

黄河上中游淤地坝高分辨率光学遥感卫星影像 样本数据集

ISSN 2096-2223

CN 11-6035/N

敏玉芳^{1,2*}, 艾鸣浩^{1,2,3}, 张耀南^{1,2}, 黄波^{1,2,3}



文献 CSTR:

32001.14.11-6035.ncdc.2024.0003.zh



文献 DOI:

10.11922/11-6035.ncdc.2024.0003.zh

数据 DOI:

10.12072/ncdc.nieer.db4212.2024

文献分类: 地球科学

收稿日期: 2024-02-29

开放同评: 2024-03-19

录用日期: 2024-11-29

发表日期: 2024-12-27

1. 中国科学院西北生态环境资源研究院, 兰州 730000
2. 国家冰川冻土沙漠科学数据中心, 兰州 730000
3. 中国科学院大学, 北京 100010

摘要: 淤地坝是黄土高原地区一种行之有效的水土保持工程措施, 在拦泥保土、减少入黄泥沙、防洪减灾、促进粮食生产和水资源合理利用等方面发挥了重要作用。定期开展淤地坝生态效益评估、淤地坝风险易发性评价等需要淤地坝分布、淤积面积等基础数据。本文基于 Google Earth 19 级遥感影像数据, 在黄河上中游区域通过目视解译标记了 714 张淤地坝样本, 并对每个样本进行了图像增强, 总产生了 3570 个淤地坝样本。基于此样本进行淤地坝识别模型训练, 以陕西省绥德县为例, 识别精度达到 91.5%。本样本集可作为淤地坝图像识别的基础数据, 促进黄河上中游区域的淤地坝自动识别研究, 节省了野外调查、遥感解译的人力物力成本, 为水土保持效益评价、淤地坝风险评价提供动态数据支撑。

关键词: 淤地坝; 遥感影像; 机器学习; 图像识别

数据库(集)基本信息简介

数据库(集)名称	黄河上中游淤地坝高分辨率光学遥感卫星影像样本数据集
数据作者	敏玉芳, 艾鸣浩, 张耀南, 黄波
数据通信作者	敏玉芳 (myf@lzb.ac.cn)
数据时间范围	2019–2021年
地理区域	黄河上中游区域, 地理范围为(36°31′–39°35′N, 106°43′–111°28′E)
空间分辨率	0.54 m
数据量	2567 MB
数据格式	*.tif, *.tifw, *.xml
数据服务系统网址	http://dx.doi.org/10.12072/ncdc.nieer.db4212.2024
基金项目	国家重点研发计划(2021YFF0704200); 国家重点研发计划(2022YFF0711700)。
数据库(集)组成	数据集共包括2个数据文件夹, 其中: (1) samples文件夹存放手动标注的样本, 样本量714, 数据量423 MB; (2) augsamples文件夹存放增强后的样本, 样本量2856, 数据量2144 MB。

* 论文通信作者

敏玉芳: myf@lzb.ac.cn

引言

淤地坝是修建在多泥沙沟道中用于固沟减蚀、蓄洪滞洪和控制泥沙输移的一种水土保持工程措施^[1-2]，具有良好的水土保持功效，在拦泥淤地、滞洪蓄水、建设农田、巩固退耕还林（草）、发展生产、减轻黄河泥沙、改善生态环境等方面均发挥了重要作用^[3-4]。

黄土高原是中国水土流失最为严重的地区之一。截至 2019 年 11 月，黄土高原淤地坝共计 58776 座，大型、中型、小型淤地坝分别有 5095 座、12169 座、40702 座，淤地坝工程建设有效抑制了黄土高原地区水土流失过程，对区域生态保护具有显著效益^[5]。黄土高原地区现状淤地坝设计总库容 110.33 亿 m^3 、淤积库容 77.50 亿 m^3 ，目前已淤积 55.04 亿 m^3 ，剩余淤积库容 22.46 亿 m^3 ，大部分淤地坝面临着淤满失效的问题^[6]。

随着淤地坝长期淤积、管理缺失等以及近年来极端降水事件频发，淤地坝面临溃坝风险越来越高，严重威胁了下游群众的生命财产安全。开展淤地坝生态效益评估、淤地坝风险易发性评价等需要淤地坝分布、淤积面积等基础数据。传统淤地坝基础数据的获得方法多以分层抽样野外调查法和人工测绘为主，存在准确性不足、效率低下和数据更新周期长的问题，难以满足当下及未来对淤地坝动态持续监测与科学布局的任务要求^[7]。

遥感技术的迅猛发展，使越来越多的研究者尝试利用遥感影像进行水土保持措施的快速、精准识别。传统遥感影像信息提取方法有人工目视解译、最小距离分类、最大似然分类等，这些方法对影像空间分辨率要求不高，但是耗费人力，无法处理大量遥感影像。近年来，深度学习发展迅速，在数据挖掘、自然语言处理、图像等领域取得成功。以机器学习、深度学习为代表的人工智能方法也成为遥感影像自动化解译的主流技术手段，在精度和效率方面取得了跨越式的进步^[8]。淤地坝通常分布少、面积小，且中低分辨率遥感影像难以提取，故提取淤地坝的研究较少，高分辨率遥感影像可为提取较小地物提供良好的数据源。

基于光学遥感影像的深度学习淤地坝检测方法本质上属于图像识别任务，丰富多样、数量充足的高质量研究样本有助于提升模型的检测精度，增强其鲁棒性。而淤地坝受到地形和地貌的影响形态多样，因其物质组成和环境差异也呈现出不同的色彩、几何纹理等特征。且与自然堆土、田坎的影像特征相似，因此，构建一个淤地坝类型多样，样本丰富的数据集，对淤地坝的快速检测具有重要意义。目前公开发布的仅有 Zeng Yi 等^[9]发布的黄土高原淤地坝坝地矢量数据集，基于面向对象的分类方法，结合超高分辨率（0.3–1.0 m）的卫星影像，通过人机交互的方式获得淤地坝的空间分布坐标、坝地面积等基础数据。

本研究基于 Google Earth 19 级遥感影像数据，在黄河中上游流域通过目视解译标记了 714 张淤地坝样本，每个样本进行了图像增强，总产生了 3570 个淤地坝样本。本样本集可作为淤地坝图像识别的基础数据，促进黄河上中游区域的淤地坝自动识别研究。

1 数据采集和处理方法

1.1 数据来源

本文建立了以黄河上中游流域淤地坝为研究对象的高分辨率光学影像样本集。样本基于 Google Earth 遥感影像数据，使用 BIGEMAP 地图下载器获取缩放级别为 19 的 Google Earth 卫星图像，影

像时间范围为 2019–2022 年，主要区域为黄河流域多沙粗沙区域和山西省，空间范围为 ($35^{\circ}55' - 40^{\circ}23'N$, $106^{\circ}39'E - 111^{\circ}49'E$)，分辨率为 0.54m/pixel，包括红、绿、蓝波段。为了减少植被引起的识别误差，影像尽量选择 10 月–4 月期间。样本数据选取淤地坝密集修建的榆林地区、山西汾西县等区域，样本分布均匀，可代表黄河上中游大部分淤地坝特征。为了消除偏差并增加样本的多样性，使用样本增强技术扩充数据集，使得图像识别模型可以充分学习目标物体的语义特征。

1.2 数据采集与处理方法

本文基于 Google Earth 卫星图像数据，使用 ArcGIS Pro 软件，通过人工的方式对遥感图像进行淤地坝的目视解译和样本标注。数据集制作流程如图 1 所示，主要步骤包括：（1）制定标注规范。明确淤地坝遥感图像解译的标志、标注方式，形成标注规范文档；（2）数据标注。人工目视解译遥感图像上的淤地坝并进行标注。通过自检、互检及终检过程保证标注的准确性；（3）样本切片导出。按制定的标注规范导出样本；（4）样本增强。通过翻转、裁剪、缩放、颜色调整等方式增强样本。

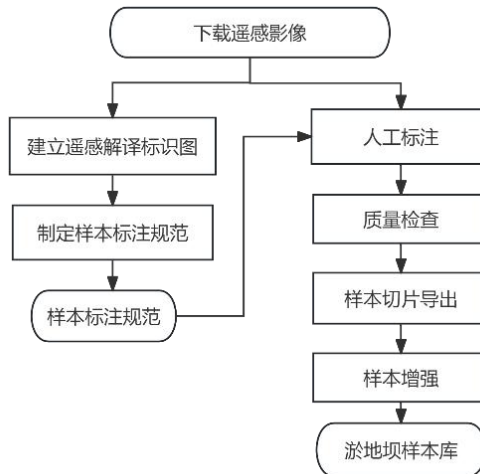


图 1 淤地坝遥感影像样本数据集制作流程

Figure 1 Flowchart for producing the remote sensing image samples of sediment storage dams

1.2.1 数据样本特征

淤地坝一般位于沟道中，形似水坝，上窄下宽；坝体至少有一侧为淤积地、耕地或水域；一条沟道中可有多个淤地坝^[10]。本文借鉴马煜栋等^[10]在遥感解译时对淤地坝的分类，样本分为两侧淤积型、单侧水体型、单侧淤积型三类样本（表 1），样本均匀分布。

表 1 淤地坝遥感解译特征表

Table 1 Features of remote sensing interpretation of sediment storage dams

序号	类型	特点	图片
1	两侧淤积型	沟道内为一细长型光滑图案；一坝体型构造横贯沟道，与沟道两侧坡体相连；两侧皆为光滑图案。	

序号	类型	特点	图片
2	单侧水体型	沟道内为一细长型蓝色水域图案；一坝体型构造横贯沟道，与沟道两侧坡体相连；一侧为水域、一侧为山体、道路等。	
3	单侧淤积型	沟道内为一细长型光滑图案；一坝体型构造横贯沟道，与沟道两侧坡体相连，一侧为淤积，一侧为山体、道路等。	

淤地坝样本与塘坑、道路、护坡、尾矿库或水库大坝等在遥感影像上有一定的相似性，为了保证目视解译的正确性，本文进一步细化了淤地坝的特征：所有淤地坝都位于沟道，坝体长度一般为20–200米范围，坝顶宽度4–6米，坝体坡比1:2.5–1:2之间；淤地坝一般拥有坝体、溢洪道、放水涵卧管三大件（部分只有坝体，或者只有坝体和溢洪道），淤地坝附近没有大型工厂。

1.2.2 数据标注

Google Earth 19级遥感影像数据量比较大，无法一次性加载到软件进行标注。对影像进行切片，切片后逐片进行标注。本文使用 ArcGIS Pro 的训练样本管理器进行样本的标记。首先加载需要标记样本的遥感影像和水系数据。根据上一节描述的三类淤地坝特征，在沟道里手动标记样本。样本包含淤地坝的坝体、卧管、溢洪道三大件设施，也包括周围的部分水体、坝地、道路等信息。数据标注形式及类别如（图2）所示。

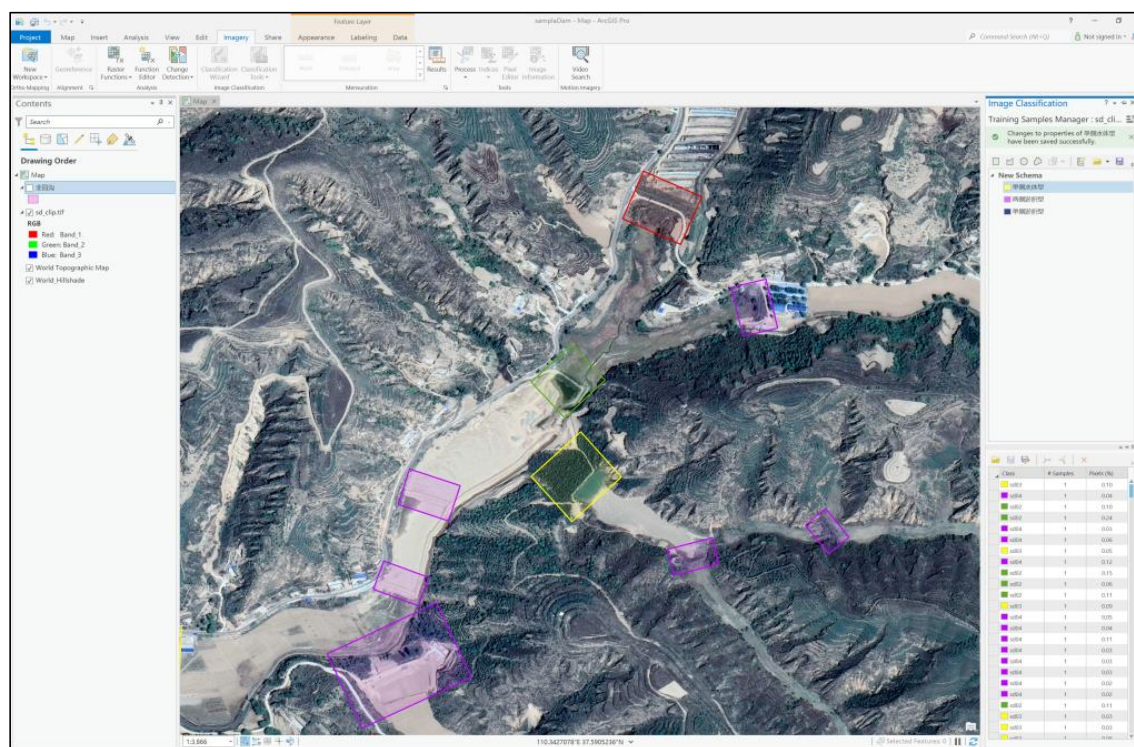


图2 淤地坝样本数据标注过程

Figure 2 The labeling process of sediment storage dam samples

在训练样本管理器中导出样本，样本为 tif 格式，样本大小 512×512 ，元数据模式采用“MetaData Mode”：“Pascal Visual Object Classes (VOC)”。“Pascal VOC 是一个广泛使用的数据集格式，用于计算机视觉任务中的对象检测和图像分割；Pascal VOC 数据集包含多种类别的带注释的图像，这些注释包括对象边界框、分割掩模以及对象类别标签。Pascal VOC 数据集通常包括一个注解文件夹、图像文件夹及其他说明文件。本文中，每个样本都包含一个 xml、tif 和 tfw 文件，其中 xml 文件用于标记基础信息，tif 文件是样本影像数据，tfw 文件存放影像坐标信息。

1.2.3 数据样本增强

在目标检测任务中，参与模型训练的数据质量会影响最终建立的模型精度，丰富、多样的样本能够使模型学习到更多的特征，提高模型的检测能力。通常使用数据增强来增加数据集的数量和多样性，常见的图片数据增强操作有翻转、裁剪、缩放、颜色调整等。本文采用图像 0.7–1.3 任意倍数缩放、 -60° – 60° 任意角度旋转、0.8–1.2 任意倍数亮度调节来增强数据。每个样本增强 4 个样本，结果如（图 3）所示，其中 A 列为原始影像；B–E 列为增强后的样本。

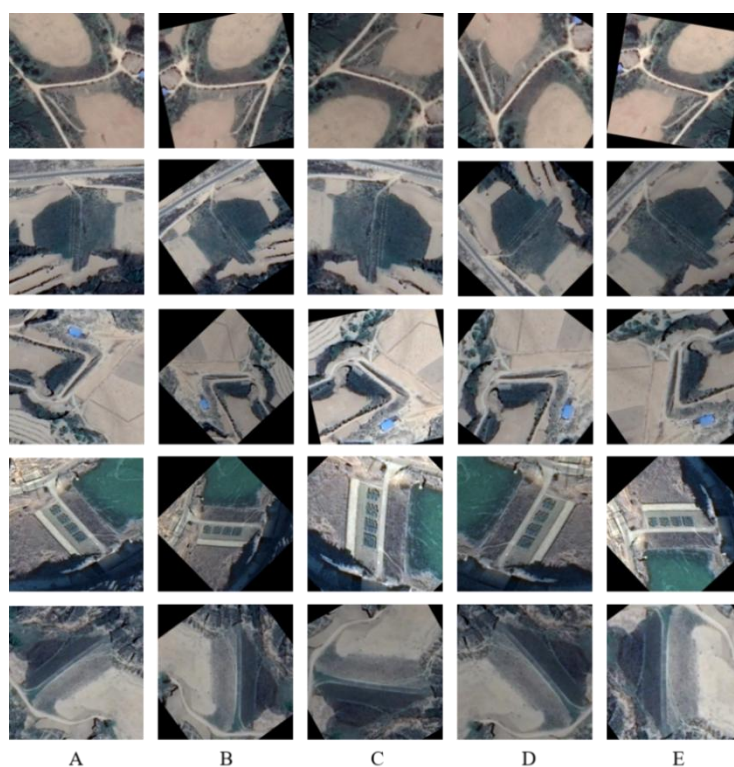


图 3 淤地坝增强样本示意图

Figure 3 Schematic diagram of enhanced sediment storage dam samples

2 数据样本描述

数据集包括手动标注的样本 sample 和增强的样本 augsample 两个文件夹。两个样本文件夹都包括 images 和与之对应的 labels 文件。每个样本包括 3 个相关的文件，其中 images 文件存放样本影像 tif 文件和影像坐标信息的 tfw 文件，labels 文件夹中存放样本影像的基础信息的 xml 文件。3 个文件具有相同的文件名（样本数据的文件组成如图 4 所示）。

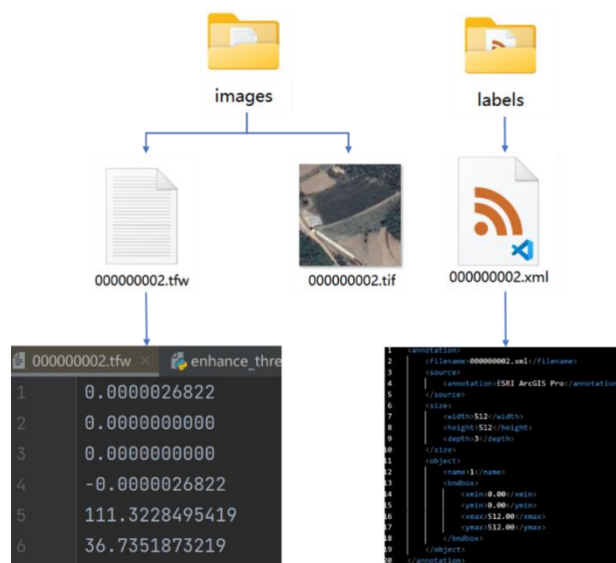


图 4 淤地坝样本数据的文件组成示意图

Figure 4 The file composition of sediment storage dam sample data

tfw 文件保存样本影像的地理坐标系信息，第 1 行：X-Scale，一个像元的 X 方向上的 X 分辨率尺度，单位为度；第 2 行：平移量；第 3 行：旋转量；第 4 行：负的 Y-Scale，一个像元在 Y 方向上的 Y 分辨率尺度负值，单位为度；第 5 行：左上角的 X 坐标；第 6 行，左上角的 Y 坐标。

xml 文件代表一张图像及其注释信息。是 Pascal VOC 数据集的 xml 格式，包含图像的名称、尺寸、边框信息。

3 数据质量控制和评估

(1) 样本地域分布均匀

样本标记来自黄河上中游淤地坝密集修建区域，主要包括陕西省榆林地区、山西省吕梁和汾西县等。据统计数据分析，这两个区域的淤地坝修建数量占全国的 60%以上^[11]，样本具有普适性。

(2) 与已有调查数据比较

选取典型区域，与已有淤地坝调查数据作对比。以陕西省绥德县王茂沟流域为例，据 2017 年调查数据显示^[12]，王茂沟共有淤地坝 22 座，其中骨干坝 2 座、中型坝 6 座、小型坝 14 座，淤地坝调查分布如图 5 (a) 所示。截取本数据集在王茂沟流域的样本如图 5 (b) 所示，共有 10 个淤地坝样本 S01-S10，与调查数据对比分析，发现本文数据标注的 10 个样本均为淤地坝。

(3) 淤地坝识别结果评价

为验证本样本集的有效性，采用 YOLOV7 深度学习算法^[13]进行淤地坝识别模型训练。YOLO (You Only Look Once) 算法是一种实时目标检测算法，它将目标检测任务转化为一个回归问题，通过单个神经网络同时预测目标的边界框和类别，实现了快速且准确的目标检测^[13-15]。以陕西省绥德县韭园沟 19 级谷歌影像为例，进行淤地坝的识别，置信度设置为 0.5，IoU 阈值设置为 0.45，输出为预选淤地坝的分布经纬度、宽度信息。输出淤地坝分布如图 6 所示，共检测出预选淤地坝 881 个。结合水文分析法，首先设置河网缓冲区来剔除非沟道内的错误识别数据；其次，针对每个坝或者坝系提取集水区，提取每个集水区的地貌信息熵、沟谷比降、流向、流量、是否有淤地坝，构建集水区淤地坝筛选模型，剔除坡梗、沟道边坡等错误识别的数据，识别淤地坝 236 座。据统计数据分析^[16]，

截至 2017 年底, 韭园沟流域共有淤地坝 217 座, 其中骨干坝 27 座, 中型坝 40 座, 小型坝 150 座。将输出的结果与统计数据对比验证, 证明基于机器学习结合水文分析的方法淤地坝识别准确率可达 91.2%, Kappa 系数达到 0.862。

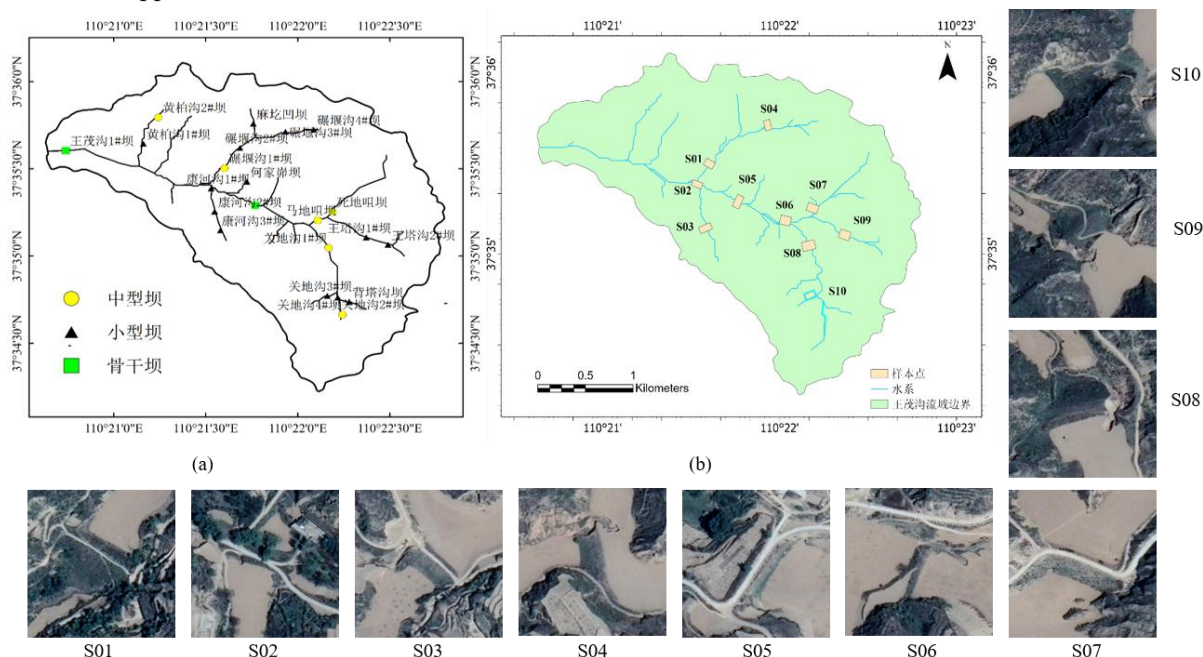


图 5 王茂沟流域淤地坝调查数据与样本数据对比

Figure 5 Comparison of survey data and sample data of sediment storage dams in Wangmaogou Basin

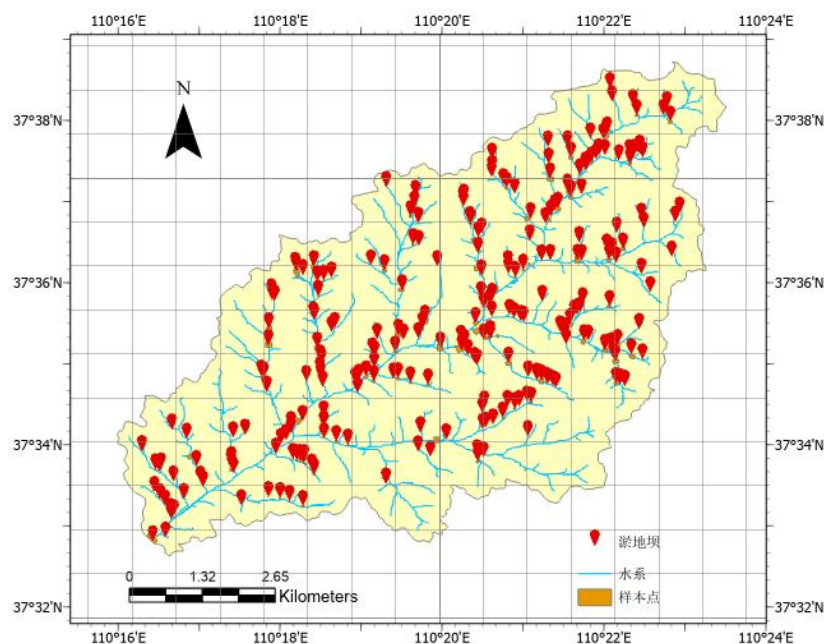


图 6 韭园沟流域淤地坝识别分布图

Figure 6 Distribution of sediment storage dams in Jiuyuangou Basin based on machine learning recognition

4 数据价值

以机器学习、深度学习为代表的人工智能方法也成为遥感影像自动化解译的主流技术手段, 在

精度和效率方面取得了跨越式的进步。遥感影像提取地物的主要对象是建筑物、水体、道路、不透水面等，淤地坝通常分布少、面积小，且中低分辨率遥感影像难以提取，故提取淤地坝的研究较少，目前尚未有基于高分辨率遥感影像的针对黄河上中游淤地坝样本集发布。本数据集可作为利用机器学习算法大范围快速识别黄河上中游淤地坝的基础样本集，支持 YOLO、SSD、Faster R-CNN 等图像识别算法进行训练，并结合水文分析等其他相关算法，建立高精度的淤地坝快速识别算法。通过定期更新影像，更换区域自动进行淤地坝识别，节省了野外调查、遥感解译的人力物力成本，为水土保持效益评价、淤地坝风险评价提供动态数据支撑。

数据作者分工职责

敏玉芳（1983—），女，甘肃省甘南州人，博士，高级工程师，研究方向为地学大数据管理与分析、基于人工智能的遥感数据信息提取。主要承担工作：数据集制作。

艾鸣浩（1985—），男，山东济宁市人，硕士，工程师，研究方向为地学大数据管理与分析。主要承担工作：数据准备及预处理。

张耀南（1966—），男，甘肃省天水人，博士，研究员，研究方向为遥感与地理信息、数据工程学及数据工程防灾减灾、多源数据融合。主要承担工作：数据处理过程指导。

黄波（1997—），男，四川资阳人，硕士在读，研究方向为淤地坝溃决风险性评估、数据工程应用。主要承担工作：数据集精度验证。

参考文献

- [1] XU Y D, FU B J, HE C S. Assessing the hydrological effect of the check dams in the Loess Plateau, China, by model simulations[J]. Hydrol. Earth Syst. Sci., 2013, 17(6): 2185–2193.
- [2] ABBASI N A, XU X, LUCAS-BORJA M E, et al. The use of check dams in watershed management projects: Examples from around the world[J]. Science of the Total Environment, 2019, 676: 683–691.
- [3] DING H, NA J, HUANG X, et al. Stability analysis unit and spatial distribution pattern of the terrain texture in the northern Shaanxi Loess Plateau[J]. Journal of Mountain Science, 2018, 15(3): 577–589.
- [4] 艾开开. 黄土高原淤地坝发展变迁研究[D]. 西北农林科技大学, 2019. [AI K K. Study on Development and Change of Warping Dams on Loess Plateau[D]. Xi'an: Northwest A & F University, 2019.]
- [5] 高健翎, 陈小科, 张佃茂. 黄土高原地区淤地坝现状分析[J]. 中国水土保持, 2023, 490(1): 1-5. [Gao J L, Chen X K, Zhang D M. Current Situation of Warping Dams in the Loess Plateau[J]. Soil and water conservation in china, 2023, 490(1): 1-5.]
- [6] 刘雅丽, 贾莲莲, 张奕迪. 新时代黄土高原地区淤地坝规划思路与布局[J]. 中国水土保持, 2020, (10): 23-27. [LIU Y L, JIA L L, ZHANG Y D. Thoughts and Layout of Warping Dam Planning of the Loess Plateau Region in the New Era[J]. Soil and water conservation in china, 2020, (10): 23–27.]
- [7] 李云飞. 融合多源数据和地形特征的黄土高原流域淤地坝识别研究[D]. 长安大学, 2023. [LI Y F. Study on Identification of Check Dams by Combining Multi-source Data and Topographic Features in Basin of the Chinese Loess Plateau[D]. Xi'an: Chang'an University, China, 2023]

- [8] 周培诚, 程堪, 姚西文, 等. 高分辨率遥感影像解译中的机器学习范式[J]. 遥感学报, 2021, 25(1): 182–197. [ZHOU P C, CHEN G, YAO X W, et al. Machine learning paradigms in high-resolution remote sensing image interpretation[J]. National Remote Sensing Bulletin, 2021, 25(1): 182–197.]
- [9] ZENG Y, JING T, XU B, et al. Vectorized dataset of silted land formed by check dams on the Chinese Loess Plateau[J]. Scientific Data, 2024, 11(1): 348.
- [10] 马煜栋, 杨帅, 韩静, 等. 陕西榆林地区无定河流域淤地坝遥感解译[J]. 干旱区地理, 2022, 45(3): 786–791. [MA Y D, YANG S, HAN J, et al. Remote sensing interpretation of check dams in Wuding River Basin in Yulin Area of Shaanxi Province[J]. Arid Land Geography, 2022, 45(3): 786–791.]
- [11] 刘雅丽, 王白春. 黄土高原地区淤地坝建设战略思考[J]. 中国水土保持, 2020, (9): 48–52. [LIU Y L, WANG B C. Pondering over Construction Strategy of Warring Dams in Loess Plateau Area[J]. Soil and water conservation in china, 2020, (9): 48–52.]
- [12] 钟少华. 王茂沟流域淤地坝防洪风险评价与除险方法研究[D]. 西安理工大学, 2021 [ZHONG S H. Study on Flood Risk Assessment and Risk Removal Method of Check Dam in WANGMOGOU Watershed[D]. Xi'an: Xi'an University of Technology, 2021.]
- [13] WANG C Y, BOCHKOVSKIY A, LIAO H. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Computer Vision & Pattern Recognition, 2022. DOI:10.48550/arXiv.2207.02696.
- [14] BOCHKOVSKIY A, WANG C, LIAO H M. Yolo v4: Optimal speed and accuracy of object detection[C]//Computer Vision & Pattern Recognition, 2020. DOI:10.48550/arXiv.2004.10934.
- [15] REDMON J, DIVVALA S, GIRSHICK R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]//Computer Vision & Pattern Recognition, 2016. DOI:10.48550/arXiv.1506.02640
- [16] 杨媛媛. 黄河河口镇-潼关区间淤地坝拦沙作用及其拦沙贡献率研究[D]. 西安理工大学, 2022. [YANG Y Y. Analysis on Sediment Retention of Check Dams and Its Contribution Rate of Sediment Reduction in Check Dams in HEKOU-TONGGUAN Section of the Yellow River, China[D]. Xi'an: Xi'an University of Technology, 2022.]

论文引用格式

敏玉芳, 艾鸣浩, 张耀南, 等. 黄河上中游淤地坝高分辨率光学遥感卫星影像样本数据集[J/OL]. 中国科学数据, 2024, 9(4). (2024-12-27). DOI: 10.11922/11-6035.ncdc.2024.0003.zh.

数据引用格式

敏玉芳, 艾鸣浩, 张耀南, 等. 黄河上中游淤地坝高分辨率光学遥感卫星影像样本数据集[DS/OL]. 国家冰川冻土沙漠科学数据中心, 2024. (2024-02-29). DOI: 10.12072/ncdc.nieer.db4212.2024.

A dataset of high-resolution optical remote sensing satellite image samples of sediment storage dams in the upper and middle reaches of the Yellow River

MIN Yufang^{1,2*}, AI Minghao^{1,2}, Zhang Yaonan^{1,2}, HUANG Bo^{1,2,3}

1. Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, P.R. China
2. National Cryosphere Desert Data Center, Lanzhou 730000, P.R. China
3. University of Chinese Academy of Science, Beijing 100010, P.R. China

*Email: myf@lzb.ac.cn

Abstract: As an effective soil and water conservation engineering measure in Loess Plateau area, sediment storage dams play an important role in trapping mud and preserving soil, reducing yellow sediment, controlling floods, mitigating disasters, increasing grain production, and promoting rational use of water resources. Basic data, such as the distribution and area of sediment storage dams, are essential for regularly carrying out ecological benefit assessments and risk susceptibility evaluations of the dams. Based on 19-level remote sensing image data of Google Earth, this study labeled 714 sediment storage dam samples through visual interpretation in the upper and middle reaches of the Yellow River. All samples have been further processed through the technique of image enhancement, resulting in a total of 3,570 sediment storage dam samples. Based on this sample set, a sediment storage dam identification model was trained. Taking Suide County of Shaanxi Province as an example, the model achieved a recognition accuracy of 91.5%. This sample set can serve as the basic data for image recognition of sediment storage dams, and promote the automatic recognition of sediment storage dams in the upper and middle reaches of the Yellow River. By regularly updating images and automatically extracting sediment storage dams based on changing areas, the dataset can help save reduce labor and material costs associated with field investigation and remote sensing interpretation. Moreover, it can also provide dynamic data support for soil and water conservation benefit assessments and sediment storage dam risk evaluations.

Keywords: Sediment storage dam; remote sensing image; machine learning; image recognition

Dataset Profile

Title	A dataset of high-resolution optical remote sensing satellite image samples of sediment storage dams in the upper and middle reaches of the Yellow River
Data authors	MIN Yufang, AI Minghao, ZHANG Yaonan, HUANG Bo
Data corresponding author	MIN Yufang (myf@lzb.ac.cn)
Time range	2019–2021
Geographical scope	Upper and middle reaches of the Yellow River (36°31′–39°35′N, 106°43′–111°28′E)
Spatial resolution	0.54 m
Data volume	2,567 MB

Data format	*.tif, *.tfw, *.xml
Data service system	< http://dx.doi.org/10.12072/ncdc.nieer.db4212.2024 >
Sources of funding	National Key Research and Development Plan (2021YFF0704200); National Key Research and Development Plan (2022YFF0711700).
Dataset composition	The dataset consists of samples and augsamples recorded in 2 subset folders. (1) The samples folder stores 714 entries of manually marked samples, with a data volume of 423 MB; (2) the augsamples folder stores 2,856 entries of enhanced samples, with a data volume of 2,144MB.