

中图法分类号: TP751 文献标识码: A 文章编号: 1006-8961(XXXX)XX-0001-17

论文引用格式: Tan Chunhui, Shi Jiao, Lei Yu, Xu Xiaohan, Zhang Jiangtao, Peng Fulun, Wang Jing. Multi-dataset collaborative hyperspectral band selection based on evolutionary multitasking optimization [J/OL]. Journal of Image and Graphics, XXXX: 1-17. DOI: 10.11834/jig.260237. (谭春晖, 侍佼, 雷雨, 徐晓翰, 张江涛, 彭富伦, 王静. 进化多任务优化的多数据集协同高光谱波段选择[J/OL]. 中国图象图形学报, XXXX: 1-17. DOI: 10.11834/jig.260237.) [DOI: 10.11834/jig.260237]

进化多任务优化的多数据集协同高光谱波段选择

【华山论剑】

谭春晖^{1,2}, 侍佼^{1,2,3}, 雷雨^{1,2,3}, 徐晓翰⁴, 张江涛⁵, 彭富伦⁵, 王静⁵

1. 西北工业大学深圳研究院, 深圳 518057; 2. 西北工业大学, 西安 710129; 3. 西北工业大学重庆科创中心, 重庆 401135; 4. 中国运载火箭技术研究院, 北京 100076; 5. 西安应用光学研究所, 西安 710000

摘要: 目的 高光谱图像具有丰富的光谱信息, 能够为低空场景下地物的精细分类提供重要支撑, 但其高维特性也带来了较大的存储与计算负担。因此, 在低空平台载荷受限且标记样本稀缺的条件下, 获取低维但表示能力优秀的波段子集成为一个值得研究的方向。现有半监督波段选择方法通常仅依赖单一数据集中的标记与未标记样本, 受标记样本不足的影响, 所获得的波段子集判别能力有限。由于同一传感器获取的高光谱图像具有相似的反射和辐射特性, 协同分析多个高光谱数据集能够有效缓解单一数据集监督信息不足的问题。**方法** 本文提出一种进化多任务优化的多数据集协同波段选择方法 (Evolutionary Multitasking Optimization-based Multi-dataset Collaborative Band Selection, EMDCBS)。该算法为每个数据集建立了各自的私有监督波段选择任务, 充分挖掘各个数据集具有丰富地物信息的波段子集。此外, 设计了一个共享监督波段选择任务, 同时利用多数据集的标记样本缓解了标记样本不足的问题, 获取判别能力强的波段子集。在优化过程中, 借助进化多任务优化机制实现不同任务之间的知识迁移与协同搜索, 其中私有监督任务以信息熵为目标, 共享监督任务以类间散度与类内散度构成的可分性准则为目标。进一步地, 针对不同任务获得的多样化最优波段子集, 设计多波段子集融合策略, 从信息量、判别性和冗余性三个方面进行综合筛选, 最终为各数据集生成综合性能更优的波段子集。**结果** 在 PaviaU、PaviaC、LongKou 和 HongHu 4 个公开高光谱数据集上的实验表明, 在每类仅使用 50 个标记样本、波段子集大小设置为 20 的条件下, 所提方法取得了优于对比波段选择方法的综合分类性能。其总体分类精度 (OA) 分别达到 90.05%、95.89%、93.39% 和 79.21%。此外, 实验部分还展示了各种参数对实验结果的影响, 对算法进行了收敛性分析, 并验证了所提出策略的有效性。**结论** 所提 EMDCBS 方法能够充分挖掘并利用同一传感器多数据集之间潜在的公共光谱知识, 在有限标记样本条件下实现更高效、更具判别性的波段选择。通过联合私有监督任务与共享监督任务, 并结合进化多任务优化和子集融合策略, 该方法有效提升了任务优化效率和最优波段子集性能。

关键词: 高光谱图像; 波段选择; 半监督学习; 多数据集协同学习; 进化多任务优化; 低空经济

Multi-dataset collaborative hyperspectral band selection based on evolutionary multitasking optimization

收稿日期: 2026-04-24; 修回日期: 2026-09-14

基金项目: 深圳市科技计划项目 (JCYJ20230807145601004); 国家自然科学基金项目 (62076204, 62472356); 重庆市自然科学基金项目 (CSTB2023NSCQ-MSX1099); 陕西省自然科学基金基础研究计划项目 (2024JC-YBMS-461); 中央高校基本科研业务费专项资金

Supported by: Shenzhen Science and Technology Program (JCYJ20230807145601004); National Natural Science Foundation of China (62076204, 62472356); Natural Science Foundation of Chongqing (CSTB2023NSCQ-MSX1099); Natural Science Basic Research Program of Shaanxi (2024JC-YBMS-461); Fundamental Research Funds for the Central Universities

© 中国图象图形学报版权所有

Tan Chunhui^{1,2}, Shi Jiao^{1,2,3}, Lei Yu^{1,2,3}, Xu Xiaohan⁴, Zhang Jiangtao⁵, Peng Fulun⁵, Wang Jing⁵

1. Northwestern Polytechnical University, Xi'an 710129, China; 2. Shenzhen Research Institute of Northwestern Polytechnical University, Shenzhen 518057, China; 3. Innovation Center NPU Chongqing, Chongqing 401135, China; 4. China Academy of Launch Vehicle Technology, Beijing 100076, China; 5. Xi'an Institute of Applied Optics, Xi'an 710000, China

Abstract: **Objective** Hyperspectral images (HSIs) capture dozens or even hundreds of narrow spectral bands, providing a wealth of spectral details that make it possible to differentiate land-cover classes even when their spectra differ only slightly. With the rapid development of the low-altitude economy, HSIs have been widely used in remote sensing applications such as land-cover mapping, agricultural monitoring, environmental analysis, and urban observation. However, the high dimensionality of hyperspectral data also introduces severe redundancy among adjacent bands, which substantially increases the burden of storage, transmission, and computation in subsequent image analysis. Band selection has become an important dimensionality reduction approach to alleviate these problems. It extracts a compact subset of informative bands directly from the original spectral space while preserving the physical meaning of the selected wavelengths. Existing band selection methods can generally be divided into unsupervised, supervised, and semi-supervised methods. Among these, semi-supervised band selection is particularly attractive because it attempts to take advantage of both labeled and unlabeled samples. However, most existing semi-supervised methods are built for a single dataset, meaning they mainly depend on the small number of labeled samples available within one image. In real-world scenarios, hyperspectral labeling is expensive and difficult to obtain, restricting the discriminative capability of the selected band subsets. In fact, hyperspectral datasets acquired by the same sensor typically exhibit similar physical properties, such as reflectance or radiance characteristics, even though they may correspond to different scenes and land-cover distributions. This indicates that multiple related datasets may contain latent common spectral knowledge that can be jointly explored. Motivated by this, the paper seeks to enhance band selection performance under limited supervision by moving away from optimization on a single dataset and instead adopting collaborative analysis across multiple datasets. **Method** To achieve this goal, an evolutionary multitasking optimization (EMTO)-based multi-dataset collaborative band selection method, termed EMDCBS, is proposed. The core idea is to jointly exploit dataset-specific spectral information and shared discriminative knowledge within a unified multitasking optimization framework. Specifically, for each hyperspectral dataset, EMDCBS establishes a private unsupervised band selection task to search for band subsets with rich spectral information. In parallel, a shared supervised band selection task is constructed by integrating the labeled samples from multiple datasets so as to alleviate the limitation caused by insufficient annotations in any single dataset and to enhance the discriminative capability of the resulting band subset. To optimize these tasks collaboratively, EMDCBS adopts an evolutionary multitasking optimization mechanism in which candidate band subsets are encoded as binary individuals and iteratively evolved through crossover and mutation. During this process, both intra-task interaction and inter-task interaction are introduced. The former enhances the search within the same task, while the latter enables knowledge transfer across different tasks, thereby facilitating the reuse of potentially effective band combinations and improving global search ability. For the private unsupervised tasks, information entropy is used as the objective function to preserve information-rich bands. For the shared supervised task, a separability criterion based on between-class scatter and within-class scatter is adopted to characterize the discriminative quality of the selected bands more directly under limited labeled samples. After task optimization, EMDCBS obtains multiple diverse optimal band subsets from different tasks. Since these subsets emphasize different properties, a multiple band subset fusion strategy is further introduced. By jointly considering information content, discriminative ability, and redundancy, the proposed strategy generates a final band subset with better comprehensive performance for each dataset. **Result** This paper evaluates the effectiveness of the EMDCBS method using four hyperspectral datasets. The overall accuracy (OA) values of the proposed EMDCBS method on the Pavia University, Pavia Center, LongKou, and HongHu datasets are 90.05%, 95.89%, 93.39%, and 79.21%, respectively. These results demonstrate that, compared to other band selection methods, the EMDCBS method can generate a better band subset with limited labeled samples, and achieves better classification performance. Notably, the improvements observed on the LongKou and HongHu datasets indicate that the advantages of collaborative learning across multiple datasets are more pronounced when labeled data are scarce and scene

complexity is high. Furthermore, in the parametric analysis, the classification accuracy of all band selection methods increases with the number of labeled samples. However, when the number of labeled samples per class exceeds 50, the improvement in accuracy for all methods tends to plateau. In addition, the impact of the size of the selected band subset on classification performance is also analyzed in the experiments. Experimental results show that increasing the number of selected bands improves classification accuracy, but the performance improvement becomes minimal when the subset size exceeds 20 bands. This indicates that selecting additional bands may introduce redundancy and increase storage and computational burden, but has limited impact on improving classification accuracy. Ablation experiments, by comparing the classification accuracy of single-dataset band selection with multi-dataset collaborative band selection, confirm the effectiveness of multi-dataset collaborative analysis. Furthermore, comparing convergence curves with and without task interaction mechanisms shows that these mechanisms not only accelerate convergence but also produce better band subsets. This indicates that knowledge sharing between tasks positively impacts the evolutionary optimization process. In addition, ablation experiments verify that by balancing information richness and class discriminative ability and reducing redundancy, the multi-band subset fusion strategy can obtain better band subsets. In conclusion, these results confirm that each major component of EMDCBS makes a significant contribution to performance improvement. **Conclusion** This study proposes a collaborative semi-supervised hyperspectral band selection approach based on evolutionary multi-task optimization, namely EMDCBS. Unlike traditional approaches that optimize band subsets independently within single datasets, EMDCBS concurrently formulates private unsupervised tasks alongside shared supervised tasks across multiple datasets. EMDCBS employs a collaborative multi-task evolutionary search combined with a multiple band subset fusion strategy. This improves both optimization efficiency and the quality of selected band subsets. As a result, it enables more informative and discriminative band selection even when label supervision is limited. Experimental results conducted on four hyperspectral datasets demonstrate that the proposed method outperforms compared techniques. Nonetheless, several areas for further enhancement remain. First, acknowledging that shared spectral knowledge among tasks is inherently dynamic rather than static, there is a need to develop an adaptive knowledge transfer mechanism that accounts for task relevance, population diversity, and optimization progress to enable more precise and effective inter-task collaboration. Furthermore, with the increase in the number of tasks and the scale of data, the time and computational resources required for evolutionary strategies will increase significantly, making it crucial to explore more efficient search strategies.

Key words: hyperspectral images; band selection; semi-supervised learning; multi-dataset collaborative learning; evolutionary multitasking optimization; low-altitude economy

论文引用格式:谭春晖,侍佼,雷雨,徐晓翰,张江涛,彭富伦,王静. 进化多任务优化的多数据集协同高光谱波段选择. 中国图象图形学报. DOI: 10.11834/jig. 260237

0 引言

随着低空经济的发展,对高分辨率与精细化感知的遥感数据需求日益增长。高光谱图像包含数十甚至上百个连续窄波段,能够反映地物细微的光谱差异,因此被广泛应用于地质勘探、农业监测和城市规划等领域(Goetz, 2009; 张兵, 2016; Patro 等, 2021)。但是,高维光谱数据也带来了数据冗余问题,为高光谱图像的分析带来了存储和计算资源的挑战,尤其在低空平台载荷和算力受限的条件下,这

一问题更加突出(张良培和李家艺, 2016; Shang 等, 2021; Feng 等, 2017)。鉴于上述问题,降维操作对高光谱图像的分析具有重要意义。

高光谱图像的降维方法主要分为两种:特征提取和特征选择。特征提取是基于某种映射关系将原始高维数据变换到低维特征空间中,并选择部分重要特征代表原始数据(Sun 和 Du, 2019)。特征提取的方法包括主成分分析法(Principal Component Analysis, PCA)(Du 和 Fowler, 2007)、独立成分分析(Independent Component Analysis, ICA)(Wang 和 Chang, 2006)和局部线性嵌入(Locally Linear Embedding, LLE)(Fang 等, 2014)等。经过上述的空间变换,保留了重要地物光谱信息,减少冗余,但是高光谱数据固有光谱结构信息丢失(Patro 等, 2021)。与特征提取不同,特征选择,即波段选择,能够从原始

波段集中选择一个弱相关且信息丰富的波段子集,实现了降维并保留原始光谱通道的光谱含义(孙伟伟等,2017;Wang等,2018)。

现有的波段选择方法主要分为无监督方法和监督方法。无监督方法(Fu等,2022;孙伟伟等,2022;Wang等,2022;路燕等,2022)不依靠任何先验性的标签信息,而是通过分析数据的统计特征从高光谱遥感影像中提取信息丰富的波段子集。Fu等(2022)设计结合波段选择和空间降噪的无监督波段选择方法,从而获得低冗余和强噪声鲁棒性的波段子集。Wang等(2022)构造与皮尔逊相关系数相适应的自适应子空间调整方法,得到相关性弱且分类精度高的波段子集。尽管无监督波段选择方法可以有效减少原始高光谱的冗余光谱波段,但是由于缺乏监督信息的引导,其往往忽略包含关键判别信息但信息量有限的波段,从而限制了下游任务性能的提升。基于标记样本,监督方法(Yang等,2011;王立国等,2013;Cao等,2016;Cao等,2017;赵亮等,2019)可以保留有利于学习标签与样本关系的重要波段,去除冗余甚至干扰的波段,提升样本的可判别性。Cao等(2016)基于局部空间信息和分类结果进行波段选择。赵亮等(2019)根据波段相关系数矩阵划分波段子空间,基于分类精度在波段子空间上选择最优波段子集。虽然监督波段选择方法通过引入标记信息,使波段选择过程有效保留判别能力更强的波段,但是这些方法通常需要大量的标记样本才能获得性能优秀的波段子集。然而,获取大量的高光谱数据时间与人力成本高。

因此,如何在标记数据有限的情况下获得性能优秀波段子集的半监督波段选择问题被研究。现有的半监督波段选择方法主要分为三类:1)基于排序的(Bai等,2015;Yu等,2022);2)基于聚类的(Su等,2011;Jiao等,2015);和3)基于启发式搜索的(Feng等,2015;Cao等,2019;Aghaei等,2022)。基于排序的方法通过量化每个波段的重要性,挑选出最重要的波段构成最优波段子集,忽略了波段之间的冗余。基于聚类的方法将原始的波段划分为几个簇,并从每个簇中选择最具代表性的波段构成波段子集。虽然基于聚类的方法获取的波段子集冗余性较低,但是可能会忽略一些重要的波段。基于启发式搜索的方法具有较强的全局搜索能力和处理波段之间的非线性关系的能力,因此,被应用于半监督波段选择问

题中。Feng等(2015)利用有限的标记样本和足够的未标记样本,提出了基于最大识别和信息的半监督准则,选择出了高区分度、高信息量和低冗余的波段子集。Cao等(2019)使用少量标记信息训练支持向量机(Support Vector Machine, SVM),生成大量可靠伪标签参与最优波段子集的搜索。Aghaei等(2022)提出的半监督波段选择方法同时基于空间邻接度和光谱相似度增加训练样本,从而搜索优秀波段子集。在这些方法中,无标记和标记数据被充分利用,在一定程度上缓解了标记信息稀缺的问题,取得了不错的分类结果。

然而,现有半监督波段选择算法大多仅依赖单幅高光谱图像中的有限标记样本,导致所获得的波段子集难以充分刻画原始图像中具有判别性的光谱特征。事实上,由同一传感器获取的高光谱数据通常具有相似的物理性质,如辐射率或反射率特性,这种共性并不依赖于具体数据集所对应的地物类别(Plaza等,2009)。因此,可利用这一性质对来自同一传感器的多幅高光谱图像进行协同波段选择,在不同数据集之间共享潜在的公共光谱知识,从而提升波段子集的搜索效率及最优波段子集的性能。

进化多任务优化(Evolutionary Multi-Task Optimization, EMTO)能够同时处理多个相关优化任务,并为各任务输出对应的最优解。近年来,EMTO已被引入高光谱波段选择领域。Shi等(2022)提出了一种多准则半监督波段选择模型,将基于标记信息和未标记信息的波段选择分别构建为两个任务,并通过任务间协同优化搜索兼具高信息量和良好分类性能的波段子集。He等(2022)提出了基于变长聚类的多任务蜂群波段选择算法,借助EMTO可同时获得多个不同规模的最优波段子集,从而提升了蜂群算法的搜索能力。由此可见,EMTO能够有效刻画不同波段选择任务之间的依赖关系,提升各任务的优化效率及最优波段子集性能。然而,现有相关方法大多仍局限于在单一高光谱数据集上构建多个波段选择任务,尚未充分挖掘多数据集之间潜在的协同优化价值。

因此,基于同一传感器获取的高光谱图像具有相似的物理特性,本文提出了进化多任务优化的多数据集协同波段选择(EMTO-based Multi-dataset Collaborative Band Selection, EMDCBS)方法。该方法充分利用多个高光谱数据集的共有物理特性,挖

掘潜在公共光谱知识,实现多数据集波段选择的协同优化。主要贡献如下:

1)本文提出了进化多任务优化的多数据集协同波段选择方法。该方法首次联合优化多个数据集的半监督波段选择问题,将传统单数据集优化扩展为多数据集协同优化,充分挖掘多幅图像的公共光谱知识,提升各个数据集最优波段子集的性能。

2)为了充分利用有限标记信息,一个共享监督波段选择任务被提出。通过合并多个数据集的标记

样本,实现不同数据集间判别知识的直接迁移,缓解单一数据集标记样本不足导致的判别能力受限问题。

3)本方法为多个数据集建立了多个私有无监督波段选择任务和一个共享监督波段选择任务。基于进化多任务优化机制,多个波段选择任务协同优化,实现了重要光谱知识在任务间共享,提升了各个任务的优化效率。

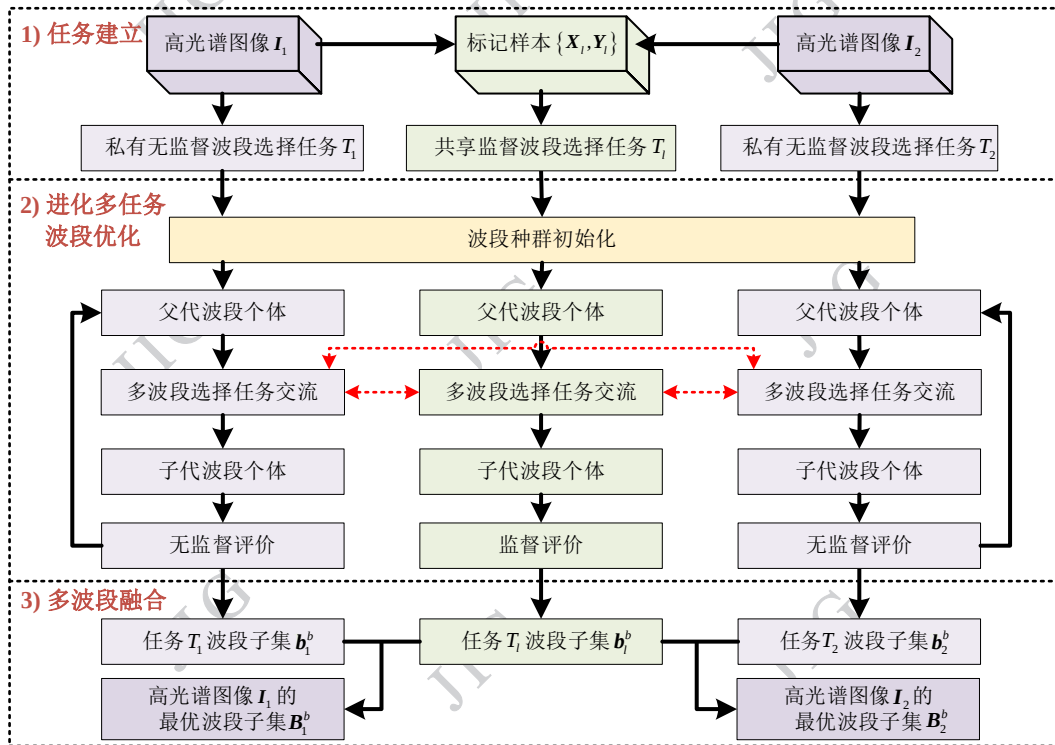


图1 进化多任务优化的多数据集协同波段选择算法框图。

Fig. 1 The framework of EMTO-based multi-dataset collaborative band selection.

4)考虑到所获取波段子集的多样性,本文提出了多波段子集融合策略,同时考虑信息熵、判别信息与冗余度,对不同任务产生的最优波段子集进行互补融合,最终为每个数据集生成综合性能优秀的最优波段子集。

1 本文方法

在本节中,将详细介绍针对高光谱分类的EMDCBS算法整体框架、波段选择任务建立、进化多任务波段优化以及多波段子集融合。

1.1 整体框架

由于单幅高光谱图像的标记样本有限,导致传统半监督波段选择方法生成的波段子集判别性能受限。因此,基于同一传感器获取的高光谱图像具有相似的物理性质,本文提出了进化多任务优化的多数据集协同波段选择方法。所提算法EMDCBS的流程如图1所示。伪代码如算法1所示。所提算法大致分为如下三个部分:

任务建立:首先,该方法为来自同一传感器的多个数据集建立了各自的私有无监督波段选择任务和一个共享监督波段选择任务。私有无监督任务用于优化特定于各个数据集的信息熵高的波段子集;共

算法 1: EMDCBS

输入: 高光谱数据 I_1 和 I_2 。

输出: 最优波段子集 B_1^b 和 B_2^b 。

- 1: 建立两个私有无监督波段选择任务 $\{T_1, T_2\}$ 和一个共享监督波段选择任务 T_i ;
- 2: 随机初始化 POP 个波段个体, 并随机分配波段个体到 3 个波段选择任务中, 组成波段种群 P_0 ;
- 3: 根据任务目标函数评价种群 P_0 所有波段个体;
- 4: **while** 不满足终止条件:
- 5: 所有波段个体进行多波段选择任务交流, 生成多个属于不同任务的子代波段个体, 组成子代种群 P_c ;
- 6: 基于任务目标函数评价所有子代波段个体, 并合并父代种群 P_0 和子代种群 P_c 形成波段种群 $P = P_0 \cup P_c$;
- 7: 从波段种群 P 选择最优 POP 个波段个体构成新的父代波段种群 P_0 ;
- 8: **end while**
- 9: 获得每个任务的最优个体 $\{p_1^b, p_2^b, p_i^b\}$ 并解码为最优波段子集 $\{b_1^b, b_2^b, b_i^b\}$;
- 10: 整合 $\{b_1^b, b_2^b, b_i^b\}$, 基于多波段子集融合策略, 为每个数据集生成最优波段子集 B_1^b 和 B_2^b 。

共享监督任务基于所有数据集的标记数据, 提取判别能力强的共享波段子集。

进化多任务波段优化: 其次, 初始化波段种群, 将种群中的波段个体随机均匀分配给私有无监督任务和共享监督任务。基于进化多任务优化, 多个任务通过遗传操作协同迭代优化种群, 挖掘波段潜在关联性, 基于不同目标函数生成偏好不一的最优个体并解码为多个最优波段子集。

多波段子集融合: 最后, 考虑到各个任务获取的波段子集的多样性, 通过多波段子集融合策略将偏好不一的波段子集进行融合重组, 为每个高光谱数据集生成冗余度低、信息量丰富且判别性能优秀的波段子集。

1.2 波段选择任务建立

给定来自同一传感器的两幅三维高光谱数据 I_1 和 I_2 , 分别表示为两个二维矩阵 $X_1 \in \mathbf{R}^{M_1 \times N}$ 和 $X_2 \in \mathbf{R}^{M_2 \times N}$, 式中 M_1 和 M_2 分别为 I_1 和 I_2 的像素数, N 为高光谱图像波段数。两个数据集中所有的标记像素组成一个二维矩阵 $X_i \in \mathbf{R}^{M_i \times N}$, 其标签表示为 $Y_i \in \mathbf{Z}^{M_i \times 1}$, 式中 M_i 为标记样本总数。若不同数据集类别具有一致地物类别标签, 则映射到同一地物类别; 若语义不同, 则映射为不同地物类别。

所提算法 EMDCBS 为每个数据集都建立了各自的特定波段选择任务 $T_i (i = 1, 2)$, 以及一个共有波段选择任务 T_i 。基于 EMTO, 三个任务协同优化, 其

数学表达式为:

$$\{b_1^b, b_2^b, b_i^b\} = \arg \min \{F_u(b_1), F_u(b_2), F_i(b_i)\}, \quad (1)$$

$$\text{s.t. } b_1, b_2, b_i \in \Omega$$

式中, Ω 表示波段子集的搜索空间, $F_u(\cdot)$ 为私有无监督任务的目标函数, $F_i(\cdot)$ 为共享监督任务的目标函数。 $b_i (i = 1, 2)$ 和 b_i 表示对应任务的可行波段子集。 $b_i^b (i = 1, 2)$ 和 b_i^b 为任务 $T_i (i = 1, 2)$ 和 T_i 在对应目标函数下的最优波段子集。

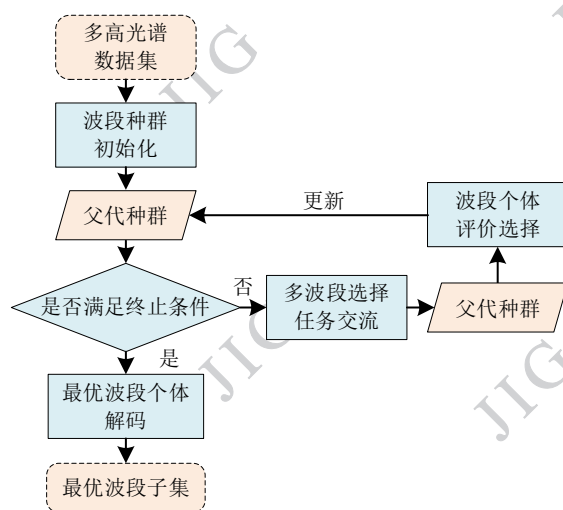


图2 进化多任务波段选择流程图。

Fig. 2 The flowchart of evolutionary multitasking band selection.

1.3 进化多任务波段优化

进化多任务波段选择的流程图如图 2 所示, 主要分为波段种群初始化, 多波段选择任务交流和波段个体评价三个部分。

1.3.1 波段种群初始化

本算法采用 0-1 编码, “0” 表示该位置序号对应的波段不被选中, “1” 表示该位置序号对应的波段被选中。因此, 波段个体的初始化可以表示为:

$$P = \{p_1, p_2, \dots, p_N\} \quad (2)$$

$$p_i \in \{0, 1\} \text{ s.t. } \sum_{i=1}^N p_i = S_b$$

式中, N 为高光谱图像波段数, S_b 表示波段子集大小。图 3 展示了一个波段个体的编码样例。随机初始化 POP 个波段个体, 并将个体随机分配到 3 个不同的波段选择任务中。

1.3.2 多波段选择任务交流机制

在本方法中, 如图 4 所示, 波段个体基因交流的方式可分为任务内交流和任务间交流。任务内交流

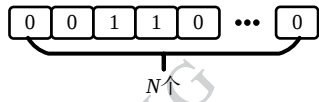


图3 个体初始化样例

Fig. 3 An example of individual initialization

是指来自同一任务的波段个体通过遗传操作生成新的子代波段个体的过程。任务间交流则是来自不同任务的优秀波段个体通过交叉操作得到跨任务波段子代,跨任务子代拥有多个种群的优秀基因,避免优化陷入局部最优。多波段选择任务交流机制充分探索整个搜索空间,挖掘潜在公共光谱知识,实现了重要光谱信息在任务内和任务间共享,促进了各个波段选择任务搜索最优波段子集的进程。

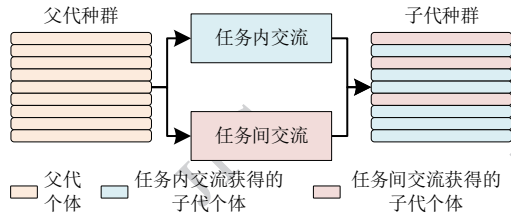
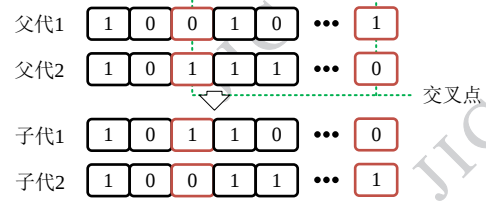


图4 多波段选择任务交流机制。

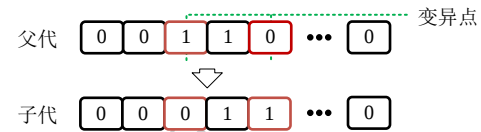
Fig. 4 The communication mechanism of multiple band selection tasks.

在多波段选择任务交流中,交叉与变异操作是实现波段个体基因重组与更新的关键遗传操作。如图5(a)所示,交叉操作通过交换不同父代波段个体中的基因产生新的波段个体。首先从波段种群中随机选取两个父代波段个体,从不同的位置中随机选择若干交叉点,并交换对应位置的基因,以形成新的子代波段个体。该操作能够将不同父代波段个体中的波段进行重新组合,从而增强对潜在优良波段组合的搜索能力。

变异操作如图5(b)所示,其主要目的是在现有波段个体的基础上引入随机扰动,以维持种群多样性。具体地,在个体编码序列中随机选择若干位置作为变异点,并将对应基因值由“0”变为“1”或由“1”变为“0”。由于“1”变为“0”分别表示波段被选中 and 未被选中,因而该操作相当于对当前波段子集进行局部替换、补充或删除。通过这种方式,变异操作能够有效打破波段子集固化问题,避免波段子集种群在演化早期陷入局部最优区域。同时,变异操作还为一些原本未被选中的潜在有效波段提供了被重新



(a) 交叉操作



(b) 变异操作

((a) crossover operation; (b) mutation operation)

图5 遗传操作

Fig. 5 Genetic operations

探索的机会,有助于进一步提升全局搜索能力。

对于高光谱波段选择问题,遗传操作能够促进信息量丰富且判别能力较强的波段不断被筛选和重组,从而提高多任务协同优化的效率,并提升最终获得的最优波段子集性能。

1.3.3 波段个体评价

在多任务优化机制中,通过对子代个体进行评价与选择能够保留优秀的个体参与下一阶段的进化。不同的评价准则所需的已知信息不同,并且对波段子集的偏好不同。在本文中,私有无监督波段选择任务的目标函数为信息熵。任务 $T_i (i = 1, 2)$ 的目标函数可表示为:

$$F_u(b_i) = 1/\sum H(x_i), i = 1, 2 \quad (3)$$

式中, $H(\cdot)$ 表示信息熵, x_i 表示波段选择后的高光谱数据。该目标旨在选择信息量丰富的波段,以提高波段子集的表达能力。

对于共享监督波段选择任务 T_s ,传统方法一般采用分类误差作为目标函数(赵亮等, 2019; Cao 等, 2019; Aghaee 等, 2022),但是在有限标记样本条件下分类误差不仅受波段子集质量影响,还容易受到分类器训练稳定性和训练样本划分波动的干扰,因而难以完全刻画不同波段子集对类别可分性的真实贡献。相比之下,散度矩阵能够从类间分离和类内聚合两个角度直接度量样本分布结构,更适合作为共享监督任务的优化目标。因此,共享监督任务的目标函数可以表示为:

$$F_l(b_l) = \frac{\text{trace}(S_w)}{\text{trace}(S_B)} \quad (4)$$

$$S_w = \frac{1}{C} \sum_{i=1}^C \frac{1}{N_i} \sum_{j=1}^{N_i} (x_{ij}^i - x_l^i)(x_{ij}^i - x_l^i)^T \quad (5)$$

$$S_B = \frac{1}{C} \sum_{i=1}^C (\bar{x}_l^i - \bar{x}_l)(\bar{x}_l^i - \bar{x}_l)^T \quad (6)$$

式中, $\text{trace}(\cdot)$ 表示矩阵的迹。 S_w 表示类内散度矩阵, 用于衡量同一类别内的样本相似度。 S_B 表示类间散度矩阵, 用于衡量不同类别之间的差异性。 x_{ij}^i 表示第 i 类别的第 j 个样本, N_i 为第 i 个类别的样本数。 $\bar{x}_l^i$ 表示第 i 个类别的均值向量, $\bar{x}_l$ 是全部的标记样本的均值向量, C 为总类别数。当 F_l 较小时, 表示类内散度矩阵 $\text{trace}(S_w)$ 较小, 类间散度矩阵 $\text{trace}(S_B)$ 较大, 即不同类别之间的差异性较大, 同一类别的差异性较小。

通过上述各个任务的目标函数对个体进行评价, 按照从小到大依次排序, 选择最优的部分个体参与下一次的进化优化, 其余个体被淘汰。当满足终止条件后, 输出不同任务对应的最优个体 $\{p_1^b, p_2^b, p_l^b\}$, 将其解码为最优波段子集 $\{b_1^b, b_2^b, b_l^b\}$ 。

1.4 多波段子集融合

通过进化多任务波段选择后, 任务 T_1 和 T_2 获取了信息量丰富的最优波段子集 b_1^b 和 b_2^b 。基于多个数据集的标记样本, 任务 T_l 获取了判别能力强的最优波段子集 b_l^b 。为了充分利用私有无监督任务和共享监督任务搜索到的多样化波段子集, 本文提出了一种多波段子集融合策略, 综合考量了波段子集的信息熵、判别能力以及冗余度, 为每个数据集构建综合性性能优秀的波段子集。

首先, 将多个私有无监督波段选择任务的最优波段子集 b_1^b, b_2^b 和共享监督波段选择任务的最优波段子集 b_l^b 结合起来构成数据集 I_1 和 I_2 的可选波段集。

$$B_t = \text{sort}(b_1^b \cup b_2^b \cup b_l^b), t = 1, 2 \quad (7)$$

式中, $\text{sort}(\cdot)$ 表示对波段按照序号进行排序。由于高光谱图像的波段具有连续性, 相邻波段具有较高的相似度, 因此, 采用间隔采样的方式生成两个波段冗余度较低的波段子集 B_{t1} 和 B_{t2} 。然后, 利用 KL 散度对两个波段子集进行信息冗余度评价。

$$KL_i = \sum_{m=1}^M \sum_{n=1, n \neq m}^M E(\log(b_m) - \log(b_n)) \quad (8)$$

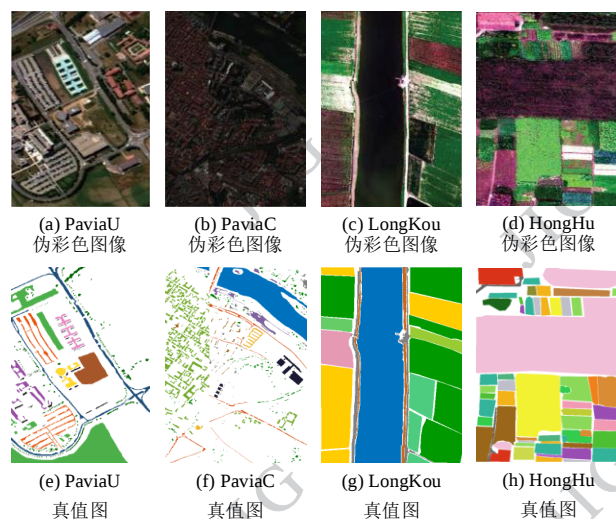
$$b_m, b_n \in B_{ti}$$

式中, b_m 和 b_n 表示波段子集 B_{ti} 中不相同的两个波段。KL 散度越大, 波段之间的冗余信息越少, 可用信息就越多。因此, 选取 KL 散度大的波段子集 B_{ti} 作为对于高光谱图像 I_i 的最优波段子集 B_i^b 。

2 实验结果与分析

为了验证本文所提算法 EMDCBS 的有效性, 在四个数据集上与其他算法进行了比较分析。此外, 在本节还展示了所提方法的参数分析、收敛性分析以及算法策略有效性验证实验结果。

2.1 实验数据集



(a) pseudo-color image of PaviaU; (b) pseudo-color image of PaviaC; (c) pseudo-color image of LongKou; (d) pseudo-color image of HongHu; (e) groundtruth map of PaviaU; (f) groundtruth map of PaviaC; (g) groundtruth map of LongKou; (h) groundtruth map of HongHu)

图6 四个数据集

Fig. 6 Four dataset

(1) ROSIS 成像传感器: 对 0.43-0.86 μm 波长范围内的 115 个波段连续成像, 空间分辨率为 1.3 m。PaviaU 数据集在意大利帕维亚大学获取, 大小为 610×340 像素, 包含 9 个类别, 伪彩色图像和真值图像如图 6(a)(e) 所示。PaviaC 数据集在意大利北部的帕维亚获取, 大小为 1096×715 像素, 包含 9 个类别, 伪彩色图像和真值图如图 6(b)(f) 所示。PaviaU 与 PaviaC 数据集存在 7 类公共地物类别。

(2) Headwall Nano-Hyperspec 成像传感器: 对 400 - 1000 nm 波长范围内的 270 个波段连续成

像。WHU-Hi-LongKou(LongKou)数据集(Zhong 等, 2020)在中国湖北省龙口镇获得, 尺寸为 550×400 像素, 空间分辨率为 0.463 m , 包含 9 个类别, 伪彩色图像和真值图如图 6(c)(g) 所示。WHU-Hi-HongHu(HongHu)数据集(Zhong 等, 2020)在中国湖北省洪湖市获得, 尺寸为 940×475 像素, 空间分辨率为 0.043 m , 包含 22 个类别, 伪彩色图像和真值图如图 6(d)(h) 所示。LongKou 与 HongHu 数据集均为农田, 但是公共类别仅有棉花与道路。

2.2 对比算法与参数设置

对比的波段选择算法分为无监督波段选择算法和半监督波段选择算法。无监督算法包括沃德发散联动策略 WaLuMI(Martinez-Usó 等, 2007)、基于秩的多目标波段选择方法(Incorporated Rank-Based Multiobjective Band Selection, IRMoBS)(Xu 等, 2017)和进化多任务协同波段选择(Evolutionary Multitask Collaborative Band Selection, EM_CBS)(Shi 等, 2022)。半监督算法为基于改进分类图的波段选择(Band Selection Using Improved Classification Map, BS_IC)(Cao 等, 2017)、基于动态分类器选择的波段选择(Band Selection Based on Dynamic Classifier Selection, BSDCS)(Cao 等, 2019)和基于进化多任务优化的多准则半监督方法(Multicriteria Semi-supervised Band Selection Method Based on Evolution-

ary Multitasking Optimization, MTBS)(Shi 等, 2022)。

对于多数据集协同方法 EM_CBS 和 EMDCBS, PaviaU 和 PaviaC 数据集为一组被同时处理, LongKou 和 HongHu 为一组同时波段选择。在所提方法 EMDCBS 中, PaviaU 与 PaviaC 数据集(LongKou 和 HongHu 数据集)的公共类别被映射到同一个标签, 不同类别则映射到不同标签。PaviaU 和 PaviaC 数据集存在 7 类公共地物类别, 因此协同处理时被分为 11 类。LongKou 数据集有 9 类地物, HongHu 数据集存在 22 类地物, 公共类别有 2 类, 因此, 这两个数据集被协同学习时的地物类别一共有 29 类。其他方法仅处理单幅图像。

所提方法中, 种群规模为 60, 迭代次数为 100 次, 波段子集大小为 20, 跨任务交互概率为 0.3, 任务数为 3, 标记样本数为 50 个每类。实验采用 SVM 作为分类器, 该 SVM 采用径向基函数作为核函数。指标为总体分类精度(Overall Accuracy, OA)、平均分类精度(Average Accuracy, AA)和 Kappa 系数。

2.3 对比实验

不同算法在 PaviaU 和 PaviaC 上的分类结果如图 7、图 8 和表 1 所示。在无监督算法中, 基于多数据集协同的 EM_CBS 在 PaviaU 和 PaviaC 的 OA 为 86.12% 和 95.43%, 高于其他无监督算法 WaLuMI 和 IRMoBS, 说明了多数据协同优化能够得到性能更优

表 1 不同算法在四个数据集上的分类结果(加粗表示该指标下的最优结果)

Table 1 The classification results of different algorithms on four datasets (Bold text indicates the optimal result for this metric)

数据集	指标	All Bands	WaLuMI	IRMoBS	EM_CBS	BS_IC	BSDCS	MTBS	EMDCBS
PaviaU	OA	93.85	85.74	85.33	86.12	88.92	85.30	85.44	90.05
	AA	94.09	88.80	88.72	89.04	90.35	87.52	88.92	89.55
	Kappa	93.19	84.17	83.71	84.59	87.71	83.68	83.83	88.99
PaviaC	OA	98.34	94.70	94.76	95.43	94.87	87.73	94.70	95.89
	AA	96.78	84.87	83.87	84.94	85.04	65.00	83.72	86.91
	Kappa	98.18	94.19	94.26	95.00	94.37	86.61	94.19	95.50
LongKou	OA	96.70	85.50	87.44	90.08	85.35	87.08	85.64	93.39
	AA	95.79	82.00	83.07	87.31	79.38	82.96	80.38	91.33
	Kappa	96.35	83.88	86.05	88.99	83.73	85.66	84.07	92.67
HongHu	OA	83.02	76.18	74.50	78.68	77.74	70.11	71.99	79.21
	AA	83.64	69.98	72.34	75.50	75.31	70.03	65.41	75.74
	Kappa	82.53	75.18	73.42	77.80	76.80	68.82	70.82	78.35

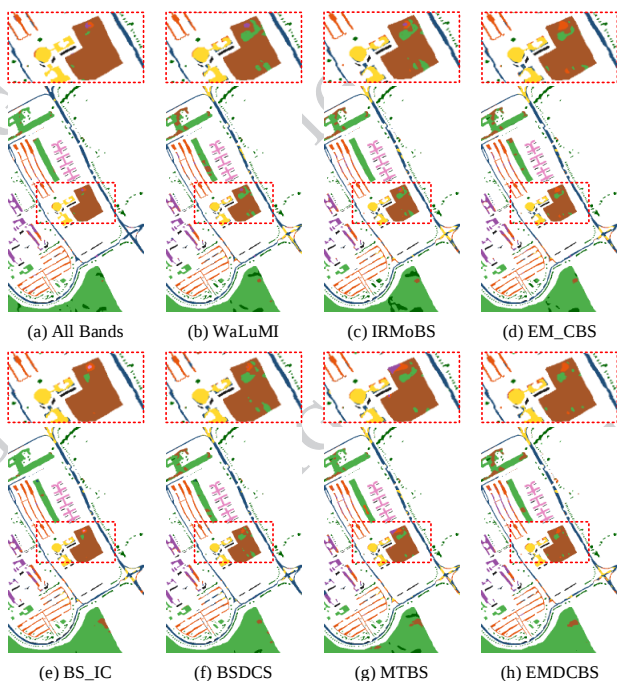


图7 不同方法在PaviaU数据集的分类图

Fig. 7 Different classification maps of PaviaU dataset

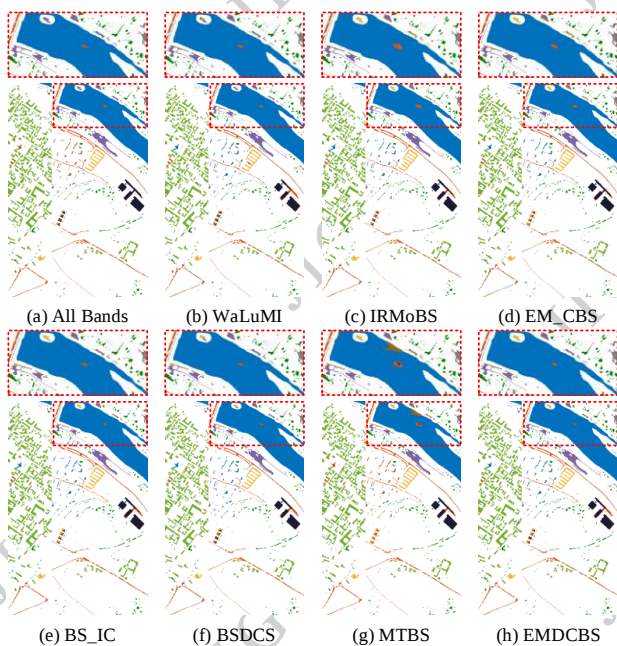


图8 不同方法在PaviaC数据集的分类图

Fig. 8 Different classification maps of PaviaC dataset

秀的波段子集。在半监督方法中,相较于BS_IC、BSDCS和MTBS,所提多数数据集协同方法EMCBS能够获取性能更优秀的波段子集,从而得到分类准确度更高的分类结果图。可以看出,相较于对比方法,所提方法在PaviaU和PaviaC上均有最好的OA,分别为90.05%和95.89%。

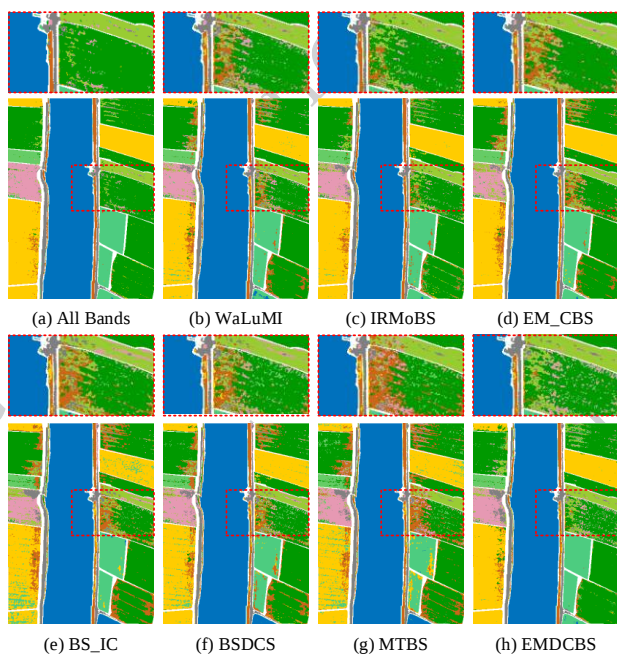


图9 不同方法在LongKou数据集的分类图

Fig. 9 Different classification maps of LongKou dataset

图9、图10和表1展示了各个算法在LongKou和HongHu数据集上的分类图和量化指标。相较于非协同无监督算法WaLuMI和IRMoBS,协同方法EM_CBS具有最好的分类性能,其在LongKou和HongHu数据集上的OA为90.08%和78.68%。在所有的半监督算法中,所提方法通过协同利用LongKou和HongHu数据集的样本,提升了样本丰富度,从而获得了所有方法中分类性能最好的波段子集。所提方法EMDCBS在LongKou和HongHu数据集上的OA值分别为93.39%和79.21%。

2.4 参数分析

本节研究标记样本数和波段子集大小对波段子集性能的影响。

2.4.1 标记样本数对分类精度的影响

图11展示了在LongKou数据集上不同算法分类精度随着样本数从每类10个到每类100个的变化情况。对于无监督算法WaLuMI、IRMoBS和EM_CBS,标记样本虽然不影响波段子集的生成,但是影响分类器性能,其分类精度也会随着标记样本数的增加而增加。在半监督方法中,BS_IC算法的分类精度随样本数变化的幅度最小。BSDCS随着标记样本数的增加其分类精度增长幅度最大,说明了BSDCS对标记样本的依赖性较强。MTBS算法充分利用了无标记信息和有标记信息,使其受标记样本数量的影

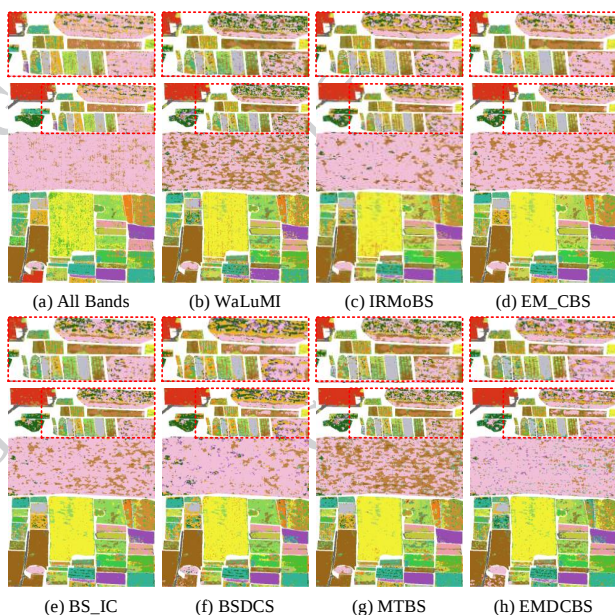


图 10 不同方法在 HongHu 数据集的分类图

Fig. 10 Different classification maps of HongHu dataset

响较小。所提方法 EMDCBS 在有限标记样本情况下分类精度高于其他波段选择方法,说明不同数据集的少量标记样本被充分利用,有效提高了波段子集的判别性能。在样本数少于 50 个每类的情况下,所提方法的分类精度增长较快;而当样本数大于 50 个每类时,分类精度增长速度放缓。因此,在本文中,标记样本数为 50 个每类。

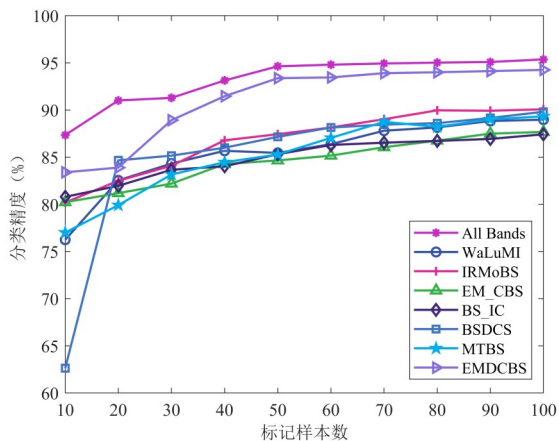


图 11 不同标记样本数对分类结果的影响

Fig. 11 The influence of different number of labeled samples on classification results

2.4.2 波段子集大小对分类精度的影响

图 12 展示了在 LongKou 数据集上不同算法的分类精度随着波段子集大小从 5 到 50 的变化情况。

可以看出,随着所选波段数量的增加,所有方法的分类精度都呈现上升趋势。当波段子集大小从 5 增加到 50 时, WaLuMI、IRMoBS、EM_CBS、BS_IC 和 BSDCS 算法的分类精度增长幅度基本一致,大约 20%。MTBS 的分类精度增长最多,达到 30%。所提方法 EMDCBS 的分类精度增加了 27% 左右。当所选波段数小于 20 时,所提方法 EMDCBS 的分类准确度上升速度较快;当所选波段数大于 20 后,分类精度的上升趋势不明显。虽然波段数越多,可用于分类的信息越多,但是其冗余的信息越多,计算和存储的代价越高。因此,考虑到分类精度和计算代价,所提方法的最优波段子集大小为 20。

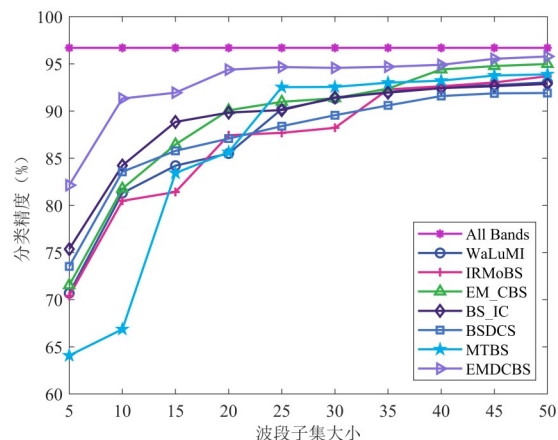


图 12 不同波段数对分类结果的影响

Fig. 12 The influence of different number of bands on classification results

2.5 算法机理与策略验证

本小节分别验证了多数据集协同有效性、共享监督任务目标函数有效性、以及多波段子集融合策略有效性。

2.5.1 多数据集协同分析有效性验证

本文算法 EMDCBS 实现了多数据集协同的半监督波段选择,克服了单一数据集光谱信息不足的问题。为验证本文所提算法策略的有效性,以 LongKou 数据集为例,验证了多数据集协同和共享监督任务的有效性。各个对比方法的介绍如下所示: 1) 单数据集半监督波段选择 (SD_SBS): 单数据集半监督波段选择; 2) 多数据集协同半监督波段选择 (MD_SBS): 无共享监督任务的多数据集协同半监督波段选择; 3) 所提方法 (EMDCBS): 基于共享监督任务多数据集协同波段选择。

实验结果如图 13 所示。可以看出,多数据集协同的半监督波段选择方法 MD_SBS 和 EMDCBS 的分类精度高于 SD_SBS,证明了同一传感器获取的多个数据集之间能够互相提供重要光谱信息,提升波段子集的性能。此外,相较于 MD_SBS,本文所提方法 EMDCBS 获取波段子集的分类性能较好,其 OA 为 93% 左右,证明了本文所设计的共享监督任务能够丰富标记样本的多样性,从而提升波段子集的判别性能。

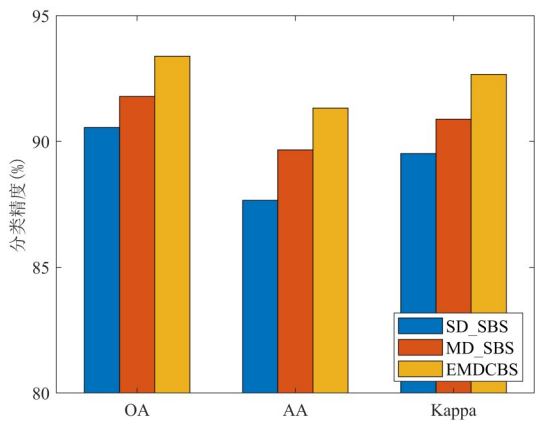


图 13 多数据集协同分析有效性验证
Fig. 13 Validation of multi-dataset collaboration analysis

2. 5. 2 共享监督任务目标函数有效性验证

传统方法采用分类器的分类精度作为目标函数,所提算法采用散度矩阵作为目标函数,体现了不同地物类别的细微差距。在 LongKou 和 HongHu 数据集上,以 SVM 和 K 近邻 KNN(K Nearest Neighbors)分类器的分类误差作为对比的监督目标函数,分类精度如表 2 所示。可以看出,当以分类器的分类误差作为目标函数时,波段子集的判别性能一致。当共享监督任务的目标函数为类间可分性倒数时,最优波段子集在 LongKou 数据集上的分类准确度提升了 2% 左右,在 HongHu 数据集上提升了 12% 左右。因此,该实验结果说明了所提目标函数能够对波段的判别能力进行精细化评估,有效提高波段子集的判别性能。

2. 5. 3 多波段子集融合策略验证

为了获取信息熵高且判别能力强的波段子集,本文提出了多波段子集融合策略。表 3 展示了所提算法在 LongKou 和 HongHu 数据集上是否进行波段融合的信息熵和分类精度。可以看出,融合后波段

表 2 在不同监督目标函数下的分类准确度
Table 2 Classification OA (%) under different supervised objective functions

监督目标函数	SVM 误差	KNN 误差	所提目标函数
LongKou	94.52	94.00	96.04
HongHu	60.57	61.37	73.28

子集的信息熵并非最大,但分类精度最高,说明高信息量并不必然对应更强的判别能力。所提出的融合策略并非单纯追求信息熵最大化,而是通过综合平衡信息量、判别性与冗余度,保留了对下游分类真正有贡献的互补波段。因此,融合后的波段子集能够在控制冗余的同时兼顾信息丰富性与类别可分性,从而获得更优的最终分类结果。

表 3 波段子集融合策略验证
Table 3 Validation of the band subset fusion strategy

评价准则	波段子集是否融合	LongKou	HongHu
信息熵	不融合 (最大/最小)	6.08/3.84	7.03/4.02
	融合	4.85	5.41
分类精度 (%)	不融合 (最大/最小)	92.23/85.75	62.13/56.86
	融合	96.04	73.28

2. 5. 4 消融实验

为系统验证 EMDCBS 方法中各核心模块的独立贡献,设计了五组消融实验:(1) 仅私有无监督任务;(2) 仅共享监督任务;(3) 无任务间交互与融合策略;(4) 有任务间交互但无融合策略;(5) 有任务间交互与融合策略。

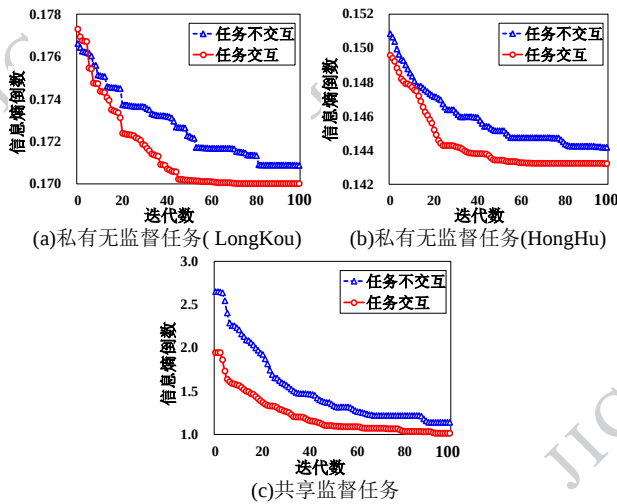
表 4 展示了在 LongKou 数据集上消融实验结

表 4 消融实验结果
Table 4 Results of ablation experiments

消融策略	波段数	OA	AA	Kappa
仅私有无监督任务	20	83.21	76.37	78.7
仅共享监督任务	20	87.38	82.17	80.41
无任务间交互与融合策略	40	92.14	91.25	89.86
有任务间交互但无融合策略	40	93.06	91.72	91.02
有任务间交互与融合策略	20	92.04	90.99	89.44

果。相较于仅私有无监督任务,共享监督任务的OA达到87.38%,提升了4.17个百分点,说明了监督信息有效增强了波段子集类别可分性。在无融合策略的情况下,对比无任务间交互且和有任务间交互,发现通过引入任务交互OA和Kappa分别提升了0.92个百分点和1.16个百分点,说明了任务间交互促进私有任务与共享任务之间的协同搜索,提升各任务的优化效率与最终解的质量。进一步地,在有任务间交互的前提下,引入融合策略后,虽然OA从93.06%小幅下降至92.04%,但是波段子集大小从40压缩至20,将存储成本降低了一半,这表明所提融合策略在保持分类精度的前提下实现了50%的降维,显著提升了波段子集的紧凑性与实用性。

2.5.5 收敛性分析

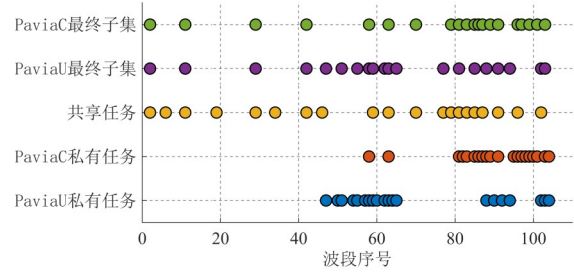


((a) private unsupervised task on LongKou; (b) private unsupervised task on HongHu; (c) shared supervised task)

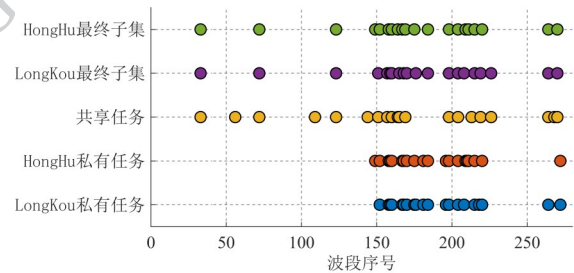
图14 收敛性分析

Fig. 14 Convergence analysis

本文算法基于两个数据集建立了三个任务,两个面向不同数据集的私有无监督波段选择任务和一个共享监督波段选择任务。私有无监督任务的目标函数值为信息熵倒数,共享监督目标函数值表示类间可分性倒数。图14展示了在LongKou和HongHu数据集上三个优化任务是否交互对各个波段选择任务收敛速度的影响。从图14(a)可以看出,在LongKou数据集上,任务交互的初始目标函数值大于任务不交互的初始目标函数值,而在100代时,任务交互的最优目标函数值小于任务不交互的最优目标函数值。从图14(b)可以看出,在HongHu数据集



(a) PaviaU和PaviaC数据集



(b) LongKou和HongHu数据集

((a) PaviaU and PaviaC datasets; (b) LongKou and HongHu datasets)

图15 不同任务的波段子集分布情况

Fig. 15 Distribution of spectral band subsets for different tasks

上,任务通过交互操作在60代时目标函数保持不变,且目标函数值小于任务不交互的目标函数值。对于共享监督任务,任务交互的初始目标函数值优于任务不交互的初始目标函数值,并且其目标函数值始终优于任务不交互。因此,该实验结果说明任务交互不仅提升了优化效率,也增强了最终波段子集的表达能力和判别能力。

2.6 波段子集分布

为了说明所提算法EMDCBS的多数据集协同优化过程中的知识共享与冗余降低机理,对各任务获得的最优波段子集进行了可视化分析,并统计了不同任务之间的波段重叠比例。

图15展示了不同任务以及不同数据集的波段子集。可以看出,在两个传感器数据集组中,私有任务之间均存在明显的波段重叠区域,反映出同一传感器下高光谱数据具有相似的波段分布特性。共享任务的波段子集不仅与两个私有监督任务有重叠,而且能提供私有监督任务并未保留的波段信息,表明共享任务不仅起到了知识迁移的桥梁作用,而且提供更多判别波段。此外,相较私有任务,融合后子集的波段分布明显更分散,有效避免了相邻波段的

表 5 不同任务的波段子集的重叠比例

Table 5 The overlapping ratio of the band subsets of different tasks

数据集	私有 任务 1	私有 任务 2	共享 任务	波段重叠比例(%)
PaviaU(1)和 PaviaC(2)	√	√		25
	√		√	15
		√	√	35
LongKou(1)和 HongHu(2)	√	√		50
	√		√	20
		√	√	15

冗余。表 5 展示了不同任务选中波段重叠比例。可以看出,在两组实验中私有任务之间的重叠比例分别为 25% 和 50%,均显著高于随机选择的期望值,证明确实存在可被挖掘的公共光谱知识。共享监督任

务与各私有任务的重叠比例在 15%~35%之间,说明共享任务在吸收多数据集判别信息的同时,也保留了与各私有任务的差异化波段,为后续融合策略提供了互补性基础。

表 6 不同方法获取的 LongKou_1 的波段子集在不同数据集上的分类准确率

Table 6 Classification OA (%) of the band subsets of LongKou_1 obtained by different methods on different datasets

数据集	WaLuMI	IRMoBS	EM_CBS	BS_IC	BSDCS	MTBS	EMDCBS
LongKou_1	89.11	88.43	90.15	88.11	91.56	93.23	93.27
LongKou_2	88.96	84.32	89.27	89.05	89.94	88.71	92.63
LongKou_3	91.37	89.04	93.50	92.28	94.15	93.87	96.63

表 7 不同方法获取的 LongKou_2 的波段子集在不同数据集上的分类准确率

Table 7 Classification OA (%) of the band subsets of LongKou_2 obtained by different methods on different datasets

数据集	WaLuMI	IRMoBS	EM_CBS	BS_IC	BSDCS	MTBS	EMDCBS
LongKou_1	86.82	90.64	91.57	89.28	80.79	92.41	93.58
LongKou_2	83.06	88.96	91.36	89.62	90.35	91.67	93.18
LongKou_3	87.39	93.47	94.35	92.99	86.85	95.26	96.48

2.7 跨数据集泛化能力验证

为了验证所提方法离线运行的有效性,将 LongKou 数据集按照地理空间位置分为三个小数据集,分别为 LongKou_1, LongKou_2 以及 LongKou_3 数据集。将 LongKou_1 以及 LongKou_2 作为训练集,获得波段子集,LongKou_3 作为离线数据集,验证各个波段选择方法的跨数据集泛化能力。

表 6 与表 7 分别展示了不同方法通过 LongKou_1 与 LongKou_2 数据集获取的波段子集在不同数据集上的分类准确率。本文提出的 EMDCBS 方法在跨数据集(LongKou_3)上取得了所有方法中的最高准确率,且相比次优方法有明显提升。EMDCBS 在训

练集自身(LongKou_1 或 LongKou_2)上也保持最优,表明其选择的波段子集具有良好的稳定性和泛化性。相比之下,部分对比方法 BSDCS、IRMoBS 在跨域场景下出现显著精度下降,例如表 7 中 BSDCS 从 90.35% 降至 86.85%,说明其波段子集过拟合于训练集分布,不具备离线泛化能力。这些结果充分证明了 EMDCBS 方法能够在仅利用源域数据训练的情况下,选出对不同地理空间目标域均有效的波段子集。

3 结 论

本文针对单幅高光谱图像标记样本有限的问题,提出进化多任务优化的多数据集协同波段选择方法。所提方法为每个数据集建立了各自的私有无监督波段选择任务和一个共享监督波段选择任务。共享监督任务综合利用多个数据集的标记样本搜索波段,缓解了标记样本匮乏的问题,提升了波段子集的判别能力。基于EMTO,多个波段候选解充分利用和共享潜在公共光谱知识,提升了任务优化效率和波段子集性能。最后,为了获取综合性能优秀的波段子集,多波段子集融合策略被提出。在四个高光谱图像数据集的试验结果表明:所提EMDCBS能够在有限标记样本的情况下为多个数据集搜索到性能优秀的波段子集。

然而,所提方法仍存在若干值得改进之处。首先,当前方法依赖同一传感器的多数据集,在实际应用中,不同高光谱数据可能来自不同传感器,因此,未来将探索跨传感器场景下的波段选择策略,通过域自适应学习传感器不变的光谱特征,从而提升方法的泛化能力。其次,在优化过程中,各任务间的潜在公共知识并非一成不变,有必要开发自适应的知识迁移机制,能够根据任务相关性与优化进度动态调整迁移策略,实现更精确、高效的任务协作。

参考文献 (References)

- Aghaee R, Momeni M. and Moallem P. 2022. Semisupervised band selection from hyperspectral images using levy flight-based genetic algorithm. *IEEE Geoscience and Remote Sensing Letters*, 19: 1-5 [DOI: 10.1109/LGRS.2022.3147272]
- Bai X, Guo Z X, Wang Y Y, Zhang Z H and Zhou J. 2015. Semisupervised hyperspectral band selection via spectral-spatial hypergraph model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8 (6): 2774-2783 [DOI: 10.1109/JSTARS.2015.2443047]
- Cao X H, Wei C C, Han J G and Jiao L C. 2017. Hyperspectral band selection using improved classification map. *IEEE Geoscience and Remote Sensing Letters*, 14 (11): 2147-2151 [DOI: 10.1109/LGRS.2017.2755541]
- Cao X H, Wei C C, Ge Y M, Feng J, Zhao J and Jiao L C. 2019. Semisupervised hyperspectral band selection based on dynamic classifier selection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12 (4): 1289-1298 [DOI: 10.1109/JSTARS.2019.2899157]
- Cao X H, Xiong T and Jiao L C. 2016. Supervised band selection using local spatial information for hyperspectral image. *IEEE Geoscience and Remote Sensing Letters*, 13 (3): 329-333 [DOI: 10.1109/LGRS.2015.2511186]
- Du Q and Fowler J E. 2007. Hyperspectral image compression using jpeg2000 and principal component analysis. *IEEE Geoscience and Remote Sensing Letters*, 4 (2): 201-205 [DOI: 10.1109/LGRS.2006.888109]
- Feng J, Jiao L C, Liu F, Sun T and Zhang X R. 2015. Mutual-information-based semi-supervised hyperspectral band selection with high discrimination, High Information, and Low Redundancy. *IEEE Transactions on Geoscience and Remote Sensing*, 53 (5): 2956-2969 [DOI: 10.1109/TGRS.2014.2367022]
- Feng S W, Itoh Y, Parente M and Duarte M. 2017. Hyperspectral band selection from statistical wavelet models. *IEEE Transactions on Geoscience and Remote Sensing*, 55 (4): 2111-2123 [DOI: 10.1109/TGRS.2016.2636850]
- Fang Y, Li H, Ma Y, Liang K, Hu Y J, Zhang S J, et al. 2014. Dimensionality reduction of hyperspectral images based on robust spatial information using locally linear embedding. *IEEE Geoscience and Remote Sensing Letters*, 11 (10): 1712-1716 [DOI: 10.1109/LGRS.2014.2306689]
- Fu H, Zhang A Z, Sun G Y, Ren J C, Jia X P, Pan Z J, et al. 2022. A novel band selection and spatial noise reduction method for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1-13 [DOI: 10.1109/TGRS.2022.3189015]
- Goetz A F. 2009. Three decades of hyperspectral remote sensing of the earth: a personal view. *Remote Sensing of Environment*, 113: S5-S16 [DOI: 10.1016/j.rse.2007.12.014]
- He C L, Zhang Y, Gong D W, Song X F and Sun X Y. 2022. A multi-task bee colony band selection algorithm with variable-size clustering for hyperspectral images. *IEEE Transactions on Evolutionary Computation*, 26 (6): 1566-1580 [DOI: 10.1109/TEVC.2022.3159253]
- Jiao L C, Feng J, Liu F, Sun T and Zhang X R. 2015. Semisupervised affinity propagation based on normalized trivariable mutual information for hyperspectral band selection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8 (6): 2760-2773 [DOI: 10.1109/JSTARS.2014.2371931]
- Lu Y, Ren Y and Cui B G. 2022. Noise robust band selection method for hyperspectral images. *Journal of Remote Sensing*, 26 (11): 2382-2398 (路燕, 任月, 崔宾阁. 2022. 噪声鲁棒的高光谱图像波段选择方法. *遥感学报*, 26 (11): 2382-2398) [DOI: 10.11834/jrs.20211128]
- Martinez-Usó A, Pla F, Sotoca J M and García-Sevilla P. 2007. Clustering-based hyperspectral band selection using information measures. *IEEE Transactions on Geoscience and Remote Sensing*, 45 (10): 2455-2465 [DOI: 10.1109/TGRS.2007.908888]

- 45(12): 4158-4171 [DOI: 10.1109/TGRS.2007.904951]
- Medjahed S A, Saadi T A, Benyetou A and Ouali M. 2016. Gray wolf optimizer for hyperspectral band selection. *Applied Soft Computing*, 40: 178-186 [DOI: 10.1016/j.asoc.2015.09.045]
- Patro R N, Subudhi S, Biswal P K and Dell'acqua F. 2021. A review of unsupervised band selection techniques: land cover classification for hyperspectral earth observation data. *IEEE Geoscience and Remote Sensing Magazine*, 9(3): 72-111 [DOI: 10.1109/MGRS.2021.3051979]
- Plaza A., Benediktsson J A, Boardman J W, Brazile J, Bruzzone L, Camps-Valls G, et al. 2009. Recent advances in techniques for hyperspectral image processing. *Remote Sensing of Environment*, 113: S110-S122 [DOI: 10.1016/j.rse.2007.07.028]
- Shang X D, Song M P, Wang Y L, Yu C Y, Yu H Y, Li F, et al. 2021. Target-constrained interference-minimized band selection for hyperspectral target detection. *IEEE Transactions on Geoscience and Remote Sensing*, 59(7): 6044-6064 [DOI: 10.1109/TGRS.2020.3010826]
- Shi J, Zhang X, Liu X D, Lei Y and Jeon G. 2022. Multicriteria semi-supervised hyperspectral band selection based on evolutionary multitask optimization. *Knowledge-Based Systems*, 240: 107934 [DOI: 10.1016/j.knsys.2021.107934]
- Shi J, Zhang X, Tan C H, Lei Y, Li N and Zhou D. 2022. Multiple datasets collaborative analysis for hyperspectral band selection. *IEEE Geoscience and Remote Sensing Letters*, 19: 1-5 [DOI: 10.1109/LGRS.2021.3126762]
- Su H J, Yang H, Du Q and Sheng Y H. 2011. Semisupervised band clustering for dimensionality reduction of hyperspectral imagery. *IEEE Geoscience and Remote Sensing Letters*, 8(6): 1135-1139 [DOI: 10.1109/LGRS.2011.2158185]
- Sun W W, Jiang M and Li W Y. 2017. Band selection using sparse self-representation for hyperspectral imagery. *Geomatics and Information Science of Wuhan University*, 42(4): 441-448 (孙伟伟, 蒋曼, 李巍岳. 2017. 利用稀疏自表达实现高光谱影像波段选择. *武汉大学学报(信息科学版)*, 42(4): 441-448) [DOI: 10.13203/j.whugis.20150052]
- Sun W W and Du Q. 2019. Hyperspectral band selection: a review. *IEEE Geoscience and Remote Sensing Magazine*, 7(2): 118-139 [DOI: 10.1109/MGRS.2019.2911100]
- Sun W W, Yang G, Peng J T and Meng X C. 2022. Robust multi-feature spectral clustering for hyperspectral band selection. *Journal of Remote Sensing*, 26(2): 397-405 (孙伟伟, 杨刚, 彭江涛, 孟祥超. 2022. 鲁棒多特征谱聚类的高光谱影像波段选择. *遥感学报*, 26(2): 397-405) [DOI: 10.11834/jrs.20209165]
- Wang L G, Wei F J. 2013. Band selection for hyperspectral imagery based on combination of genetic algorithm and ant colony algorithm. *Journal of Image and Graphics*, 18(2): 235-242 (王立国, 魏芳洁. 2013. 结合遗传算法和蚁群算法的高光谱图像波段选择. *中国图象图形学报*, 18(2): 235-242) [DOI: 10.11834/jig.20130216]
- Wang J and Chang C. 2006. Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 44(6): 1586-1600 [DOI: 10.1109/TGRS.2005.863297]
- Wang J, Tang C, Li Z L, Liu X W, Zhang W, Zhu E, et al. 2022. Hyperspectral band selection via region-aware latent features fusion based clustering. *Information Fusion*, 79: 162-173 [DOI: 10.1016/j.inffus.2021.09.019]
- Wang Q, Zhang F H and Li X L. 2018. Optimal clustering framework for hyperspectral band selection. *IEEE Transactions on Geoscience and Remote Sensing*, 56(10): 5910-5922 [DOI: 10.1109/TGRS.2018.2828161]
- Xu X, Shi Z W, and Pan B. 2017. A new unsupervised hyperspectral band selection method based on multiobjective optimization. *IEEE Geoscience and Remote Sensing Letters*, 14(11): 2112-2116 [DOI: 10.1109/LGRS.2017.2753237]
- Yang H, Du Q, Su H J and Sheng Y H. 2011. An efficient method for supervised hyperspectral band selection. *IEEE Geoscience and Remote Sensing Letters*, 8(1): 138-142 [DOI: 10.1109/LGRS.2010.2053516]
- Yu C Y, Zhou S J, Song M P and Chang C. 2022. Semisupervised hyperspectral band selection based on dual-constrained low rank representation. *IEEE Geoscience and Remote Sensing Letters*, 19: 1-5 [DOI: 10.1109/LGRS.2021.3049267]
- Zhang B. 2016. Advancement of hyperspectral image processing and information extraction. *Journal of Remote Sensing*, 20(5): 1062-1090 (张兵. 2016. 高光谱图像处理与信息提取前沿. *遥感学报*, 20(5): 1062-1090) [DOI: 10.11834/jrs.20166179]
- Zhang L P and Li J Y. 2016. Development and prospect of sparse representation-based hyperspectral image processing and analysis. *Journal of Remote Sensing*, 20(5): 1091-1101 (张良培, 李家艺. 2016. 高光谱图像稀疏信息处理综述与展望. *遥感学报*, 20(5): 1091-1101) [DOI: 10.11834/jrs.20166050]
- Zhao L, Wang L G and Liu D F. 2019. A subspace band selection method for hyperspectral imagery. *Journal of Remote Sensing*, 23(5): 904-910 (赵亮, 王立国, 刘丹凤. 2019. 高光谱图像子空间的波段选择. *遥感学报*, 23(5): 904-910) [DOI: 10.11834/jrs.20197508]
- Zhong Y F, Hu X, Luo C, Wang X Y, Zhao J and Zhang L P. 2020. WHU-Hi: UAV-borne hyperspectral with high spatial resolution (H2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with CRF. *Remote Sensing of Environment*, 250: 112012 [DOI: 10.1016/j.rse.2020.112012]

作者简介

谭春晖,女,博士研究生,研究方向为智能遥感图像解译等。

E-mail: chunhuitan@mail.nwpu.edu.cn

© 中国图象图形学报版权所有

侍佼,女,教授,研究方向为计算机视觉,遥感图像解译,智能计算等。E-mail: shijiao@nwpu.edu.cn
雷雨,男,副教授,研究方向为计算机视觉,遥感图像解译,智能计算等。E-mail: leiy@nwpu.edu.cn
徐晓翰,男,博士研究生,工程师,研究方向为大模型和强化学习在指挥控制中的应用等。E-mail: 424180922@qq.com

张江涛,男,博士研究生,高级工程师,研究方向为侦察情报系统,多源信息融合等。E-mail: zjt0108@foxmail.com
彭富伦,男,博士研究生,研究员,研究方向为多源数据处理体系,光电侦察系统等。E-mail: pcoolfy@163.com
王静,女,硕士研究生,研究员,研究方向为多源数据处理系统,人工智能等。E-mail: 13598946@qq.com