取不同模态(如合成孔径雷达数据、激光雷达数据、高光谱数据等)的空间特征与通道特征, 并构建对应的神经网络, 通过设计特定的证据融合函数, 实现基于证据信息的有效融合. 在融合过程中, 当预测结果存在冲突时, 设计基于冲突感知的动态权重调整机制, 通过折扣因子自适应地降低冲突模态的权重, 动态重分配证据质量, 从而有效降低多模态遥感数据分类方法的不确定性. 在此基础上, 为进一步预测结果的差异, 本文在模型参数优化过程中引入一致性损失函数, 以约束各模态预测结果的一致性. 实验在 3 种国际公开的多模态遥感数据集上进行, 并与 6 种主流方法进行对比, 结果表明本文所提方法在分类性能上均取得了显著提升.

**关键词:** 多模态遥感数据; 证据深度学习; 不确定性; 像素级遥感图像分类

**基金项目:** 国家自然科学基金(No.62025107, No.62371169); 黑龙江省自然科学基金(No.LH2024F015)

**中图分类号:** TP751 **文献标识码:** A **文章编号:** 0372-2112(2025)12-4267-12

**电子学报 URL:** <http://www.ejournal.org.cn> **DOI:** 10.12263/DZXB.20250882

## Uncertainty Fusion-Based Multimodal Remote Sensing Data Classification

HE Xin, CHEN Yu-shi, GU Yan-feng\*, LIU Tian-zhu

(School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin, Heilongjiang 150001, China)

**Abstract:** Classification is a key technique and a hot topic in the interpretation of multimodal remote sensing data. In recent years, deep learning methods have achieved significant progress in pixel-level classification of multimodal remote sensing data. However, the different modalities contained in multimodal remote sensing data exhibit variability in their predicted results after feature extraction, which is referred to as predictive uncertainty. This uncertainty negatively impacts the classification accuracy of multimodal remote sensing data classification methods. To reduce the prediction uncertainty, this paper proposes an uncertainty-aware fusion framework for multimodal remote sensing data classification. From the perspective of evidence quality, the framework jointly extracts spatial and channel features from different modalities (e.g., synthetic aperture radar data, light detection and ranging data, hyperspectral image, etc.) and constructs corresponding evidential neural networks. A specifically designed evidential fusion function is employed to effectively integrate multimodal evidential information. During the fusion process, when conflicting predictions arise from different modalities, a conflict-aware dynamic weight adjustment mechanism is introduced. This mechanism adaptively reduces the weight of conflicting modalities using a discount factor and dynamically reallocates the quality of evidence, thereby effectively reducing model uncertainty. Furthermore, to further minimize the discrepancy among predictions from different modalities, a consistency loss function is incorporated into the model parameter optimization process to constrain the consistency of predictions across modalities. Experiments are conducted on three publicly available international multimodal remote sensing datasets, and the proposed method is compared with six state-of-the-art approaches. The results demonstrate that the proposed method achieves significant improvements in classification performance.

**Key words:** multimodal remote sensing data; evidential deep learning; uncertainty; pixel-level remote sensing image classification

**Foundation Item(s):** National Natural Science Foundation of China (No.62025107, No.62371169); Natural Science Foundation of Heilongjiang Province of China (No.LH2024F015)

## 1 引言

随着对地观测技术的迅速发展,不同成像方式、不同波段和分辨率的数据并存,形成了同一场景和目标在不同遥感成像手段下的多模态遥感数据<sup>[1]</sup>.与单模态遥感数据相比,多模态遥感数据为地物提供不同的描述信息.因此,多模态遥感数据的综合利用为城市规划、灾害监测、军事侦察等领域提供更准确、全面的地物表达信息<sup>[2]</sup>.

为充分挖掘多模态遥感数据,需要有效的多模态遥感信息智能解译方法.其中,分类是多模态遥感信息智能解译的关键技术.遥感领域的多模态数据分类是对每幅场景中每个像素赋予一个标签,是一个像素级分类任务<sup>[3,4]</sup>.

近年来,基于深度学习的多模态遥感数据分类方法已成为主流<sup>[5]</sup>.作为深度学习领域最具代表性的模型之一,卷积神经网络(Convolutional Neural Networks, CNN)凭借局部感受野和权重共享机制,在遥感数据的特征提取中展现出卓越性能<sup>[6]</sup>.以高光谱数据-激光雷达数据分类为例,一种方式是建立双通道网络分别提取高光谱数据的光谱和空间信息,在利用级联网络提取激光雷达数据的信息<sup>[7]</sup>.另一种方式是与形态学属性相结合并使用复合核或拼接等方式融合不同特征<sup>[8]</sup>.基于CNN及其衍生模型已成为多模态遥感数据分类方法的主流,但受限于卷积核的感受野,难以提取全局特征<sup>[9,10]</sup>.以上简单的拼接和加和的方式融合不同模态的过程中,只是对特征进行简单融合,缺少考虑推理阶段不同模态预测结果间的差异.

Transformer通过自注意力机制建立输入数据间的交互机制,可以对全局关系建模,扩大图像的感受野.因此,众多学者的研究从CNN模型转向Transformer<sup>[11,12]</sup>.Transformer中的交叉注意力机制的出现使得模型能够自动学习不同模态特征的重要性和相关性,通过该机制融合后提升多模态遥感数据分类方法的效果<sup>[13]</sup>.但是,交叉注意力机制会将高的注意力值判别为不确定低的特征.在云雾污染的高光谱或相干斑严重的区域,注意力权重反而可能被噪声激活到最大,导致错误预测被高权重放大,影响分类精度.为减少模型的复杂度,基于Mamba的方法也逐渐应用于多模态遥感数据分类中,通过多尺度特征融合提升模型的计算效率<sup>[14-16]</sup>.

上述多模态遥感数据融合方法提升了多模态遥感

数据的分类精度,但是,大多数多模态遥感数据分类方法在融合过程中均未考虑不确定性,包含数据不确定性、模型不确定性和预测结果不确定性<sup>[17-20]</sup>.多模态遥感数据本身固有的不确定性,即数据不确定性,是导致分类精度受限的重要因素之一.不同模态的数据在成像过程中受到大气条件、传感器噪声以及数据后处理过程等因素的干扰程度存在显著差异.例如,云层覆盖会导致云下区域的信息缺失,传感器故障可能引发条带状数据空缺,而成像过程中的噪声也会干扰模态数据的质量<sup>[21]</sup>.这些问题引发的数据不确定性会导致分类模型输入的特征信息不准确,进而降低分类精度.为解决不确定性问题,Han等人<sup>[22]</sup>首次将动态证据融合引入多视图分类,同时提出伪视图增强的可信多视图分类框架,在保持可信性的同时提升了视图间互补信息交互,突破了传统后期融合仅依赖决策层交互的局限,为多视图学习提供了兼顾可信与性能的新基准.为解决早期-中期混合融合中的可信问题,他们在样本层面同时建模特征级与模态级信息度的动态变化,引入稀疏门控和真实类概率网络,可直接剔除高维噪声特征,在高维异构医学数据中取得了优异效果<sup>[23]</sup>.相比于聚焦于单模态分类中的置信度校准,强调对训练样本的区分处理,最近Han等人<sup>[24]</sup>关注多模态数据融合中的信息选择,提出了一种新的动态正则化方法,在训练过程中动态区别分布内外数据,对分布外数据施加正则化以降低其预测置信度.

在多模态遥感数据分类领域,数据处理方法本身也存在不确定性.以基于深度学习的多模态遥感数据分类方法为例,深度学习模型的参数随机性是导致模型不确定性的一个重要因素.在相同的训练数据集上,由于模型初始化条件的不同、优化算法的选择以及训练过程中随机扰动的影响,模型参数可能会出现差异.这种差异使得模型对同一模态的认知存在偏差,从而引发了模型不确定性<sup>[25]</sup>.此外,不同模态数据的模型不确定性程度不同,在融合过程中,这种差异可能会进一步放大模型不确定性,进而影响多模态遥感数据的分类精度效果.

在多模态遥感数据分类中,数据不确定性和模型不确定性最终体现在预测结果上,即预测结果不确定性<sup>[26]</sup>,从而影响分类精度.例如,在云雾天气下,高光谱图像(HyperSpectral Image, HSI)易受到云层遮挡和大气散射的影响,导致部分区域光谱信息缺失或失真,从而增加分类的不确定性.合成孔径雷达(Synthetic Aper-

ture Radar, SAR)虽然具备穿透云雾的能力,但其固有的相干斑噪声和低信噪比特性,也会导致边缘模糊或地物结构信息不稳定. 这些模态间的物理差异使得它们在像素级分类任务中产生不一致甚至相互矛盾的预测结果,即预测结果不确定性. 不同模态对同一像素的预测概率分布存在显著差异,若直接融合这些结果,可能引入错误降低整体分类精度. 因此,本文致力于解决预测结果不确定性问题. 以图 1 中的 Houston 数据集为

例,在 HSI 模态的右侧存在云遮挡区域. 现有的深度学习学习方法通常将不同模态的数据分别输入到深度神经网络(如 CNN、Transformer 架构或 Mamba)中进行训练. 在这一过程中,不同模态的数据可以采用相同或不同的深度神经网络架构. 然而,受云遮挡区域的高光谱模态经常预测错误,与 LiDAR 模态的预测结果不符合. 因此,如何有效识别并缓解这些冲突性预测,是多模态遥感数据融合中的关键挑战.

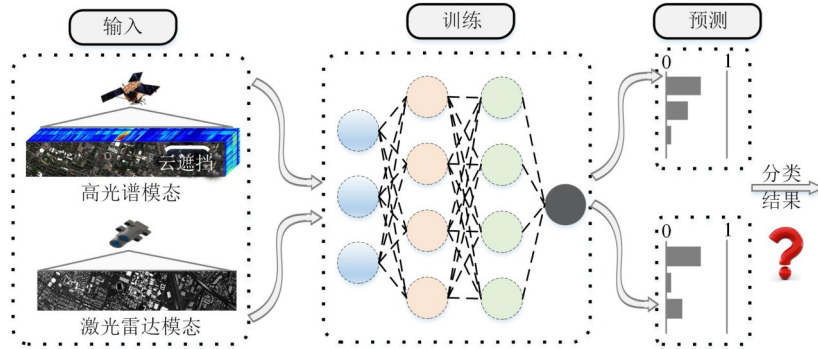


图 1 多模态遥感数据分类方法中预测结果不确定性示意图

为解决以上问题,本文提出面向不确定性融合的多模态遥感数据分类(Uncertainty Fusion-based Multimodal Remote Sensing Data, US-MRSD)方法. 首先,输入不同模态分别提取空间和通道特征. 然后,构建证据深度网络估计空间特征和通道特征输出的预测结果不确定性,对高不确定性的空间或通道特征赋予高折扣因子进行动态权重分配. 最后,建立基于不确定估计的损失函数获取预测结果的证据值输出分类结果.

本研究的主要贡献总结如下:(1)提出一种基于证据深度学习的的多不确定性融合框架,首次将证据理论引入多模态遥感数据分类任务,并量化预测结果的不确定性;(2)设计冲突感知的动态权重调整机制,通过构建冲突矩阵与一致性矩阵,自适应分配证据质量,有效缓解高冲突模态对融合结果的影响,从而提高多模态遥感数据的分类精度;(3)引入一致性损失函数,在参数优化阶段约束各模态预测结果的偏差,进一步提升分类结果的一致性.

## 2 不确定性引导的融合框架

### 2.1 空间-通道特征提取

本文提出的面向不确定性估计的多模态遥感数据分类方法的整体框架如图 2 所示. 首先,在特征提取的过程中,将输入的多模态遥感数据分别提取空间特征和通道特征. 假设输入的多模态遥感数据集为  $(\mathbf{x}_n^v, \mathbf{y}_n^v)_{n=1}^N, v \in [1, 2, \dots, V]$  表示第  $v$  个模态;  $\mathbf{x}_n^v \in X$  表示第  $v$  种遥感模态的第  $n$  个样本;  $\mathbf{y}_n^v \in [0, 1]^K$  为对应的 one-hot 编码;  $N$  是总样本数;  $K$  是多模态遥感数据集类别的

数量.

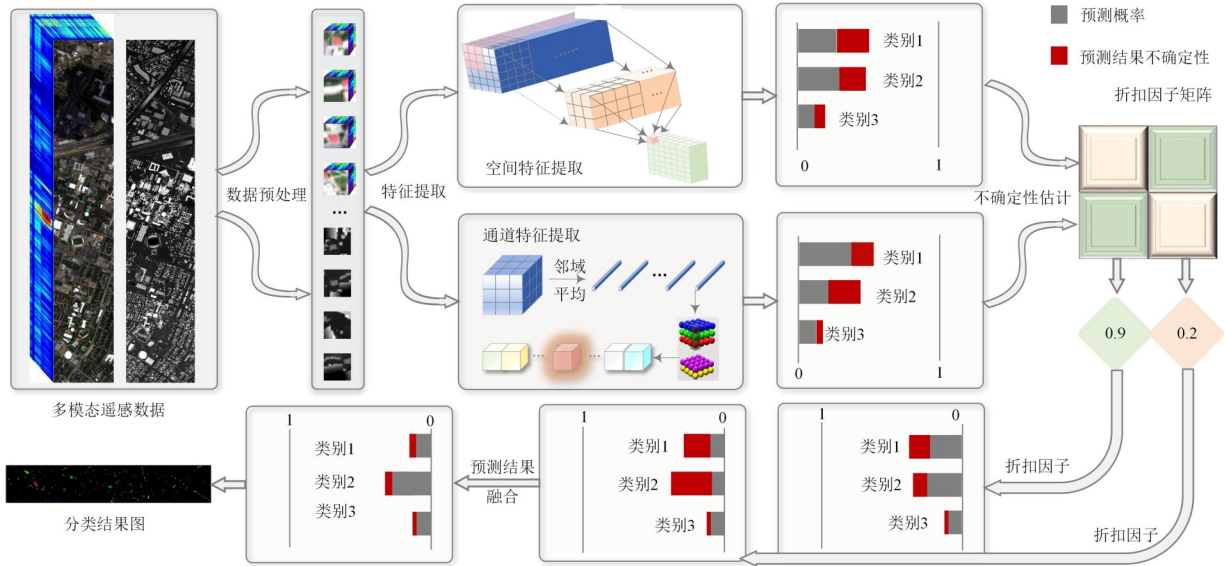
以 Berlin 数据集为例,包含高光谱模态  $\mathbf{x}_n^1$  和合成孔径雷达模态  $\mathbf{x}_n^2$ , 首先,将二者输入包含  $1 \times 1$  的卷积神经网络、批归一化操作和 ReLU 激活函数的网络  $\text{PL}(\cdot)$  分别进行预处理,输出高光谱模态特征  $\mathbf{F}_n^1 = \text{PL}(\mathbf{x}_n^1)$  和合成孔径雷达特征  $\mathbf{F}_n^2 = \text{PL}(\mathbf{x}_n^2)$ . 由于不同遥感模态同时包含空间特征和通道特征,为便于后续协同提取多模态遥感数据集的空间和通道特征,将输出不同模态的特征  $\mathbf{F}_n^1$  和  $\mathbf{F}_n^2$  组合输出  $\mathbf{F} = \text{Concat}(\mathbf{F}_n^1, \mathbf{F}_n^2)$ . 接下来,将多模态遥感数据提取特征  $\mathbf{F}$  分别输入跨模态空间特征提取模块和跨模态通道特征提取模块.

对于跨模态空间特征提取模块,不同地物的分布面积占比不同. 比如对于 Berlin 数据集,地物住宅区这一种类的占比面积较大,而耕地面积较小. 为兼容不同区域的特征提取,在跨空间特征提取模块设计的过程中选用  $3 \times 3$  和  $9 \times 9$  的卷积层为主要特征提取模块,其输出特征为  $\mathbf{S}$ .

对于跨模态通道特征提取模块,对于每一个样本输入的多模态特征  $\mathbf{F}_i$ ,通过聚合领域内所有像素值计算其均值的通道特征  $\mathbf{C}_i = \frac{1}{l^2} \sum_{j=1}^l \mathbf{F}_i^j$ . 其中,  $j$  表示邻域  $l$  内的第  $j$  个像素点. 接下来,将  $\mathbf{C}_i$  输出多层感知器中进行通道特征提取,并将领域内中心像素作为该样本的通道特征  $\mathbf{C}'_i$ , 则输出的通道特征为  $\mathbf{C} = \mathbf{C}'_1; \mathbf{C}'_2; \dots; \mathbf{C}'_N$ .

### 2.2 不确定性融合机制

在多模态遥感数据分类领域,特征融合是影响分



注:首先,对输入的多模态数据分别进行预处理;其次,将预处理结果送入跨模态空间-通道特征提取模块,同步获取各模态的空间特征与通道特征;最后,引入不确定性融合机制,对各模态特征进行不确定性估计并计算折扣因子,并据此完成基于预测结果不确定性的融合机制,输出最终分类结果。

图2 基于不确定性融合的多模态遥感数据分类方法整体框图

类效果的关键因素。然而,现有融合方法大多集中于特征拼接或模态间复交互,对模态融合过程中的预测结果不确定性研究不足。针对该问题,本文提出了一种基于证据理论与主观逻辑相结合的不确定性融合机制,通过量化分析不同模态预测结果的冲突关系,为不同遥感模态特征的融合提供依据。

一种简单的衡量融合特征的不确定性方法是将各种特征输入 softmax 后取调和平均值作为不确定性。然而,当不同模态的预测结果存在差异时,若其输出不确定性的值仍较小,此时融合后的不确定性在错误情况下仍较低,不符合真实情况。为解决该问题,本文从证据理论的角度对多模态数据融合中的不确定性建模。

首先,从证据深度学习角度对获取的空间特征和通道特征进行不确定性的估计。证据深度学习通过引入主观逻辑建模不确定性。在单次前向传播中,神经网络输出满足狄利克雷分布的预测结果,同时提供类别不确定性估计,此估计的逻辑值作为主观意见的证据,从而降低预测结果不确定性。

主观逻辑通过主观意见表示信任程度,用三元组  $\omega = (\mathbf{b}, \mu, \mathbf{a})$  表示。假设多模态遥感数据集共  $k$  类,  $\mathbf{b} = (\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k)^T$  表示每类的证据质量,  $\mu$  表示不确定性即证据不足的概率。狄利克雷参数分布  $\mathbf{a}$  提供概率基础,  $\mathbf{a} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k)^T$  表示类别先验概率即无证据时的概率。通过主观逻辑与狄利克雷分布的映射关系建模不确定性,主观意见的期望概率  $P_k$  满足  $P_k = \mathbf{b}_k + \mathbf{a}_k \mu, \forall k \in [1, 2, \dots, K]$ 。

其次,将特征提取阶段输出的空间特征  $\mathbf{S}$  和通道特征  $\mathbf{C}$  输入证据神经网络  $f(\mathbf{S})$  和  $f(\mathbf{C})$ , 输出每种特征的主观意见  $\omega^S = f(\mathbf{S})$  和  $\omega^C = f(\mathbf{C})$ 。

最后,将这些意见融合成一个最终意见:  $\omega^2 =$

$$\omega^S \circ \omega^C = (\mathbf{b}^{\circ 2}, \mu^{\circ 2}, \mathbf{a}^{\circ 2}), \text{ 其中, 证据融合算子为 } \mathbf{e}^{\circ 2} = \sum_{v=1}^2 \mathbf{e}^v.$$

$\mathbf{e} = (\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_K)^T = \mathbf{a} - 1$  表示支持每个类别的累积证据。然而,证据理论在处理证据时会遇到证据冲突的情况,即在证据合成过程中,由于证据间存在矛盾导致合成结果出现不合理的情况。考虑到多模态遥感数据中各模态固有的异质性特征,本文构建基于冲突感知的动态权重调整机制。当空间和通道特征的预测结果存在较大差异时,二者证据冲突。证据冲突是指不同信度均确信不同的类别。为量化各模态证据冲突程度,构建冲突矩阵  $\mathbf{M}$  及一致性矩阵  $\mathbf{A}$ , 并计算第  $v$  个模态的折扣因子  $\eta^v$ 。以两个模态为例,冲突矩阵  $\mathbf{M}_{12}$  表示为

$$\mathbf{M}_{12} = \frac{\sum_{k=1}^K |p_k^1 - p_k^2|}{2} \cdot (1 - \mu^1)(1 - \mu^2) \quad (1)$$

$$\eta^v = \prod_{i=1}^2 \mathbf{A}_{vi}, \forall v \in [1, 2] \quad (2)$$

其中,  $p_k^1$  和  $p_k^2$  分别表示在模态 1 和模态 2 下,通过狄利克雷分布的期望计算输出类别  $k$  的概率;  $\mathbf{A}_{vi}$  为一一致性矩阵中第  $v$  个模态对应的第  $i$  个元素;  $\mu^1$  和  $\mu^2$  分别表示模态 1 的不确定性和模态 2 的不确定性。

由于在计算过程中折扣因子乘积运算的累积效应,可能会导致折扣因子过小,也可能导致融合后的不确定性显著增加.为解决该问题,在一致性矩阵  $A$  中,增加可调节的超参数  $\lambda$ ,控制折扣因子相对于冲突的强度,即

$$A = (1 - M_{12}^\lambda)^{\frac{1}{\lambda}} \quad (3)$$

本文的多模态遥感数据的不确定性融合机制从证据深度学习的角度,设计折扣因子动态分配证据质量,不确定性高的特征折扣因子更大,从而降低其在融合过程中的权重.

### 2.3 损失函数的优化

在多模态遥感数据分类过程中,空间提取网络和通道提取网络转换为证据网络需满足网络输出为非负的证据值.为此,将输出层中的 softmax 激活函数替换为非负且单调递增的激活函数,如 ReLU、softplus 和指数函数.由于 softplus 函数因上升速度较慢,会导致证据值较小,空间和通道特征不确定性较高.US-MRSD 采用指数激活函数获取证据值.网络整体损失函数  $L$  由交叉熵损失函数  $L_{acc}$ 、KL(Kullback-Leibler)散度损失函数  $L_{KL}$  和一致性损失函数  $L_{con}$  组成.

交叉熵损失函数  $L_{acc}$  用于衡量网络预测输出与真实标签间的差异,即

$$L_{acc}(\alpha_n) = \int \left[ \sum_{j=1}^K -y_{nj} \log p_{nj} \right] \frac{\prod_{j=1}^K p_{nj}^{\alpha_{nj}-1}}{B(\alpha_n)} dp_n \quad (4)$$

其中,  $\alpha_n$  是第  $n$  个样本的狄利克雷参数;  $p$  是狄利克雷参数分布中概率单纯形中的点表示每个样本的概率.

为使预测错误标签的证据更低,US-MRSD 通过最小化修改后的狄利克雷分布参数和均匀狄利克雷分布间的 KL 散度.损失函数  $L_{KL}$  旨在促使网络探索参数,避免错误分类样本过早收敛到均匀分布,其数学表达式为

$$L_{KL}(\alpha_n) = KL[D(\mathbf{p}_n | \tilde{\alpha}_n) || D(\mathbf{p}_n | \mathbf{1})] \quad (5)$$

其中,  $\tilde{\alpha}_n = y_n + (1 - y_n) \odot \alpha_n$  是移除非误导性证据后的狄利克雷参数.

第 3 个一致性损失函数  $L_{con}$  是减小不同模态间在融合过程中的冲突程度,通过优化该损失函数,可增强多模态遥感数据在分类任务中的协同性和一致性,表达式为

$$L_{con} = \frac{1}{V-1} \sum_{v=1}^V \left( \sum_{i \neq v}^V M_{ij}^n \right) \quad (6)$$

最终,基于不确定融合的多模态遥感数据分类方法的总体损失函数为

$$L = L_\alpha(\alpha_n) + \beta \sum_{v=1}^V L_\alpha(\alpha_n^v) + \gamma L_{con} \quad (7)$$

其中,交叉熵参数  $\alpha_n$  表示第  $n$  个样本的狄利克雷参数,

$$\alpha_n > 0. L_\alpha(\alpha_n) = L_{acc}(\alpha_n) + \theta L_{KL}(\alpha_n).$$

通过合理设计以上损失函数及其组合方式,可减少因遥感模态异质性引起的预测结果不确定性,有效提升多模态遥感数据分类的准确性.

## 3 实验结果与分析

### 3.1 数据集介绍

本文使用国际上标准的多模态遥感数据集,包括 Houston、Berlin 和 MUUFL,验证所提方法的有效性. Houston 数据集于 2012 年通过 ITRES CASI-1500 传感器在美国得克萨斯州休斯敦大学校园及其周边城市区域采集,包含高光谱数据和 Light Detection and Ranging (LiDAR) 数据,空间尺寸为  $349 \times 1\,905$ ,分辨率为  $2.5 \text{ m/像素}$ .其中,高光谱数据集包含 144 个光谱波段,波长范围为  $0.38 \sim 1.05 \mu\text{m}$ ,光谱分辨率为  $10 \text{ nm}$ ,LiDAR 数据由数字表面模型经栅格处理后获取. Houston 数据集包含 15 种不同的地物类别,共计 15 029 个样本,如图 3 和表 1 所示.

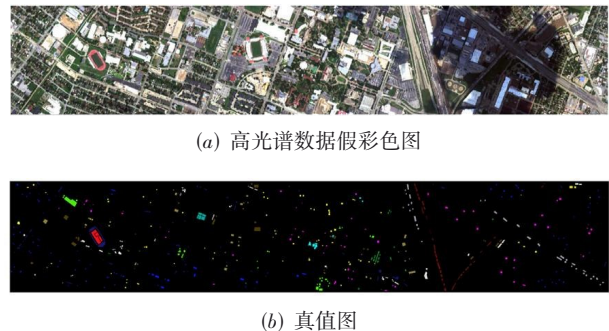





图 3 Houston 数据集伪彩色图及其真值图

Berlin 数据集包含高光谱数据和合成孔径数据.高光谱数据的空间分辨率是  $30 \text{ m}$ ,包含  $797 \times 220$  像素,包含 244 个光谱波段,波长范围为  $0.4 \sim 2.5 \mu\text{m}$ . SAR 数据是从欧洲航天局获取的 Sentinel-1 双极化单视复数产品,其空间尺寸为  $1\,723 \times 476$  像素,空间分辨率为  $13 \text{ m}$ .为匹配 SAR 图像的尺寸,对高光谱图像采用了最近邻插值技术,使其与 SAR 图像具有相同的空间分辨率. Berlin 数据集共包含 8 种不同的地物类别,共计 464 671 个样本,如图 4 和表 2 所示.

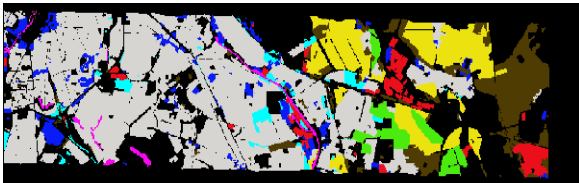
MUUFL 数据集于 2010 年在美国密西西比州长滩市的密西西比大学海湾公园校区获取.该数据集包含高光谱数据和激光雷达数据.高光谱数据的空间尺寸为  $325 \times 220$  像素,包含 64 个可用的光谱波段,波长范围为  $375 \sim 1\,050 \text{ nm}$ ,光谱间隔为  $10 \text{ nm}$ .激光雷达数据提供了两个波段的高程信息,用于描述地形起伏. MUUFL 数据集共包含 11 种不同的地物类别,共计 53 687 个样本,如图 5 和表 3 所示.

表1 Houston数据集类别信息

序号	颜色	类别名称	样本数
1		健康草地	1 251
2		受压草地	1 254
3		合成草地	697
4		树木	1 244
5		土壤	1 242
6		水	325
7		住宅区	1 268
8		商业区	1 244
9		道路	1 252
10		高速公路	1 227
11		铁路	1 235
12		停车场 1	1 233
13		停车场 2	469
14		网球场	428
15		跑道	660
样本总数			15 029





(a) 假彩色图



(b) 真值图

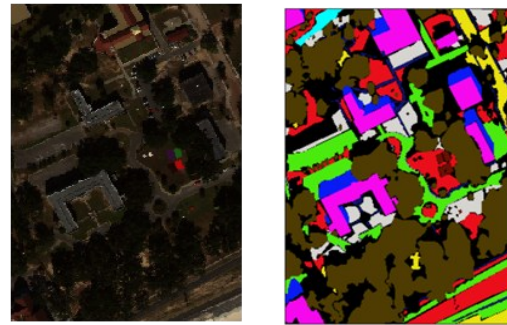
图4 Berlin数据集伪彩色图及其真值图

表2 Berlin数据集类别信息

序号	颜色	类别名称	样本数
1		森林	51 954
2		住宅区	268 642
3		工业区	19 566
4		低矮植物	59 282
5		土壤	17 426
6		耕地	13 305
7		商业区	24 824
8		水	6 672
样本总数			464 671

### 3.2 实验设置

在实验中,随机从每类的训练样本选择200个,余下的所有数据作为测试集.在训练阶段,模型的迭代次












(a) 假彩色图

(b) 真值图

图5 MUUFL数据集伪彩色图及其真值图

表3 MUUFL数据集类别信息

序号	颜色	类别名称	样本数
1		树木	23 246
2		草地	4 270
3		地面	6 882
4		泥土和沙子	1 826
5		道路	6 687
6		水	466
7		建筑物阴影	2 233
8		建筑物	6 240
9		人行道	1 385
10		黄色路缘	183
11		布制面板	269
样本总数			53 687

数设置为500,批大小为128.为全面验证提出的US-MRSD的有效性,选用近年来6种公开的对比方法,包括基于Transformer、CNN和Mamba的多模态遥感数据分类方法,其中包含的融合策略如下.

基于CNN的方法为CCRNNet(Cross-Channel Reconstruction Network)<sup>[7]</sup>和NCGLF2(Network Combining Global and Local Features)<sup>[9]</sup>.CCRNNet利用级联卷积层提取层次化特征,并通过交叉重建模块实现不同模态间的信息交换,从而通过通道相关性约束学习融合特征.NCGLF2通过结合CNN提取高频特征和Transformer捕捉低频信息,有效融合了多模态遥感数据中的局部和全局特征,显著提升了特征的表达能力.

基于Transformer的方法为MFT(Multimodal Fusion Transformer)<sup>[12]</sup>,ExViT(Extended Vision Transformer)<sup>[13]</sup>和FrIT(Fractional fourier Image Transformer)<sup>[2]</sup>.MFT引入了一种多头跨模态注意力机制,将高光谱数据与LiDAR数据进行拼接后进行特征提取,提升多模态遥感数据分类性能.ExViT采用并行的共享位置信息的Transformer分支和可分离卷积模块处理多模态遥感数据,并通过跨模态注意力模块融合不同模态特征,提升融合效果.FrIT在像素级融合HSI和LiDAR后,利用双

分支 CNN 提取局部特征,在多数数域捕获全局上下文,并通过残差跨域连接减少深浅层信息冗余.

基于 Mamba 的方法 S<sup>2</sup>cMamba (cross-State-Space Mamba) 引入了一个包含光谱 Mamba 和空间 Mamba 的双分支结构用于特征提取<sup>[15]</sup>. 该双分支结构通过 Cross-Mamba 模块实现多模态特征的动态融合,利用双向状态空间交互增强特征,实现了分类精度和计算效率的平衡.

### 3.3 参数分析

在实验中有两个主要参数影响多模态遥感数据的分类效果,分别是一致性损失函数的系数 $\gamma$ 和控制折扣因子强度 $\lambda$ 的值. 为评估在实验过程中将证据理论融入模型训练过程中的影响,采用变量控制法. $\gamma$ 值的范围为 $\{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ , $\lambda$ 值的范围为 $\{1, 10, 30, 60, 90\}$ . 图6和图7分别显示 $\gamma$ 和系数 $\lambda$ 与分类精度的关系. 从图6中可以观察到,在不同的多模态遥感数据集上,一致性损失函数的系数 $\gamma$ 对分类精度的影响较大,Houston 和 MUUFL 数据集的波动趋势较大,Berlin 数据集相对较小. 在实验中, $\gamma$ 的值在 Houston、Berlin 和 MUUFL 数据集上分别设置为 0.4、0.6 和 0.6. 从图7中实验结果可得,折扣因子强度 $\lambda$ 的值对分类结果的影响也较大. 根据参数分析的实验结果,在后续的实验,  $\lambda$  的值在 Houston、Berlin 和 MUUFL 数据集上分别设置为 30、90 和 60.

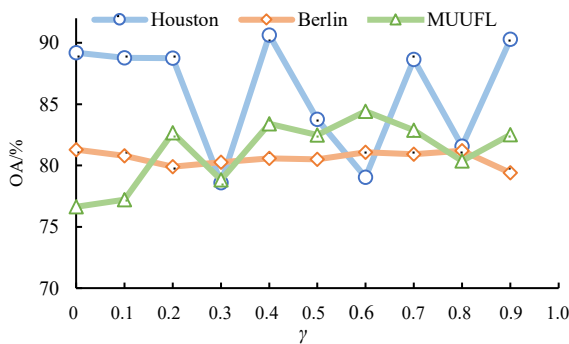


图6 一致性损失函数的系数 $\gamma$ 对分类精度的影响

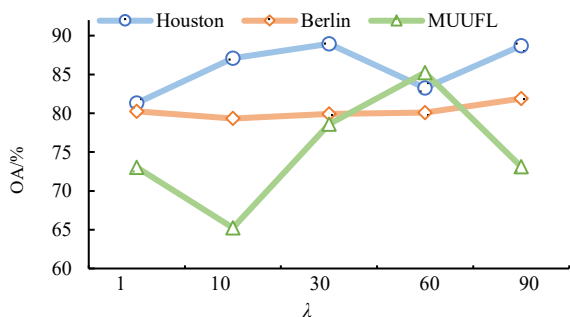


图7 折扣因子强度 $\lambda$ 对分类精度的影响

### 3.4 消融实验

#### 3.4.1 不同融合方式比较

本文是将不同遥感模态的空间信息和通道信息融合,一种简单的方式是将不同模态分别进行提取后融合,为验证空间特征和通道特征提取后融合方式的有效性,将本文的方法与单模态进行融合(Single Modality Fusion, SMF)进行比较. 此外,为验证本文提出的面向不确定性融合的多模态遥感数据分类方法在融合时的有效性,选用相同的特征提取方式以及两种不同的融合方式(简单的拼接方式 Concat、特征加和的方式 Add),在相同实验条件下评估不同方法的分类精度和不确定性估计的性能,实验结果如图8和图9所示.

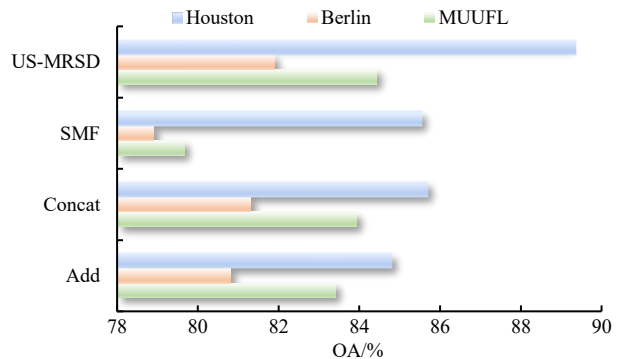


图8 不同数据集不同融合方式的分类精度变化

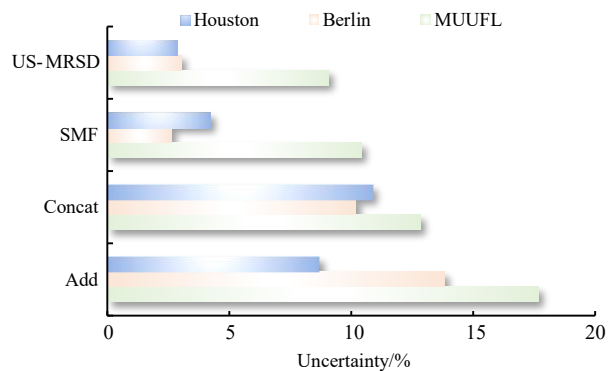


图9 不同数据集不同融合方式的不确定性变化

从实验结果可得,本文方法在3个多模态遥感数据集上的分类精度更高,与以上3种融合方式相比,在 Houston 数据集上效果较明显,约提升4%,与 SMF 方法相比,分类精度提升效果较明显,说明本文的空间特征提取和通道特征提取方法优于不同模态单独特征提取再融合的后融合方式. 同理,在其他多模态遥感数据集上也得到了类似的结果.

从不确定角度分析,图9展示了不同多模态遥感数据集上不同融合方式的不确定性值的变化. 如图9所示,基于拼接和加和的融合方式的不确定性值较高,在

10%以上.然而,本文提出的基于不确定性融合方式不确定性较低.无论是单模态特征提取后基于不确定性的融合还是空间特征和通道特征基于不确定性的融合方式,两种方法的不确定性均在10%左右.综合以上实验结果分析可得,本文基于不确定性融合的方式可以取得较高的分类精度且可以降低不确定性.

### 3.4.2 单/多模态效果对比

为验证模态融合的效果,本节实验分析每个单模态的分类精度和不确定性,对每个遥感模态进行消融,验证所提融合方法的有效性,实验结果如表4所示.表4报告了各单模态及其基于不确定性融合策略的实验结果,包括分类精度与不确定性估计.由表4可见,相较于单一模态,本文方法通过空间和通道特征的不确定性融合,能够充分挖掘多模态遥感数据的互补信息,显著提升整体分类精度.在多个多模态数据集中,高光谱模态的分类性能均优于LiDAR与SAR模态;而在相同实验设置下,经本文融合后,分类精度至少提升1%,且不确定性估计最低.上述结果验证了本文基于不确定性融合的多模态遥感数据分类方法在分类精度和不确定性融合上的优势.

表4 多模态消融实验结果

数据集	HSI	LiDAR	SAR	OA	Uncertainty
Houston	√	—	—	0.86	0.04
	—	√	—	0.60	0.05
	√	√	—	0.89	0.03
Berlin	√	—	—	0.81	0.03
	—	—	√	0.49	0.03
	√	—	√	0.82	0.02
MUUFL	√	—	—	0.82	0.13
	—	√	—	0.66	0.28
	√	√	—	0.84	0.09

### 3.4.3 不同损失函数比较

为更好地验证不同损失函数对不确定性融合框架的影响,表5报告了不同损失函数对分类精度影响的相关消融实验结果.从表5中可得,使用3个损失函数时分类精度更高.在交叉熵损失函数 $L_{acc}$ 的基础上,单独增加 $L_{KL}$ 和 $L_{con}$ 时,分类精度也有一定的提升.增加 $L_{con}$ 后,分类精度提升较为明显,如在MUUFL数据集上,提升约10%,充分说明一致性损失函数在约束各模态预测结果的一致性上效果明显.

### 3.5 实验结果及分析

表6~表8分别报告了3个多模态遥感数据集Houston、Berlin和MUUFL在不同分类方法中的分类结果.前3行表示不同方法在总体精度(OA)、平均精度(AA)和Kappa(K)系数3项评价指标上的实验结果.其中,OA表示整个测试集中被正确分类的样本比例,AA表

表5 不同损失函数的消融实验

	$L_{acc}$	$L_{KL}$	$L_{con}$	OA	AA	K
Hou Ston	√	—	—	89.14 ± 0.95	90.63 ± 1.09	88.27 ± 1.02
	√	√	—	89.19 ± 1.97	91.00 ± 1.80	88.31 ± 2.12
	√	—	√	89.23 ± 0.71	91.42 ± 0.66	88.36 ± 0.77
	√	√	√	89.37 ± 0.97	91.28 ± 5.65	88.52 ± 1.05
Berlin	√	—	—	80.93 ± 0.59	61.93 ± 5.69	66.15 ± 0.53
	√	√	—	81.29 ± 0.14	63.43 ± 5.82	66.50 ± 0.70
	√	—	√	81.67 ± 0.23	66.98 ± 4.15	67.48 ± 0.04
	√	√	√	81.89 ± 0.30	61.44 ± 5.16	70.18 ± 1.02
MUUFL	√	—	—	75.03 ± 3.45	67.99 ± 0.19	63.33 ± 6.33
	√	√	—	76.63 ± 3.69	67.59 ± 0.88	66.03 ± 6.59
	√	—	√	84.80 ± 0.77	64.19 ± 1.74	79.16 ± 1.06
	√	√	√	85.25 ± 0.58	62.75 ± 4.13	80.15 ± 0.72

示各类别精度的平均值,Kappa系数用于衡量模型预测与真实标签之间的一致性水平.实验采用5个不同的随机种子重复运行,结果为均值±标准差的形式.

从分类精度结果分析,与其他方法相比,表6~表8中所提多模态遥感数据不确定性融合方法在不同多模态遥感数据集的分类精度较高,分别达到89.37%、81.89%和85.25%.与其他经典基于CNN、Transformer和Mamba的深度学习方法相比,包括MFT、CCRNNet、ExViT、NCGLF2和FrIT,所提方法的分类精度最高.在Houston数据集上,与采取CNN特征提取方法CCRNNet和NCGLF2相比,提出的US-MRSD方法至少提升2%.在基于Transformer和Mamba的方法也得到了类似的结论.其中,Berlin数据集中类别6(耕地)、7(商业区)、8(水)分类精度较低的原因是各类别的原始样本数不均衡.如表2所示,类别6与类别8的像素数仅13305与6672,而类别2高达268642,相差两个数量级.对于样本数多的类别,比如类别2,选取的训练样本相对较多,而类别6、7、8在训练集中仅为几个,导致模型训练过程中对这些类别的学习能力不足.另一方面,第6类(浅蓝色)、第7类(深蓝色)、第8类(粉色)分布的范围比较大,且常以狭长条带或离散小块形式出现在类别的边缘.这种分布方式使得这些样本在感受野内极易混入邻域其他类别的特征,常存在边缘混淆问题,导致模型分类精度不高.而对于第1类森林(褐色)、第2类住宅区(灰色)、第3类工业区(红色)等类别,每类的分布比较密集,类内同质性高、类间边界清晰,模型更容易提取判别性特征,因此分类精度显著高于6、7、8类的精度下降.

由于不同的多模态遥感数据分类方法采取的融合策略不同,在未采取不确定性融合的分类方法中难以计算不确定性.因此,在实验结果中未报告所有对比方

法的不确定性. 一般而言, 为提升模型降低不确定性一般会降低分类精度. 然而, 从本文实验结果分析, 提出的 US-MRSD 方法在保证甚至略提升分类精度的同时不确定性显著降低. 同时, 在图 10~12 的代表性区域可视化分类结果图中也得到了类似的结论.

图 10 中, 最上面粉色小目标区域应为类别 8 商业区, 但从不同方法的分类结果图看出, 该区域经常与蓝色相混淆, 即类别 7 住宅区. 因为两类地物屋顶材

质光谱高度相似, LiDAR 高程差异亦不大. 若分类结果倾向于高光谱模态, 会导致两个类被经常混淆分类错误. 通过本文的冲突感知机制, 提高高光谱模态的不确定性, 同时降低 LiDAR 模态的不确定性, 从而得到较准确的分类结果, 分类正确. 类似地, 在 MUFFL 数据集中, 粉色周围为混合像素区域, 它应为类别 8 建筑物, 经常与建筑物阴影混淆, 而本文方法取得更好的分类结果.

表 6 Houston 数据集实验结果

单位: %

方法	MFT	CCNet	ExViT	NCGLF2	FrIT	S <sup>2</sup> cMamba	US-MRSD
OA	88.62 ± 0.54	81.54 ± 0.53	88.88 ± 0.71	87.16 ± 1.27	87.14 ± 0.82	84.61 ± 1.82	89.37 ± 0.97
AA	87.24 ± 1.07	76.19 ± 0.76	88.59 ± 0.59	88.11 ± 1.28	86.49 ± 0.11	84.28 ± 1.93	91.28 ± 5.65
K	87.69 ± 0.59	80.00 ± 0.58	87.98 ± 0.77	86.11 ± 1.38	86.08 ± 0.88	83.36 ± 1.97	88.52 ± 1.05
类别 1	91.70 ± 4.63	86.27 ± 0.54	85.80 ± 3.90	89.24 ± 6.93	89.69 ± 5.79	82.98 ± 4.36	90.72 ± 1.90
类别 2	89.18 ± 4.70	94.88 ± 0.68	94.55 ± 1.98	96.34 ± 2.62	96.62 ± 2.21	77.97 ± 8.97	94.61 ± 5.64
类别 3	98.99 ± 0.61	93.66 ± 2.44	96.18 ± 2.59	98.66 ± 1.06	99.70 ± 0.32	97.43 ± 1.95	95.10 ± 7.57
类别 4	95.52 ± 1.95	91.79 ± 0.84	88.63 ± 2.62	93.27 ± 3.82	92.47 ± 3.39	91.20 ± 2.77	91.78 ± 2.26
类别 5	97.35 ± 2.11	95.04 ± 0.48	99.44 ± 0.45	100.00 ± 0.00	93.26 ± 8.60	97.02 ± 1.14	98.34 ± 0.60
类别 6	60.68 ± 6.45	0.99 ± 0.50	81.66 ± 7.59	88.32 ± 5.82	60.40 ± 9.08	68.19 ± 13.14	97.31 ± 4.25
类别 7	88.79 ± 2.13	83.72 ± 1.10	90.18 ± 5.56	79.50 ± 5.73	88.87 ± 2.41	78.58 ± 13.60	88.88 ± 3.97
类别 8	91.94 ± 2.70	81.97 ± 0.03	89.89 ± 6.80	79.72 ± 5.40	87.68 ± 2.23	84.13 ± 7.75	92.64 ± 1.67
类别 9	72.39 ± 1.55	61.27 ± 2.63	83.59 ± 5.61	66.09 ± 10.13	83.55 ± 5.11	76.48 ± 10.01	82.08 ± 11.83
类别 10	87.93 ± 3.53	90.76 ± 4.35	77.01 ± 15.62	87.82 ± 11.45	66.35 ± 8.81	78.60 ± 4.46	77.96 ± 8.41
类别 11	85.76 ± 1.59	82.62 ± 0.58	89.82 ± 3.02	87.3 ± 6.70	89.95 ± 1.75	94.83 ± 2.93	88.79 ± 6.45
类别 12	86.77 ± 4.08	73.82 ± 0.79	88.06 ± 11.46	78.04 ± 10.81	77.42 ± 5.74	78.27 ± 3.58	87.96 ± 8.12
类别 13	65.60 ± 10.72	35.79 ± 5.05	68.41 ± 9.02	77.61 ± 7.08	73.31 ± 9.78	68.23 ± 4.40	88.60 ± 14.20
类别 14	96.44 ± 2.52	74.93 ± 10.81	99.71 ± 0.47	100.00 ± 0.00	99.19 ± 0.11	91.88 ± 4.20	96.80 ± 4.10
类别 15	99.59 ± 0.58	95.35 ± 3.43	95.97 ± 5.82	99.72 ± 0.55	98.89 ± 1.46	98.46 ± 1.23	97.72 ± 2.33

表 7 Berlin 数据集实验结果

单位: %

方法	MFT	CCNet	ExViT	NCGLF2	FrIT	S <sup>2</sup> cMamba	US-MRSD
OA	81.78 ± 0.16	77.61 ± 0.97	79.82 ± 3.05	81.42 ± 1.36	81.30 ± 0.74	78.10 ± 3.71	81.89 ± 0.30
AA	61.81 ± 2.98	50.74 ± 2.88	53.68 ± 8.14	56.24 ± 2.47	62.22 ± 3.00	55.79 ± 4.67	61.44 ± 5.16
K	69.39 ± 0.35	62.03 ± 1.81	65.56 ± 7.32	69.54 ± 1.83	69.18 ± 1.18	65.6 ± 3.93	70.18 ± 1.02
类别 1	73.42 ± 1.63	78.40 ± 2.40	79.63 ± 11.70	84.97 ± 5.51	79.16 ± 3.39	75.91 ± 5.30	85.51 ± 4.37
类别 2	94.49 ± 0.21	91.93 ± 0.44	93.61 ± 2.66	92.97 ± 2.79	92.46 ± 1.22	88.24 ± 8.12	81.47 ± 3.03
类别 3	56.56 ± 3.58	14.56 ± 0.88	31.78 ± 16.85	34.21 ± 23.38	38.46 ± 8.43	41.05 ± 15.94	77.74 ± 9.83
类别 4	82.22 ± 3.31	75.73 ± 1.25	75.32 ± 15.07	82.55 ± 3.48	81.42 ± 3.41	83.51 ± 9.19	83.23 ± 2.00
类别 5	80.83 ± 4.89	80.66 ± 7.16	69.79 ± 16.70	79.31 ± 4.88	85.35 ± 6.04	80.32 ± 8.42	82.37 ± 18.79
类别 6	29.22 ± 16.45	7.86 ± 2.86	5.49 ± 10.46	21.03 ± 12.76	29.01 ± 10.29	24.77 ± 26.16	12.86 ± 26.03
类别 7	14.03 ± 1.62	19.05 ± 2.07	36.41 ± 7.96	30.91 ± 16.94	27.22 ± 4.19	29.61 ± 18.58	39.35 ± 8.66
类别 8	63.67 ± 8.08	37.76 ± 10.24	37.43 ± 25.18	23.99 ± 19.65	64.68 ± 16.96	22.87 ± 23.91	23.46 ± 42.37

### 3.6 模型计算复杂度分析

为衡量不同方法的模型复杂度, 从参数量、FLOPs 和推理时间 3 个维度与 5 种代表性对比方法进行横向对比, 结果如表 9 所示.

所提出的 US-MRSD 方法在 3 个数据集的参数量低

于基于全局-局部和基于频域的 Transformer 方法, 如 NCGLF2 和 FrIT, 与结构简单 ExViT 和 MFT 方法参数量相近. 参数量和复杂度高于基于 Mamba 的 S<sup>2</sup>cMamba 方法. 因为基于 Mamba 的方法本质可以降低模型复杂度, 因此, 本文提出的方法比 S<sup>2</sup>cMamba 的参数量要高.

表 8 MUUFL数据集实验结果

单位:%

Methods	MFT	CCRNet	ExViT	NCGLF2	FrIT	S <sup>2</sup> cMamba	US-MRSD
OA	83.35 ± 1.36	81.00 ± 0.33	83.28 ± 0.69	81.40 ± 1.50	83.66 ± 2.18	84.24 ± 0.85	85.25 ± 0.58
AA	52.48 ± 5.47	56.20 ± 2.19	52.43 ± 4.13	54.62 ± 6.54	57.20 ± 3.50	58.19 ± 4.24	62.75 ± 4.13
K	77.57 ± 1.89	74.41 ± 0.48	77.49 ± 1.06	75.31 ± 1.87	78.04 ± 2.88	78.87 ± 1.19	80.15 ± 0.72
类别1	95.91 ± 1.66	95.35 ± 0.05	95.70 ± 1.36	93.21 ± 2.42	96.18 ± 1.92	94.61 ± 1.44	94.03 ± 0.25
类别2	66.38 ± 12.95	64.78 ± 4.15	66.47 ± 8.37	73.25 ± 13.32	73.34 ± 7.92	79.24 ± 7.44	77.69 ± 5.65
类别3	76.40 ± 9.62	75.90 ± 0.56	71.73 ± 12.16	70.56 ± 9.83	74.66 ± 2.47	69.46 ± 10.15	71.71 ± 1.17
类别4	43.01 ± 18.54	57.33 ± 12.97	41.70 ± 8.56	55.23 ± 27.51	75.43 ± 9.13	58.14 ± 21.77	70.23 ± 35.63
类别5	87.25 ± 2.90	83.82 ± 2.52	89.77 ± 2.37	81.23 ± 8.51	84.62 ± 6.76	90.78 ± 2.76	73.91 ± 1.74
类别6	11.30 ± 14.63	55.91 ± 0.26	6.78 ± 9.21	23.31 ± 38.29	7.12 ± 14.23	17.58 ± 32.76	38.62 ± 47.36
类别7	61.10 ± 10.54	31.94 ± 20.15	59.11 ± 11.17	65.34 ± 8.62	72.89 ± 9.16	65.75 ± 5.05	80.72 ± 3.39
类别8	91.47 ± 3.93	86.88 ± 0.40	94.31 ± 1.44	90.91 ± 3.51	81.53 ± 9.54	93.23 ± 3.41	93.56 ± 3.27
类别9	30.68 ± 19.07	7.72 ± 0.84	34.27 ± 8.11	12.50 ± 8.16	25.81 ± 7.61	19.04 ± 10.36	47.58 ± 9.77
类别10	2.28 ± 3.68	4.92 ± 4.31	2.15 ± 2.82	1.44 ± 1.80	10.91 ± 8.74	5.81 ± 6.86	20.74 ± 39.66
类别11	11.48 ± 19.22	8.57 ± 29.28	14.75 ± 20.36	33.79 ± 32.65	26.75 ± 25.18	46.23 ± 26.08	59.09 ± 48.28

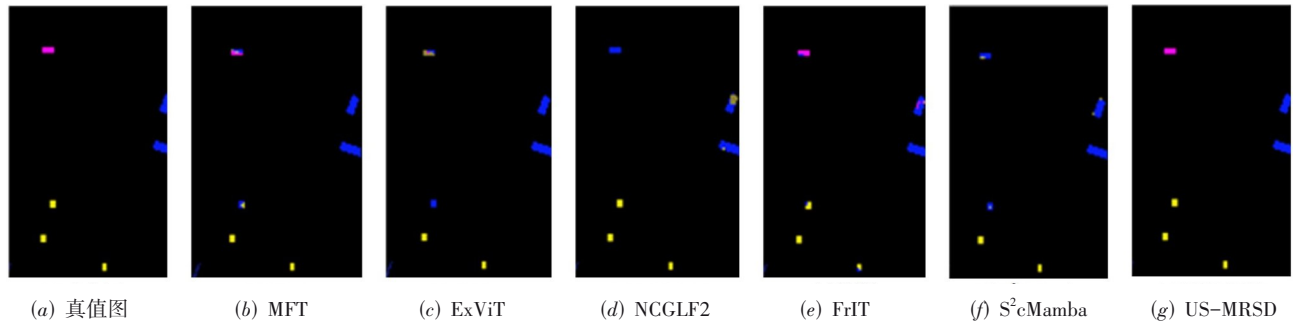


图 10 Houston数据集代表性区域的分类结果

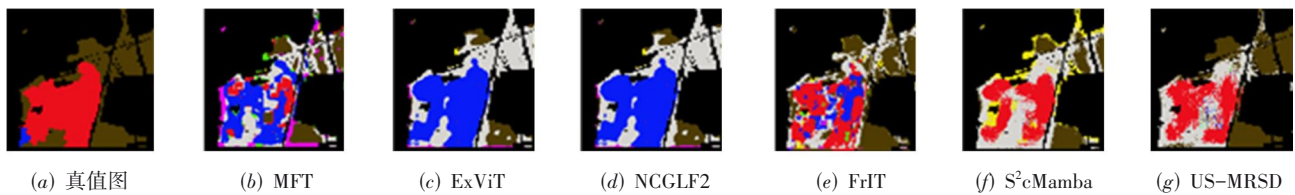


图 11 Berlin数据集代表性区域的分类结果

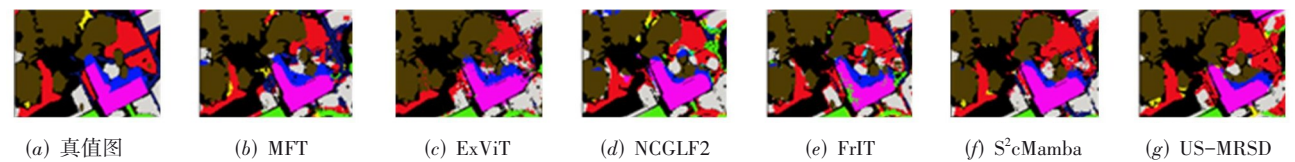


图 12 MUUFL数据集代表性区域的分类结果

但从推理时间角度分析,在Berlin数据集上的实测推理时间仅为406.34 s,较S<sup>2</sup>cMamba方法的525.74 s缩短约23%。这表明,本文通过证据融合机制降低不确定性的同时,仍维持了中等复杂度,权衡了精度、时间和参数量,兼具分类结果的准确性与实用性。

#### 4 结论

针对现有多模态遥感数据分类方法存在的预测结

果不确定性,本文提出了一种面向不确定性融合的多模态遥感数据分类框架。US-MRSD从证据深度学习的角度出发,联合提取不同模态(如合成孔径雷达数据、激光雷达数据、高光谱数据等)的空间特征与通道特征,并构建对应的证据神经网络。通过设计特定的证据融合函数,实现了基于证据信息的有效融合。在融合过程中,引入冲突感知的动态权重调整机制,利用折扣因子自适应地降低冲突模态的权重,动态重分配证据质

表 9 三个多模态遥感数据集上不同方法的计算复杂度

方法	计算复杂度	MFT	ExViT	NC-GLF2	FrIT	S <sup>2</sup> c Mamba	US-MRSD
Houston	参数量/M	0.32	0.28	0.64	4.33	0.12	0.29
Berlin		0.40	0.28	0.64	4.32	0.12	0.31
MUUFLL		0.26	0.27	0.64	4.33	0.12	0.29
Houston	FLOPs/G	0.27	1.04	0.9	3.5	0.26	1.64
Berlin		0.45	1.05	0.9	4.5	0.26	1.84
MUUFLL		0.12	1.04	0.9	2.7	0.26	1.50
Houston	推理时间/s	4.44	11.81	7.04	10.16	11.18	7.82
Berlin		224.39	450.15	201.37	501.72	525.74	406.34
MUUFLL		7.32	0.95	21.92	22.75	22.09	9.16

量,从而有效降低预测结果不确定性.此外,本文在模型参数优化过程中引入一致性损失函数,约束各模态预测结果的一致性.实验结果表明,本文提出的面向不确定性融合的多模态遥感数据分类在不降低分类精度的情况下,与其他融合方式相比,US-MRSD方法在分类精度略微提升的情况下,不确定性至少降低5%.实验结果证明了US-MRSD的有效性,为多模态数据融合提供了一种新的思路.未来,本文将进一步深化对多模态数据融合机制的研究,探索更高效的证据融合策略,以应对更复杂的多模态遥感数据,进一步提升模型对不同模态数据差异的适应能力.

#### 参考文献

- [1] SUN X, TIAN Y, LU W X, et al. From single- to multimodal remote sensing imagery interpretation: A survey and taxonomy[J]. *Science China Information Sciences*, 2023, 66(4): 140301.
- [2] ZHAO X D, ZHANG M M, TAO R, et al. Fractional Fourier image transformer for multimodal remote sensing data classification[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(2): 2314-2326.
- [3] KIEU N, NGUYEN K, NAZIB A, et al. Multimodal co-learning meets remote sensing: Taxonomy, state of the art, and future works[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024, 17: 7386-7409.
- [4] HONG D F, LI C Y, ZHANG B, et al. Multimodal artificial intelligence foundation models: Unleashing the power of remote sensing big data in earth observation[J]. *The Innovation Geoscience*, 2024, 2(1): 100055.
- [5] 贾永红, 李德仁. 多源遥感影像像素级融合分类与决策级分类融合研究[J]. *武汉大学学报(信息科学版)*, 2001, 26(5): 430-434.
- [6] 张良培, 沈焕锋. 遥感数据融合的进展与前瞻[J]. *遥感学报*, 2016, 20(5): 1050-1061.
- [7] WU X, HONG D F, CHANUSSOT J. Convolutional neural networks for multimodal remote sensing data classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 5517010.
- [8] 付琨, 王佩瑾, 冯瑛超, 等. 遥感跨模态智能解译: 模型、数据与应用[J]. *中国科学(信息科学)*, 2023, 53(8): 1529-1559.
- [9] TU B, REN Q, LI J, et al. NCGLF2: Network combining global and local features for fusion of multisource remote sensing data[J]. *Information Fusion*, 2024, 104: 102192.
- [10] 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.
- [11] YANG B, WANG X, XING Y, et al. Modality fusion vision transformer for hyperspectral and LiDAR data collaborative classification[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024, 17: 17052-17065.
- [12] 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: 5515620.
- [13] YAO J, ZHANG B, LI C Y, et al. Extended vision transformer (ExViT) for land use and land cover classification: A multimodal deep learning framework[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 5514415.
- [14] LI Y S, LI D X, XIE W Y, et al. Semi-mamba: Mamba-driven semi-supervised multimodal remote sensing feature classification[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2025, 35(10): 9837-9849.
- [15] ZHANG G L, ZHANG Z X, DENG J W, et al. S<sup>2</sup>Cross-

- Mamba: Spatial-spectral cross-mamba for multimodal remote sensing image classification[J]. IEEE Geoscience and Remote Sensing Letters, 2024, 21: 5510705.
- [16] MA M R, ZHAO J X, MA W P, et al. A mamba-aware spatial-spectral cross-modal network for remote sensing classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2025, 63: 4402515.
- [17] WEN J G, WU X D, YOU D Q, et al. The main inherent uncertainty sources in trend estimation based on satellite remote sensing data[J]. Theoretical and Applied Climatology, 2023, 151(1): 915-934.
- [18] WOODCOCK C E. Uncertainty in Remote Sensing[M]// Uncertainty in Remote Sensing and GIS. Hoboken: Wiley, 2002: 19-24.
- [19] KOLLER C, SHAHRAD M, ZHU X X. Uncertainty-guided representation learning in local climate zone classification[C]//IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium. Piscataway: IEEE, 2022: 183-186.
- [20] XU Y H, BAI T, YU W K, et al. AI security for geoscience and remote sensing: Challenges and future trends[J]. IEEE Geoscience and Remote Sensing Magazine, 2023, 11(2): 60-85.
- [21] ZINNER T, HAUSMANN P, EWALD F, et al. Ground-based imaging remote sensing of ice clouds: Uncertainties caused by sensor, method and atmosphere[J]. Atmospheric Measurement Techniques, 2016, 9(9): 4615-4632.
- [22] HAN Z B, ZHANG C Q, FU H Z, et al. Trusted multi-view classification with dynamic evidential fusion[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(2): 2551-2566.
- [23] HAN Z B, YANG F, HUANG J Z, et al. Multimodal dynamics: Dynamical fusion for trustworthy multimodal classification[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 20675-20685.
- [24] HAN Z B, YANG Y F, ZHANG C Q, et al. Selective learning: Towards robust calibration with dynamic regularization[EB/OL]. (2024-07-14) [2025-11-10]. <https://arXiv.org/abs/2402.08384>.
- [25] RUSSELL R L, REALE C. Multivariate uncertainty in deep learning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(12): 7937-7943.
- [26] TYRALIS H, PAPACHARALAMPOUS G. A review of predictive uncertainty estimation with machine learning[J]. Artificial Intelligence Review, 2024, 57(4): 94.

## 作者简介



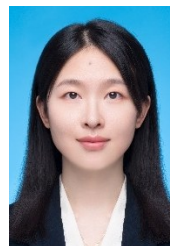
何欣女, 1994年1月出生于黑龙江省齐齐哈尔市. 现为哈尔滨工业大学副研究员. 主要研究方向为遥感图像智能处理.  
E-mail: hexin1@hit.edu.cn



陈雨时男, 1978年9月出生于河南省信阳市. 现为哈尔滨工业大学教授、博士生导师. 主要研究方向为遥感图像处理及深度学习理论和应用.  
E-mail: cheniyushi@hit.edu.cn



谷延锋男, 1977年12月出生于黑龙江省佳木斯市. 现为哈尔滨工业大学电子与信息工程学院教授、博士生导师. 主要研究方向为空天智能信息处理、高光谱遥感、多维度信息探测与处理系统.  
E-mail: guyf@hit.edu.cn



刘天竹女, 1990年12月出生于黑龙江省哈尔滨市. 现为哈尔滨工业大学电子与信息工程学院研究员、博士生导师. 主要研究方向为多模遥感信息智能处理. 中国电子学会会员编号: E190025027M.  
E-mail: tzliu@hit.edu.cn