



文献 CSTR:

32001.14.11-6035.noda.2022.0002.zh



文献 DOI:

10.11922/11-6035.noda.2022.0002.zh

数据 DOI:

10.11922/sciencedb.j00001.00371

文献分类: 地球科学

收稿日期: 2022-01-19

开放同评: 2022-02-12

录用日期: 2022-05-18

发表日期: 2023-06-30

1. 西北农林科技大学, 机械与电子工程学院, 陕西杨凌 712100
2. 西北农林科技大学, 信息工程学院, 陕西杨凌 712100
3. 西安电子科技大学, 先进遥感技术研究院, 西安 710071

摘要: 卫星遥感能够及时获取大尺度下的地物分布, 为经济作物种植结构的信息获取提供良好的数据与技术支撑。本数据集以杨凌农业示范区为研究区域, 由遥感数据、地面真值数据、杨凌边界以及分类结果 4 部分组成。遥感数据由 2021 年 4 月到 9 月的哨兵 2 号、高分 1 号 (包括高分 1 号 C 星)、高分 2 号以及高分 6 号等卫星数据构成, 经过辐射校正、大气校正、正射校正、图像融合以及影像配准等遥感影像处理。通过实地调查、Google Earth 目视解译、小区域的无人机近地遥感等多种方式, 建立地面真实分布验证区。在质量控制方面, 遥感数据整体含云量很少、颜色均匀、空间分辨率为 2 m; 地面真值图通过实地调查进行绘制, 真实可靠。本数据集采用随机森林算法验证, 总体分类精度为 86.17%。本数据集可为相关算法在经济作物种植结构获取方面的研究及应用提供训练样本, 也可为杨凌示范区的土地利用分类及变化、农作物长势监测等方面提供数据支撑。

关键词: 杨凌示范区; 经济作物; 种植结构; 多源多时相; 遥感数据

数据库 (集) 基本信息简介

数据库 (集) 名称	杨凌农业示范区经济作物种植结构多源多时相遥感数据集
数据作者	郭交
数据通信作者	郭交 (gjiao@nwfau.edu.cn)
数据时间范围	2021年4月1日—2021年9月30日
地理区域	地理范围为107°59′–108°08′E, 34°14′–34°20′N之间。地理区域范围包括陕西省杨凌示范区。
空间分辨率	2 m
数据量	54.7 GB (解压后)
数据格式	*.tif, *.shp, *.xml, *.png
数据服务系统网址	http://doi.org/10.11922/sciencedb.j00001.00371
基金项目	国家对地观测科学数据中心开放基金 (NODAOP2021013); 国家自然科学基金 (41301450、62071350)。
数据库 (集) 组成	数据集包含4个文件夹: 遥感数据、地面真值、杨凌边界、分类结果。 (1)遥感数据包含14张遥感影像, 共54.4 GB, 以tif格式存储; (2)地面真值下有2个文件, label.xml是通过实地调查绘制的样本; label.tif将地面真值以图片形式存储; 总共52 MB; (3)杨凌边界总共23.1 KB; (4)分类结果数据量57.3 MB, 包含通过随机森林提取的示范区经济作物种植结构。

* 论文通信作者

郭交: gjiao@nwfau.edu.cn

引言

农业种植结构信息是现代农业生产的重要数据支撑,经济作物已逐渐成为我国农业农村经济发展的支柱产业^[1]。通过卫星遥感技术能够及时获取大尺度区域下的地物分布信息,为经济作物种植结构的信息获取提供了良好的数据支撑^[2]。

本数据集以位于陕西省关中地区西部的杨凌农业高新技术产业示范区(杨陵区或杨凌示范区)为研究区域。该地区是我国第一个国家级农业高新技术产业示范区,在我国现代农业发展中有重要地位。近年来,示范区大力发展高效农业和特色农业。园林水果栽植主要有猕猴桃、苹果、葡萄和樱桃等;设施农业以蔬菜、果树、花卉种植和设施养殖为主。目前在作物类型识别方面,针对不同类型的果园提取研究整体偏少,随着果园面积不断扩大,遥感识别技术在更多作物类型(如果园)提取中的研究值得进一步探讨^[3]。

本数据集中包含的经济作物在关中平原地区也具有典型代表。周至县、眉县是陕西省猕猴桃种植基地;武功县、扶风县盛产苹果,与杨凌示范区相接壤;郿邑区葡萄栽种历史悠久,种植葡萄品种众多。基于本数据集,可快速获取示范区经济作物的种植结构,对于实现合理规划经济作物、促进产业布局优化以及可持续发展具有重要意义;也可为后续对关中平原地区的经济作物分布的实时监测奠定基础。

1 数据采集和处理方法

1.1 研究区概况

研究区域如图1所示。杨凌示范区($107^{\circ}59'-108^{\circ}08'E$, $34^{\circ}14'-34^{\circ}20'N$)位于陕西省关中西部地区,该地区与武功、扶风、周至等水果种植基地县相接壤;全区地势北高南低,海拔435–563 m,土地总面积约为134 km²,属暖温带半湿润半干旱季风气候区,全年平均日照时数2163.8小时,年平均气温12.9°C,年降水量635 mm,属猕猴桃、苹果等经济作物的种植适宜区^[4]。

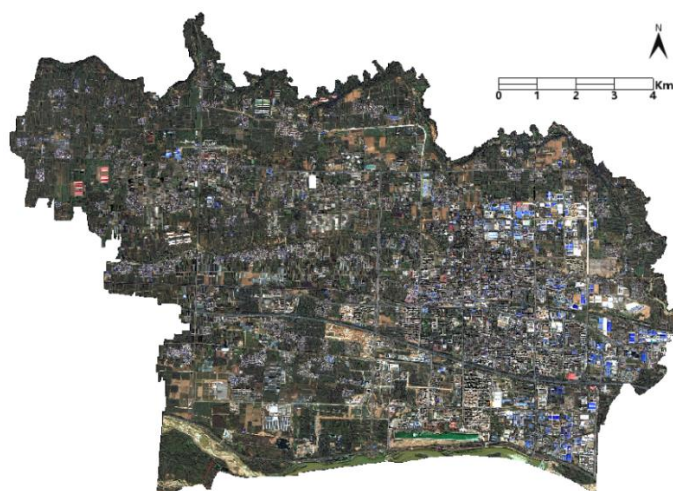


图1 杨凌示范区高分1号示意图

Figure 1 Schematic diagram of GF-1 in Yangling Demonstration Zone

1.2 数据源

本数据集包含哨兵 2 号、高分 1 号（包括高分 1 号 B/C）、高分 2 号以及高分 6 号卫星数据。哨兵 2 号从 GEE 云平台（<https://earthengine.google.com/>）、欧空局哥白尼数据中心（<https://scihub.copernicus.eu/dhus/#/home>）以及 EarthExplorer（<https://earthexplorer.usgs.gov/>）免费下载获取，主要获取的是 L1C 级产品以及 L2A 级产品；高分系列卫星则是通过代理商购买获取，高分卫星数据均是由各自卫星所携带的高分相机获取，均包含一幅 4 波段的多光谱影像和一幅单波段的全色影像。

杨凌区内果园多为已进入结果期的成龄果园。苹果、猕猴桃、桃、樱桃等果树每年从 3–4 月开始萌芽生长到 9–10 月果实成熟（桃、樱桃成熟期为 5–7 月），11 月落叶。每年 6–8 月果树处于稳定生长期。果园内植被覆盖度较大。结合不同果园的生长周期以及实地调查结果，对卫星影像获取时间进行选择。每颗卫星影像获取时间请参见表 1。

表 1 影像获取时间
Table 1 Image acquisition data

卫星	获取时间
高分 1 号	2021-04-14, 2021-07-29, 2021-08-27
高分 2 号	2021-05-04
高分 6 号	2021-05-18, 2021-06-28
哨兵 2 号	2021-04-18, 2021-04-28, 2021-05-08, 2021-06-22, 2021-08-01, 2021-08-06, 2021-09-10, 2021-09-20

1.3 数据处理方法和流程

1.3.1 遥感数据

哨兵 2 号主要获取的是 L1C 级产品以及 L2A 级产品。L1C 级产品是经正射校正和亚像元级几何精校正后的大气表观反射率产品，并没有进行大气校正，可以通过 SNAP 软件的 Sen2cor 插件进行辐射定标以及大气校正；L2A 级产品在 L1C 级基础上进行了辐射定标以及大气校正，可以直接和处理后的高分卫星进行配准。高分卫星数据均是由各自卫星所携带的高分相机获取，对于每颗卫星的多光谱影像进行辐射定标、大气校正和正射校正；对全色影像进行辐射定标以及正射校正；然后将处理好后的多光谱影像和全色影像进行 Gram-Schmidt 图像融合。由于影像是不同卫星获取，同一地区内的地物会存在偏移现象，因此需要对影像进行配准。在图像配准时，以 2021 年 7 月 29 日获取的高分 1 号图像为基准，此图像坐标系为 WGS1984，投影坐标为 UTM zone 49 North WGS84,空间分辨率为 2 m。在图像配准过程中，哨兵 2 号影像空间分辨率自动上采样到 2 m。

1.3.2 地面真值

为更准确地获取影像区域的经济作物种植结构，同时使得数据集类别平衡，选取杨凌示范区经济作物集中地区开展实地调研，在实地调查中结合 RTK 定位（Real Time Kinematic，载波相位差分）以及无人机近地遥感方式记录采样地块的类型属性。结合实地调查结果以及 Google Earth 目视解译，通过 ENVI 软件对实地调查区域进行人工数字化，制作地面真值图。数据采集和处理的总体技术路线见图 2。

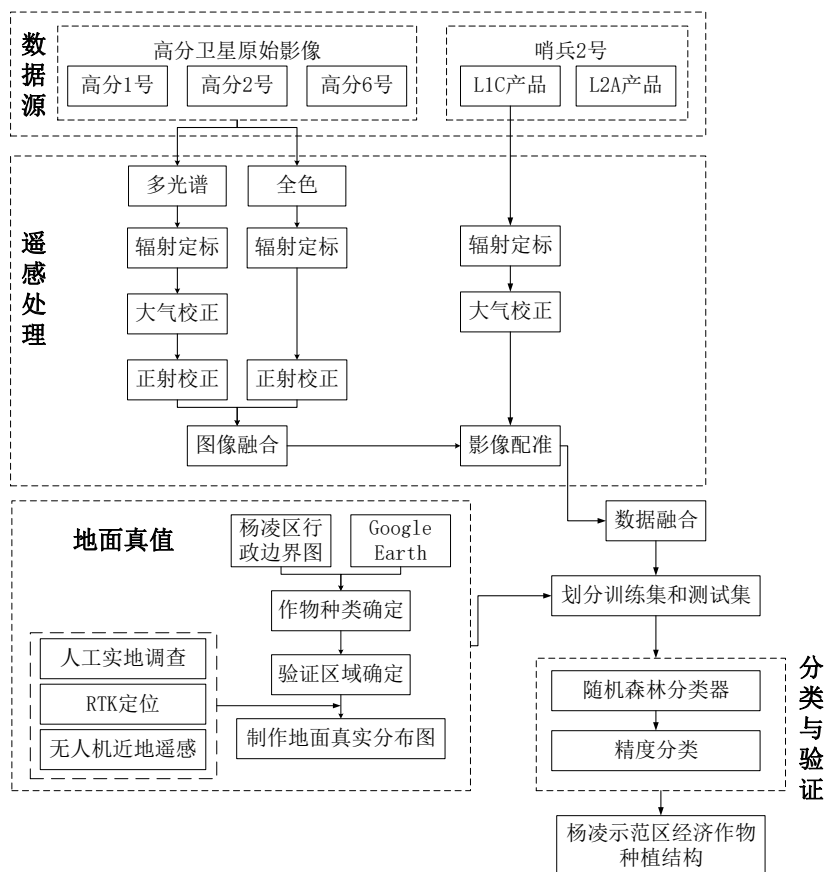


图2 总体技术路线图

Figure 2 Overall technology roadmap

2 数据样本描述

上述数据保存在“杨凌农业示范区经济作物种植结构多源多时相遥感数据集”文件夹内，文件夹具体结构见图3。

本数据集主要包含两部分：遥感数据以及地面真值。遥感数据由哨兵2号、高分1号、高分2号以及高分6号卫星获取，共14张影像，空间分辨率均为2 m，坐标系均为WGS1984，投影坐标均为UTM zone 49 North WGS84。遥感数据总大小约为54.4 GB，保存在“遥感数据”文件夹下，数据格式为tif。高分系列卫星影像均包含4个波段，分别为蓝、绿、红、近红外。哨兵2号卫星影像包含16个波段。遥感数据命名方式为“卫星名_获取时间_地名”，如2021年7月29日获取到的高分1号卫星命名为“GF1_20210729_yangling”。各影像的具体信息可以通过ENVI、ArcGIS等遥感处理软件打开查看。

在“地面真值”文件夹下，xml文件可直接用ENVI软件打开并叠加到遥感数据上，方便观察样本所在区域，后续分类可在本文件基础上利用ENVI软件进行新的样本标注；tif文件为单通道图像，图像数值为0到4，每一个数值代表类别为“0代表非果园类别；1代表苹果园；2代表其他果园；3代表猕猴桃园”，主要作用是方便利用Python等编程语言将地面真值进行读取，以此能够进行相关算法的研究与验证。

除了上述文件外，“杨凌区经济作物种植结构数据集”文件夹内还包含杨凌示范区的行政边界矢量图以及杨凌示范区经济作物种植结构图，分别在“杨凌边界”以及“分类结果”子文件夹下。

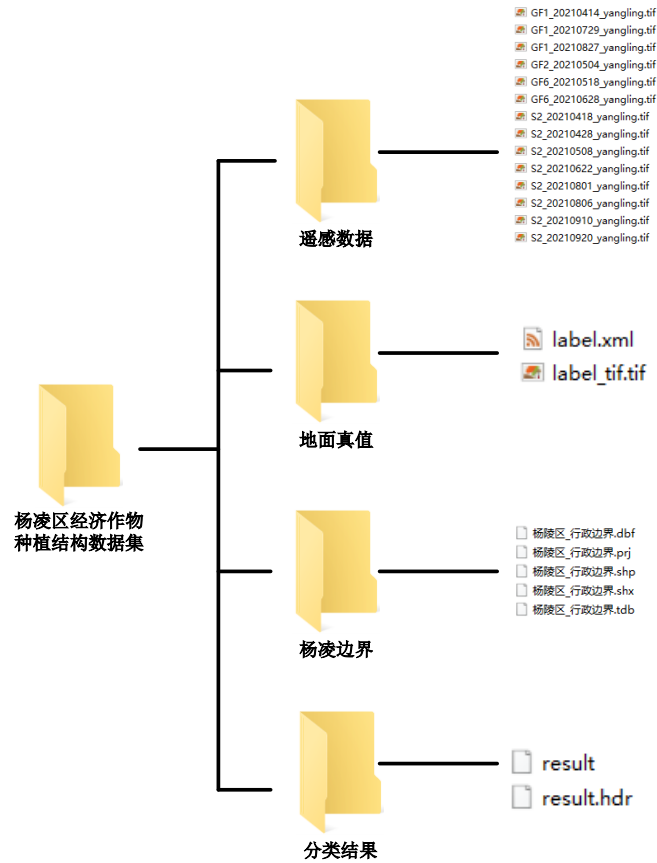


图3 数据文件构成

Figure 3 Data file composition

3 数据质量控制和评估

本文数据质量控制包括遥感影像层面以及地面真值层面，图4为本文数据质量控制的技术路线。

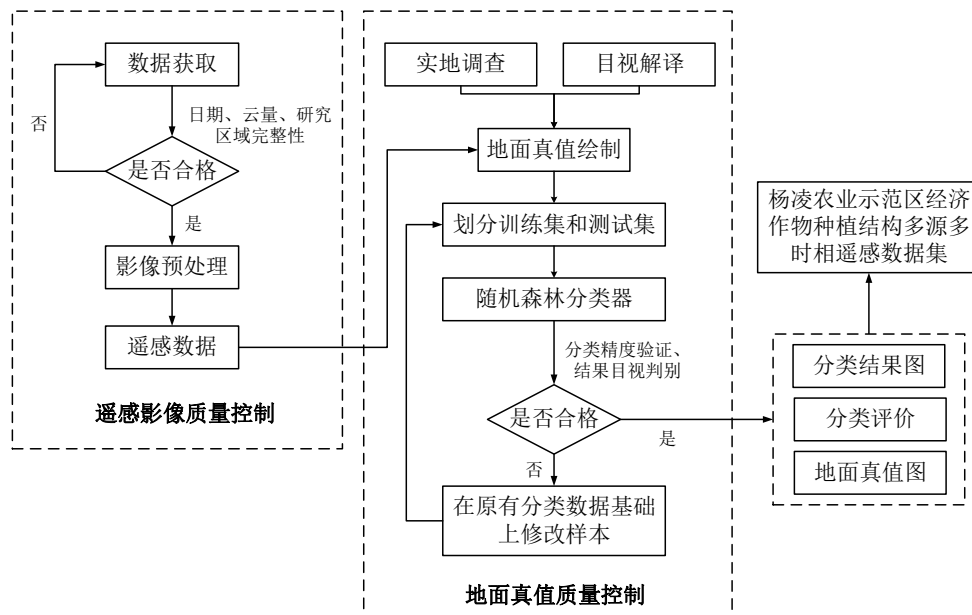


图4 数据控制流程

Figure 4 Flow chart of data control

3.1 遥感影像质量

影像获取阶段：结合多数果园生长周期的交汇时间段，选择 2021 年 4 月 1 日到 2021 年 9 月 30 日之间的卫星影像。由于数据集包含的影像均是光学卫星，因此在获取的时候选择研究区没有云层覆盖的卫星影像。若是不可避免有云覆盖，也是选取了研究区内覆盖少量薄云以及云层不遮挡果园的影像。同时为了避免图像镶嵌可能带来的黑边、颜色不均匀等问题，选取一景图像就能包含杨凌示范区的卫星影像；对于一景影像中研究区域有少部分缺失的现象，将该景影像和相邻影像进行拼接，再从中利用杨凌示范区行政边界进行裁剪得到研究区域。最终获取的遥感影像质量良好、包含少量云层、色彩均衡，空间分辨率为 2 m。

3.2 地面真值与分类结果质量

模型的搭建需要样本进行训练，模型的正确性或者精度也需要样本的验证与评估，因此对于地面真值的质量控制也是对样本的质量控制。首先根据实地调查结果以及 Google Earth 目视解译，在 ENVI 上进行果园样本选取。然后采用随机森林^[5]对样本进行训练，训练样本数量为每类像素数目的 10%，并利用训练好的模型对杨凌示范区的果园进行分类。为定量评估样本质量以及果园提取精度，基于混淆矩阵计算总体分类精度（OA）以及 Kappa 系数^[6]。结合具体的评价指标以及将分类结果与遥感图像叠加对比，如果分类结果不满足制图要求，重新绘制地面真值图，直到满足分类要求为止。

分类流程如图 5 所示，其中 R、G、B、NIR 分别表示红、绿、蓝和近红外波段。对光谱特征以及纹理特征进行叠加构建分类特征。其中，纹理特征选择常用的 8 个 GLCM 纹理量（均值、方差、同质性、对比度、不相似性、熵、角二阶矩阵、相关性）作为 GLCM 纹理特征^[7]；光谱特征选取农情监测常用波段：红、绿、蓝和近红外波段^[8]。随机森林进行训练时，设置随机森林中树的数目为 50，采用随机可重复的取样（bootstrap）策略有放回地从原始样本集中选取训练样本子集；树中每个节点分裂时，从所有特征中随机选取 N 个特征，其中 N 为分类特征数量的平方根，并按照 Gini 指数最小原则^[7]选取最优特征作为分裂属性，生成新分支。

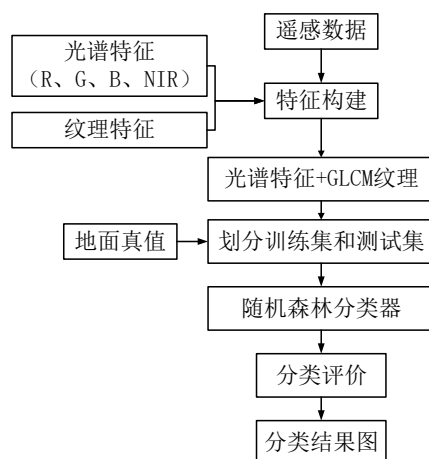


图 5 分类流程图

Figure 5 Classification flow chart

在完成了地面真值控制后，利用随机森林算法对完成质量控制的数据集随机划分训练集和测试集；训练样本数量为每一类像素数目的 10%，使用测试样本进行定量评估，最终获得的结果如表 2 所示。数据集总体良好，适合相关政府部门对杨凌示范区果园分布数据的需求，也为后续对关中平

原地区的经济作物分布的实时监测奠定基础。

表 2 分类结果
Table 2 Classification results

分类方法	F1-score			OA (%)	Kappa
	苹果园	其他果园	猕猴桃园		
随机森林	0.83	0.87	0.92	86.17	0.72

4 数据使用方法和建议

数据使用方法如图 6 所示。本数据集针对杨陵区经济作物种植结构制作，可为相关算法在果园提取方面的应用及研究提供训练和测试样本，从而为经济作物长势监测提供基础。数据集中包含杨陵区的遥感影像以及猕猴桃园、苹果园以及其他果园相关的地面标签，用户可以根据算法特点以及对训练样本的需求，通过相关编程语言或者 GIS 相关软件对数据集划分训练样本、测试样本；比如一些深度学习方法在训练模型中会产生大量参数，因此在模型输入的时候需要对大尺寸的遥感影像以及地面标签进行裁剪，分批读取到模型中。对于地面标签中的“其他果园”类别，可以结合更加高清遥感影像目视解译以及实地考察等途径进一步细化。数据集结合相关算法能够获取到杨陵地区的果园空间分布，再通过实地调查的形式可以进一步验证获取的空间分布是否正确。对于如房屋、道路、水体等容易区分的地物，可以结合高清遥感影像直接进行目视解译，并在相关 GIS 软件中绘制地面标签；而对于一些难区分的地物，如不同种类的农作物，草地和林地等，在结合目视解译的同时，也需要进行实地调查来提高地面标签的真实可靠性。若模型的训练需要大量标注样本，可以利用 ENVI 软件编辑“地面真值”文件夹下的 label.xml 文件，以此来扩充样本量。

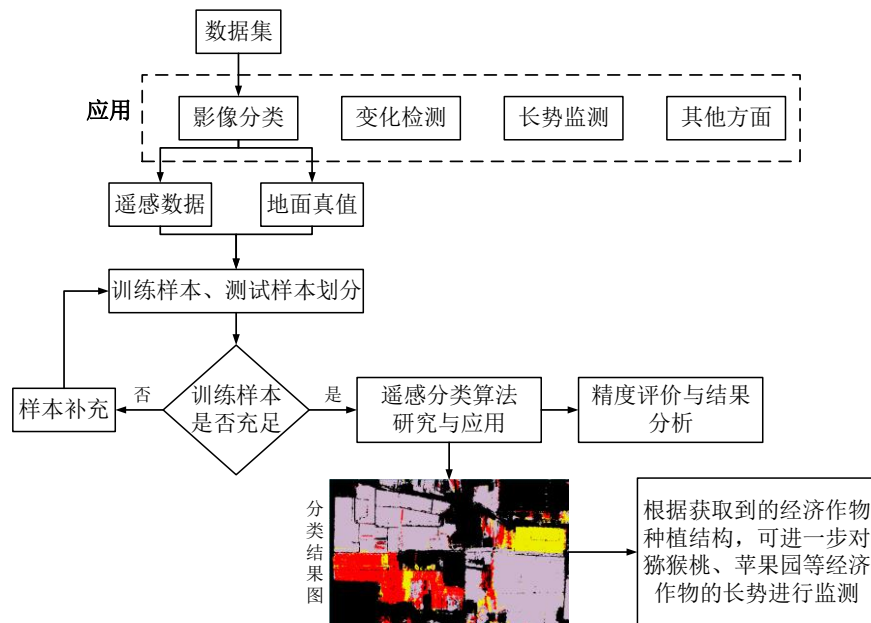


图 6 数据使用方法
Figure 6 Data application

数据集所选取的经济作物（如苹果、猕猴桃等）在陕西特别是关中平原也具有广泛的代表性，通过本文数据集可以探寻高效的果园提取算法，然后将训练好的模型应用到关中平原的其他地区，

为后续更大面积遥感应用奠定基础，也为我国卫星遥感产业链的发展起到积极的推动作用。数据集也包含不同时间段的影像，可为植被覆盖度变化监测、违章建筑检测等方面提供数据支持。

作者分工职责

郭交（1984—），男，陕西西安人，博士，副教授，研究方向为农业遥感、雷达遥感。主要承担工作：算法构建，论文撰写。

白静远（1998—），男，广东省东莞人，硕士，硕士研究生，研究方向为遥感图像分类。主要承担工作：遥感数据获取与处理，实地调查与地面真值绘制，数据质量控制。

叶永凯（2001—），男，河南信阳人，本科，本科生，研究方向为农业机器人。主要承担工作：样本收集。

韩超越（2001—），男，山东济南人，本科，本科生，研究方向为无人机遥感。主要承担工作：数据质量控制。

张伟涛（1983—），男，陕西西安人，博士，副教授，研究方向为深度学习、遥感数据处理。主要承担工作：数据质量控制。

参考文献

- [1] 梁书民. 中国农业种植结构及演化的空间分布和原因分析[J]. 中国农业资源与区划, 2006, 27(2): 29 - 34. DOI:10.7621/cjarrp.1005-9121.20060229. [LIANG S M. Space distribution and reason analysis of the changes in agriculture planting structure of China[J]. Chinese Journal of Agricultural Resources and Regional Planning, 2006, 27(2): 29 - 34. DOI:10.7621/cjarrp.1005-9121.20060229.]
- [2] 苏伟, 姜方方, 朱德海, 等. 基于决策树和混合像元分解的玉米种植面积提取方法[J]. 农业机械学报, 2015, 46(9): 289 - 295, 301. DOI:10.6041/j.issn.1000-1298.2015.09.042. [SU W, JIANG F F, ZHU D H, et al. Extraction of maize planting area based on decision tree and mixed-pixel unmixing methods[J]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(9): 289 - 295, 301. DOI:10.6041/j.issn.1000-1298.2015.09.042.]
- [3] 宋荣杰, 宁纪锋, 刘秀英, 等. 基于纹理特征和 SVM 的 QuickBird 影像苹果园提取[J]. 农业机械学报, 2017, 48(3): 188 - 197. DOI:10.6041/j.issn.1000-1298.2017.03.024. [SONG R J, NING J F, LIU X Y, et al. Apple orchard extraction with quick bird imagery based on texture features and support vector machine[J]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(3): 188 - 197. DOI:10.6041/j.issn.1000-1298.2017.03.024.]
- [4] 贺文丽, 李星敏, 朱琳, 等. 基于 GIS 的关中猕猴桃气候生态适宜性区划[J]. 中国农学通报, 2011, 27(22): 202 - 207. DOI:10.3969/j.issn.1000-6362.2011.01.022. [HE W L, LI X M, ZHU L, et al. Climate ecological applicability regionalization for kiwifruit based on GIS in Guanzhong of Shaanxi Province[J]. Chinese Agricultural Science Bulletin, 2011, 27(22): 202 - 207. DOI:10.3969/j.issn.1000-6362.2011.01.022.]
- [5] 杨蕙宇, 王征强, 白建军, 等. 基于多特征提取与优选的冬小麦面积提取[J]. 陕西师范大学学报(自然科学版), 2020, 48(1): 40 - 49. DOI:10.15983/j.cnki.jsnu.2020.04.005. [YANG H Y, WANG Z Q, BAI J J, et al. Winter wheat area extraction based on multi-feature extraction and feature selection[J]. Journal of

Shaanxi Normal University (Natural Science Edition), 2020, 48(1): 40 - 49.
DOI:10.15983/j.cnki.jsnu.2020.04.005.]

[6] FANG P, ZHANG X W, WEI P P, et al. The classification performance and mechanism of machine learning algorithms in winter wheat mapping using sentinel-2 10 m resolution imagery[J]. Applied Sciences, 2020, 10(15).

[7] 杨彦荣, 宋荣杰, 胡国强, 等. 基于随机森林和纹理特征的苹果园遥感提取[J]. 现代电子技术, 2020, 43(3): 40 - 44. DOI:10.16652/j.issn.1004-373x.2020.03.010. [YANG Y R, SONG R J, HU G Q, et al. Apple orchard remote sensing extraction based on random forest and texture features[J]. Modern Electronics Technique, 2020, 43(3): 40 - 44. DOI:10.16652/j.issn.1004-373x.2020.03.010.]

[8] 李长春, 陈伟男, 王宇, 等. 基于多源 Sentinel 数据的县域冬小麦种植面积提取[J]. 农业机械学报, 2021, 52(12): 207 - 215. [LI C C, CHEN W N, WANG Y, et al. Extraction of winter wheat planting area in County based on multi-sensor sentinel data[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(12): 207 - 215.]

论文引用格式

郭交, 白静远, 叶永凯, 等. 杨凌农业示范区经济作物种植结构多源多时相遥感数据集[J/OL]. 中国科学数据, 2023, 8(2). (2023-06-30). DOI: 10.11922/11-6035.noda.2022.0002.zh.

数据引用格式

郭交. 杨凌农业示范区经济作物种植结构多源多时相遥感数据集[DS/OL]. Science Data Bank, 2023. (2023-06-30). DOI: 10.11922/sciencedb.j00001.00371.

A dataset of multi-source and multi-temporal remote sensing data of cash crop planting structure in Yangling Agricultural Demonstration Zone

GUO Jiao^{1*}, BAI Jingyuan², YE Yongkai¹, HAN Chaoyue¹, ZHANG Weitao³

1. College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi 712100, P.R. China

2. College of Information Engineering, Northwest A&F University, Yangling, Shaanxi 712100, P.R. China

3. Research Institute of Advanced Remote Sensing Technology, Xidian University, Xi'an 710071, P.R. China

*Email: gjiao@nwfau.edu.cn

Abstract: Satellite remote sensing technology can obtain the distribution of ground objects on a large scale in a timely manner, and provide great data and technical support for the acquisition of information on the planting structure of cash crops. Taking Yangling Agricultural Demonstration Area as the research area, this

dataset is composed of four parts: remote sensing data, ground truth data, Yangling boundary and classification results. The remote sensing data consist of satellite data, such as Sentinel-2, Gaofen-1 (including Gaofen-1C satellite), Gaofen-2, and Gaofen-6 from April to September in 2021 after radiation correction, atmospheric correction, and remote sensing image processing such as orthorectification, image fusion, and image registration. Through on-the-spot investigation, visual interpretation of Google Earth, and near-ground remote sensing of UAVs in small areas, we established the ground truth distribution verification area. In terms of quality control, the remote sensing data are characteristic of little overall cloud content, uniform color, and a spatial resolution of 2m; the ground truth map, authentic and reliable, is drawn through field surveys. The dataset has been verified by random forest algorithm, and the overall classification accuracy is 86.17%. It can provide training samples for the research and application of related algorithms in the acquisition of cash crop planting structure, and can also provide data support for land use classification and changes as well as crop growth monitoring in Yangling Demonstration Zone.

Keywords: Yangling Demonstration Zone; cash crops; planting structure; multi-source and multi-temporal; remote sensing data

Dataset Profile

Title	A dataset of multi-source and multi-temporal remote sensing data of cash crop planting structure in Yangling Agricultural Demonstration Zone
Data corresponding author	GUO Jiao (gjiao@nwafu.edu.cn)
Data authors	GUO Jiao
Time range	April 1—September 30, 2021
Geographical scope	Yangling Demonstration Zone in Shaanxi Province (34°14'-34°20' N, 107°59'–108°08' E)
Spatial resolution	2 m
Data volume	54.7 GB (after decompression)
Data format	*.tif , *.shp , *.xml , *.png
Data service system	< http://doi.org/10.11922/sciencedb.j00001.00371 >
Sources of funding	Open Research Fund of National Earth Observation Data Center (NODAOP2021013); National Natural Science Foundation (41301450, 62071350).
Dataset composition	The dataset is composed of 4 folders: Remote Sensing Data, Ground Truth, Yangling Boundary, and Classification Results. (1) The remote sensing data contain 14 remote sensing images with a data volume of 54.4GB in “tif” format. (2) There are 2 files under the ground truth with a data volume of 52MB: one is “label.xml”, a sample drawn through field investigation; the other is “label_tif.tif” in the form of pictures. (3) The Yangling boundary has a total of volume of 23.1KB. (4) The data volume of classification results is 57.3MB, containing the planting structure map of cash crops in Yangling Demonstration zone randomly extracted from forests.