引用格式: 初壮, 钱育蓉, 范迎迎, 等. 基于对比学习的多模态遥感图像融合分类研究[J]. 微电子学与计算机, 2025, 42(1): 35-44.

CHU Z, QIAN Y R, FAN Y Y, et al. Multimodal remote sensing image fusion classification based on contrastive learning network[J]. Microelectronics & Computer, 2025, 42(1): 35-44.

DOI: 10.19304/J.ISSN1000-7180.2023.0941

基于对比学习的多模态遥感图像融合分类研究

初 壮[1,2,3], 钱育蓉[1,2,3], 范迎迎[2,3,4], 刘怡然[5]

(1 新疆大学 软件学院, 新疆 乌鲁木齐 830091;
2 信号检测与处理重点实验室(新疆维吾尔自治区), 新疆 乌鲁木齐 830046;
3 新疆大学 软件工程重点实验室, 新疆 乌鲁木齐 830091;
4 新疆财经大学 信息管理学院, 新疆 乌鲁木齐 830012;
5 新疆大学 计算机科学与技术学院, 新疆 乌鲁木齐 830046)

摘 要: 高光谱与激光雷达图像融合分类技术能够实现对地物的高精度分类。目前,有监督的传统与深度学习方法取得了较好的分类结果,但往往需要大量的标记样本。基于自监督的多模态遥感融合分类研究相对较少,现有的自监督对比学习框架使用数据增强来生成正样本对,并不适用多模态遥感图像,会破坏多模态数据之间的空间分布与语义相似性,且模型过于复杂不利于下游微调任务的泛化。由此,提出了一种基于对比学习的多模态遥感图像融合分类网络(MMCLNet)。与传统的对比学习网络不同,该网络在预训练阶段无需数据增强操作,即可充分利用大量未标记的数据来学习判别特征表示。同时,精心设计的双分支网络降低了网络的复杂性。此外,在微调阶段采用多层次特征融合分类网络,充分整合两个模态数据的异构特征。使用 3 个真实的多模态遥感图像融合分类数据集进行了大量实验,证明了提出的研究方法在少量标记样本的数据集上具有一定的优势。

关键词: 对比学习; 多模态遥感分类; 高光谱; 激光雷达; 多层次特征融合; 自监督学习

中图分类号: TP391 **文献标识码:** A **文章编号:** 1000-7180(2025)01-0035-10

Multimodal remote sensing image fusion classification based on contrastive learning network

CHU Zhuang[1,2,3], QIAN Yurong[1,2,3], FAN Yingying[2,3,4], LIU Yiran[5]

(1 School of Software, Xinjiang University, Urumqi 830091, China;
2 Key Laboratory of Signal Detection and Processing, Xinjiang Uygur Autonomous Region, Urumqi 830046, China;
3 Key Laboratory of Software Engineering, Xinjiang University, Urumqi 830091, China;
4 School of Information Management, Xinjiang University of Finance and Economic, Urumqi 830012, China;
5 School of Computer Science and Technology, Xinjiang University, Urumqi 830046, China)

Abstract: Hyperspectral and LiDAR fusion classification techniques can realize high-precision classification of features. Currently, supervised traditional and deep learning methods have achieved better classification results, but often require a large number of labeled samples. There are relatively few studies on self-supervised multimodal remote sensing fusion

收稿日期: 2023-12-16; **修回日期:** 2024-01-10

基金项目: 新疆维吾尔自治区自然科学基金(2022D01B123); 国家自然科学基金(62266043, 61966035, 62261053); 国防科工局高分辨率对地观测系统重大专项(95-Y50G37-9001-22/23); 天山创新团队(2023D14012); 新疆维吾尔自治区杰出青年科学基金(2023D01E01)

classification based on self-supervision, and the existing self-supervised contrast learning framework uses data augmentation to generate positive sample pairs, which is not applicable to multimodal remote sensing images, and will destroy the spatial distribution and semantic similarity between multimodal data, and the model is too complex to be conducive to the generalization of the downstream fine-tuning task. Therefore, a multimodal remote sensing image fusion classification based on contrastive learning network (MMCLNet) is proposed, which is different from the traditional contrast learning network in that it can fully utilize a large amount of unlabeled data in the pre-training stage without data enhancement operations to learn the discriminative feature representations. At the same time, the well-designed two-branch network reduces the complexity of the network. Moreover, it adopts the multilevel feature fusion in the fine-tuning stage. In addition, a multi-level feature fusion network is used in the fine-tuning stage to fully integrate the heterogeneous features of the two modal data. Extensive experiments using three real multimodal remote sensing image fusion classification datasets demonstrate the advantages of the proposed research method on datasets with a small number of labeled samples.

Key words：contrastive learning；multimodal remote sensing classification；HSI；LiDAR；multilevel feature fusion；self-supervised learning

1 引言

高光谱图像(Hyperspectral Image, HSI)可以捕获地面物体的空间与光谱信息，而光探测和测距(Light Detection and Ranging, LiDAR)主动遥感成像技术可以区分光谱特征相似但高程信息不同的物体，为高光谱图像提供了有价值的补充数据。因此，融合不同传感器获取的不同数据[1-3],可以综合各个模态的优势，提高遥感图像分类的精度。多模态遥感影像协同应用正成为遥感科学领域新的研究热点[4]。

传统的多模态遥感数据融合分类方法通常致力于挖掘和融合这两类数据的特征[5-9]，这些方法依赖于手工特征，深层特征挖掘不足，无法很好地拟合HSI 和 LiDAR 数据中地物特征的复杂非线性关系[10]。深度学习方法在一定程度上可以弥补传统融合方法的不足，可以自动从多模态遥感数据中提取特征，学习丰富的语义信息。基于有监督方法中，Xu 等[11]设计了一个双分支网络，采用单个卷积神经网络提取多模态特征，然后通过直接连接操作进行特征融合。Hang 等[12] 提出了耦合卷积神经网络来融合HSI 和 LiDAR。一个 CNN 分支用于从 HSI 数据中学习空间光谱特征，另一个分支用于从 LiDAR 数据中学习高程信息；在融合阶段，采用特征级和决策级的融合策略来融合多源特征。Hong 等[13] 提出了一种基于编码器-解码器的结构，通过重构多模态输入来合并多模态特征。虽然有监督方法在分类准确度方面有了很大的提升，但容易受到人为主观因素的影响，而且标记样本的获取通常需要耗费大量的时间和人力资源。

随后，也有研究者提出了无监督的方法，侧重于从无标记样本的数据中自动学习潜在特征。例如，Wang 等[14] 利用生成对抗网络生成人工样本，在标记样本有限的情况下进行数据增强，从而提高了HSI 分类性能。Lu 等[15] 提出了一种新的基于耦合对抗学习的分类方法，用于 HSI 和 LiDAR 数据的融合分类，通过优化对抗损失和分类损失组成的联合损失函数，对网络进行协同训练，从而提高分类精度。生成对抗学习网络解决了样本生成和数据分布匹配的挑战，而对比学习则致力于通过学习样本之间的关系，提高模型在分类任务中的性能。Jia 等[16] 设计了一种协同对比学习方法，通过两阶段的协作策略在没有标记样本的情况下实现两模态遥感数据之间的协调特征表示和匹配。类似地，Wang 等[17] 提出了一种基于最近邻的无监督学习对比学习网络，使用大量未标记数据学习判别特征表示，并结合基于最近邻的数据增强操作来捕获准确的模态间语义对齐。

然而基于自监督的 HSI 和 LiDAR 融合分类的研究很少被探索，仍处于起步阶段。在自监督学习中，对于自然图像进行数据增强操作可以从原始数据中生成正样本对，但是随机颜色抖动、随机水平翻转和随机灰度转换等方式[18-19] 破坏了多源遥感数据之间的空间分布与语义相似性等信息，使模型训练了额外的虚假特征而忽略了丰富的原始细节信息；此外，过于复杂的模型结构使得模型不利于下游微调任务的泛化；同时，多模态数据分布不一致所导致的异构差距可能使特征表示变得更加复杂，选择合适的融合策略变得至关重要，错误的融合策略可能导致信息丢失或者引入噪音，影响最终分类性能。

基于以上分析，本文提出了一种基于对比学的多模态遥感图像融合分类网络，该网络在预训练阶

段通过输入原始的 HSI 与 LiDAR 图像样本对，充分利用大量未标记数据来学习多模态数据之间的空间分布、语义相似性和判别特征表示；此外，特征提取编码器部分采用极为简洁的双分支卷积结构，可以适用于大部分下游任务；同时在微调阶段采用多层次特征融合分类网络，充分融合了两个模态数据的异构特征。在 Houston2013、Trento、MUUFL 这 3 个真实的多模态遥感图像融合分类数据集上的实验结果验证了方法的有效性。

2　基于对比学习的多模态遥感图像融合分类

基于对比学习的分类任务在预训练阶段的目标是从大量的无标记样本中自动学习特征表示，通过比较数据增强后样本对的相似性，生成一个预先训练好的模型，该模型得到的编码器可以在微调阶段用于下游任务。早期自监督对比学习模型[18-21]避免坍塌的关键在于负样本的构建。Chen 等[22]首次提出 SimSiam 网络框架，无需负样本也可避免模型的坍塌，得益于模型结构中的三大设计：梯度停止(Stop-gradient)、投影层(Predictor)以及余弦相似度(Cosine Similarity)损失函数。考虑到从 HSI 和 LiDAR 数据中提取多源异构特征，需要比较来自同一区域的不同模态的样本相似性。受 SimSiam 启发，本文提出了一种基于对比学习的多模态遥感图像融合分类网络(MMCLNet)。该网络训练过程分为两阶段，在预训练阶段使用高光谱与激光雷达对比学习网络学习多模态数据之间的空间分布、语义相似性和判别特征表示，在微调阶段使用预训练阶段的编码器提取特征，并用多层次特征融合分类网络整合多模态异构特征，得到最终的分类结果。

2.1　高光谱与激光雷达对比学习网络

在预训练阶段，设计了一个高光谱和激光雷达对比学习网络，模型详细设计如图 1 所示。为了便于下游任务的泛化，编码器采用了一个相对简单的双分支网络结构来提取多模态特征。具体来说，选取同一位置为中心的 HSI 与 LiDAR 数据记为 $X_h \in R^{h\times w\times b_h}$ 与 $X_l \in R^{h\times w\times b_l}$，$h\times w$ 代表图像的尺寸 11×11，b_h 与 b_l 分别代表 HSI 与 LiDAR 的通道数，由于 HSI 数据中存在大量的冗余光谱信息，采用主成分分析(PCA)对 X_h 进行降维处理得到 $X_h' \in R^{h\times w\times b_p}$，$b_p$ 为对 X_h 降维后的通道数 30。X_h' 与 X_l 分别通过 HSI 编码器与 LiDAR 编码器进行特征提取，两个编码器的卷积块都采用 2D 卷积+批归一化+ReLU 激活操作的结构，卷积类型、卷积核数量与尺寸、输出尺寸的详细设计如表 1 所示。

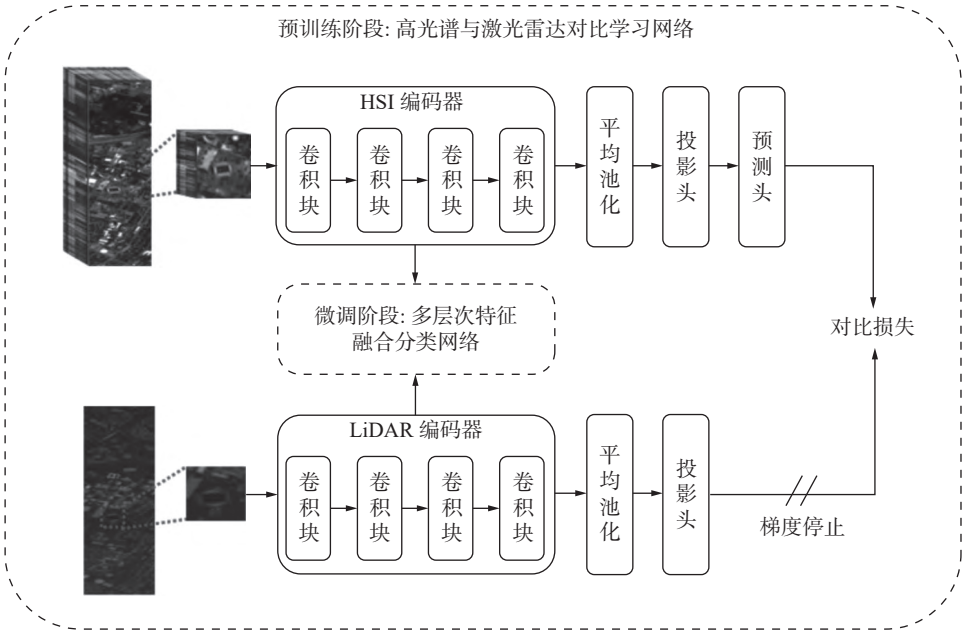

图 1　高光谱和激光雷达对比学习网络

Fig. 1 Schematic illustration of the hyperspectral and LiDAR contrast learning networks

经过编码器得到的特征在通过平均池化层与投影头后，特征尺寸降为 1×1，得到两个表征向量 z_1 和 z_2，首先使用预测头对 z_1 做预测得到 p_1，将最小化 p_1 与 z_2 的负余弦相似度作为目标函数去学习。与此同时，需要将 z_2 对称地做一次预测得到 p_2，再计算 p_2 与 z_1 的负余弦相似度。由于在反向传播过程中

需要对其中一个分支进行梯度停止操作，因此最终的对比损失函数L定义为

$$L=\frac{1}{2}D(p_1,sg(z_2))+\frac{1}{2}D(p_2,sg(z_1)) \quad (1)$$

$$D(p_i,z_j)=-\frac{p_i}{\|p_i\|_2}\cdot\frac{z_j}{\|z_j\|_2} \quad (2)$$

式中：sg为梯度停止策略，表示孪生网络其中一个分支不进行梯度更新。

表 1 编码器结构的详细设计

Tab. 1 Detailed design of encoder structure

卷积类型	核的数量@尺寸	输出尺寸
输入	–	(11, 11, 1/30)
二维卷积	32@3×3	(9, 9, 32)
二维卷积	64@3×3	(7, 7, 64)
二维卷积	128@3×3	(5, 5, 128)
二维卷积	256@3×3	(5, 5, 256)

2.2 多层次特征融合网络

在微调阶段，设计了一个多层次特征融合网络，使用标记样本对网络参数进行微调，并对预训练阶段中提取的无监督特征进行多级连接和残差连接，帮助模型更好地学习特征，从而进一步提高模型的表达能力和性能。

在深度神经网络中，编码器的低层和中层包含光谱、纹理和结构信息，而高层代表更多的语义特征，从多层编码器[23-24]中学习到的多层次特征之间存在着互补性，充分利用这些互补特征能够进一步提升网络性能。如图 2 所示，由预训练阶段得到的 HSI 编码器与 LiDAR 编码器中的 4 个卷积块依次记为$HCB_i,i\in\{1,2,3,4\}$与$LCB_i,i\in\{1,2,3,4\}$，经过卷积块得到的特征依次为$h_i=HCB_i(X_h^i)$与$l_i=LCB_i(X_l)$，先通过残差运算得到$h_4=h_1+h_4$与$l_4=l_1+l_4$，分别获取更精准的 HSI 和 LiDAR 数据的特征表示；为进一步融合 HSI 和 LiDAR 数据的特征，将各分支的特征根据权重相加，通过这种多级连接方式获取多层次的融合特征表示，即

$$f_j=\alpha\cdot h_{j+1}+\gamma\cdot l_{j+1} \qquad j\in\{1,2,3\} \quad (3)$$

式中：α和γ是两个可学习的参数张量，能够自动地被优化器所更新，从而使模型能够学习和适应地将两个模态的特征相加。接下来f_j依次经过 3 个结构相同的特征融合分类模块 FFC。分别对 3 个层次的融合特征分类，FFC 的结构如图 3 所示。

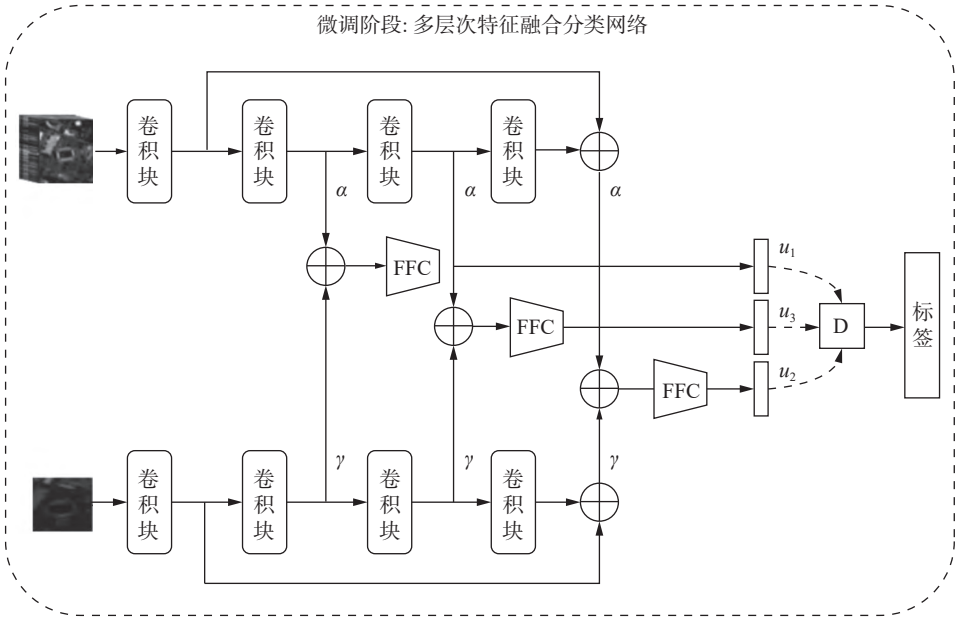

图 2 多层次特征融合分类网络

Fig. 2 Schematic illustration of the multi-level feature fusion classification network

每个 FFC 的每个卷积块依次由卷积核大小为 1×1 的二维卷积、批处理归一化和 ReLU 激活操作组成，并在第二个卷积块后添加了平均池化操作，用于将特征尺寸减小到 1×1。紧接着是一个独立的卷积层，卷积核大小为 1×1，用于将特征减少到C维度，其中C表示每个数据集中对象类别的数量。然后通过 Sigmoid 运算将C维特征处理为 0 到 1 的C维概率，该概率反映了目标像素的对象类别。最后，对 3 个层次的输出概率自适应决策级融合[12]。

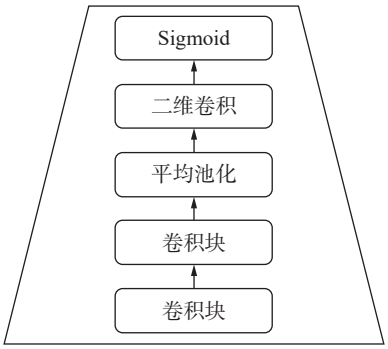

图 3　特征融合分类模块结构图

Fig. 3 Feature fusion classification module

具体来说，通过前向传播过程，每个训练样本可以得到 3 个输出$\hat{y}_j = FFC_j(f_j)$，并由交叉熵损失函数计算损失值$L_j = \mathrm{CELoss}(\hat{y}_j)$，$L_1$、$L_2$和$L_3$分别监督低层、中层和高层融合特征的学习过程，最终的损失值L表示为

$$L = \lambda(L_1 + L_2) + L_3 \tag{4}$$

式中：λ为L_1和L_2的权重参数，根据经验将其设置为 0.01，进而整体损失L可以使用反向传播进行优化。

在训练与测试的过程中，首先使用训练集计算每个层次的融合特征的权重u_j，即

$$u_j = \frac{oa_j + 10^{-5}}{\sum_{k=1}^{3} oa_k + 10^{-5}} \tag{5}$$

式中：oa_j表示每个层次的融合特征在训练集上的总体分类精度。最终在测试集上采用加权求和得到分类概率O，表达式为

$$O = \sum_{j=1}^{3} u_j \cdot \hat{y}_j \tag{6}$$

3　实验与结果分析

为了验证本文提出的 MMCLNet 的有效性，在 3 个 HSI 和 LiDAR 融合分类数据集上进行了大量的实验。

3.1　实验细节

实验在 Window 操作系统中进行，使用的 CPU 为 Intel(R) Xeon(R) Gold 5117，内存为 256 GB，GPU 为 NVIDIA Tesla V100-PCIE-16 GB。开发环境为 Pytorch 1.10.0、Cuda 10.2、Python 3.7.2。对于预训练阶段，使用学习率为 0.001、动量为 0.9、权重衰减为 0.000 5 的 SGD 优化器；在微调阶段，使用学习率为 0.000 5 的 Adam 优化器；两个阶段的 Batch 大小为 64，Patch 大小为 11×11，迭代次数设置为 200。为了客观地评价分类性能，实验选取 3 个常用的评价指标来评估分类性能，即总体精度(Overall Accuracy, OA, 记为 A_{O})、平均精度(Average Accuracy, AA, 记为 A_{A})和 Kappa 系数(记为 K)。

$$A_{\mathrm{O}} = \sum_{i=1}^{C} M_{ii} / N \tag{7}$$

$$A_{\mathrm{A}} = \frac{\sum_{i=1}^{C} \left(M_{ii} \Big/ \sum_{j=1}^{C} M_{ij} \right)}{C} \tag{8}$$

$$K = \frac{A_{\mathrm{O}} - P_{\mathrm{e}}}{1 - P_{\mathrm{e}}} \tag{9}$$

其中：

$$P_{\mathrm{e}} = \frac{\sum_{i=1}^{C} \left(\sum_{j=1}^{C} M_{ij} \sum_{j=1}^{C} M_{ji} \right)}{N^2} \tag{10}$$

式中：$\boldsymbol{M}$为混淆矩阵；M_{ii}是第i类被识别为第i类的数量；M_{ij}为第i类被识别为第j类的数量；C为类别总数；N为测试集总数；P_{e}为机会一致性的假设概率。

3.2　实验数据集

Houston2013 数据集是 2012 年在休斯顿大学校园及其周边地区通过航空传感器捕获的航拍图像，包含了总共 15 029 个地面真实样本，分为 15 个类别，并划分为包含 2 832 个样本的训练集和包含了 12 197 个样本的测试集。

Trento 数据集是意大利特伦托南部的一个农村地区，包含了总共 30 214 个地面真实样本，分为 6 个类别，并划分为包含 819 个样本的训练集和包含 29 395 个样本的测试集。

MUUFL 数据集是在南密西西比海湾公园大学获得的，共包含 53 687 个地面真实样本，分为 11 个类别并划分为包含 1 650 个样本的训练集和包含 53 687 个样本的测试集。

3.3　分类结果与分析

为了充分验证 MMCLNet 在多模态遥感图像融合分类上的优越性，选择了 5 种先进的多模态融合分类方法进行比较，包括基于编码器-解码器结构的融合网络 EndNet[13]、基于双注意力的光谱空间融合网络 FusAtNet[25]、自校正卷积网络 SCNet[26]、空间光谱跨模态增强网络 S2ENet[27]、耦合对抗学习分类网络 Calc[15] 等。

3.3.1　定量分析

表 2～表 4 分别给出了不同方法在 Houston2013、

Trento 和 MUUFL 这 3 个数据集上的 OA、AA、Kappa 和不同类别的精度结果，粗体数值表示每行对应性能指标的最优值。为了公平比较，在所有方法中均采用了官方建议的训练集和测试集划分方法。

表 2 不同算法在 Houston2013 数据集上的分类精度

Tab. 2 Classification accuracy on Houston2013 dataset of different algorithm

类别	EndNet	FusAtNet	SCNet	S2ENet	Calc	Ours
Healthy grass	90.46	91.15	91.22	**91.44**	89.74	90.88
Stressed grass	91.96	93.70	93.50	**93.74**	92.53	93.33
Synthetic grass	99.61	93.46	95.28	97.40	97.74	**100**
Tree	97.22	97.58	98.69	**98.78**	94.51	98.60
Soil	97.89	96.65	99.11	99.81	**99.88**	99.81
Water	91.03	79.88	97.58	98.96	95.60	**100**
Residential	86.55	93.35	90.59	96.17	93.93	**97.19**
Commercial	83.77	78.79	87.90	94.16	91.39	**94.41**
Road	84.77	85.85	87.62	87.82	93.15	**95.62**
Highway	89.76	74.68	79.31	89.56	94.42	**95.85**
Railway	89.74	91.09	91.55	93.68	95.17	**99.53**
Parking lot 1	84.58	81.83	92.83	91.57	89.32	**95.65**
Parking lot 2	74.51	90.87	**95.83**	95.43	94.53	**95.83**
Tennis court	89.82	90.31	**100**	**100**	**100**	99.40
Running track	99.15	92.90	**100**	**100**	96.93	**100**
OA	90.15	88.71	92.15	94.33	94.05	**96.54**
AA	91.15	89.42	93.49	95.21	94.56	**96.96**
Kappa	89.32	87.75	91.48	93.85	93.57	**96.24**

表 3 不同算法在 Trento 数据集上的分类精度

Tab. 3 Classification accuracy on Trento dataset of different algorithm

类别	EndNet	FusAtNet	SCNet	S2ENet	Calc	Ours
Apple trees	74.37	95.21	98.21	**99.67**	99.05	99.25
Buildings	**96.00**	94.29	94.53	95.29	93.00	95.38
Ground	86.33	63.22	93.26	**97.78**	88.04	91.57
Wood	99.25	99.25	99.93	99.91	99.99	**100**
Vineyard	85.76	96.51	99.42	99.88	99.87	**100**
Roads	92.96	92.55	94.60	95.36	91.70	**95.40**
OA	89.63	95.47	98.25	98.81	98.11	**99.12**
AA	92.40	94.05	96.23	98.18	94.61	**98.31**
Kappa	86.40	93.97	97.66	98.41	97.48	**98.83**

表 4 不同算法在 MUUFL 数据集上的分类精度

Tab. 4 Classification accuracy on MUUFL dataset of different algorithm

类别	EndNet	FusAtNet	SCNet	S2ENet	Calc	Ours
Trees	92.03	91.83	94.22	94.19	93.71	**97.04**
Mostly Grass	73.26	80.52	83.37	81.72	77.36	**85.65**
Mixed Ground Surface	77.74	78.64	83.51	81.45	81.77	**87.55**
Dirt and Sand	84.93	75.80	85.87	85.76	86.89	**91.61**
Road	91.95	88.38	92.22	91.62	85.91	**93.22**
Water	87.39	85.40	82.66	84.35	82.33	**91.91**
Building Shadow	67.32	84.14	80.99	82.19	82.21	**87.85**
Building	92.99	92.35	94.97	94.30	93.22	**97.46**
Sidewalk	67.06	68.93	79.26	**80.61**	67.41	67.05
Yellow Curb	**78.28**	29.45	38.16	36.42	20.93	38.47
Cloth Panels	**94.35**	91.62	86.13	87.01	62.26	80.90
OA	86.12	83.20	87.80	87.34	86.80	**92.40**
AA	89.86	84.40	89.03	88.74	88.69	**92.64**
Kappa	82.16	78.66	84.31	83.72	83.02	**90.06**

通过观察表 2～表 4 中实验结果数据，在所有基于深度学习的方法中，本文提出的 MMCLNet 获得了最好的分类性能，OA，AA 和 Kappa 指标都高于其他比较算法。与其他方法相比，不难发现，EndNet 中基于编码器-解码器结构的特征学习能力是有限的，FusAtNet 虽然使用交叉注意方法实现对另一种模态增强，但是缺乏更先进的融合策略，S2Enet 提出在特征融合之前的跨模态交互学习，以增强每个模态的信息表示，且都没有考虑到遥感图像中的多尺度信息，缺乏有效的特征融合方法。Calc 使用耦合生成式对抗网络提取了高级语义信息，但是耦合的结构无法很好的提取多模态的异构信息。本文提出的 MMCLNet 在大多数陆地对象类别上取得了更高的分类精度，因为其保留了多源遥感数据之间的空间分布与语义相似性，多层次特征融合分类网络能够充分整合两个模态数据的异构特征。

3.3.2 计算量对比分析

一般来说，模型结构越复杂，计算量与参数量越大。从表 5 中可以看出，MMCLNet 的计算量与参数量较大，在所有比较方法中仅次于 FusAtNet，这是因为在微调阶段引入了多层次特征融合的结构。然而，这些代价是可以接受的，因为该模型算法取得了更好的分类精度。

表 5　不同算法在 3 个数据集上的计算量与参数量比较

Tab. 5 Comparation of the flops and number of parameters on the three datasets of different algorithm

数据集	指标	EndNet	FusAtNet	SCNet	S2ENet	Calc	Ours
Houston 2013	计算量/10^6	5.8	221 535.1	411.4	847.8	28.7	1 468.2
	参数量/10^3	92.0	36 905.7	150.2	270.8	284.1	860.4
Trento	计算量/10^6	5.6	216 476.3	378.9	555.1	16.8	1 468.1
	参数量/10^3	88.8	36 243.3	139.5	177.2	236.8	859.3
MUUFL	计算量/10^6	5.6	221 535.1	379.3	559.0	17.1	1 469.7
	参数量/10^3	89.2	36 905.7	139.8	178.6	238.3	860.2

3.3.3　消融实验

设计了一系列消融实验，以深入探索预训练权重和多层次融合网络对多模态遥感图像融合分类模型性能的影响。从表 6 中实验结果显示，在未加载预训练权重和未使用多层次特征融合网络时，模型在 Houston2013、Trento 和 MUUFL 数据集上分别达到了 OA 为 90.73%、93.50% 和 84.44% 的基准性能。通过单独加载预训练权重或应用多层次融合网络，模型在不同数据集上均取得了显著的性能提升。更为引人注目的是，同时引入这两种策略时，模型在 3 个数据集上的 OA 分别提升至 96.54%、99.12% 和 92.40%，表明这两种策略具有协同效应，能够有效提高遥感图像分类的性能。

表 6　不同模块组合在 3 个数据集上的消融实验

Tab. 6 Ablation experiments about different module combination on three datasets

预训练权重	多层次融合	Houston2013	Trento	MUUFL
		90.73	93.50	84.44
√		92.65	98.45	86.63
	√	94.65	97.32	91.40
√	√	96.54	99.12	92.40

3.3.4　训练样本数量分析

为了验证 MMCLNet 在标记样本较少的情况下的分类效果，通过在微调阶段将训练样本数量设置为 17%、33%、…、100%，并保持测试样本数量不变，在 3 个数据集上进行了实验。图 4 展示了不同设置下的性能表现，其中 100% 代表了所有训练样本的数量。观察到随着训练样本数量的增加，性能指标也呈现增加的趋势。特别是在 Houston2013 与 Trento 数据集上，性能指标在不同训练样本数量下的变化相对较小，这验证了本文提出的 MMCLNet 在标记样本较少的情况下同样能够取得较好的成绩。

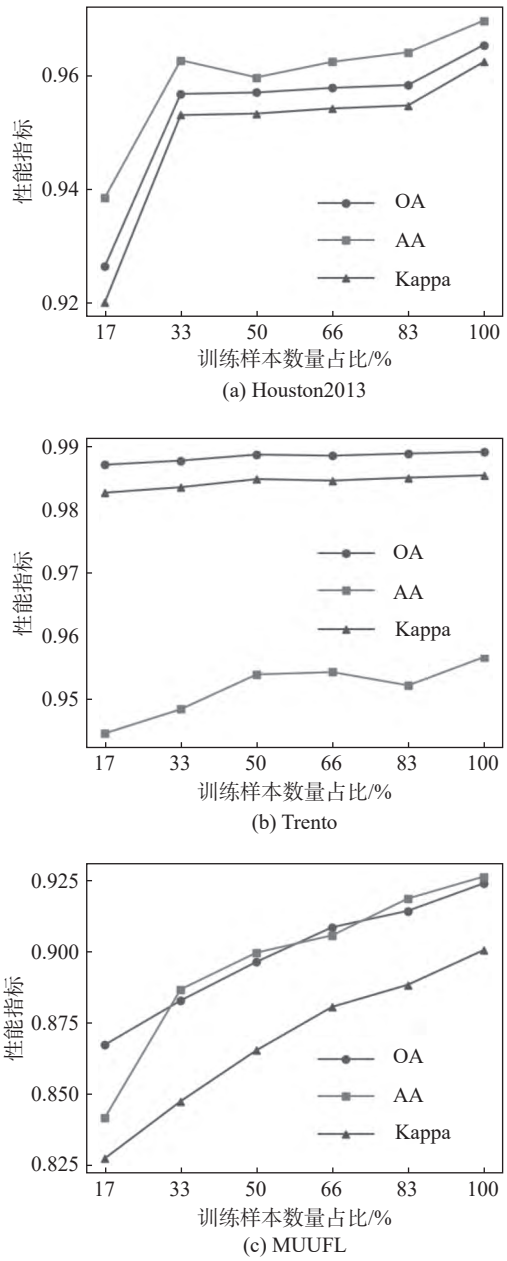

图 4　3 个数据集随训练样本数量变化的性能指标对比图

Fig. 4 Comparison of performance indicators of three data sets using different numbers of training samples

3.3.5 可视化分析

图 5 ~ 图 7 展示了 6 种不同的模型算法分别在 Houston2013、Trento、MUUFL 数据集上预测的可视化结果图，为了便于比较，还提供了地面真实标签，括号中的数值表示 OA 指标。观察实验结果可以得出，MMCLNet 在 3 个数据集上表现出色，其平均准确率分别比其他算法高出 4.65%、2.82% 和 6.15%，远超其他比较算法。从预测结果图来看，MMCLNet 获得的分类预测图与地面真实标签误差更小，能够更准确地预测出接近于样本真实标签的结果，进一步验证了本文模型算法的优势。

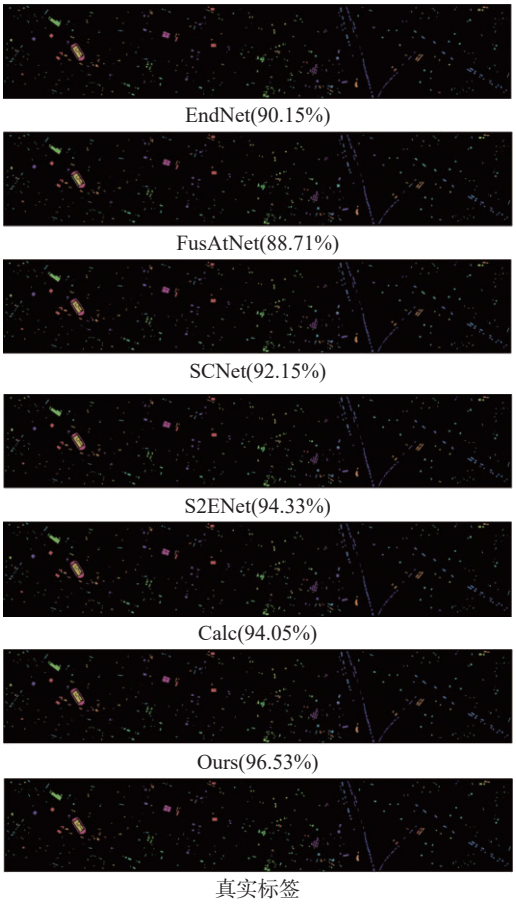

图 5 所有方法在 Houston2013 数据集上的分类图与真实标签

Fig. 5 Classification maps of all the methods and ground truth on the Houston2013 dataset

4 结束语

本文提出的基于对比学习的多模态遥感图像融合分类网络(MMCLNet)，充分利用了大量未标记数据来学习判别特征表示；无需数据增强操作，保留了多源遥感数据之间的空间分布与语义相似性；采

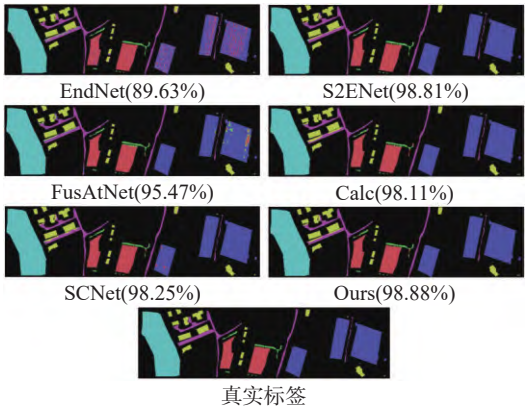

图6 所有方法在 Trento 数据集上的分类图与真实标签

Fig. 6 Classification maps of all the methods and ground truth on the Trento dataset

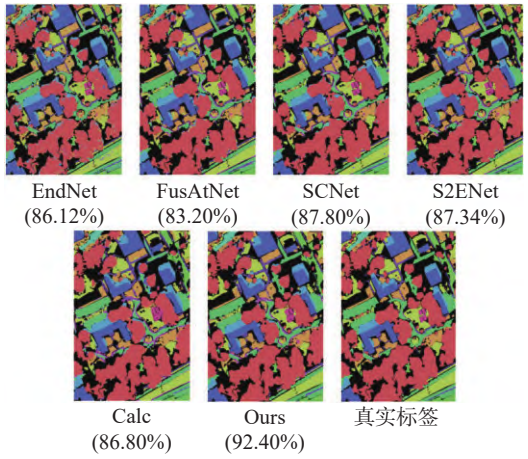

图7 所有方法在 MUUFL 数据集上的分类图与真实标签

Fig. 7 Classification maps of all the methods and ground truth on the MUUFL dataset

用较为简洁的双分支网络，可以很方便的应用于各种下游微调任务，为后续研究者提供思路；此外，在微调阶段采用多层次特征融合分类网络，充分整合两个模态数据的异构特征，弥补了源数据分布不一致所导致的异构差距大的不足。通过大量的实验，表明了本文提出的 MMCLNet 在 3 个真实的 HSI 和 LiDAR 融合数据集上具有良好的性能，显著提高了分类精度。后续目标是探索 HSI 和 LiDAR 数据之间更深层次的语义和空间关系，并探讨如何进一步增强 HSI 和 LiDAR 数据之间的特征交互作用。

参考文献:

[1] LUO F L, ZHANG L P, ZHOU X C, et al. Sparse-adaptive hypergraph discriminant analysis for hyperspectral image classification[J]. IEEE Geoscience and Remote Sensing Letters, 2020, 17(6): 1082-1086. DOI: 10.1109/LGRS.2019.2936652.

[2] JIANG W, CAO Y, DENG X Y. A novel *Z*-network model based on Bayesian network and *Z*-number[J]. IEEE Transactions on Fuzzy Systems, 2020, 28(8): 1585-1599. DOI: 10.1109/TFUZZ.2019.2918999.

[3] LUO F L, HUANG H, MA Z Z, et al. Semisupervised sparse manifold discriminative analysis for feature extraction of hyperspectral images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(10): 6197-6211. DOI: 10.1109/TGRS.2016.2583219.

[4] HONG D F, GAO L R, YOKOYA N, et al. More diverse means better: Multimodal deep learning meets remote-sensing imagery classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(5): 4340-4354. DOI: 10.1109/TGRS.2020.3016820.

[5] PEDERGNANA M, MARPU P R, MURA M D, et al. Classification of remote sensing optical and LiDAR data using extended attribute profiles[J]. IEEE Journal of Selected Topics in Signal Processing, 2012, 6(7): 856-865. DOI: 10.1109/JSTSP.2012.2208177.

[6] RASTI B, GHAMISI P, GLOAGUEN R. Hyperspectral and LiDAR fusion using extinction profiles and total variation component analysis[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(7): 3997-4007. DOI: 10.1109/TGRS.2017.2686450.

[7] LIAO W Z, PIŽURICA A, BELLENS R, et al. Generalized graph-based fusion of hyperspectral and LiDAR data using morphological features[J]. IEEE Geoscience and Remote Sensing Letters, 2015, 12(3): 552-556. DOI: 10.1109/LGRS.2014.2350263.

[8] GAO L R, HONG D F, YAO J, et al. Spectral superresolution of multispectral imagery with joint sparse and low-rank learning[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(3): 2269-2280. DOI: 10.1109/TGRS.2020.3000684.

[9] LI W, ZHANG Y X, LIU N, et al. Structure-aware collaborative representation for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(9): 7246-7261. DOI: 10.1109/TGRS.2019.2912507.

[10] XIONG F C, ZHOU J, TAO S Y, et al. SMDS-Net: Model guided spectral-spatial network for hyperspectral image denoising[J]. IEEE Transactions on Image Processing, 2022, (31): 5469-5483. DOI: 10.1109/TIP.2022.3196826.

[11] XU X D, LI W, RAN Q, et al. Multisource remote sensing data classification based on convolutional neural network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(2): 937-949. DOI: 10.1109/TGRS.2017.2756851.

[12] HANG R L, LI Z, GHAMISI P, et al. Classification of hyperspectral and LiDAR data using coupled CNNs[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(7): 4939-4950. DOI: 10.1109/TGRS.2020.2969024.

[13] HONG D F, GAO L R, HANG R L, et al. Deep encoder–decoder networks for classification of hyperspectral and LiDAR data[J]. IEEE Geoscience and Remote Sensing Letters, 2022, (19): 5500205. DOI: 10.1109/LGRS.2020.3017414.

[14] WANG W Y, LI H C, DENG Y J, et al. Generative adversarial capsule network with ConvLSTM for hyperspectral image classification[J]. IEEE Geoscience and Remote Sensing Letters, 2021, 18(3): 523-527. DOI: 10.1109/LGRS.2020.2976482.

[15] 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. DOI: 10.1016/j.inffus.2022.12.020.

[16] JIA S, ZHOU X, JIANG S G, et al. Collaborative contrastive learning for hyperspectral and LiDAR classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 5507714. DOI: 10.1109/TGRS.2023.3263511.

[17] WANG M, GAO F, DONG J Y, et al. Nearest neighbor-based contrastive learning for hyperspectral and LiDAR data classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 5501816. DOI: 10.1109/TGRS.2023.3236154.

[18] HE K M, FAN H Q, WU Y X, et al. Momentum contrast for unsupervised visual representation learning[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 9729-9738. DOI: 10.1109/CVPR42600.2020.00975.

[19] CHEN T, KORNBLITH S, NOROUZI M, et al. A simple framework for contrastive learning of visual representations[C]//Proceedings of the 37th International Conference on Machine Learning. New York: ACM, 2020: 1597-1607.

[20] GRILL J B, STRUB F, ALTCHÉ F, et al. Bootstrap your own latent-a new approach to self-supervised learning[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2020: 21271-21284.

[21] CARON M, MISRA I, MAIRAL J, et al. Unsupervised learning of visual features by contrasting cluster assignments[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc. , 2020: 831.

[22] CHEN X L, HE K M. Exploring simple siamese representation learning[C]//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 15750-15758. DOI: 10.1109/CVPR46437.2021.01549.

[23] XU Y H, ZHANG L P, DU B, et al. Spectral-spatial unified networks for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(10): 5893-5909. DOI: 10.1109/TGRS.2018.2827407.

[24] ZHANG M Y, GONG M G, MAO Y S, et al. Unsupervised feature extraction in hyperspectral images based on Wasserstein generative adversarial network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(5): 2669-2688. DOI: 10.1109/TGRS.2018.2876123.

[25] MOHLA S, PANDE S, BANERJEE B, et al. FusAtNet: Dual attention based spectrospatial multimodal fusion network for hyperspectral and LiDAR classification[C]// Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE, 2020: 92-93. DOI: 10.1109/CVPRW50498.2020.00054.

[26] HAN K, REZENDE R S, HAM B, et al. SCNet: Learning semantic correspondence[C]//Proceedings of 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 1831-1840. DOI: 10.1109/ICCV.2017.203.

[27] FANG S, LI K Y, LI Z. S²ENet: Spatial–spectral cross-modal enhancement network for classification of hyperspectral and LiDAR data[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 6504205. DOI: 10.1109/LGRS.2021.3121028.

作者简介：

初　壮　硕士研究生, 1617812854@qq.com

钱育蓉(通信作者)　博士, 教授, qyr@xju.edu.cn