

珠海一号高光谱场景分类数据集

刘渊^{1,2}, 郑向涛³, 卢孝强³

1. 中国科学院西安光学精密机械研究所 光谱成像技术实验室, 西安 710119;

2. 中国科学院大学, 北京 100049;

3. 福州大学 物理与信息工程学院, 福州 350108

摘要: 高空间分辨率、高光谱分辨率、大幅宽与大数据量是高光谱卫星数据发展趋势, 传统高光谱影像的像素级分类面临难以处理海量数据、无法高效获取复杂海量影像中隐含信息的困境。已有研究开始关注高光谱影像的场景级分类, 并逐步建立完善高光谱遥感场景分类数据集。然而, 目前的数据集制作过程多参考高空间分辨率可见光遥感场景数据集的制作方法, 主要采用遥感影像的空间信息进行场景类别解译, 忽视了高光谱场景的光谱信息。因此, 为构建高光谱影像的遥感场景分类数据集, 本文利用“珠海一号”高光谱卫星拍摄的西安地区高光谱数据, 使用无监督光谱聚类辅助定位、裁剪与标注待选场景样本, 结合 Google Earth 高分影像进行目视筛选, 构建6类场景类型和737幅场景样本的珠海一号高光谱场景分类数据集。并基于光谱与空间两个视角开展场景分类实验, 通过视觉词袋、卷积神经网络等方法的基准测试结果, 对不同算法在现有多光谱和高光谱遥感场景分类数据集下的性能进行深入分析。本研究可为后续的高光谱影像解译研究提供了有力的数据支撑。

关键词: 高光谱遥感, 珠海一号, 场景分类, 数据集, 特征提取

中图分类号: TP701/P2

引用格式: 刘渊, 郑向涛, 卢孝强. 2024. 珠海一号高光谱场景分类数据集. 遥感学报, 28(1): 306-319

Liu Y, Zheng X T and Lu X Q. 2024. Hyperspectral scene classification dataset based on Zhuhai-1 images. National Remote Sensing Bulletin, 28(1):306-319[DOI:10.11834/jrs.20233283]

1 引言

高光谱遥感影像具有“图谱合一”的特点, 同时具有观测地物的连续光谱与几何空间信息, 能够反演目标的真实类型, 在城市规划 (Kuras 等, 2021)、生态调查 (张成业 等, 2022)、农业建设 (钟亮 等, 2023) 等领域中获得了持续关注与广泛应用。伴随航空航天与光学成像技术的发展, 国内外高光谱遥感观测卫星的稳步增多, 高光谱影像向着高空间分辨率、高光谱分辨率、大幅宽与大数据量方向发展 (李德仁 等, 2014) (刘银年, 2021)。然而, 如何快速从结构复杂, 体量巨大的高光谱数据中挖掘出地物信息, 实现高光谱影像智能解译一直是当前的研究难点。

高光谱影像分类利用纹理、光谱等特征识别图像或像素类别, 是遥感影像解译的重要组成部分

分, 也是遥感领域活跃的研究课题之一 (杜培军 等, 2016)。早期高光谱影像分类集中于像素级的光谱特征, 使用人为设计的谱特征, 例如归一化植被指数 NDVI (Normalized Difference Vegetation Index)、增强植被指数 EVI (Enhanced Vegetation Index) 等光谱指数, 或经过一维神经网络、堆叠自编码器等提取的光谱特征, 实现高光谱影像的像素级分类。这类方法没有利用影像的空间信息, 易受大气光照与传感器的影响, 具有泛化能力弱、分类精度低的缺点, 也难以应对“同谱异物”现象与“同物异谱”现象。高光谱影像的空间信息对提升分类精度具有积极作用, 获得研究学者的持续关注, 例如形态学剖面获得的影像几何特征 (Benediktsson 等, 2005)、图像边缘特征 (Kang 等, 2014) 等。通过同时考虑光谱和空间的联合信息, 有助于提高特征的表达能力, 从而有效提

收稿日期: 2023-01-06; 预印本: 2023-01-06

基金项目: 国家自然科学基金(编号: 62271484); 国家杰出青年科学基金(编号: 61925112); 陕西省重点研发计划(编号: 2023-YBGY-225)

第一作者简介: 刘渊, 研究方向为遥感图像数据处理。E-mail: liuyuan213@mailsucas.ac.cn

通信作者简介: 郑向涛, 研究方向为遥感图像处理。E-mail: xiangtaoz@gmail.com

升分类性能。随着深度学习的迅猛发展，大量深度学习方法已经被引入到高光谱影像分类领域（Li等，2019）。Liu等（2022a）提出中心注意力网络CAN（Central Attention Network），设计了中心注意力模块着重提取目标像素光谱信息与相邻相似像素间的有效空间信息，并利用密集连接策略，将浅层提取的空谱特征与相邻像素间的相似度权重输入到更深的特征层，有效保留了目标像素的空谱信息。Yang等（2022）将Vision Transformer引入到高光谱影像分类领域，利用双分支3D卷积模块分别提取高光谱影像的局部与全局空谱信息，聚合后输入到Transformer编码器中捕获特征间的相关性。同时，在编码过程中额外增加空间维与光谱维编码，减少特征传递时局部信息的丢失，实现了有效的局部、全局空谱特征提取。上述深度学习方法同时考虑光谱和空间特征，能够捕捉高层次的信息，获得稳定的分类结果。但是，现有方法一般比较复杂，分类效率较低，无法适应高光谱影像大幅宽大数据量的发展趋势。同时，像素级分类方法中的空间信息主要用于减少类别边缘的错分或降低分类结果的椒盐噪声（Chen等，2022），丢失了空间要素、结构和格局信息，缺乏像素、目标与场景关系的认知（李德仁等，2014）。为构建“像素—目标—场景”的层次性认知，宜采用更大的观测单元——场景，对高光谱影像进行解译分析。

遥感场景是指遥感影像中不同类型地表覆盖对象组成具有明确语义的区域，例如林地、居住区、公园等。同类别的遥感场景内部具有相似的

地物目标以及相近的空间分布模式，而相近的地物目标可以通过不同的空间语义关系形成不同的场景类别（张继贤等，2021）。遥感场景分类不仅需要研究场景内部的地物目标组成，还需要利用知识与规则，分析目标间的空间分布模式，实现对场景的高层语义信息提取。相对像素级分类，场景级分类具有更大的视野，能够从“像素—目标—场景”层次化的反映地物的分布规律，也更符合人类的认知规律。然而，现有遥感场景级分类主要研究对象是3个波段的高空间分辨率可见光影像（高分影像），通过构建遥感场景的空间表达实现场景识别。这种场景分类方法，仅仅利用地物的空间信息，忽视了地物本身的光谱信息。高分场景具有丰富的地物纹理信息，能够在精细尺度上反映地物目标的类型、数量与分布结构；而高光谱遥感影像的空间分辨率通常较低，空间信息的丰富度远不及高分影像，尽管如此，高光谱影像具有独特的连续光谱，从光谱角度拓展了地物的辨识能力，并获得了研究人员的关注。Liu等（2022b）提出了跨模态知识蒸馏框架，将三波段影像中学习到的先验知识通过教师网络迁移到多光谱场景分类问题，实验证明提出的模型在面对不同输入波段组合时同样有效，并且波段组合的改变可以影响各个类别的分类精度。遥感场景分类数据集多基于高分影像，例如航空图像数据集AID（Aerial Image Dataset）（Xia等，2017）、UC-Merced数据集（Yang和Newsam，2010）等，多光谱和高光谱场景分类数据集逐步增加（表1）。

表1 现有多光谱和高光谱遥感场景分类数据集

Table 1 Existing multispectral and hyperspectral remote sensing scene classification datasets

数据集名称	场景类别数	影像总数	波段总数	光谱范围/nm	空间分辨率/m	影像大小/像素	平台
天宫一号高光谱遥感场景分类数据集(刘康等,2020)	9	204	1(全色谱段)	500—800	5	512×512	卫星
			54(可见红外)	400—1000	10	128×128	
			52(短波红外)	1000—2500	20	256×256	
天宫二号自然场景分类数据集(Zhou等,2021)	10	20000	14	400—1040	100	128×128	卫星
面向场景分类的高光谱图像数据集(徐科杰等,2021)	5	1385	48	380—1050	1	256×256	航拍
				490—665,842	10		
EuroSAT数据集(Helber等,2019)	10	27000	13	705—783,865,1610,2190	20	64×64	卫星
				443,945,1375	60		
本文提出数据集	6	737	32	400—1000	10	64×64	卫星

(1) 天宫一号高光谱遥感场景分类数据集 TG1HRSSC (Tiangong-1 Hyperspectral Remote Sensing Scene Classification): 由刘康等 (2020) 利用天宫一号获得的可见近红外、短波红外、全色影像构建的数据集, 包括9个类别, 共204幅影像。

(2) 面向场景分类的高光谱图像数据集 HSRS-SC (Hyperspectral Remote Sensing data set for Scene Classification): 徐科杰等 (2021) 基于航拍高光谱影像构建的高光谱场景分类数据集。该数据集的场景影像兼具高空间分辨率与高光谱分辨率, 具有丰富地物纹理细节与高维光谱信息。

(3) 天宫二号自然场景分类数据集 NaSC-TG2 (Natural Scene Classification with Tiangong-2): 由Zhou等 (2021) 利用天宫二号获得的多光谱影像, 目视标注而成。该数据集包括10个场景类别, 共20000幅遥感场景影像, 具有样本均衡, 场景视野大等特点。

(4) EuroSAT数据集: Helber等 (2019) 利用哨兵2A (Sentinel-2A) 号卫星的多光谱数据构造了EuroSAT数据集。该团队将各波段图像重采样至10 m空间分辨率, 借助《欧洲城市地图集 (European Urban Atlas)》裁剪并标注10个类别, 共27000幅场景影像。

以上多光谱和高光谱场景分类数据集中, EuroSAT数据集与TG1HRSSC数据集中的场景影像由不同空间分辨率传感器获取的图像组合而成, 光谱连续性差, 波段间相关性受到限制, 不利于场景光谱信息的研究; NaSC-TG2数据集中, 空间分辨率低, 空间几何、纹理等细节特征少, 无法适应高光谱场景的空间信息挖掘; HSRS-SC数据集则源于航拍影像, 无法直接应用于空间尺度、细节纹理差异巨大的卫星场景分类任务。此外, 在这些数据集的构建过程中, 或是仅利用遥感影像的空间信息, 忽略了易造成特征解译不确定的场景内部光谱差异, 或是参考图过于特殊 (例如EuroSAT数据集参考的《欧洲城市地图集 (European Urban Atlas)》), 难以扩展到其他数据集的构造过程。

针对上述问题, 本文构建了珠海一号高光谱场景分类数据集 HSCD-ZH (Hyperspectral Scene Classification Dataset from Zhuhai-1), 并基于空间与光谱视角提供了现有多光谱和高光谱场景分类数据集的实验基准测试结果。具体来说, 数据集

来源于“珠海一号”卫星拍摄的高光谱影像, 制作过程中将像元光谱分类结果图与假彩色图像作为参考, 辅助定位、裁剪与标注待选场景影像, 再结合谷歌地图 (Google Earth) 高分影像进行人工目视筛选, 充分利用高光谱影像“图谱合一”的特点, 将影像的光谱特性结合到高光谱遥感场景分类数据集制作过程中, 选择内部光谱构成与空间纹理代表性强的场景作为数据集样本, 构建了具有统一空间分辨率与光谱分辨率的数据集, 并利用视觉词袋 BoVW (Bag of Visual Words) (Sivic 和 Zisserman, 2003)、卷积神经网络 CNN (Convolutional Neural Network) 等方法从空间角度与光谱角度对现有多光谱和高光谱场景分类数据集进行评估, 为进一步研究高光谱场景分类算法提供数据支持。

2 数据集介绍

2.1 数据集来源

原始高光谱影像 (图1) 来自欧比特公司“珠海一号”遥感维纳卫星星座的高光谱卫星 OHS-3A (Orbita Hyperspectral Satellites-3A)。该卫星采用3颗互补金属氧化物半导体 CMOS (Complementary Metal Oxide Semiconductor) 图像传感器拼接设计, 本文采用的影像由CMOS2号传感器, 在2020年7月2日获取于西安地区。OHS-3A卫星使用推扫成像方式, 空间分辨率10 m, 光谱分辨率2.5 nm, 波长范围400—1000 nm, 共有32个波段 (表2), 单颗卫星的重访时间约为6 d, 若联合星座中的8颗高光谱卫星, 则综合重访时间可缩短至1 d, 能够实现地表的动态监控与短时重访。选取的西安地区高光谱影像尺寸像素为5056×5056像素, 覆盖了西安北部、中部城区以及部分西部郊区, 其中城市区域占比最大, 其次是农田、荒地, 同时也包含丰富的林地、水体等。

2.2 数据集制作流程

数据集的制作流程如图2所示。原始高光谱影像为经过几何校正、相对辐射校正与全谱段配准等处理的L1B级标准产品。为消除传感器、大气等因素的影响, 遵循欧比特公司官方手册中的预处理方法, 根据元数据文件中的信息, 使用ENVI遥感图像处理平台对原始高光谱影像进行了辐射定标与大气校正。

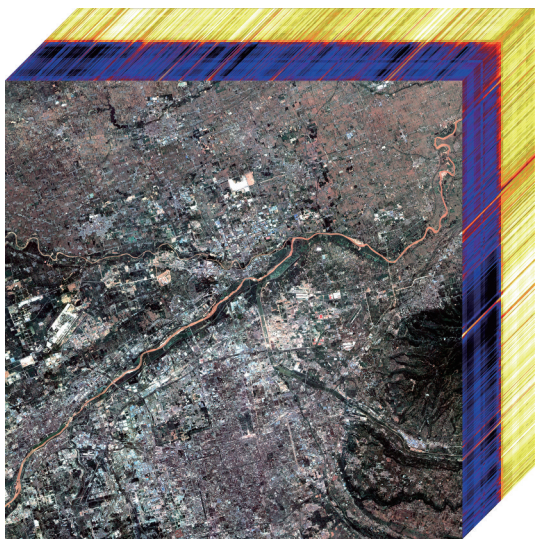


图 1 经过对比度增强的原始高光谱影像预览图
Fig. 1 Original hyperspectral image after contrast enhancement

表 2 珠海一号高光谱场景分类数据集谱段信息

Table 2 Band information of hyperspectral scene classification dataset from Zhuhai-1

波段号	中心波长/ nm	波长范围/ nm	波段号	中心波长/ nm	波长范围/ nm	波段号	中心波长/ nm	波长范围/ nm	波段号	中心波长/ nm	波长范围/ nm
B1	443	7	B9	580	13	B17	709	6	B25	833	8
B2	466	8	B10	596	13	B18	730	7	B26	850	8
B3	490	8	B11	620	5	B19	746	6	B27	865	8
B4	500	9	B12	640	6	B20	760	7	B28	880	8
B5	510	9	B13	665	5	B21	776	7	B29	896	8
B6	531	10	B14	670	6	B22	780	7	B30	910	8
B7	550	11	B15	686	6	B23	806	7	B31	926	10
B8	560	12	B16	700	6	B24	820	7	B32	940	9

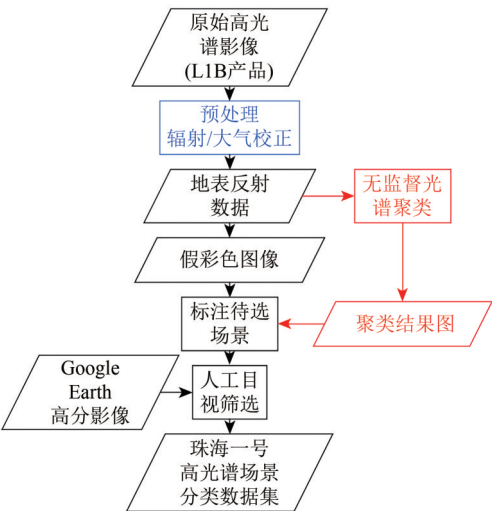


图 2 预处理与数据集制作流程
Fig. 2 Process of preprocessing and dataset creation

(2) 结合假彩色图像与聚类结果图进行样本标记。依据假彩色图像，可以相对容易地定位城区、乡村等场景，但在低分辨率卫星影像的复杂

现有多光谱和高光谱遥感场景分类数据集的标注多参考高分影像的场景分类数据集制作方法，依赖人工目视定位、标注场景影像，忽视了高光谱影像蕴含的多谱段信息，本数据集在构造过程中同时考虑光谱信息和空间信息：

(1) 采用 K-Means 对高光谱影像进行聚类。设置聚类中心为 16，遵循类内相似度大和类间距离大的算法原则，将各个像素划分到不同的类中，获得反映地物覆盖差异的聚类结果图（图 3）。较假彩色图像或 Google Earth 高分影像，获得的聚类结果图能够从光谱角度反映地物特性，且背景单一，可以快速确定地物（图 3 中标注为红色的水体与蓝色的林地）的分布与具体位置，提高定位待选场景位置的效率。

背景下精准定位水体、林地、农田等零星分布且极易混淆的场景是项具有挑战性的任务，聚类结果能够反映地表覆盖的差异，可为辨别地物所属类别、定位待选场景区域提供便利。本数据集制作过程中，首先寻找少量具有代表性的区域作为标准场景样本，依据假彩色图像与聚类结果图确定该类样本内部不同地物目标的形状、组成与分布情况；然后，遵循具有相近地物目标组成与分布规律的场景属于同一类别的原则，以标准场景样本为参考，寻找特征相近的区域，进行影像裁剪与标注，获取待选场景样本，形成初始的 HSCD-ZH 数据集。

(3) 参考 Google Earth 高分影像人工目视筛选场景影像。在上述样本标记过程中，易受原始影像低分辨率与聚类错误的影响出现标注混淆，故参考 Google Earth 影像进行目视筛选，排除标注错误或光谱组成与空间纹理、分布、结构等代表性

差无法分辨具体类别的场景影像，形成最终的HSCD-ZH数据集。

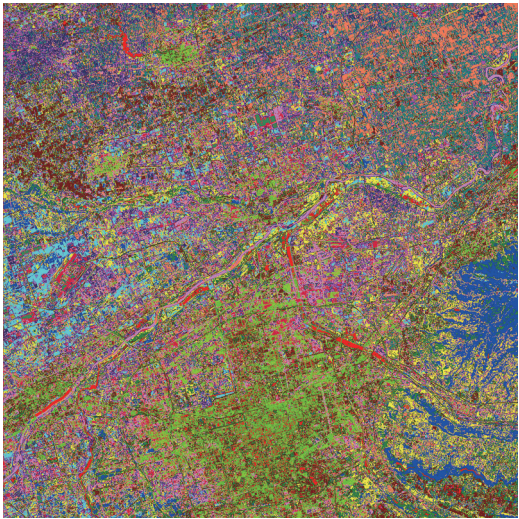


图3 高光谱影像无监督聚类结果图
Fig. 3 Results of unsupervised clustering of hyperspectral images

2.3 数据集分析

本文提出的数据集包括城市 (Urban)、乡村 (Rural)、农田 (Agriculture)、林地 (Forest)、未

利用地 (Unused land) 以及水体 (Water) 6类场景，共737幅高光谱遥感场景影像。每幅影像像素尺寸为64×64像素，空间分辨率为10 m，共32个波段，覆盖400—1000 nm波长范围。各个类别的数量分布与示例见图4与图5，其中图5的伪彩色图由B14 (670 nm)、B8 (560 nm)、B6 (531 nm) 3个波段组合而成，聚类图设置为16个聚类中心时获得。

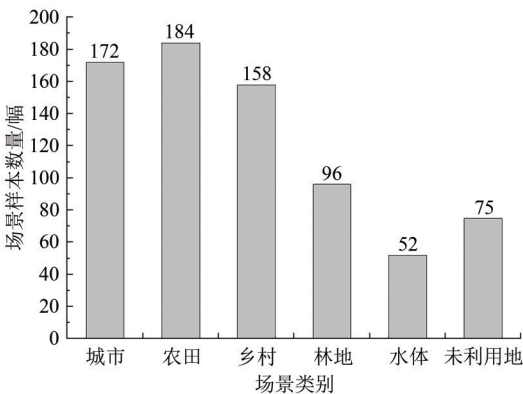


图4 数据集场景样本数量分布统计
Fig. 4 Distribution statistics of the number of scene samples in the dataset

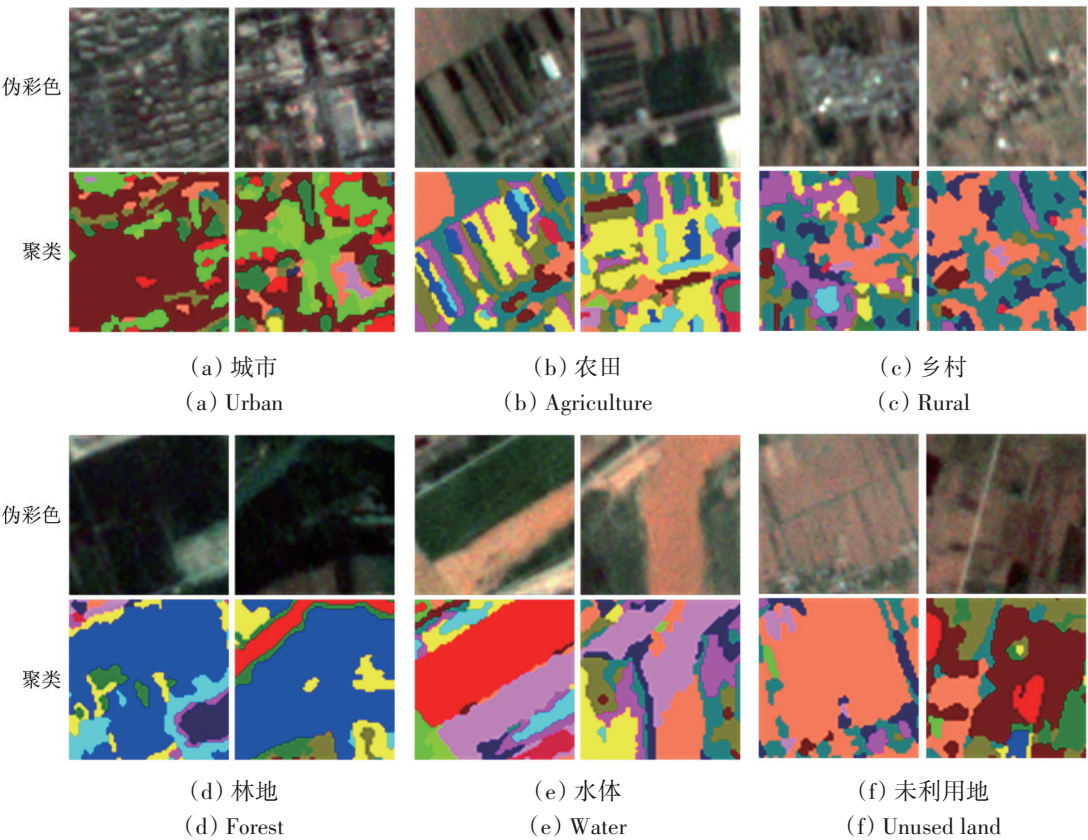


图5 珠海一号高光谱场景分类数据集样例及对应的聚类图
Fig. 5 Samples and corresponding clustering maps of HSCD-ZH

相较现有高光谱遥感场景分类数据集, 本数据集具有如下特点与优势:

(1) 构造过程考虑高光谱场景的地物组成和空间分布, 根据“像素—目标—场景”的逐层关联获取场景影像, 符合人类认知规律与高光谱影像智能解译需求。在构造过程中, 我们结合光谱聚类结果图与假彩色图像, 从光谱与空间2个视角, 识别像素与地物目标所属类别、提取目标间空间结构特征, 逐层推断场景类别, 降低裁剪的不确定性, 减少场景标注的遗漏与混淆, 可为场景与地物目标关系的研究提高良好数据支持。

(2) 保证类内场景多样, 具有丰富的视觉与光谱表达; 同时部分类间场景具有相似特征, 可为鲁棒性高、泛化性能强的算法研究提供良好的数据支撑。本数据集的制作过程中, 参考了光谱聚类结果并进行目视筛选, 获得的各类场景或具有代表性的光谱组成, 或具有特定的空间要素、结构与格局, 或二者兼具, 例如图6中水体1与水体2所属同类, 但因为地理位置与所含悬浮物的差异, 空间结构与光谱构成存在差异; 图6中的水体2与林地分属不同类别, 空间特征相似, 但光谱特性截然不同。

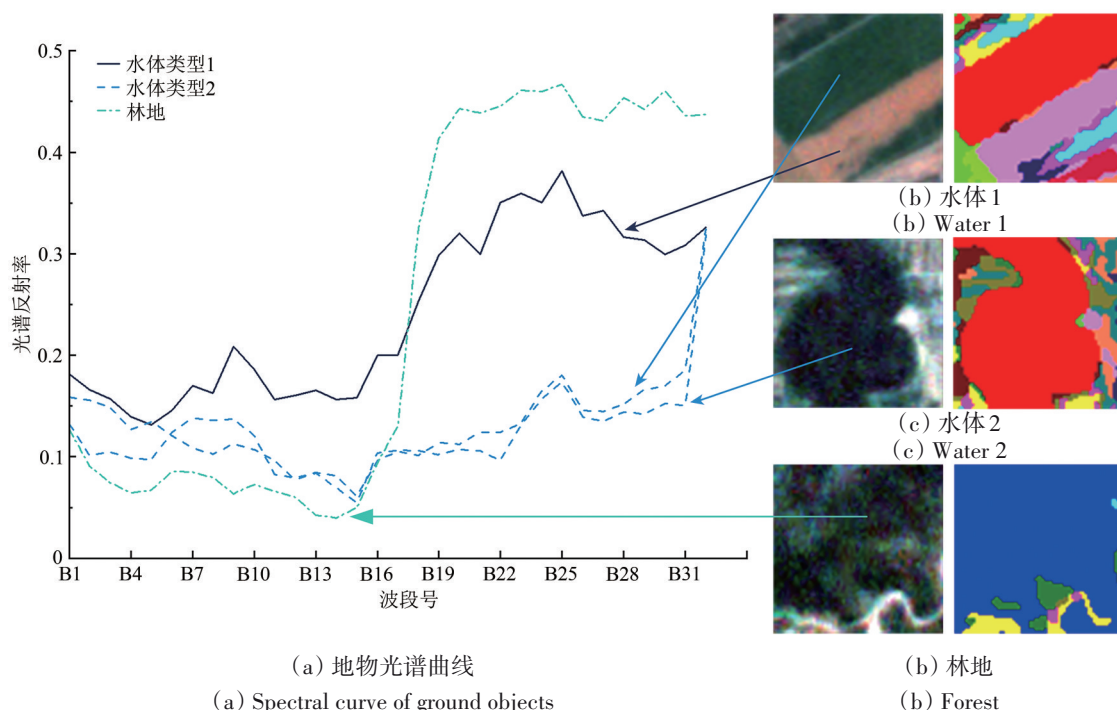


图6 场景样本典型地物光谱曲线
Fig. 6 Spectral curve of typical objects in scene samples

(3) 原始数据来源于欧比特公司开放的免费科研数据, 涵盖多个时相、地区和场景的高光谱数据易于获取。且数据集构造过程中仅依赖Google Earth高分影像来验证标注的正确性, 所需外来数据少, 使得场景的种类和数量更易扩展。

3 实验结果与分析

3.1 数据集描述

实验分别在现已公开的多光谱和高光谱场景分类数据集EuroSAT、NaSC-TG2与本数据集上进行。现有公开数据集介绍如下:

(1) EuroSAT数据集包括Industrial Buildings、

Residential Buildings、Annual Crop、Permanent Crop、River、Sea & Lake、Herbaceous Vegetation、Highway、Pasture 和 Forest 10类场景。每类场景有2000—3000幅样本, 共27000幅场景影像。每幅影像尺寸均为64×64像素, 空间分辨率为10 m。

(2) NaSC-TG2包括Beach、Circle Farmland、Cloud、Desert、Forest、Mountain、Rectangle Farmland、Residential、River、Snowberg 10类场景。每个类别具有2000幅样本, 共20000幅场景影像。每幅影像空间分辨率为100 m, 尺寸为128×128像素。具有单幅场景的地表覆盖范围大、各类数量分布均衡等特点。

3.2 评价指标

使用总体精度 OA (Overall Accuracy) 作为实验结果评价指标, 计算方法如下:

$$OA = \frac{N_{True}}{N} \quad (1)$$

式中, N_{True} 代表正确分类的测试样本数量, N 代表测试集样本总数量。实验中所有数据集均以 6:4 随机划分为训练集与测试集, 在实验结果中, 给出 5 次实验的 OA 均值与标准差的无偏估计 STD (Unbiased Estimation of Standard Deviation), 计算公式如下:

$$STD = \sqrt{\frac{1}{4} \sum (x - \bar{x})^2} \quad (2)$$

式中, x 代表单次实验的 OA 值, $\bar{x}$ 代表 5 次实验的 OA 均值。统一使用支持向量机 SVM (Support Vector Machine) 作为分类器。

3.3 实验方法

高光谱场景具有“图谱合一”的特点, 高维光谱提供地物特定物理信息的同时, 大幅提升了处理图像立方体的难度, 无法直接采用基于 3 波段图像的高分场景分类方法。本文从 2 种不同角度获取整幅场景的描述性编码, 验证不同视角下的分类精度:

(1) 基于光谱视角: 将各像素的高维光谱信号作为原始基础特征, 采用 BoVW、改进 Fisher 核 IFK (Improved Fisher Kernel) (Perronnin 等, 2010) 或局部约束线性编码 LLC (Locality-constrained Linear Coding) (Wang 等, 2010) 编码方法, 对原始光谱信号进行空间维度聚合, 形成兼具光谱信息与空间信息的全局场景特征。

(2) 基于空间视角: 将高光谱影像的光谱维降至 3 个维度, 本文根据高光谱数据的真实波长选择 3 个光谱波段组成假彩色图像, 如 HSCD-ZH、EuroSAT 与 NaSC-TG2 分别选择 B14(670 nm)、B8(560 nm)、B6(531 nm), 4(665 nm)、3(560 nm)、2(490 nm), 8(665 nm)、10(565 nm)、12(490 nm) 号波段。在假彩色图像基础上, 可借用高分场景分类方法提取特征, 进行场景识别。本文分别使用非深度与深度 2 类方法从降维后的场景中提取空间特征, 进行基于空间视角的实验。

采用非深度学习方法提取特征时, 首先从局部场景中提取颜色直方图 CH (Color Histogram) (Swain 和 Ballard, 1991)、局部二值模式 LBP

(Local Binary Pattern) (Ojala 等, 2002) 或尺度不变特征变换 SIFT (Scale Invariant Feature Transform) (Lowe, 2004) 基础空间特征, 然后使用流行的 BoVW、IFK 或 LLC 方法进行编码, 构建场景样本的全局描述符 (Xia 等, 2017)。

深度学习方面, 采用卷积神经网络和 Swin Transformer (Liu 等, 2021) 模型提取高光谱影像的空间特征。本文选择 ResNet34 (He 等, 2016) 与 VGG16 (Simonyan 和 Zisserman, 2014) 作为卷积神经网络的代表, 应用到场景分类的基准测试中。

基于空间视角的方法在光谱降维过程中损失了大量光谱信息, 基于光谱视角的方法则只考虑了不同地物在整幅场景中的组成比例, 遗漏了空间结构、分布信息。为进一步研究使用空谱特征完成分类任务, 我们将获得的全局光谱特征直接与 CNN、Swin Transformer 提取的空间特征进行级联, 生成兼具光谱信息与空间信息的场景全局描述符, 进行了第 3 类实验。

3.4 实验设置

3.4.1 方法设置

光谱视角的实验中, 采用像元光谱信号作为基础特征。基于空间视角的实验中, 分别将 CH、LBP 与 SIFT 特征作为图像的基本空间特征。

(1) CH (Swain 和 Ballard, 1991): 分别在图像块的 3 个通道中, 统计各区间数值出现的频次, 形成 3 个直方图, 拼接后进行归一化, 即可获得图像的 CH 特征。本文设置各个通道具有 32 个数值区间, 最后获得的特征向量具有 96 维。

(2) LBP (Ojala 等, 2002): 将中心像素与 8 个邻域像素灰度值比较, 获得的 8 位二进制数即为中心像素的 LBP 描述符。统计图像块中 LBP 描述值的直方图, 可以得到代表图像空间纹理特征的特征向量。

(3) SIFT (Lowe, 2004): SIFT 特征是通过计算图像块的梯度直方图得到的。实验中, 将场景图像划分为 16 个子区域, 计算子区域的梯度直方图, 拼接形成 128 维特征向量。

基于上述基础特征, 分别使用 BoVW、LLC 与 IFK 方法对基础特征进行编码形成图像描述符。

(1) BoVW (Sivic 和 Zisserman, 2003) 方法采用聚类从 SIFT 等基础特征中学习字典, 编码过程中, 统计场景中各个基础词汇出现的频次, 形成

可以代表全局特征的直方图；

(2) 类似 BoVW, LLC (Wang 等, 2010) 方法的字典也是通过聚类获得, 不同点是, LLC 目标是通过字典内的局部基获取基础特征的稀疏性表示, 最后对场景中获得的所有编码特征进行池化形成场景全局描述符；

(3) 不同于以上 2 种方法, IFK (Perronnin 等, 2010) 方法使用混合高斯模型对特征的分布进行建模, 基于构建的对数似然函数, 针对单个参数, 使用单幅场景中提取的所有基础特征计算梯度向量, 将所有参数获得的梯度向量拼接形成代表图像的描述幅 (具有 $2 \times K \times D$ 维, K 为高斯分布的数量即字典数, D 为基本特征向量的维度)。

本文选用 VGG16、ResNet34 与 Swin Transformer 网络模型进行基于空间视角的场景分类实验。

(1) VGG 模型 (Simonyan 和 Zisserman, 2014) 使用了较小的 3×3 卷积核与深层网络结构, 广泛应用于包括 TG1HRSSC、HSRS-SC 与 NaSC-TG2 在内的验证实验中。本文采用的 VGG16 模型具有 6 个块状结构, 前 5 部分分别具有 2, 2, 3, 3, 3 个卷积层, 各层之间均由最大池化层 (Maxpool) 分离, 完成分类任务的第 6 层由 3 个全连接层与 1 个 Softmax 层构成, 本文将第 1 个全连接层输出的 4096 维向量作为该网络提取的单幅影像特征。

(2) ResNet 模型 (He 等, 2016) 使用残余网络模块缓解了训练过程中出现的梯度爆炸或消失问题, 极大的提升了可训练网络的深度。本文选择具有 34 层的 ResNet-34 网络作为基准模型, 该模型具有 6 部分, 前 5 部分进行特征提取, 分别具有 1, 6, 8, 12, 6 个卷积层, 第 6 部分对特征进行平均池化, 将获得的 512 维特征向量输入到最后的全连接层中完成场景影像的分类。

(3) 2018 年提出的 Transformer (Vaswani 等, 2017), 凭借自注意力机制与前馈神经网络结构, 成为了与 CNN、循环神经网络 RNN (Recurrent Neural Network) 并列的流行深度特征提取模型。本文采用在计算机视觉任务中展现优异性能的 Swin Transformer (Liu 等, 2021) 作为提取深度特征的对比模型。其具有 4 个 CNN 层次化设计的 Stage, 分别具有 2、2、6、2 个 Swin Transformer Block, 其中多头自注意力的 Heads 数设置为 3、6、12、24, 首个 Stage 具有 96 个隐层, 第 4 个 Stage 通过一个 LayerNorm 层与 Global Average Pooling 层输出了 768 维的图像特征, 最后通过 1 个全连接层完

成场景分类。

3.4.2 参数设置

进行基于空间视角的非深度学习实验时, 设置步长为 8, 将假彩色图像划分为多个 16×16 大小的网格, 分别提取 CH、LBP 或 SIFT 基本空间特征。为减少后续的计算量, 会对每幅场景提取的 CH、LBP、SIFT 与光谱特征进行随机采样, 保证基本特征的数量不超过 200000。为探索字典大小对分类精度的影响, 将 BoVW 与 LLC 的字典大小设置为 128 到 4096, 过大的字典对 IFK 没有意义, 其字典大小设置为 16 到 512。进行基于光谱角度的实验时, 字典大小分别设置为 128 到 16384 与 16 到 2048。

使用深度网络模型在 3 个数据集上进行训练时, Batch Size 统一设置为 16, Epoch 设置为 200。测试不同起始学习率 $\{0.1, 0.05, 0.01, 0.005, 0.001, 0.0005\}$ 对 HSCD-ZH 数据集分类精度的影响, 选择获得最高精度的 0.001 与 0.0005 作为 ResNet34 与 VGG16 的初始学习率; 学习率调整策略采用随机梯度下降算法 SGD (Stochastic Gradient Descent), 依据经验, 权重衰减设置为 0.0005, Momentum 设置为 0.9, 在 Epoch 为 60、120、160 时调整学习率为当前的 0.2 倍。进行 Swin Transformer 训练时, 设置训练集大小为 512 时, 基础学习率为 0.0005, 训练过程中依据数据集大小自适应调整学习率。在进行深度学习实验时, 所有图像尺寸重采样为 224×224 像素。

3.5 结果及分析

3.5.1 基于光谱视角的不同分类方法对比结果

图 7 展示了字典大小对基于光谱视角方法的分类精度影响。可以发现, 使用 K-Means 聚类获取字典的 BoVW 与 LLC 方法具有相近的变化趋势: 小字典无法分离用于分类的局部特征, 所以字典增加时, 分类精度逐步提升; 字典过大则可能导致相近特征分属两个簇类, 影响分类性能, 所以字典过大后, 时间与空间复杂度增加, 精度的提升幅度却不大, 甚至有所有下降。字典大小的增加对 IFK 的总体精度影响较小。结合表 3, 可以发现, 在 EuroSAT 与 HSCD-ZH 数据集上, IFK 方法获得的精度最高, BoVW 最差, 在数据集 NaSC-TG2 中则表现相反。NaSC-TG2 数据集的场景内部多为大尺度的同质区域, 异类场景间地物组成相差较大, 如 Beach、Cloud、Forest 等, 因而使用词

典内单词组合表示图像的 BoVW 与 LLC 可获得更易于辨识、分离的特征。相比 NaSC-TG2, 其他 2 个数据集中场景的空间分辨率更高, 地物组成更为复杂多样, IFK 展现了优异的性能和稳定, 反映出其对复杂场景具有较强的建模能力。

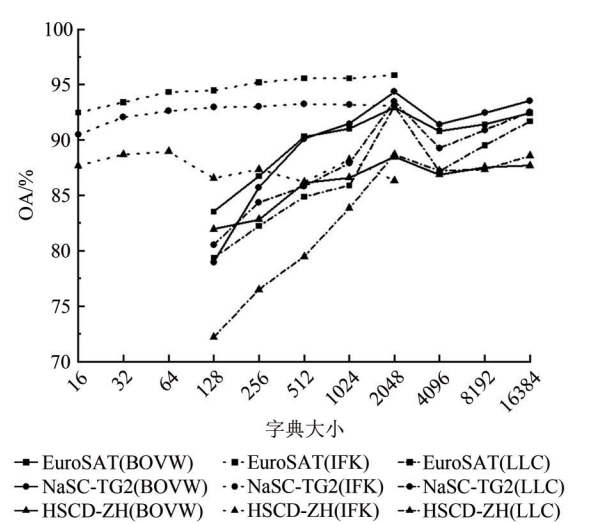


图7 在EuroSAT,NaSC-TG2与HSCD-ZH数据集上,字典大小对基于光谱视角方法性能的影响

Fig. 7 Impact of dictionary size on the performance of spectral-based methods on EuroSAT, NaSC-TG2, and HSCD-ZH

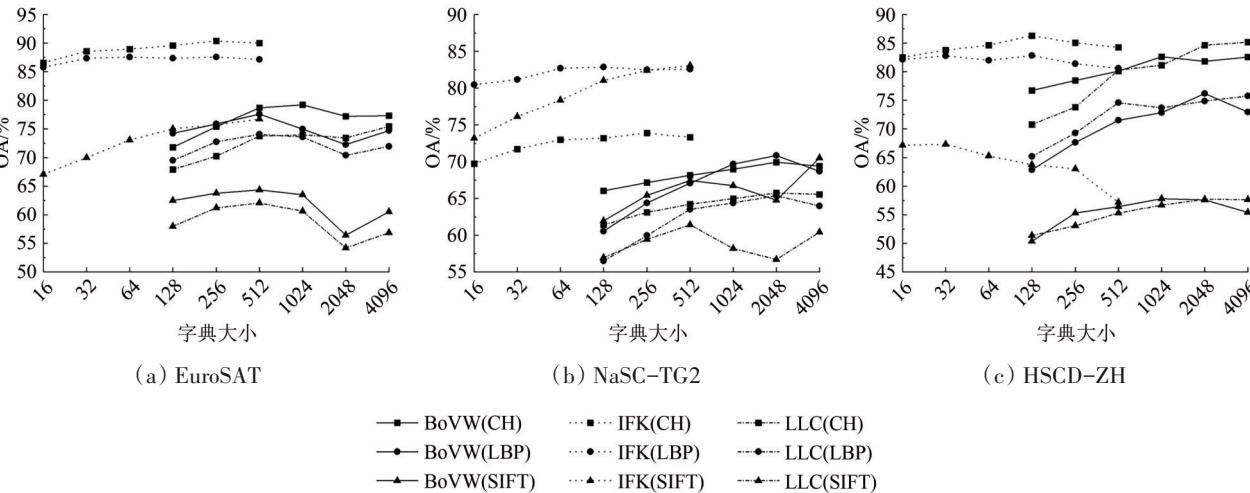


图8 不同数据集上字典大小对基于空间视角的非深度学习方法性能的影响

Fig. 8 Impact of dictionary size on the performance of non-deep learning spatial-based methods on different datasets

表4中展示了基于空间视角,使用不同基本特征提取方法对总体精度的影响。

结合图8可以发现,进行EuroSAT与HSCD-ZH数据集的分类实验时,采用空间纹理作为特征的SIFT与LBP没有展现良好性能,这主要由于高光谱影像为卫星拍摄,空间分辨率较低,纹理信息不丰富,且这2个数据集在制作过程中不完全依

表3 在EuroSAT,NaSC-TG2和HSCD-ZH数据集上,基于光谱视角的不同方法的分类总体准确率均值及其无偏估计

Table 3 Mean of overall accuracy of classification and its unbiased estimation of standard deviation for different spectral-based methods on EuroSAT, NaSC-TG2, and HSCD-ZH

对比方法	数据集		
	EuroSAT	NaSC-TG2	HSCD-ZH
BoVW (Spectral)	92.92(0.2725)	94.34(0.1683)	88.48(1.5864)
IFK (Spectral)	95.85(0.1205)	93.26(0.3721)	88.96(1.4944)
LLC (Spectral)	93.03(0.1783)	93.48(0.1740)	88.69(1.6598)

注:表中粗体数值为数据集上获得的最优分类精度均值;括号中的数字为5次实验总体精度的无偏估计。

3.5.2 基于空间视角的不同分类方法对比结果

基于空间视角,使用不同大小的字典对3个数据集分类的结果如图8所示。可见,字典大小对基于光谱与空间视角方法的精度影响相近:一定区间内,字典增大分类精度增加;字典过大时,精度的提升较缓甚至略有下降。

明显，此外，数据集中存在颜色相近的但不属于同类别的场景，所以使用CH特征没有展现较好性能，但在所有基本特征提取方法中，CH特征最具鲁棒性，这主要因为3个数据集中，大量同类高光谱场景在所选的3个波段具有相近的反射值，比如，林地多反射绿光，荒地多呈现黄色等。

表 4 在 EuroSAT,NaSC-TG2 和 HSCD-ZH 数据集上,基于空间视角的不同方法的分类总体准确率均值及其无偏估计

Table 4 Mean of overall accuracy of classification and its unbiased estimation of standard deviation for different spatial -based methods on EuroSAT, NaSC-TG2, and HSCD-ZH

方法类型	对比方法	数据集			/%
		EuroSAT	NaSC-TG2	HSCD-ZH	
非深度学习(SVM)	BoVW (CH)	78.96(0.5155)	69.12(0.7788)	82.63(2.1108)	
	IFK (CH)	90.12(0.2507)	74.23(0.2961)	86.26(0.8715)	
	LLC (CH)	75.70(0.3695)	65.88(0.5732)	85.19(0.9816)	
	BoVW (LBP)	77.32(0.3008)	70.32(0.2768)	76.16(0.4390)	
	IFK (LBP)	87.80(0.1878)	82.81(0.1732)	82.83(0.4124)	
	LLC (LBP)	74.24(0.1495)	65.29(0.5535)	75.76(1.2371)	
	BoVW (SIFT)	64.41(0.3930)	70.94(0.4190)	57.78(3.4386)	
	IFK (SIFT)	76.54(0.3875)	83.22(0.4939)	67.34(0.4762)	
	LLC (SIFT)	62.11(0.4083)	60.62(0.4380)	57.71(0.8780)	
深度学习(End to End)	ResNet34	97.23(0.1810)	93.45(0.1282)	93.56(0.8644)	
	VGG16	97.69(0.1498)	93.87(0.2297)	90.17(0.7207)	
	Swin Transformer	97.49(0.1380)	94.78(0.2371)	91.32(1.5272)	
深度学习(SVM)	ResNet34	96.98(0.1688)	93.19(0.1358)	92.34(1.3259)	
	VGG16	97.59(0.1263)	93.70(0.2427)	90.03(0.6158)	
	Swin Transformer	97.47(0.1397)	94.74(0.1393)	90.17(1.590)	

注：表中粗体数值为最优结果。

从表4可以发现，在3个数据集上，采用IFK对CH、LBP与SIFT编码获得的精度均高于另外2个方法，其在基于光谱与空间视角的实验中表现出较高的鲁棒性，BoVW 优于 LLC，但总体精度相差较小。比较表3与表4中的结果，基于空间视角的非深度学习方法的总体精度不及基于光谱视角的方法，一方面是因为图像降维过程中损失了大量信息，另一方面采用的方法特征提取能力较差，难以获得较高精度。

表4中展示了使用VGG16、ResNet34与Swin Transformer从空间角度对各数据集进行分类的实验结果。总体上，使用SVM作为分类器的分类精度略小于端到端分类结果，深度学习方法较非深度学习方法展现了更为优异和稳定的性能。使用3种层数差异较大的模型对3个数据集进行分类，获得的最终分类精度相近，说明采用的方法接近了仅使用所选波段完成分类的极限。

3.5.3 基于光谱与空间特征级联的不同分类方法对比结果

本文将获得的光谱特征与端到端网络提取的

空间特征级联，作为新的空谱特征，进行场景的分类实验。图9与表5展示了级联不同特征的分类结果，可以发现，光谱特征的加入，增加了互补光谱信息，分类精度与稳定性具有明显提升。在EuroSAT与HSCD-ZH数据集上，使用LLC编码的光谱特征与深度空间特征融合获得的提升最大，可以看出其采用的稀疏性编码降低了拼接后特征的复杂度，在进行分类时更为有效。NaSC-TG2数据集的实验中，LLC与BoVW提取的光谱特征对分类精度的提升增益较小，字典增加时也没有明显的提升，说明场景内部具有大尺度的均一、同质区域时，2种方法与基于空间角度的方法可分辨的场景有很大交叉，提供的互补光谱信息很少，IFK则可以分辨组成成分复杂的场景，并提供相应互补光谱信息。综合以上实验，可以了解从光谱与空间单一角度完成场景立方体的分类任务均具有局限性，二者合理结合能够有效利用立方体所蕴含的信息弥补各自的缺陷，提升分类性能。

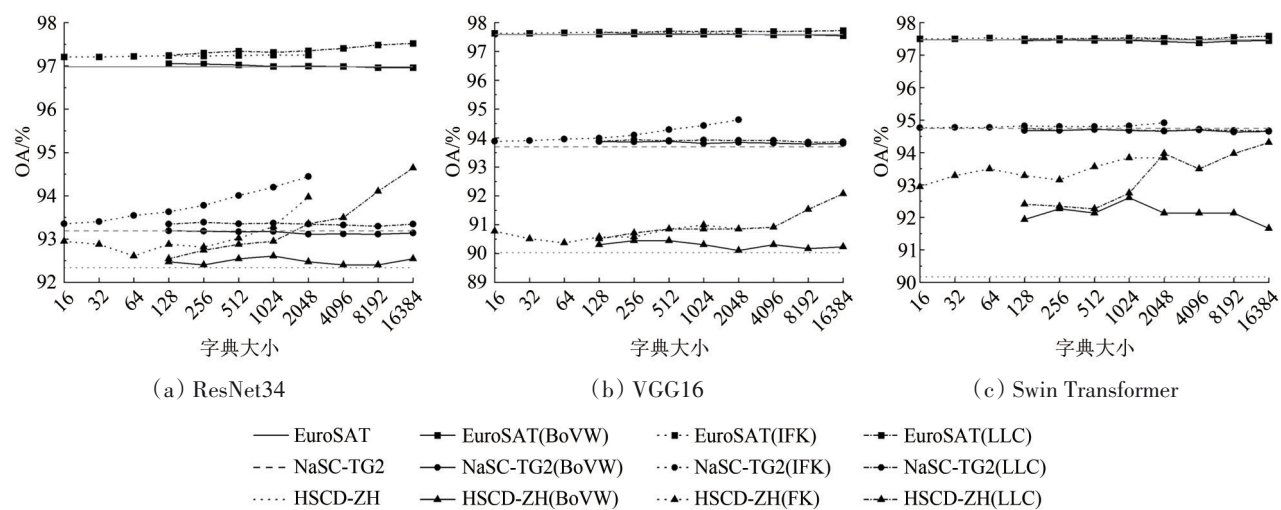


图9 不同数据集上,字典大小对级联不同光谱与空间特征的分类总体精度的影响,空间特征分别获取自 ResNet34, VGG16, Swin Transformer

Fig. 9 Impact of dictionary size on the overall accuracy of cascading different spectral and spatial features for classification on different datasets, spatial features obtained by ResNet34, VGG16, Swin Transformer

表 5 在 EuroSAT, NaSC-TG2 和 HSCD-ZH 数据集上,基于不同级联特征的分类总体准确率均值及其无偏估计
Table 5 Mean of overall accuracy of classification and its unbiased estimation of standard deviation for different cascade features on EuroSAT, NaSC-TG2, and HSCD-ZH

		EuroSAT		NaSC-TG2		HSCD-ZH	
空间特征提取算法	光谱特征提取算法	空间特征精度	级联特征精度	空间特征精度	级联特征精度	空间特征精度	级联特征精度
ResNet34	BoVW		97.05(0.1690)	93.19(0.1369)		92.61(1.7516)	
	IFK	96.98(0.1688)	97.25(0.1952)	93.19(0.1358)	94.44(0.2846)	92.34(1.3259)	93.97(1.5046)
	LLC		97.52(0.2234)		93.39(0.1706)		94.64(0.9999)
VGG16	BoVW		97.61(0.1340)		93.89(0.2356)		90.44(0.5570)
	IFK	97.59(0.1263)	97.67(0.1442)	93.70(0.2427)	94.63(0.1597)	90.03(0.6158)	90.98(0.5141)
	LLC		97.72(0.1448)		93.94(0.2094)		92.07(0.7029)
Swin Transformer	BoVW		97.46(0.1496)		94.69(0.1595)		92.61(1.3432)
	IFK	97.47(0.1397)	97.53(0.1146)	94.74(0.1393)	94.91(0.1786)	90.17(1.5900)	93.83(1.4659)
	LLC		97.59(0.1285)		94.74(0.1631)		94.31(1.8782)

注: 表中粗体数值为最优结果。

4 结 论

针对高光谱像素级分类面临难以适应当前智能快速解译需求的现状, 本文结合现有高光谱场景分类数据集的缺陷与优势, 以“珠海一号”高光谱卫星拍摄的西安地区影像为基础, 采用“像素—目标—场景”的层次化构造方法, 构建了具有 6 个类别, 737 幅场景样本的珠海一号高光谱场景分类数据集 HSCD-ZH (Hyperspectral Scene Classification Dataset from Zhuhai-1)。通过充分利用高光谱影像“图谱合一”的特点, 将光谱聚类结果结合到数据集构造过程中, 通过构建全局像

素间的相关性, 有效辅助判别地物类型并高效定位待选场景, 从而降低了数据集构造过程中的不确定性, 再结合 Google Earth 高分影像减少场景的混淆与误标。由此获得的高光谱遥感场景分类数据集, 具有符合高光谱影像层次化智能解译需求、类内场景多样、扩展性强等特点。同时, 探讨了基于光谱视角、空间视角与空谱联合的高光谱场景分类方法, 结果表明, 合理地结合空间与光谱互补信息可以有效提升分类性能。具有连续光谱, 通过层次化构建的 HSCD-ZH 在研究地物目标与地物目标间关系、地物光谱与高层语义间关系、空谱互补信息提取等方面具有广泛的应用价值。

进一步的研究应在当前数据集的基础上开展高光光谱场景空谱特征提取算法的研究。同时, 为了使研究更具广泛适应性, 逐步扩展数据集的类别数与样本数, 此外, 可以考虑引入像素级标注或对应的高分场景影像形成更为层次化、适应多种分类任务的数据集。

志 谢 本研究数据集基于欧比特公司提供的高光谱卫星影像构建, 在此表示衷心感谢!

参考文献(References)

- Benediktsson J A, Palmason J A and Sveinsson J R. 2005. Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3): 480-491 [DOI: 10.1109/TGRS.2004.842478]
- Chen R, Pu D, Tong Y and Wu M H. 2022. Image-denoising algorithm based on improved K-singular value decomposition and atom optimization. *CAAI Transactions on Intelligence Technology*, 7(1): 117-127 [DOI: 10.1049/cit2.12044]
- Du P J, Xia J S, Xue Z H, Tan K, Su H J and Bao R. 2016. Review of hyperspectral remote sensing image classification. *Journal of Remote Sensing*, 20(2): 236-256 (杜培军, 夏俊士, 薛朝辉, 谭琨, 苏红军, 鲍蕊. 2016. 高光光谱遥感影像分类研究进展. *遥感学报*, 20(2): 236-256) [DOI: 10.11834/jrs.20165022]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- Helber P, Bischke B, Dengel A and Borth D. 2019. EuroSAT: a novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(7): 2217-2226 [DOI: 10.1109/JSTARS.2019.2918242]
- Kang X D, Li S T and Benediktsson J A. 2014. Spectral-spatial hyperspectral image classification with edge-preserving filtering. *IEEE Transactions on Geoscience and Remote Sensing*, 52(5): 2666-2677 [DOI: 10.1109/TGRS.2013.2264508]
- Kuras A, Brell M, Rizzi J and Burud I. 2021. Hyperspectral and lidar data applied to the urban land cover machine learning and neural-network-based classification: a review. *Remote Sensing*, 13(17): 3393 [DOI: 10.3390/rs13173393]
- Li D R, Zhang L P and Xia G S. 2014. Automatic analysis and mining of remote sensing big data. *Acta Geodaetica et Cartographica Sinica*, 43(12): 1211-1216 (李德仁, 张良培, 夏桂松. 2014. 遥感大数据自动分析与数据挖掘. *测绘学报*, 43(12): 1211-1216) [DOI: 10.13485/j.cnki.11-2089.2014.0187]
- Li S T, Song W W, Fang L Y, Chen Y S, Ghamisi P and Benediktsson J A. 2019. Deep learning for hyperspectral image classification: an overview. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9): 6690-6709 [DOI: 10.1109/TGRS.2019.2907932]
- Liu H, Li W, Xia X G, Zhang M M, Gao C Z and Tao R. 2022a. Central attention network for hyperspectral imagery classification. *IEEE Transactions on Neural Networks and Learning Systems*, 34(11): 8989-9003 [DOI: 10.1109/TNNLS.2022.3155114]
- Liu H, Qu Y and Zhang L Q. 2022b. Multispectral scene classification via cross-modal knowledge distillation. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1-12 [DOI: 10.1109/TGRS.2022.3174352]
- Liu K, Zhou Z, Li S Y, Liu Y F, Wan X, Liu Z W, Tan H and Zhang W F. 2020. Scene classification dataset using the Tiangong-1 hyperspectral remote sensing imagery and its applications. *Journal of Remote Sensing*, 24(9): 1077-1087 (刘康, 周壮, 李盛阳, 刘云飞, 万雪, 刘志文, 谭洪, 张万峰. 2020. 天宫一号高光光谱遥感场景分类数据集及应用. *遥感学报*, 24(9): 1077-1087) [DOI: 10.11834/jrs.20209323]
- Liu Y N. 2021. Development of hyperspectral imaging remote sensing technology. *National Remote Sensing Bulletin*, 25(1): 439-459 (刘银年. 2021. 高光光谱成像遥感载荷技术的现状与发展. *遥感学报*, 25(1): 439-459) [DOI: 10.11834/jrs.20210283]
- Liu Z, Lin Y T, Cao Y, Hu H, Wei Y X, Zhang Z, Lin S and Guo B N. 2021. Swin transformer: hierarchical vision transformer using shifted windows//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. Montreal: IEEE: 10012-10022 [DOI: 10.1109/ICCV48922.2021.00986]
- Lowe D G. 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2): 91-110 [DOI: 10.1023/B:VISI.0000029664.99615.94]
- Ojala T, Pietikainen M and Maenpää T. 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7): 971-987 [DOI: 10.1109/TPAMI.2002.1017623]
- Perronnin F, Sánchez J and Mensink T. 2010. Improving the fisher kernel for large-scale image classification//*Proceedings of the 11th European Conference on Computer Vision*. Heraklion: Springer: 143-156 [DOI: 10.1007/978-3-642-15561-1_11]
- Simonyan K and Zisserman A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv: 1409.1556* [DOI: 10.48550/arXiv.1409.1556]
- Sivic J and Zisserman A. 2003. Video Google: a text retrieval approach to object matching in videos//*Proceedings of the 9th IEEE International Conference on Computer Vision*. Nice: IEEE: 1470-1470 [DOI: 10.1109/ICCV.2003.1238663]
- Swain M J and Ballard D H. 1991. Color indexing. *International Journal of Computer Vision*, 7(1): 11-32 [DOI: 10.1007/BF00130487]
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Ł and Polosukhin I. 2017. Attention is all you need// *Pro-*

- ceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc. [DOI: 10.48550/arXiv.1706.03762]
- Wang J J, Yang J C, Yu K, Lv F J, Huang T and Gong Y H. 2010. Locality-constrained linear coding for image classification//Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco: IEEE: 3360-3367 [DOI: 10.1109/CVPR.2010.5540018]
- Xia G S, Hu J W, Hu F, Shi B G, Bai X, Zhong Y F, Zhang L P and Lu X Q. 2017. AID: a benchmark data set for performance evaluation of aerial scene classification. IEEE Transactions on Geoscience and Remote Sensing, 55(7): 3965-3981 [DOI: 10.1109/TGRS.2017.2685945]
- Xu K J, Deng P F and Huang H. 2021. HSRS-SC: a hyperspectral image dataset for remote sensing scene classification. Journal of Image and Graphics, 26(8): 1809-1822 (徐科杰, 邓培芳, 黄鸿. 2021. HSRS-SC: 面向遥感场景分类的高光谱图像数据集. 中国图象图形学报, 26(8): 1809-1822) [DOI: 10.11834/jig.200835]
- Yang X F, Cao W J, Lu Y and Zhou Y C. 2022. Hyperspectral image transformer classification networks. IEEE Transactions on Geoscience and Remote Sensing, 60: 1-15 [DOI: 10.1109/TGRS.2022.3171551]
- Yang Y and Newsam S. 2010. Bag-of-visual-words and spatial extensions for land-use classification//Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems. San Jose: ACM: 270-279 [DOI: 10.1145/1869790.1869829]
- Zhang C Y, Li J, Lei S G, Yang J Z and Yang N. 2022. Progress and prospect of the quantitative remote sensing for monitoring the eco-environment in mining area. Metal Mine, (3): 1-27 (张成业, 李军, 雷少刚, 杨金中, 杨楠. 2022. 矿区生态环境定量遥感监测研究进展与展望. 金属矿山, (3): 1-27) [DOI: 10.19614/j.cnki.jsks.202203001]
- Zhang J X, Gu H Y, Yang Y, Zhang H and Li H T. 2021. Research progress and trend of high-resolution remote sensing imagery intelligent interpretation. National Remote Sensing Bulletin, 25(11): 2198-2210 (张继贤, 顾海燕, 杨懿, 张鹤, 李海涛. 2021. 高分辨率遥感影像智能解译研究进展与趋势. 遥感学报, 25(11): 2198-2210) [DOI: 10.11834/jrs.20210382]
- Zhong L, Qian J W, Chu X Y, Qian Z H, Wang M and Li J L. 2023. Monitoring heavy metal contamination of wheat soil using hyperspectral remote sensing technology. Transactions of the Chinese Society of Agricultural Engineering, 39(5): 265-270 (钟亮, 钱家炜, 储学远, 钱志红, 王森, 李建龙. 2023. 利用高光谱遥感技术监测小麦土壤重金属污染. 农业工程学报, 39(5): 265-270) [DOI: 10.11975/j.issn.1002-6819.202207160]
- Zhou Z, Li S Y, Wu W, Guo W L, Li X, Xia G S and Zhao Z F. 2021. NaSC-TG2: natural scene classification with tiangong-2 remotely sensed imagery. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14: 3228-3242 [DOI: 10.1109/JSTARS.2021.3063096]

Hyperspectral scene classification dataset based on Zhuhai-1 images

LIU Yuan^{1,2}, ZHENG Xiangtao³, LU Xiaoqiang³

1. Key Laboratory of Spectral Imaging Technology, Xi'an Institute of Optics and Precision Mechanics,
Chinese Academy of Sciences, Xi'an 710119, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China;

3. College of Physics and Information Engineering, Fuzhou University, Fuzhou 350108, China

Abstract: Hyperspectral remote sensing is a key technology for remotely obtaining the physical parameters of ground objects and realizing fine identification. It can not only get geometrical properties of the target scenes but also obtain radiance that reflects the characteristics of ground objects. With the development of hyperspectral remote sensing data to unprecedented spatial, spectral, temporal resolution and large data volume, how to adapt to the requirements of massive data and achieve efficient and rapid processing of hyperspectral remote sensing data has become the current research focus. Researchers are introducing scene classification into hyperspectral image classification, integrating the spatial and spectral information to obtain semantic information oriented to larger observation units. However, almost all existing multispectral/hyperspectral scene classification datasets have a number of limitations, including inconsistent spectral and spatial resolutions or spatial resolutions too large to meet the needs of fine-grained classification. Based on the hyperspectral images of Xi'an taken by the "Zhuhai-1" constellation, we combine the result of unsupervised spectral clustering and Google Earth to establish a hyperspectral satellite image scene classification dataset named HSCD-ZH (Hyperspectral Scene Classification Dataset from Zhuhai-1). It consists of 737 images divided into six categories: urban, agriculture, rural, forest, water, and unused land. Each image with a size of 64×64 pixels consists of 32 bands covering the wavelength in the range of 400—1000 nm. In addition, we conduct spatial-based and spectral-based experiments to analyze the performance of existing datasets, and the benchmark results are reported as a valuable baseline for subsequent research. We choose false-color image for the spatial-based experiments and then use popular deep and non-deep learning scene classification techniques. In the experiments based on spectral, the spectral vectors at the pixel are directly used as local spectral features, and BoVW, IFK, and LLC

are used to encode them to generate global representations for the scene. Using SVM as the classifier, the optimal overall classification achieved by the two experiments on the proposed dataset is 92.34% and 88.96%, respectively. Considering that those methods have a large amount of information loss, we cascade the features extracted by the two approaches to generate spatial-spectral features. The highest overall accuracy obtained reaches 94.64%, which is the highest improvement in overall accuracy compared to the other datasets. We construct HSCD-ZH by effectively exploiting both spectral and spatial features of hyperspectral images, selecting various scenes that either have representative spectral compositions, clear spatial textures, or both. It has the advantages of big intraclass diversity, strong scalability, and adapting to satellite hyperspectral intelligent information extraction requirements. Both dataset and experiments can provide effective data support for remote sensing scene classification research in the hyperspectral field. Meanwhile, experiments can indicate that extracting features based on spatial or spectral misses a large amount of available information, and integrating the features extracted by the two methods can compensate for this deficiency. In our future work, we aim to expand the number of categories and images of HSCD-ZH and continue to explore algorithms for integrating spatial and spectral information that can accelerate the interpretation and efficient exploitation of hyperspectral scene cubes.

Key words: hyperspectral remote sensing, Zhuhai-1, scene classification, dataset, feature extraction

Supported by National Natural Science Foundation of China (No. 62271484); National Science Fund for Distinguished Young Scholars (No. 61925112); Key Research and Development Program of Shaanxi (No. 2023-YBGY-225)