arXiv:2301.03335v1 [eess.IV] 9 Jan 2023

面向高光谱和激光雷达数据分类的基于最近邻的对比学习

Meng Wang、Feng Gao、Junyu Dong、Heng-Chao Li、Qian Du

摘要—联合高光谱图像（HSI）和激光雷达数据分类旨在以更详细、更精确的程度解译地面目标。尽管深度学习方法在多源数据分类任务中已取得显著成功，但自监督学习却很少受到探索。由于现有对比学习框架未利用邻域区域的语义相似性，为多源数据分类构建稳健的自监督学习模型通常并非易事。此外，多源数据分布不一致所导致的异质性差距阻碍了分类性能。为克服这些缺点，我们提出了一种基于最近邻的对比学习网络（NNCNet），充分利用大量无标签数据来学习判别性特征表示。具体而言，我们提出了一种基于最近邻的数据增强方案，以利用邻近区域之间增强的语义关系，从而更准确地捕获跨模态语义对齐。此外，我们设计了一个双线性注意力模块，以挖掘高光谱图像与激光雷达数据之间的二阶乃至高阶特征交互。在四个公开数据集上的大量实验表明，我们的NNCNet优于最先进的方法。源代码可在 https://github.com/summitgao/NNCNet获取。

索引词—高光谱图像、自监督学习、光探测与测距、对比学习、图像分类。

I. 引言

RECENTLY 随着卫星传感器的快速发展，每天都能获取越来越多的多模态图像（光学、合成孔径雷达、高光谱和激光雷达）[1]。在这些多模态数据中，高光谱图像（HSI）提供详细的光谱信息，用于识别地面上的特定目标；而激光雷达数据则提供该区域的高程信息 [2] [3] [4]。这些高光谱图像和激光雷达传感器在成像机制、空间分辨率乃至覆盖范围上各不相同。因此，两种传感器捕获了地球的不同属性，例如光谱辐射亮度和高度信息。例如，地面上的“树木”与山丘上的“树木”在光谱域中没有显著差异，但可以通过激光雷达数据 [5]将它们区分开来。因此，联合利用高光谱图像和激光雷达数据，可以更详细、更准确地解译地面目标，而单模态数据 [6]很难做到这一点。因此，跨模态数据分类受到了广泛关注，并已广泛应用于多源图像解译 [7] [8]。

本研究部分得到了中国国家重点研发计划项目（项目编号：2018AAA0100602）和国家自然科学基金委员会项目（项目编号：42106191）的资助。Meng Wang、Feng Gao 和 Junyu Dong 就职于中国青岛266100，中国海洋大学信息科学与工程学院。（通讯作者：*Feng Gao*。）H. -C. Li 就职于中国成都610031，西南交通大学四川省信息编码与传输重点实验室。Qian Du 就职于美国密西西比州斯塔克维尔39762，密西西比州立大学电气与计算机工程系。

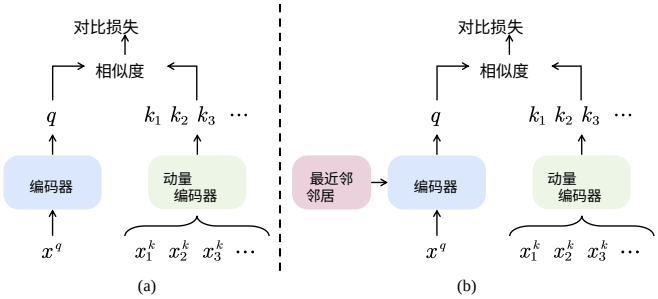


图1. MoCo与所提出的基于最近邻的对比学习框架的概念比较。在所提出的框架中，最近邻被视为正样本。该框架利用了邻域区域之间的语义相似性。

为解决高光谱图像和激光雷达联合分类问题，人们付出了大量努力。传统上，人们提出了特征级融合模型，这些模型通常将高光谱图像和激光雷达特征拼接起来进行分类 [9] [10] [11]。除特征级融合外，决策级融合也是高光谱图像和激光雷达分类的另一种常见方案。人们分别为高光谱图像和激光雷达数据设计了若干分类器。通常采用投票策略来获得最终分类图 [12]。随后，为了进一步利用高层语义特征，人们将卷积神经网络（CNN）用于多源数据分类 [13]。编码器-解码器网络 [14]，耦合卷积神经网络 [15],Gabor 卷积神经网络 [16]，交叉注意力 [17]，和 Transformer [18]被用于提取具有代表性的多源特征，这些方法取得了良好的性能。

在实践中，深度学习模型已在各种多源数据联合分类任务中取得了显著成功。然而，构建有效的高光谱图像和激光雷达分类模型并非易事。关键原因之一是，基于深度学习的模型通常需要大量标注样本才能达到令人满意的准确率，而在地物建模中，标注样本成本高昂且数量有限。近期的自监督学习研究促使深度网络学习更具代表性且可解释性更强的

特征 [19] [20]，应用于自然语言处理 [21] [22]和计算机视觉任务。自监督学习挖掘大规模无标签数据的内在属性和语义，以获得有益的语义表示，且无需人工标注数据 [23]。自监督训练完成后，所学特征可作为预训练模型迁移至分类任务（尤其是在训练数据较少时），从而提升分类性能并缓解过拟合 [24] [25] [26] [27]。在高光谱图像和激光雷达联合分类中，自监督学习鲜少得到探索。本文旨在构建一种有效的自监督模型来解决这一问题。

由于以下原因，为高光谱图像和激光雷达联合分类任务构建稳健的自监督学习模型通常并非易事：**1）数据增强方案**。在用于自监督学习的动量对比（MoCo）中[22]，，采用随机颜色抖动、随机水平翻转和随机灰度转换进行数据增强。然而，这种数据增强方案未考虑正负样本之间的空间距离，也没有利用邻域区域的语义相似性。因此，如何恰当地利用邻近区域之间的语义相似性是一项重大挑战。**2）异质性差距**。高光谱图像和激光雷达联合分类要求同时全面理解复杂的异质数据。然而，多源数据分布不一致所导致的异质性差距会极大地阻碍其实现。因此，弥合这一差距对于实现更稳健的多源数据分类至关重要。

为应对上述挑战，我们提出了一种**N**最近邻**N**邻居式**C**对比学习**Net**网络，**NNCNet**简称 NNCNet，旨在学习一种编码器，将同类相似数据编码为相似表示，并使不同类别数据的编码结果尽可能不同。具体而言，我们提出了一种基于最近邻的框架，以利用邻近区域之间增强的语义关系。如图 1 所示，将正样本的最近邻输入编码器进行对比学习。通过鼓励同一样本及其在空间域中的最近邻的不同视图之间保持接近，从而学习特征表示。因此，该对比学习框架能够泛化到数据增强可能未涵盖的新特征嵌入。此外，我们设计了一个双线性注意力融合模块，以利用高光谱图像与激光雷达数据之间的二阶乃至更高阶特征交互，并更灵活地控制信息流。

本文的贡献如下：

- 我们提出了一种自监督对比学习方法 NNCNet，该方法融合了基于最近邻的数据增强方案。该方案能够利用邻域区域之间的语义相似性，从而更准确地捕捉跨模态语义对齐。据我们所知，我们首次将自监督对比学习应用于高光谱图像和激光雷达的联合分类，这在理论和实践上都具有重要意义。
- 我们提出了一种双线性注意力融合模块，旨在增强高光谱图像和激光雷达数据的上下文表示。该模块能够捕捉多源数据之间的二阶特征交互。
- 我们在四个基准数据集上进行了大量实验，以验证 NNC-Net 的有效性。此外，我们还公开了代码和参数，以便其他研究人员使用。

II. 相关工作

A. 基于形态学滤波器的方法：高光谱图像和激光雷达分类

高光谱图像与激光雷达的联合使用已应用于多种场景，例如照明校准 [28],森林区域分析 [29],灌木丛火灾监测 [30], 和城市蔓延建模 [31]。研究人员已投入大量精力，致力于挖掘多源数据之间的互补信息，尤其是在基于形态学滤波器的方法中。形态学滤波器被广泛用于减弱冗余的空间细节，同时保留几何结构。Pedergnana 等人 [9] 将形态学扩展属性剖面应用于高光谱图像和激光雷达数据分类，并将从两类数据中提取的特征堆叠后用于分类。Liao 等人 [32]从高光谱图像和激光雷达数据中计算形态学属性剖面，并使用广义图方法融合这些属性剖面。Khodadadzadeh 等人 [33]指出，简单堆叠多源数据的形态学属性剖面可能会包含冗余特征。为解决这一问题，他们提出了一种基于多项逻辑回归分类器的多特征学习方法，可自适应地利用空间和光谱派生特征。后来，人们发现属性剖面的阈值初始化过程复杂且耗时，因此提出了消亡剖面 [34] 来解决这一问题。Ghamisi 等人 [35] 提出了一种基于消亡剖面和深度学习的分类框架。

B. 基于卷积神经网络的高光谱图像和激光雷达分类方法

近年来，深度卷积神经网络在遥感数据融合领域受到广泛关注，许多基于卷积神经网络的模型已被提出用于多源数据分类。Xu 等人 [36] 提出了一种双分支卷积神经网络模型，由一个二维卷积网络和一个一维卷积网络组成。Zhang 等人 [37] 提出了一种块到块卷积神经网络，用于高光谱图像和激光雷达数据的联合特征提取。Chen 等人 [38] 提出了一种用于多源特征提取的卷积神经网络与深度神经网络混合模型。卷积神经网络用于从多源数据中提取信息丰富的特征，深度神经网络则用于融合这些异构特征，以实现稳健分类。Li 等人 [39] 提出了一种用于多源数据分类的双通道空间、光谱和多尺度注意力卷积神经网络。Hang 等人 [15] 使用耦合卷积神经网络进行多源数据分类。耦合层减少了参数数量，并引导两个网络相互学习。Zhao 等人 [16] 提出了一种分数阶 Gabor 卷积神经网络，

并着重于高效的特征融合。分数阶 Gabor 卷积核用于多尺度和多方向特征提取，并能生成抵御语义变化的稳健特征表示。在 [40], 一种多源图融合网络被提出，用于将特征提取与融合整合到单一网络中。该网络构建多模态图，以指导多模态图像特征提取。Gao 等人[17] 提出了一种用于多源遥感图像分类的深度特征交互网络。该网络设计了一致性损失、判别损失和分类损失，以进行参数优化。

尽管卷积神经网络已成功应用于高光谱图像与激光雷达联合分类，但其性能对于实际应用而言仍不尽如人意。一个重要因素可能是缺乏充足的标注数据。在实际应用中，遥感数据标注成本高昂，难以获得稳健的深度学习模型。为解决这一问题，我们旨在构建一种简单而有效的自监督方法，用于多源数据联合分类。该方法从大规模无标签数据中提取内在属性和语义，以获取有益的特征表示。此外，还采用基于最近邻的数据增强方案，以利用邻近区域之间的语义关系。

III. 方法

如图2所示，提出的 NNCNet 由三个部分组成：基于最近邻的数据增强、基于双线性注意力的编码器以及对比损失计算。考虑到提出的 NNCNet 基于自监督对比学习框架，我们首先介绍基于最近邻的对比学习框架，随后详细阐述基于双线性注意力的编码器。

A. 基于最近邻的动量对比学习框架

鉴于无标签数据没有监督信息，我们旨在从大规模无监督高光谱图像和激光雷达数据中提取监督信息。我们的目标是训练一个编码器，使同一样本经过不同变换后在特征空间中的表示尽可能接近，而不同样本的表示尽可能远离。为解决这一问题，He 等人 [22] 提出了用于自监督学习的动量对比（MoCo）。具体而言，从数据中选取一个小批量的样本。对每个样本进行随机数据增强（高斯模糊、翻转或光谱失真），生成一个查询样本和一个键样本。查询样本和键样本分别经过编码，得到嵌入 q和 k。计算 q 与 k 之间的余弦相似度，以进行表征学习。来自同一图像的嵌入定义为正键，来自不同图像的嵌入定义为负键。在 MoCo 中，通过队列和动态编码器构建动态字典。

对于遥感数据分类，我们认为随机增强几乎无法为同一对象表征提供正样本对。为了涵盖某一类别中更多的方差，我们提出了基于最近邻的对比

算法 1 基于最近邻的 PyTorch 风格对比学习伪代码。

```
# f_q：查询的编码器网络 # f_k：键的编


f_k.param = f_q.param# 初始化参数

# 加载包含 N 个样本的小批量 x
对于 loader 中的 x:
    nn   =邻居(x)# 生成 x 的邻居
    x_查询  =数据增强(x)# 随机进行查询数据增强
    x_键   =数据增强(x)# 键的随机数据增强
    x_n   =数据增强(nn)# 邻居的随机数据增强
    # 将 x_k 中一半的样本随机替换为 x_n
    x_k  =替换(x_k, x_n)

    q = f_q.forward(x_q)# 查询：NxC
    k = f_k.forward(x_k)# 键：NxC
    k = k.detach() # 键无梯度

    # 正逻辑值：Nx1
    # bmm：批量矩阵乘法
    l_pos = bmm(q.view(N,1,C),      k.view(N,C,1))
    # 负逻辑值：NxK
    # mm：矩阵乘法
    l_neg = mm(q.view(N,C), 队列.view(C,K))

    # 逻辑值：Nx(1+K)
    逻辑值  =cat([l_pos, l_neg], dim=1)

    # 对比损失计算
    标签  =zeros(N)
    损失 = CrossEntropyLoss(logits/t,          标签)

    # 反向传播，仅更新查询网络
    loss.backward()
    更新(f_q.param)

    # 字典更新
    f_k.param = r*f_k.param+(1-r)*f_q.param
    将 k 入列（queue）# 将当前键入栈
    将 queue 出列 # 将最早的键出栈
```

对比学习框架。在遥感图像中，某个指定位置的邻居标签往往相同。如图3所示，红框内的区域是屋顶，绿色框表示其最近邻。蓝色框表示远离红框的区域。从可视化的特征空间中可以观察到，最近邻的特征彼此接近。本文采用一种基于最近邻的对比学习框架，利用邻域区域之间的语义相似性。因此，跨模态语义对齐得以加强。

基于最近邻的对比学习。算法1 给出了所提出的基于最近邻的对比学习方法的伪代码。在所提出的框架中，从高光谱图像和激光雷达数据中选取一组位于相同位置的样本对。训练期间，对每个样本对进行随机数据增强，以生成查询样本 x^q 和键样本 x^k。它们分别被编码为嵌入 q 和 k。来自同一图像的嵌入定义为正键，来自不同图像的嵌入定义为负键。大量负键嵌入存储在字典 $\{k_1, k_2, k_3, \ldots\}$中，而一个正键 k_+则单独存储。此外，我们随机选取 q 的一些最近邻以生成嵌入，记为 k_{n+}。接着，在一个小批量中，将一半的正键 k_+替换为 k_{n+} ，以形成新的正键。因此，最近邻起到了微小语义扰动的作用。在我们的

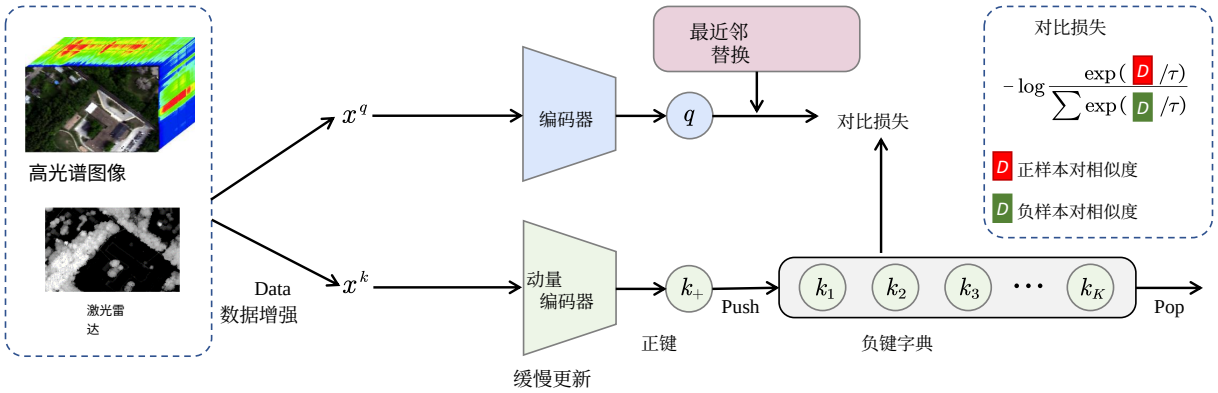


图2. 基于最近邻的对比学习示意图。该方法由三个部分组成：1）基于最近邻的数据增强。对输入样本进行随机数据增强，以生成查询样本和键样本。在一个小批量中，将一半的正键样本替换为其最近邻，从而形成新的正键样本。这些最近邻可作为微小的语义扰动。2）基于双线性注意力的特征编码器。将查询样本和键样本输入编码器进行特征提取。采用双线性注意力融合模块来捕获多源数据之间的二阶特征交互。3）对比损失计算。正键和负键存储在动态字典中，并通过计算对比损失，为正键分配高分，为负键分配低分。

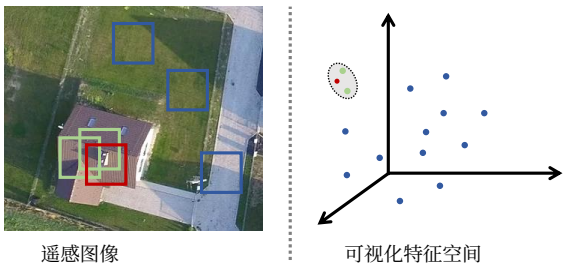


图3. 遥感图像中的典型区域及其对应的可视化特征。红色方框内的区域是屋顶，绿色方框表示其最近邻。蓝色方框表示距离红色方框较远的区域。在可视化特征空间中，可以观察到最近邻的特征彼此接近。因此，上下文信息对于遥感图像分类中的对比学习至关重要。

实现中，最近邻指的是与 x^q 的重叠面积大于 80% 的区域。

我们计算 q 与各个键（包括正键和负键）之间的余弦相似度。然后，将结果存储为 $\{\mathcal{D}_+, \mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \ldots, \mathcal{D}_K\}$。其中，$\mathcal{D}_+$ 是 q 与正键 k_+之间的相似度。其余的是 q 与负键之间的相似度。K 是负键的数量。

对比学习的目标是促使查询与正键匹配，并与负键保持较大距离。具体来说，当 q 与正键 k_+ 相似且与所有负键都不相似时，对比损失的值较低。因此，对比损失函数设计如下：

$$\mathcal{L} = -\log \frac{\exp(\mathcal{D}_+/\tau)}{\sum_{i=1}^{K} \exp(\mathcal{D}_i/\tau)} \tag{1}$$

其中，τ 是温度超参数。直观来说，Softmax 分类器和交叉熵损失可以结合到上述公式中。

字典更新与移动平均编码器。与 MoCo 类似 [22], 我们将字典维护为一个队列，其中其中存储了许多小批量的负样本。字典中的负样本会逐步更新。具体而言，在训练集期间，当一个新的小批量入栈到字典中时，最旧的小批量会被移除。字典的长度可灵活设为一个超参数。

此外，字典编码器的参数更新较慢。与 MoCo 类似，我们为键样本使用一个独立的移动平均编码器。在训练集上，不对键编码器执行反向传播。键编码器的参数按如下方式更新：

$$\theta_k = r\theta_k + (1-r)\theta_q, \tag{2}$$

其中，θ_k 表示键编码器的参数，θ_q表示查询编码器的参数。r 是一个动量系数，用于控制键编码器的更新速度。在训练过程中，只有 θ_q 通过反向传播进行更新。在我们的实现中，r 设为 0.9，因为缓慢演变的键编码器对于稳健的特征学习至关重要。

打乱批归一化。编码器中采用批归一化（BN），以加快收敛速度并提高网络的泛化能力。与 MoCo 类似，我们采用打乱批归一化，以获得更好的特征表示。具体而言，我们会打乱当前小批量中样本的顺序，以供键编码器使用。查询编码器所用小批量中的样本顺序则保持不变。

B. 基于双线性注意力的多源编码器 *er*

在本研究中，对比学习旨在生成一个预训练模型，该模型可用于分类任务。为了获得较高的分类准确率，编码器是对比学习框架中的关键部分。我们设计了一种用于高光谱图像和激光雷达特征建模的多源编码器，如图 4所示。该编码器包含三个部分：高光谱图像特征提取、激光雷达特征提取和双线性注意力融合。

编码器的层类型、卷积核大小和输出尺寸的详细摘要如表I所示。

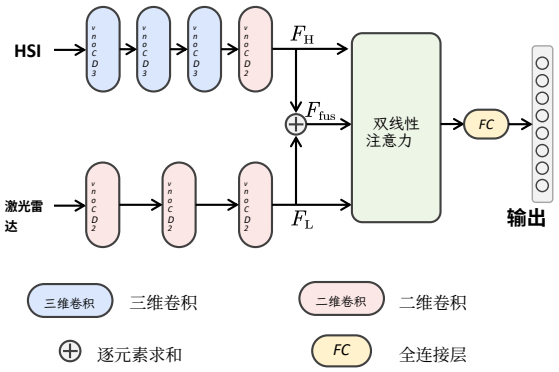


图 4. 基于双线性注意力的多源编码器。

表 I 所提出的多源编码器摘要

高光谱图像特征提取子网络			
#	层类型	卷积核数量@尺寸	输出尺寸
	输入	—	(11, 11, 30, 1)
1	三维卷积	8@3×3×9	(9, 9, 22, 8)
2	三维卷积	16@3×3×7	(7, 7, 16, 16)
3	三维卷积	32@3×3×5	(5, 5, 12, 32)
4	重塑	—	(5, 5, 384)
5	二维卷积	256@3×3	(5, 5, 256)
6	重塑	—	(25, 256)
激光雷达特征提取子网络			
#	层类型	卷积核数量@尺寸	输出尺寸
	输入	—	(11, 11, 1)
1	二维卷积	64@3×3	(9, 9, 64)
2	二维卷积	128@3×3	(7, 7, 128)
3	二维卷积	256@3×3	(5, 5, 256)
4	重塑	—	(25, 256)

高光谱数据采用了与 HybridSN 类似的网络 [41]，该网络使用三维卷积和二维卷积进行特征提取。通过三个三维卷积层和一个二维卷积层提取高光谱图像特征 F_H。与此同时，使用三个二维卷积层生成激光雷达特征F_L。接下来，F_H 和 F_L 组合形成融合特征。使用二维卷积进行特征嵌入。然后，融合特征 F_{fus} 与 F_H 和 F_L具有相同的维度。为有效减少高光谱图像中固有的冗余，从而减少分类过程中需要处理的数据量，使用主成分分析（PCA）为高光谱图像特征提取选取最佳的 30 个光谱波段。

最后，将 F_H、F_L 和 F_{fus} 分别作为$\mathbf{Q}$、$\mathbf{K}$和$\mathbf{V}$输入双线性注意力融合模块。该模块的输出被送入全连接层，以生成用于分类的最终特征。

C. 双线性注意力融合模块

注意力机制在深度神经网络领域取得了重要突破，并已成功应用于

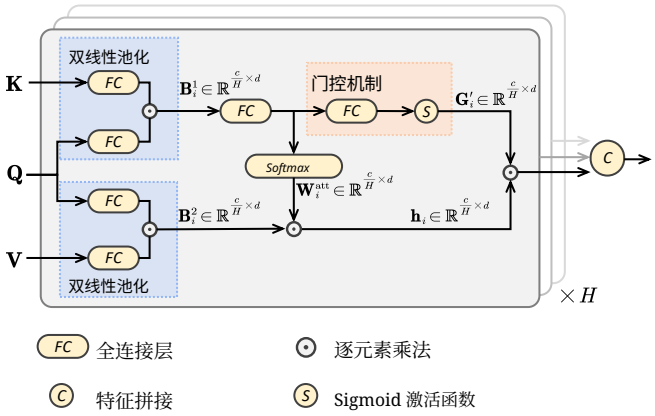


图5. 双线性注意力融合模块。它可以捕获多源数据之间的二阶特征交互。

跨模态任务（例如视觉问答 [42], 图像描述 [43],和图文匹配 [44]）。这促使近期方法采用注意力机制，触发多模态遥感数据之间的交互 [45] [46] [47] [48]。在传统注意力机制中，注意力权重通过对输入进行线性融合来估计。然而，我们认为，传统注意力机制利用的是一阶特征交互，因此在复杂的多源特征推理方面存在局限。

为此，我们提出了一种双线性注意力融合模块，以利用高光谱数据与激光雷达数据之间的二阶特征交互。如图5所示，该模块主要包含两个部分：多头双线性注意力和门控机制。

多头双线性注意力。假设查询$\mathbf{Q} \in \mathbb{R}^{c\times d}$、键$\mathbf{K} \in \mathbb{R}^{c\times d}$和值$\mathbf{V} \in \mathbb{R}^{c\times d}$，其中$d$ 表示特征维度，c 表示通道数。为了增强特征表示能力，我们采用多头方案，从不同子空间对特征交互进行建模，如下所示：

$$\mathbf{h}_i = \text{BiAttention}(\mathbf{Q}_i, \mathbf{K}_i, \mathbf{V}_i), \tag{3}$$

其中，$\mathbf{h}_i$ 表示第 i 个头的输出，BiAttention 表示双线性注意力。头的数量记为 H。

双线性注意力首先将 $\mathbf{Q}_i \in \mathbb{R}^{\frac{c}{H}\times d}$ 和 $\mathbf{K}_i \in \mathbb{R}^{\frac{c}{H}\times d}$ 映射到一个联合空间中：

$$\mathbf{B}_i^1 = \sigma(\mathbf{Q}_i\mathbf{W}_q^1) \odot \sigma(\mathbf{K}_i\mathbf{W}_k), \tag{4}$$

其中，$\mathbf{W}_q^1 \in \mathbb{R}^{d\times d}$ 和 $\mathbf{W}_k \in \mathbb{R}^{d\times d}$ 是权重矩阵，σ 表示 ReLU 激活，$\odot$ 表示逐元素乘法。因此，$\mathbf{B}_i^1 \in \mathbb{R}^{\frac{c}{H}\times d}$ 表示查询$\mathbf{Q}_i$ 与键 $\mathbf{K}_i$之间的二阶表示。

类似地，我们计算$\mathbf{Q}_i$ 与$\mathbf{V}_i$ 之间的双线性表示：

$$\mathbf{B}_i^2 = \sigma(\mathbf{Q}_i\mathbf{W}_q^2) \odot \sigma(\mathbf{V}_i\mathbf{W}_v), \tag{5}$$

其中，$\mathbf{W}_q^2 \in \mathbb{R}^{d\times d}$ 和$\mathbf{W}_v \in \mathbb{R}^{d\times d}$ 为权重矩阵。$\mathbf{B}_i^2 \in \mathbb{R}^{\frac{c}{H}\times d}$ 表示查询$\mathbf{Q}_i$ 与值$\mathbf{V}_i$之间的二阶表示。

接下来，双线性表示$\mathbf{B}_i^1$通过线性层和 Softmax 层投影为注意力权重$\mathbf{W}_i^{\text{att}} \in \mathbb{R}^{\frac{c}{H} \times d}$，如下所示：

$$\hat{\mathbf{B}}_i^1 = \sigma(\mathbf{W}_B \mathbf{B}_i^1), \tag{6}$$

$$\mathbf{W}_i^{\text{att}} = \text{softmax}(\hat{\mathbf{B}}_i^1), \tag{7}$$

其中，$\mathbf{W}_B \in \mathbb{R}^{\frac{c}{H} \times \frac{c}{H}}$是权重矩阵。接下来，通过增强注意力权重，得到注意力特征$\mathbf{h}_i \in \mathbb{R}^{\frac{c}{H} \times d}$，如下所示：

$$\mathbf{h}_i = \mathbf{W}_i^{\text{att}} \odot \mathbf{B}_i^2 \tag{8}$$

门控机制。上述双线性注意力利用了$\mathbf{Q}_i$、$\mathbf{K}_i$和$\mathbf{V}_i$之间的特征交互。然而，查询和键中可能包含噪声信息。为自适应地增强有用部分并抑制无用部分，我们设计了一种门控机制。具体而言，对于第 i个头，将$\hat{\mathbf{B}}_i^1$输入线性层，再通过Sigmoid函数处理，以计算权重掩码$\mathbf{G}_i \in \mathbb{R}^{\frac{c}{H} \times 1}$，如下所示：

$$\mathbf{G}_i = \text{sigmoid}(\hat{\mathbf{B}}_i^1 \mathbf{W}_{B'}), \tag{9}$$

其中，$\mathbf{W}_{B'} \in \mathbb{R}^{d \times 1}$是权重矩阵。接下来，将$\mathbf{G}_i$扩展为$\mathbf{G}'_i \in \mathbb{R}^{\frac{c}{H} \times d}$。然后，应用得到的门控掩码来控制$\mathbf{h}_i \in \mathbb{R}^{\frac{c}{H} \times d}$的信息流，如下所示：

$$\hat{\mathbf{h}}_i = \mathbf{G}'_i \odot \mathbf{h}_i. \tag{10}$$

最后，将多个头的结果拼接起来，即可得到多源数据的融合表示。在本文中，$\mathbf{Q}$、$\mathbf{K}$ 和$\mathbf{V}$ 的大小为 25×256。头数 H issetto5。

IV. 实验结果与分析

为验证所提出的 NNCNet 的有效性，我们在四个广泛使用的基准数据集上进行了大量实验：休斯敦2013数据集、特伦托数据集、MUUFL数据集和休斯敦2018数据集。我们首先将所提出的 NNCNet 与最先进的方法进行比较。随后，我们进行了补充评估，以考察所提方法各组成部分的有效性。

A. 数据集和评估指标

休斯敦2013数据集：该数据集由国家机载激光测绘中心采集，并作为一项挑战赛用于2013年GRSS数据融合竞赛。高光谱图像由CASI传感器采集（144个光谱波段，分辨率为2.5米）。此外，还提供了配准后的激光雷达数据，其分辨率相同。共计15029个地面真实样本，分布于15个类别中。这些样本被划分为训练集和测试集，分别包含2832和12197个像素。我们采用标准的训练集和测试集，表II 列出了训练样本和测试样本的数量。

特伦托数据集：该数据集采集于意大利特伦托以南的一个乡村地区。高光谱图像由 63 个波段组成，波长范围为 0.42-0.99 μm。数据集大小为 166×660 像素，空间分辨率为 1.0 m。共分布有 30214 个地面真实样本

表 II 休斯敦2013数据集的训练集与测试集样本分布。

No.	类别名称	训练集	Test
1	健康草地	198	1053
2	受胁迫草地	190	1064
3	人工草地	192	505
4	Tree	188	1056
5	Soil	186	1056
6	水体	182	143
7	住宅区	196	1072
8	商业区	191	1053
9	Road	193	1059
10	高速公路	191	1036
11	铁路	181	1054
12	停车场1	192	1041
13	停车场2	184	285
14	网球场	181	247
15	跑道	187	473
	总计	2832	12197

表III 特伦托数据集的训练集与测试集样本分布。

No.	类别名称	训练集	Test
1	苹果树	129	3905
2	建筑物	125	2778
3	地面	105	374
4	Wood	154	8969
5	葡萄园	184	10317
6	道路	122	3052
	总计	819	29595

表 IV MUUFL数据集的训练集与测试集样本分布。

No.	类别名称	训练集	Test
1	树木	150	23096
2	以草地为主	150	4120
3	混合地表	150	6732
4	泥土和沙地	150	1676
5	Road	150	6537
6	水体	150	316
7	建筑物阴影	150	2083
8	建筑物	150	6090
9	人行道	150	1235
10	黄色路缘	150	33
11	布料面板	150	119
	总计	1650	52037

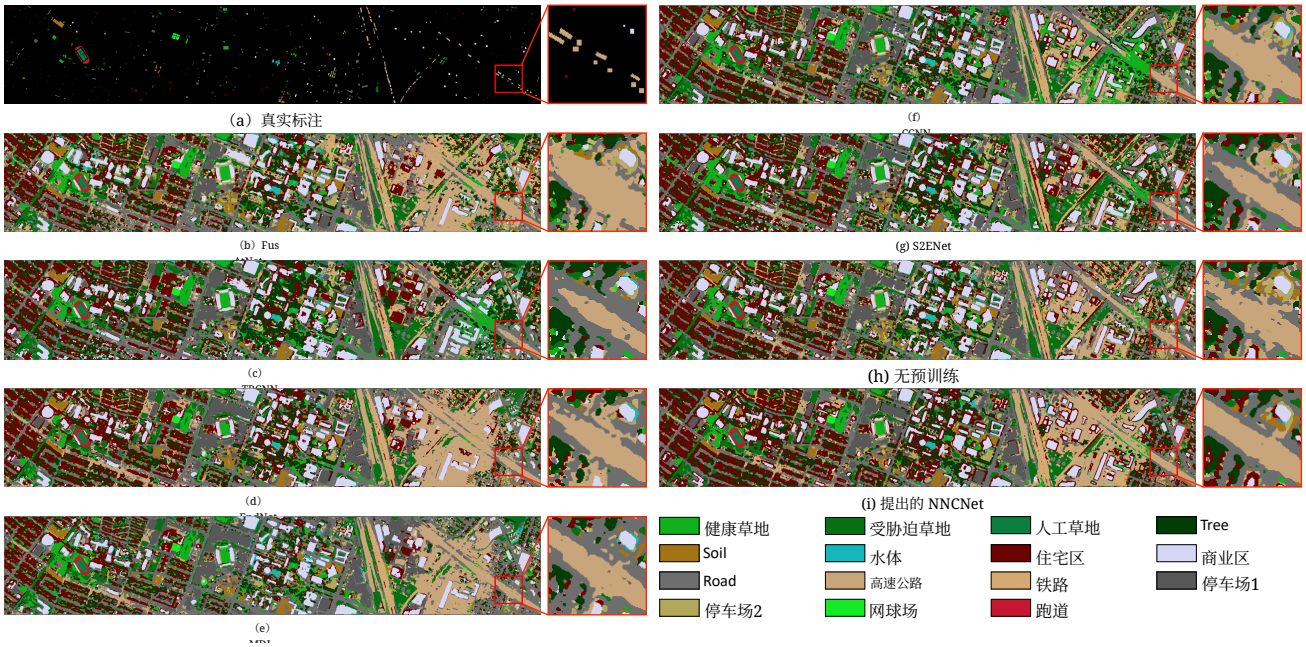


图 6. 休斯敦2013数据集的分类图。(a) 真实标签。(b) FusAtNet。(c) TBCNN。(d) EndNet。(e) MDL。(f) CCNN。(g) S2ENet。(h) 未经预训练的提出的NNCNet。(i) 提出的 NNCNet。

表 V 休斯敦2018数据集的训练集与测试集样本分布。

No.	类别名称	训练集	Test
1	健康草地	500	9299
2	受胁迫草地	500	32002
3	人工草坪	68	616
4	常绿树	500	13095
5	落叶树	500	4521
6	裸地	451	4065
7	水体	26	240
8	住宅建筑	500	39272
9	非住宅建筑	500	223252
10	道路	500	45366
11	人行道	500	33529
12	人行横道	151	1367
13	主干道	500	45848
14	高速公路	500	9365
15	铁路	500	6437
16	铺装停车场	500	11000
17	未铺装停车场	14	132
18	Cars	500	6047
19	火车	500	4869
20	体育场座椅	500	6324
	总计	8210	496646

于 6 个类别中。表 III 列出了特伦托数据集的训练集和测试集样本分布。

MUUFL数据集：MUUFL数据集采集于2010年11月的南密西西比大学墨西哥湾沿岸校区。HSI包含72个光谱波段，但为进行噪声去除，去除了最前和最后的四个波段，剩余64个波段用于分类。该数据集的总大小为 325×220 个像素。表IV 列出了该数据集可用的训练集和测试集样本。在我们的实验中，预训练阶段使用全部数据；而在训练验证阶段，仅使用训练集中具有标签的部分。

休斯敦2018数据集：该数据集由休斯敦大学高光谱图像分析实验室和国家机载激光测绘中心（NCALM）采集。该数据集最初为2018年IEEE GRSS数据融合竞赛发布。高光谱数据覆盖380–1050 nm光谱范围，包含48个波段，地面采样距离为1.0 m。该数据集共包含 4768×1202个像素，其中一部分被划分为训练集，规模为 2384×601 个像素。表V 列出了休斯敦2018数据集训练样本和测试样本的分布。

模型性能通过总体精度（OA）、平均精度（AA）和Kappa系数进行评估。OA表示模型在所有测试样本中的正确预测数量与样本总数之比。AA表示每个类别中的正确预测数量与该类别样本总数之比，最后取各类别精度的平均值。Kappa表示经校正后的协议百分比，其中校正依据是纯粹由偶然因素所预期的协议数量。

B. 实现细节

所提出的对比学习架构用于生成预训练模型。在对比学习阶段，我们使用 Adam优化器。小批量大小设为 64，学习率设为 0.0005。从数据集中随机裁剪出大小为

表VI 休斯敦2013数据集上的分类准确率（%）

类别	FusAtNet [49]	TBCNN [36]	EndNet [14]	MDL [5]	CCNN [15]	S2ENet [50]	NNCNet（本文方法）
健康草地	79.20	81.01	78.54	83.00	**91.55**	82.72	81.84
受胁迫草地	96.71	97.93	96.33	98.68	**99.72**	100.0	**99.72**
人工草地	97.82	99.60	**100.0**	99.80	99.60	99.60	99.80
Tree	97.63	94.13	88.26	93.94	97.63	95.74	**99.43**
Soil	**100.0**	98.86	**100.0**	99.05	**100.0**	99.81	**100.0**
水体	91.61	97.90	**100.0**	**100.0**	95.80	97.20	**100.0**
住宅区	76.31	80.50	83.02	79.66	83.12	91.23	**94.87**
商业区	74.17	87.46	79.96	80.44	94.49	91.55	**94.78**
Road	89.05	86.50	93.30	84.70	93.20	95.94	**96.03**
高速公路	92.86	64.86	92.28	94.88	89.96	84.75	**99.81**
铁路	94.21	93.74	85.86	85.67	96.39	94.31	**99.34**
停车场1	87.32	74.93	**99.81**	98.75	99.71	97.79	**99.81**
停车场2	84.21	85.96	83.16	82.46	89.82	89.47	**90.88**
网球场	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
跑道	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
OA	89.70	87.57	90.71	90.80	94.98	93.99	**96.77**
AA	90.73	89.55	92.03	92.06	95.40	94.67	**97.06**
Kappa	88.81	86.50	89.92	90.01	94.56	93.48	**96.49**

表VII 特伦托数据集上的分类准确率（%）

类别	FusAtNet [49]	TBCNN [36]	EndNet [14]	MDL [5]	CCNN [15]	S2ENet [50]	NNCNet（我们的方法）
苹果树	99.95	99.87	**99.90**	**99.90**	**99.90**	**99.90**	99.13
建筑物	98.92	98.81	99.03	**99.10**	**99.10**	98.88	98.92
地面	85.56	81.02	85.83	86.36	86.90	86.36	**99.73**
Wood	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
葡萄园	99.68	98.40	99.31	99.61	99.67	99.21	**100.0**
道路	**92.07**	89.35	90.83	91.12	91.25	91.32	91.88
OA	98.77	97.96	98.52	98.66	98.71	98.53	**98.92**
AA	96.03	94.57	95.81	96.01	96.13	95.94	**98.26**
Kappa	98.35	97.27	98.01	98.21	98.27	98.03	**98.55**

11×11 像素的图像块作为训练样本。

获得预训练模型后，使用数据集中的训练样本对模型进行微调。在微调阶段，小批量大小设为 128，优化器的设置与对比学习阶段相同。

C. 分类准确率与讨论

提出的 NNCNet 在休斯敦2013、特伦托、MUUFL 和 Houston 2018 数据集上实现。为验证提出的 NNCNet 的有效性，我们将其与六种最先进的方法进行比较，包括 FusAtNet [49],TBCNN [36], EndNet [14], MDL [5],CCNN [15], 和 S2ENet [50]。具体而言，FusAtNet [49] 通过交叉注意力利用高光谱图像和激光雷达特征，并结合注意力加权的光谱与空间表征，以计算模态特定的特征嵌入。TBCNN [36]采用双分支卷积神经网络，用于高光谱图像和激光雷达特征提取。在 EndNet [14], 中，采用深度编码器 – 解码器网络进行多模态信息融合与分类。MDL [5]提出了一种通用的多模态深度学习框架。CCNN [15]提出了一种用于多模态信息融合的耦合网络。该方法整合特征级融合与决策级融合，以实现异构特征表征。S2ENet [50]是一种空谱增强网络，可同时改善空间和光谱特征表示。为确保比较公平，所有参与比较的方法均采用其论文中提供的默认参数。

表 VI展示了休斯敦2013数据集上的分类结果。提出的 NNCNet 在总体精度、平均精度和 Kappa 系数方面均取得了最佳表现。与竞争方法（CCNN 和 S2ENet）相比，我们的 NNCNet 在总体精度上分别高出 1.78% 和 2.78%。这表明，所提出的自监督框架能够有效地建模多源样本之间的相关性。此外，“高速公路”类别的准确率（99.81%）显著

表八 MUUFL数据集上的分类准确率（%）

类别	FusAtNet [49]	TBCNN [36]	EndNet [14]	MDL [5]	CCNN [15]	S2ENet [50]	NNCNet（本文方法）
树木	**95.31**	91.18	90.86	90.95	92.40	93.91	93.09
以草地为主	79.83	83.98	83.30	84.54	83.52	**88.28**	86.82
混合地表	83.69	83.72	84.27	83.01	84.34	81.85	**86.29**
泥土和沙地	**97.73**	96.12	96.00	96.42	96.72	97.32	96.18
Road	84.86	91.23	91.11	90.44	91.68	91.28	**92.35**
水体	**99.68**	**99.68**	**99.68**	**99.68**	**99.68**	**99.68**	**99.68**
建筑物阴影	83.01	**92.85**	92.61	92.75	92.61	88.29	92.75
建筑物	94.70	96.86	**96.90**	96.70	96.80	95.99	96.21
人行道	89.80	87.85	88.34	87.85	89.23	88.50	**91.34**
黄色路缘	87.88	**90.91**	**90.91**	**90.91**	**90.91**	87.88	84.85
布料面板	**99.16**	**99.16**	**99.16**	**99.16**	**99.16**	**99.16**	**99.16**
OA	90.68	90.53	90.39	90.27	91.20	91.61	**92.07**
AA	90.51	92.14	92.10	92.03	92.45	92.01	**92.61**
Kappa	87.65	87.59	87.42	87.26	88.44	88.93	**89.56**

表 IX 休斯敦2018数据集上的分类准确率（%）

类别	FusAtNet [49]	TBCNN [36]	EndNet [14]	MDL [5]	CCNN [15]	S2ENet [50]	NNCNet（本文方法）
健康草地	89.40	90.91	89.55	**93.99**	93.39	91.29	93.36
受胁迫草地	90.65	88.83	89.51	88.95	90.84	**91.97**	91.95
人工草坪	98.54	84.90	75.97	**98.86**	98.38	96.59	98.38
常绿树	85.30	71.97	67.97	90.60	**94.25**	88.91	92.00
落叶树	73.15	70.87	66.98	76.55	**80.62**	79.12	76.86
裸地	**100.0**	99.78	**100.0**	**100.0**	99.48	99.78	99.98
水体	**99.17**	96.67	92.92	98.75	95.83	93.75	96.25
住宅建筑	**97.29**	94.93	92.54	86.31	91.43	91.31	87.58
非住宅建筑	94.36	95.78	96.78	**97.75**	93.93	95.22	97.04
道路	62.29	53.26	42.71	69.65	**73.14**	70.95	71.25
人行道	64.00	72.67	71.00	68.30	**78.85**	76.82	70.63
人行横道	40.53	41.84	03.66	49.82	52.38	**56.18**	38.33
主干道	69.77	78.48	71.08	60.56	76.08	76.25	**81.58**
高速公路	97.16	98.55	96.11	96.18	**98.70**	98.24	98.54
铁路	99.43	99.19	98.91	97.84	99.52	99.67	**99.94**
铺装停车场	85.68	78.73	75.27	82.50	87.32	91.12	**95.25**
未铺装停车场	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
Cars	56.13	72.65	24.77	47.87	89.37	77.11	**92.16**
火车	91.29	65.06	60.67	90.37	76.38	92.75	**99.01**
体育场座椅	99.62	99.49	99.34	99.59	**99.83**	99.65	99.78
OA	85.98	86.33	83.84	86.70	88.64	88.87	**89.89**
AA	84.68	82.72	75.78	84.72	88.48	88.33	**88.99**
Kappa	81.46	81.83	78.06	82.15	85.19	85.38	**86.65**

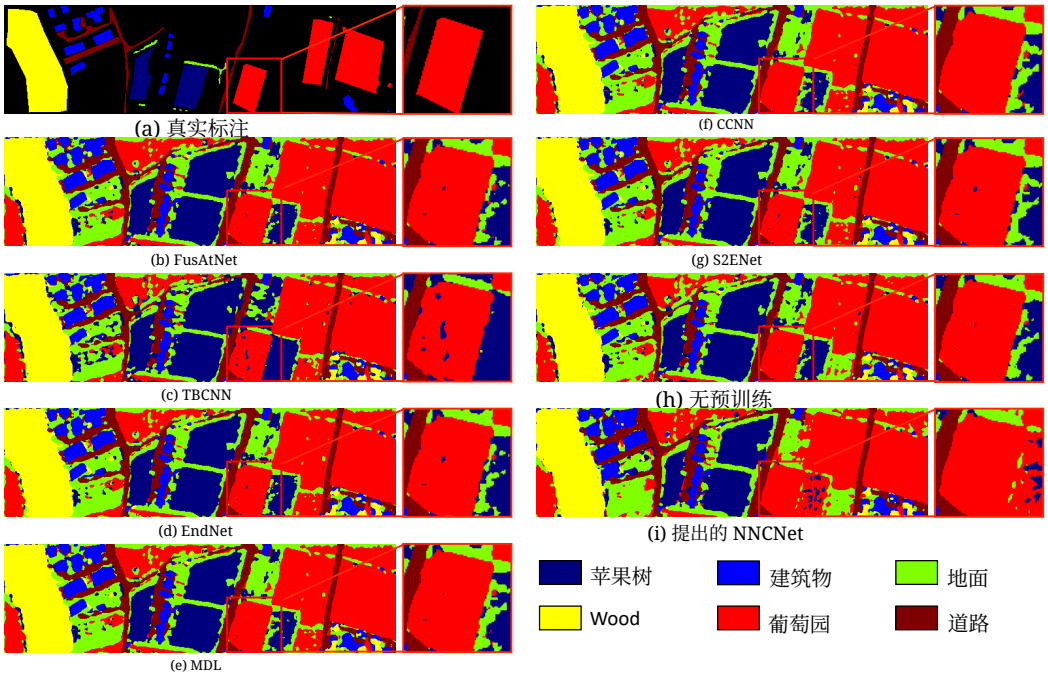


图7. 特伦托数据集的分类图。(a) 真实标签。(b) FusAtNet。(c) TBCNN。(d) EndNet。(e) MDL。(f) CCNN。(g) S2ENet。(h) 未进行预训练的提出的 NNCNet。(i) 提出的 NNCNet。

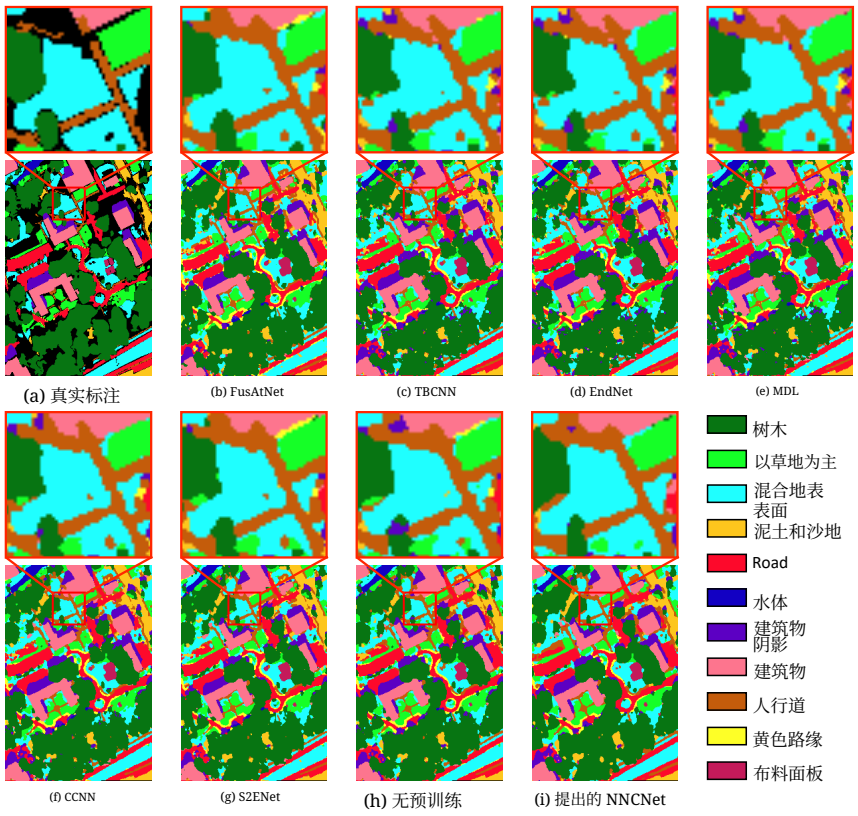


图8. MUUFL数据集的分类图。(a) 真实标签。(b) FusAtNet。(c) TBCNN。(d) EndNet。(e) MDL。(f) CCNN。(g) S2ENet。(h) 未进行预训练的提出的 NNCNet。(i) 提出的 NNCNet。

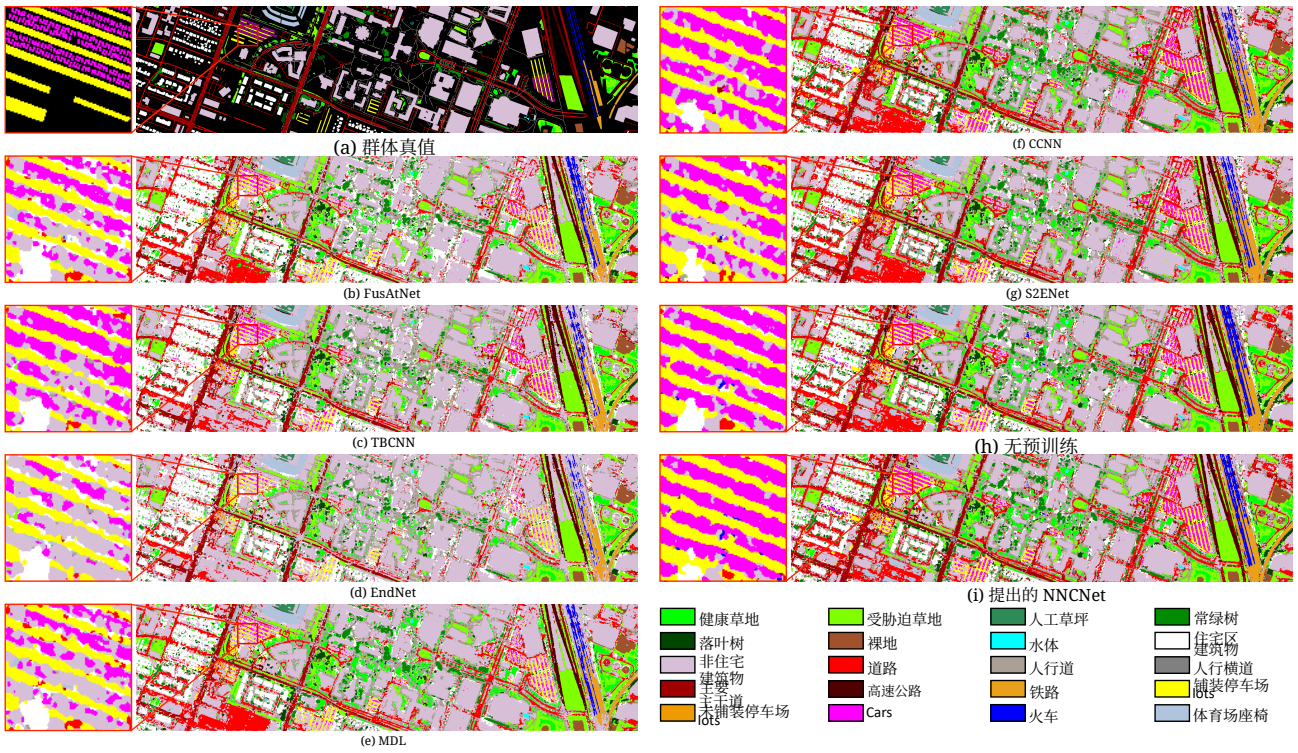


图9. 休斯敦2018数据集的分类图。(a) 真实标签。(b) FusAtNet。(c) TBCNN。(d) EndNet。(e) MDL。(f) CCNN。(g) S2ENet。(h) 未进行预训练的提出的 NNCNet。(i) 提出的 NNCNet。

得益于我们的 NNCNet，分类结果得到了改进。休斯敦2013数据集中存在许多无标签的高速公路区域。因此，我们的 NNCNet 通过从无标签数据中进行对比学习，捕获了高速公路的纹理和光谱特征。分类图如图 6所示。可以观察到，在未进行预训练的情况下，一些高速公路区域被错误地分类为道路。相比之下，所提出的 NNCNet 通过对比学习取得了更好的效果。

表 VII展示了不同方法在特伦托数据集上的分类结果。分类图如图7所示。可以看出，在未进行预训练的情况下，一些葡萄园区域被错误地分类为苹果树。此外，所提出的 NNCNet 在总体精度、平均精度和 Kappa 方面取得了最佳性能。所提方法在“地面”类别上获得了最佳总体精度。该类别中只有少量标注数据，但其仍占据了整幅图像的较大部分。显然，在训练样本有限的情况下，我们的 NNCNet 能够学习到稳健的特征表示。

表VIII 展示了不同方法在 MUUFL 数据集上的分类结果。与其他方法相比，所提出的 NNCNet 性能最佳。具体而言，所提方法的总体精度最高（92.07%），并且在五个类别（混合地表、道路、水体、人行道和布料面板）上取得了最高精度。所提方法对主要为建筑物和建筑物阴影的分类结果也颇具竞争力。因此，这些比较结果证明了所提出的 NNCNet 在 MUUFL 数据集上的优越性能。图中展示了所提出的 NNCNet 在预训练和未预训练情况下的分类图

图 8，可以观察到，预训练有效提升了分类性能。

表IX展示了不同方法在休斯敦2018数据集上的分类结果。与其他方法相比，提出的 NNCNet 表现最佳。尤其是在“汽车”和“铺装停车场”类别上，我们的方法分别达到 92.16% 和 95.25%，远超其他方法。分类图如图 9所示。可以看出，其他方法对汽车分类的结果不够平滑，而提出的 NNCNet 能够清晰地描绘汽车和铺装停车场的边界。显然，提出的 NNCNet 具备强大的细粒度特征表示能力。

我们发现，提出的 NNCNet 在休斯敦2013数据集和休斯敦2018数据集上的表现远优于其在特伦托和 MUUFL 数据集上的表现。我们认为，这是因为这两个数据集的图像分辨率更高（348×1905 和 2384×601 像素）。因此，提出的 NNCNet 可以通过对比学习，在较大的数据集上获得更好的特征表示。由此，我们认为，提出的 NNCNet 在实际应用中能够取得更好的分类结果，因为实际应用中可以获取更多无标签数据。

D. 消融研究

为了评估 NNCNet 中不同组件的有效性，我们进行了一系列消融研究。通过一系列消融实验验证了各个所提模块在提高分类准确率方面的有效性，具体实验结果列于表 X。

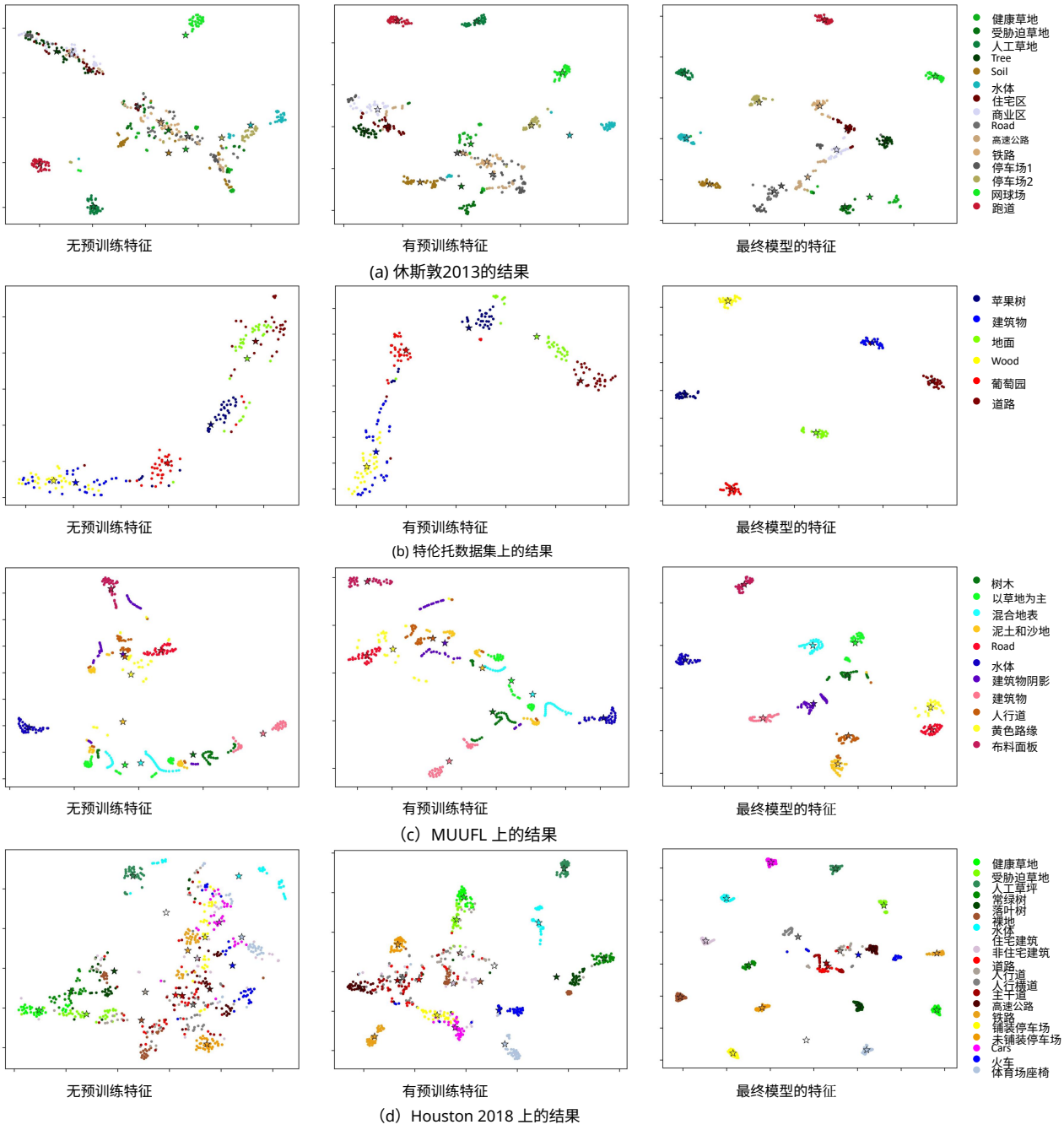


图10. 不同数据集上的特征可视化。(a) 休斯敦2013数据集上的结果。(b) 特伦托数据集上的结果。(c) MUUFL数据集上的结果。(d) 休斯敦2018数据集上的结果。第一列表示无预训练特征，第二列表示有预训练特征，最后一列表示最终模型的特征。星号表示每类特征的聚类中心。

预训练和最近邻学习的有效性。我们采用一个未经预训练、不含双线性注意力融合和最近邻对比学习的普通卷积神经网络作为基线模型。如表 X所示，与基线模型相比，预训练在四个数据集上均能在一定程度上有效提升分类性能。这表明我们的预训练方案能够提供有助于提高分类准确率的参数初始化。

我们进一步考察了基于最近邻的对比学习方案。如表X所示，采用最近邻学习的模型显著提升了分类性能。这是因为考虑了邻域区域的语义相似性，并增强了模态间的语义对齐。

为了进一步证明预训练和最近邻学习的有效性，我们在图10中将预训练前后的特征进行了可视化。我们分别可视化了未进行预训练和进行预训练时的特征，以及最终模型中的特征。在休斯敦2013、Houston 2018和特伦托数据集上，我们发现预训练后，同一类别的特征彼此分布得更加接近，而

表X 所提出模型不同变体在不同数据集上的性能比较

变体	预训练	双线性注意力	Gate 机制	最近邻	休斯敦2013	特伦托	MUUFL	Houston 2018
1	✖	✖	✖	✖	95.20	98.74	91.38	88.21
2	✔	✖	✖	✖	95.57	98.80	91.60	88.72
3	✔	✖	✖	✔	96.30	98.88	91.83	89.41
4	✔	✔	✖	✖	95.64	98.86	91.68	88.79
5	✔	✔	✖	✔	96.47	98.90	92.01	89.76
6	✔	✔	✔	✖	95.84	98.86	91.68	88.83
7	✔	✔	✔	✔	96.77	98.92	92.07	89.89

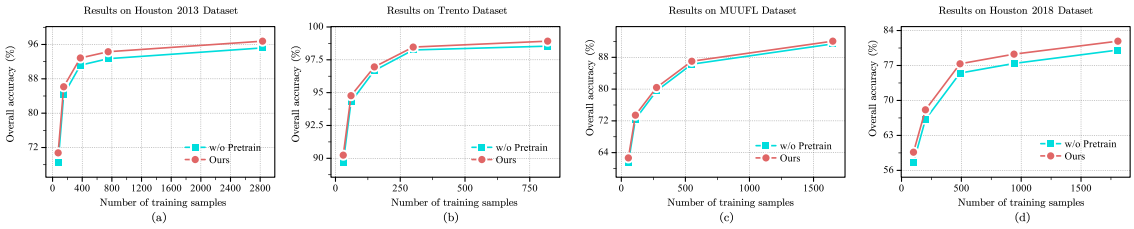


图 11. 不同样本数量下的分类准确率。

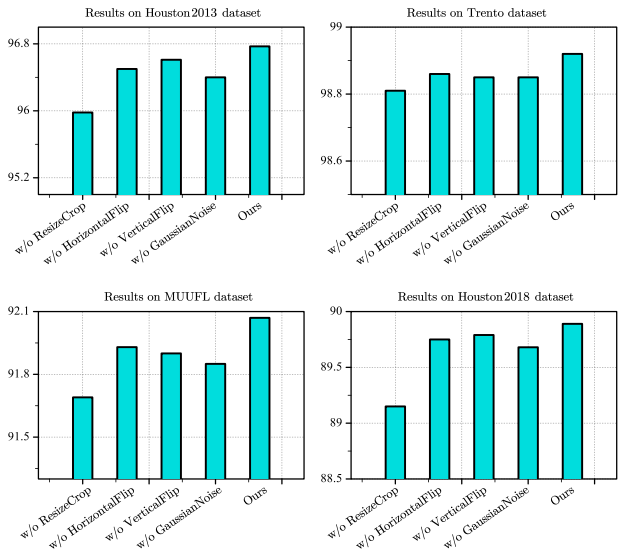


图 12. 使用不同数据增强方法时模型的性能比较。

不同类别的特征则彼此远离。显然，我们的无监督框架在休斯敦2013和特伦托数据集上是有效的。此外，我们观察到，在MUUFL数据集上，预训练后的特征并未得到显著改善。其原因可能是，休斯敦2013、Houston 2018和特伦托数据集中包含更多无标签数据。这些无标签数据在对比学习中发挥着关键作用。因此，在拥有更多无标签数据时，所提出的对比学习框架表现更好。在实际应用中，当有大量无标签数据可用时，该方法更加便捷。

训练样本数量。自监督学习策略的优势之一是，在训练样本数量较少时仍具有出色的性能。因此，我们尝试在训练过程中逐步减少样本数量，结果如图 11 所示。在休斯敦2013数据集上，仅使用 375 个训练样本（每个类别 25 个样本）时，所提方法的总体精度值为 91.86，结果令人满意且鼓舞人心。此外，在使用较小训练集时，经过预训练的模型在四个数据集上的表现始终优于未经预训练的模型。显然，所提出的 NNCNet 的对比学习策略对于小型训练集尤其有效。此外，我们观察到，预训练在休斯敦2013和2018数据集上带来的性能提升优于在特伦托和 MUUFL 数据集上的提升。如前所述，休斯敦2013和2018数据集包含更多无标签数据。因此，所提出的最近邻策略能够在这两个数据集上利用丰富的特征表示。

数据增强的有效性。数据增强的目的是扩大正样本与负样本之间的差异，从而促进编码器的训练。在提出的 NNCNet 中，我们采用了四种数据增强方案，包括 RandomResizedCrop、RandomHor- izontalFlip、RandomVerticalFlip 和 RandomGaussianNoise。相应结果如图所示。12。我们发现，RandomResizedCrop 是数据增强的关键。由于图像块被裁剪为 11×11 像素，如果比例设置得太小，语义信息就很容易受损。因此，在我们的实现中，比例设为 (0.7, 1)。

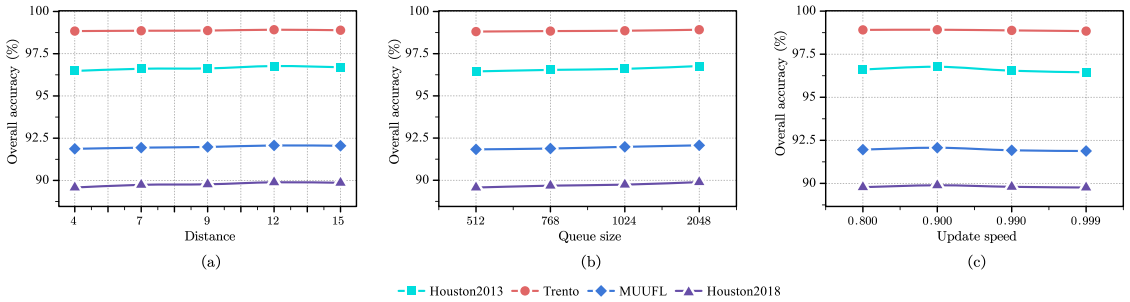


Fig. 13. 不同数据集上不同空间距离、队列大小和小批量大小对应的分类准确率。

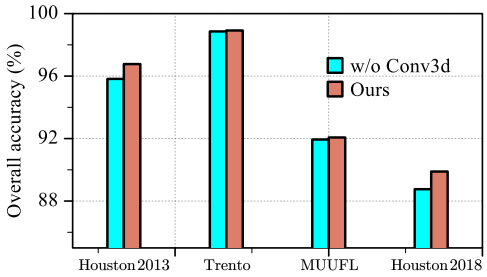


图14. 使用或不使用三维卷积时我们的方法的性能比较。

E. 参数敏感性

正负样本之间的最小空间距离。为避免正负样本之间的相似度过高，我们定义了它们之间的最小距离 s （即正负样本之间的距离需要大于 s）。结果如图 13(a) 所示。在我们的实现中，每个样本的大小为 11×11 像素。当 $4 \leqslant s \leqslant 12$时，分类性能略有提升。采用较大的距离有助于增加正负样本之间的差异。因此，在我们的实现中， s 设为 12。

负键字典大小。图13(b)展示了负键字典大小对分类性能的影响。实验表明，增大字典大小有助于预训练，这与我们先前的假设一致。我们认为，所提方法在有更多无标签数据可用时效果更好。

键编码器更新速度。我们测试了不同的键编码器更新速度 r 在预训练期间的表现。实验结果如图13(c)所示。我们发现，当 r 设为0.9时，分类性能最佳。

三维卷积的有效性。受 HybridSN 启发[41], , 我们首先使用主成分分析进行通道降维。随后，结合三维卷积和二维卷积进行特征提取。为验证三维卷积的有效性，我们设计了一个将三维卷积替换为二维卷积的网络（图14中的“w/o Conv2d”）。实验结果如图14所示。我们发现三维卷积可以在一定程度上提升分类性能。尽管主成分分析会破坏高光谱数据的光谱连续性，但我们认为，与二维卷积相比，三维卷积仍能从光谱维度生成更具判别性的特征图。三维卷积生成的这些判别性特征可以提升分类性能。

V. 结论与未来工作

本文提出了一种自监督 NNCNet 模型，用于解决高光谱图像与激光雷达数据的联合分类问题。具体而言，我们将一种基于最近邻的数据增强方案融入对比学习框架，并利用邻域区域之间的语义相似性，从而更准确地捕获跨模态语义对齐。此外，我们提出了一种双线性注意力融合模块，用于捕获高光谱图像与激光雷达数据之间的二阶特征交互。因此，该模块能够有效提升多源数据的上下文表示。在休斯敦2013、特伦托、MUUFL 和 Houston 2018 数据集上的大量实验表明，与多种最先进的方法相比，我们的模型具有优越性。

未来，我们计划明确探索高光谱图像与激光雷达数据之间的语义关系和空间关系。此外，我们还将探索如何进一步增强高光谱图像与激光雷达数据之间的特征交互。



Meng Wang 于2020年获得中国济南暨南大学计算机科学理学学士学位。目前，他正在中国青岛中国海洋大学信息科学与技术学院攻读计算机科学与应用遥感理学硕士学位。他的

计算机视觉和遥感图像处理includ
当前研究 兴趣



Feng Gao（IEEE 会员）于 2008 年在中国重庆的重庆大学获得软件工程学士学位，并于 2015 年在中国北京的北京航空航天大学获得计算机科学与技术哲学博士学位。

他现任中国海洋大学信息科学与工程学院副教授。研究方向包括遥感图像分析、模式识别和机器学习。



Junyu Dong（IEEE会员）于1993年和1999年分别获得中国海洋大学应用数学系理学学士和理学硕士学位；2003年获得英国爱丁堡赫瑞瓦特大学计算机科学系图像处理方向哲学博士学位。

他现任中国海洋大学计算机科学与技术学院教授、院长。研究兴趣包括视觉信息分析与理解、机器学习和水下图像处理。



Heng-Chao Li（IEEE高级会员）于2001年和2004年分别获得西南交通大学理学学士和理学硕士学位，并于2008年获得中国科学院研究生院哲学博士学位。

他目前是西南交通大学信息科学与技术学院正教授。其研究兴趣包括合成孔径雷达（SAR）图像的统计分析、遥感图像处理和模式识别。

李博士是《西南交通大学学报》和《雷达学报》的编委。他是IEEE应用地球观测与遥感专题选刊的副编辑。



Qian Du（IEEE会士）于2000年在美国马里兰州巴尔的摩的马里兰大学巴尔的摩分校获得电气工程博士学位。

她目前是美国密西西比州立大学电气与计算机工程系Bobby Shackouls教授，任职于美国密西西比州斯塔克维尔。她的研究兴趣包括高光谱遥感图像分析与应用，以及机器学习。

杜博士曾获2010年度最佳审稿人奖，该奖由IEEE地球科学与遥感学会（GRSS）颁发。2009年至2013年，她担任IEEE GRSS数据融合技术委员会联合主席；2010年至2014年，她担任国际模式识别协会遥感与制图技术委员会主席；2012年，她担任第四届IEEE GRSS高光谱图像与信号处理：遥感演进研讨会的总主席，该研讨会在中国上海举行。她曾担任《模式识别》和《IEEE地球科学与遥感汇刊》的副编辑。2016年至2020年，她担任《IEEE应用地球观测与遥感专题选刊》主编。她目前是IEEE期刊评审与咨询委员会以及SPIE出版委员会的成员。她是国际光学与光子学会（SPIE）会士。

参考文献

[1] A. Ma、A. M. Filippi、Z. Wang、Z. Yin、D. Huo、X. Li 和 B. G¨uneralp，“用于基于循环神经网络的高光谱图像分类的快速序列特征提取”，*IEEE*地球科学与遥感汇刊，第 59 卷，第 7 期，第 5920–5937 页，2021年。[2] M. Khodadadzadeh、J. Li、S. Prasad 和 A. Plaza，“利用多特征学习融合高光谱与激光雷达遥感数据”，*IEEE*应用地球观测与遥感精选主题期刊，第 8 卷，第 6 期，第 2971–2983 页，2015年。[3] B. Rasti、P. Ghamisi、J. Plaza 和 A. Plaza，“利用稀疏与低秩成分分析融合高光谱与激光雷达数据”，*IEEE*地球科学与遥感汇刊，第 55 卷，第 11 期，第 6354–6365 页，2017年。[4] X. Zheng、H. Sun、X. Lu 和 W. Xie，“用于高光谱图像分类的旋转不变注意力网络”，*IEEE*图像处理汇刊，第 31 卷，第 4251–4265 页，2022年。[5] D. Hong、L. Gao、N. Yokoya、J. Yao、J. Chanussot、Q. Du 和 B. Zhang，“更多样化意味着更好：多模态深度学习与遥感图像分类相结合”，*IEEE*地球科学与遥感汇刊，第 59 卷，第 5 期，第 4340–4354 页，2021年。[6] L. G´omez-Chova、D. Tuia、G. Moser 和 G. Camps-Valls，“遥感图像的多模态分类：综述与未来方向”，*IEEE*会议录，第 103 卷，第 9 期，第 1560–1584 页，2015年。[7] C. Ge、Q. Du、W. Li、Y. Li 和 W. Sun，“基于核协同表示残差融合的高光谱与激光雷达数据分类”，*IEEE*应用地球观测与遥感精选主题期刊，第 12 卷，第 6 期，第 1963–1973 页，2019年。[8] W. Li、J. Wang、Y. Gao、M. Zhang、R. Tao 和 B. Zhang，“用于高光谱与多光谱数据分类的图特征增强选择分配网络”，*IEEE*地球科学与遥感汇刊，第 60 卷，第 1–14 页，2022年。[9] M. Pedergnana、P. R. Marpu、M. Dalla Mura、J. A. Benediktsson 和 L. Bruzzone，“利用扩展属性剖面对光学遥感与激光雷达数据进行分类”，*IEEE*信号处理精选主题汇刊，第 6 卷，第 7 期，第 856–865 页，2012年。[10] R. Huang 和 J. Zhu，“利用随机森林融合激光雷达数据与高光谱图像进行土地覆盖分类”，载于*IEEE*国际地球科学与遥感研讨会（*IGARSS*），2013年，第 3978–3981 页。[11] C. Demirkesen、M. Teke 和 U. Sakarya，“基于高光谱图像与激光雷达的DEM融合：一种多模态土地利用分类策略”，载于*IEEE*国际地球科学与遥感研讨会（*IGARSS*），2014年，第 2942–2945 页。[12] J. Xia、N. Yokoya 和 A. Iwasaki，“一种利用形态学特征对高光谱与激光雷达数据进行分类的新型集成分类器”，载于*IEEE*国际声学、语音与信号处理会议（*ICASSP*），2017年，第 6185–6189 页。

[13] M. Zhang、W. Li、R. Tao、H. Li 和 Q. Du，“使用 IP-CNN 对高光谱与激光雷达数据进行分类的信息融合”，*IEEE*地球科学与遥感汇刊，第 60 卷，12，2022 年 1 页码 – 。[14] D. Hong 、 L. Gao 、 R. Hang 、 B. Zhang 和 J. Chanussot，“用于高光谱与激光雷达数据分类的深度编码器-解码器网络”，*IEEE*地球科学与遥感快报，第 19 卷，第 1–5 页，2022年。[15] R. Hang、Z. Li、P. Ghamisi、D. Hong、G. Xia 和 Q. Liu，“使用耦合卷积神经网络对高光谱与激光雷达数据进行分类”，*IEEE*地球科学与遥感汇刊，第 58 卷，第 7 期，第 4939–4950 页，2020年。[16] X. Zhao、R. Tao、W. Li、W. Philips 和 W. Liao，“用于多源遥感数据分类的分数阶Gabor卷积网络”，*IEEE*地球科学与遥感汇刊，第 60 卷，第 1–18 页，2022年。[17] Y. Gao、W. Li、M. Zhang、J. Wang、W. Sun、R. Tao 和 Q. Du，“利用深度特征交互网络对滨海湿地进行高光谱与多光谱分类”，*IEEE*地球科学与遥感汇刊，第 60 卷，第 1–15 页，2022年。[18] Z. Xue、X. Tan、X. Yu、B. Liu、A. Yu 和 P. Zhang，“用于高光谱与激光雷达数据分类的深度分层视觉 Transformer”，*IEEE*图像处理汇刊，第 31 卷，第 3095–3110 页，2022年。[19] A. K. Sarkar、Z.-H. Tan、H. Tang、S. Shon 和 J. Glass，“基于时间对比学习的深度瓶颈特征，用于文本相关说话人验证”，*IEEE/ACM*音频、语音与语言处理汇刊，第 27 卷，第 8 期，第 1267–1279 页，2019年。[20] A. T. Liu、S.-W. Li 和 H.-Y. Lee，“TERA：用于语音的Transformer编码器表示自监督学习”，*IEEE/ACM*音频、语音与语言处理汇刊，第 29 卷，2366，2021年 2351页码 – 。[21] H. Xu、H. Xiong 和 G.-J. Qi，“通过多实例增强进行视觉特征的K-shot对比学习”，*IEEE*模式分析与机器智能汇刊，2021年，doi：10.1109/T-PAMI.2021.3082567。[22] K. He、H. Fan、Y. Wu、S. Xie 和 R. Girshick，“面向无监督视觉表征学习的动量对比”，载于*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*），2020年，第 9726–9735 页。[23] L. Jing 和 Y. Tian，“利用深度神经网络进行自监督视觉特征学习：综述”，*IEEE*模式分析与机器智能汇刊，第 43 卷，第 11 期，第 4037–4058 页，2021年。[24] B. Ren、Y. Zhao、B. Hou、J. Chanussot 和 L. Jiao，“一种基于互信息的极化合成孔径雷达土地覆盖分类自监督学习模型”，*IEEE*地球科学与遥感汇刊，第 59 卷，第 11 期，第 9224–9237 页，2021年。[25] H. Jung、Y. Oh、S. Jeong、C. Lee 和 T. Jeon，“使用平滑表征进行遥感对比式自监督学习”，*IEEE*地球科学与遥感快报，第 19 卷，第 1–5 页，2022年。[26] J. Yue、L. Fang、H. Rahmani 和 P. Ghamisi，“采用自适应蒸馏进行高光谱图像分类的自监督学习”，*IEEE*地球科学与遥感汇刊，第 60 卷，13，2022年 1页码 – 。[27] X. Zheng、T. Gong、X. Li 和 X. Lu，“利用多任务学习从小规模数据集中进行广义场景分类”，*IEEE*地球科学与遥感汇刊，第 60 卷，第 1–11 页，2022年。[28] M. Brell、K. Segl、L. Guanter 和 B. Bookhagen，“高光谱与激光雷达强度数据融合：严格校正光照、各向异性效应和交叉校准的框架”，*IEEE*地球科学与遥感汇刊，第 55 卷，2810，2017年 5，2799 编号 页码 – 。[29] M. Dalponte、L. Bruzzone 和 D. Gianelle，“融合高光谱与激光雷达遥感数据，对复杂森林区域进行分类”，*IEEE*地球科学与遥感汇刊，第 46 卷，第 5 期，第 1416–1427 页，2008年。[30] B. Koetz、F. Morsdorf、S. van der Linden、T. Curt 和 B. Allg¨ower，“基于成像光谱测量与激光雷达数据的森林火灾管理多源土地覆盖分类”，森林生态与管理，第 256 卷，第 3 期，第 263–271 页，2008年。[31] U. Heiden、W. Heldens、S. Roessner、K. Segl、T. Esch 和 A. Mueller，“利用高光谱遥感与高度信息进行城市结构类型表征”，《景观与城市规划》，第 105 卷，第 4 期，第 361–375 页，2012年。[33] M. Khodadadzadeh、J. Li、S. Prasad 和 A. Plaza，“利用多特征学习融合高光谱与激光雷达遥感数据”，[32] W. Liao、A. Pi˘zurica、R. Bellens、S. Gautama 和 W. Philips，“利用形态学特征对高光谱和激光雷达数据进行通用的基于图的融合”，*IEEE*地球科学与遥感快报，第 12 卷，第 3 期，第 552–556 页，2015 年。*IEEE*应用地球观测与遥感精选主题期刊，第 8 卷，第 6 期，第 2971–2983 页，2015 年。[34] P. Ghamisi、R. Souza、J. A. Benediktsson、X. X. Zhu、L. Rittner 和 R. A. Lotufo，“用于遥感数据分类的消亡剖面”，*IEEE*地球科学与遥感汇刊，第 54 卷，第 10 期，第 5631–5645 页，2016 年。[35] P. Ghamisi、B. H¨ofle 和 X. X. Zhu，“利用消亡剖面和深度卷积神经网络融合高光谱与激光雷达数据”，*IEEE*应用地球观测与遥感精选主题期刊，第 10 卷，第 6 期，第 3011–3024 页，2017 年。[36] X. Xu、W. Li、Q. Ran、Q. Du、L. Gao 和 B. Zhang，“基于卷积神经网络的多源遥感数据分类”，*IEEE*地球科学与遥感汇刊，第 56 卷，第 2 期，第 937–949 页，2018 年。[37] M. Zhang、W. Li、Q. Du、L. Gao 和 B. Zhang，“利用块到块卷积神经网络提取高光谱和激光雷达数据分类特征”，*IEEE*控制论汇刊，第 50 卷，第 1 期，第 100–111 页，2020年。[38] Y. Chen、C. Li、P. Ghamisi、X. Jia 和 Y. Gu，“遥感数据深度融合以实现精确分类”，*IEEE*地球科学与遥感快报，第 14 卷，第 8 期，第 1253–1257 页，2017 年。[39] H.-C. Li、W.-S. Hu、W. Li、J. Li、Q. Du 和 A. Plaza，“A3clnn：用于多源遥感数据分类的空间、光谱和多尺度注意力卷积长短期记忆神经网络”，*IEEE*神经网络与学习系统汇刊，第 33 卷，第 761 页，2022 年第 2 期，第 747 页。[40] X. Du 、 X. Zheng、X. Lu 和 A. A. Doudkin，“采用图融合网络的多源遥感数据分类”，*IEEE*地球科学与遥感汇刊，第 59 卷，第 10 072 页，2021 年第 12 期，第 10 062 页。[41] S. K. Roy、G. Krishna、S. R. Dubey 和 B. B. Chaudhuri，“HybridSN：探索用于高光谱图像分类的三维—二维卷积神经网络特征层级”，*IEEE*地球科学与遥感快报，第 17 卷，第 2 期，第 277–281 页，2020年。[42] Z. Yu、J. Yu、Y. Cui、D. Tao 和 Q. Tian，“用于视觉问答的深度模块化协同注意力网络”，载于*2019 IEEE/CVF*计算机视觉与模式识别会议 *(CVPR)*，2019 年，第 6274–6283 页。[43] C. Yan、Y. Hao、L. Li、J. Yin、A. Liu、Z. Mao、Z. Chen 和 X. Gao，“用于图像描述的任务自适应注意力”，*IEEE*视频技术电路与系统汇刊，第 32 卷，第 51 页，2022 年第 1 期，第 43 页。[44] X. Xu、T. Wang、Y. Yang、L. Zuo、F. Shen 和 H. T. Shen，“用于图文匹配的语义一致跨模态注意力”，*IEEE*神经网络与学习系统汇刊，第 31 卷，第 12 期，第 5412–5425 页，2020年。[45] X. Zheng、B. Wang、X. Du 和 X. Lu，“用于遥感视觉问答的互注意力 Inception 网络”，*IEEE*地球科学与遥感汇刊，第 60 卷，第 1–14 页，2022 年。[46] C. Liu、R. Zhao 和 Z. Shi，“基于多层聚合变换器的遥感图像描述”，*IEEE*地球科学与遥感快报，第 19 卷，第 1–5 页，2022 年。[47] S. Zhuang、P. Wang、G. Wang、D. Wang、J. Chen 和 F. Gao，“结合网格特征与 Transformer 改进遥感图像描述”，*IEEE*地球科学与遥感快报，第 19 卷，第 1–5 页，2022 年。[48] Z. Zhang、W. Zhang、M. Yan、X. Gao、K. Fu 和 X. Sun，“全局视觉特征与语言状态引导的遥感图像描述注意力机制”，*IEEE*地球科学与遥感汇刊，第 60 卷，第 1–16 页，2022 年。[49] S. Mohla、S. Pande、B. Banerjee 和 S. Chaudhuri，“FusAtNet：基于双重注意力的光谱空间多模态融合网络，用于高光谱与激光雷达分类”，载于*2020 IEEE/CVF*计算机视觉与模式识别会议研讨会 *(CVPRW)*，2020 年，第 416–425 页。[50]S. Fang、K. Li 和 Z. Li，“S2ENet：用于高光谱和激光雷达数据分类的空间—光谱跨模态增强网络”，*IEEE*地球科学与遥感快报，第 19 卷，第 1–5 页，2022 年。