

基于机器学习的生产建设扰动范围识别技术研究

——以上海市浦东新区为例

甘海燕

上海市浦东新区水文水资源管理事务中心, 上海

收稿日期: 2025年11月24日; 录用日期: 2025年12月15日; 发布日期: 2025年12月26日

摘要

为准确掌握生产建设扰动范围, 提升水土保持监督管理工作效率, 以上海市浦东新区为研究案例, 探索了基于机器学习的生产建设扰动范围识别技术研究。该研究开发了基于遥感指数计算、轮廓识别的面向对象人工智能识别方法。通过调用ENVI/IDL和ArcGIS Engine软件的图像处理功能, 结合实地调研数据对比, 实现水土保持生产建设项目识别的自动化。所识别的内容包括生产建设项目扰动区域的地理位置、形状轮廓和面积范围。研究结果表明该方法总体精度正确率达86.2%, 具有巨大的应用潜力。

关键词

机器学习, 生产建设扰动范围, 图像处理

Research on the Identification Technology of Production and Construction Disturbance Range Based on Machine Learning

—Taking Pudong New Area, Shanghai as an Example

Haiyan Gan

Shanghai Pudong New Area Hydrology and Water Resources Management Affairs Center, Shanghai

Received: November 24, 2025; accepted: December 15, 2025; published: December 26, 2025

Abstract

To accurately grasp the disturbance range of production and construction activities and enhance

the efficiency of soil and water conservation supervision and management, this study takes Pudong New Area in Shanghai as a case study and explores the technology of identifying the disturbance range of production and construction activities based on machine learning. The research utilizes an object-oriented artificial intelligence recognition method that combines remote sensing index calculation and contour recognition. By utilizing the image processing functions of ENVI/IDL and ArcGIS Engine software, and comparing with field survey data, it achieves automation in the identification of soil and water conservation production and construction projects. The identified content includes the geographical location, shape contour, and area range of the disturbance areas caused by production and construction projects. The research results show that the overall accuracy of this method reaches 86.2%, indicating its great potential for application.

Keywords

Machine Learning, Production and Construction Disturbance Range, Image Processing

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

该研究旨在探索一种基于机器学习的自动化技术，用于识别生产建设项目造成的地表扰动范围，以提高水土保持监管工作的效率。研究以上海市浦东新区为案例区，利用 2010 年和 2020 年两个时期的 Landsat 遥感影像作为主要数据源。其技术路线核心是采用一种结合遥感指数计算(如 NDVI、NDBI)和轮廓识别(Sobel 算子)的面向对象识别方法。通过调用 ENVI/IDL 和 ArcGIS Engine 软件功能，该研究实现了对生产建设项目扰动区域的地理位置、形状轮廓和面积范围的自动化识别。研究通过时序对比分析，筛选出十年间新增的扰动图斑，并将其与官方备案的防治责任范围矢量数据进行空间比对，以进行合规性分析。研究结果表明，该方法的总体精度达到 86.2%，证明了其在实际监管工作中的应用潜力。

2. 项目总体架构

本项目从地物识别、扰动识别、结果分析三个部分构成，服务于水土保持监管辅助决策。总体组织架构如图 1 所示。

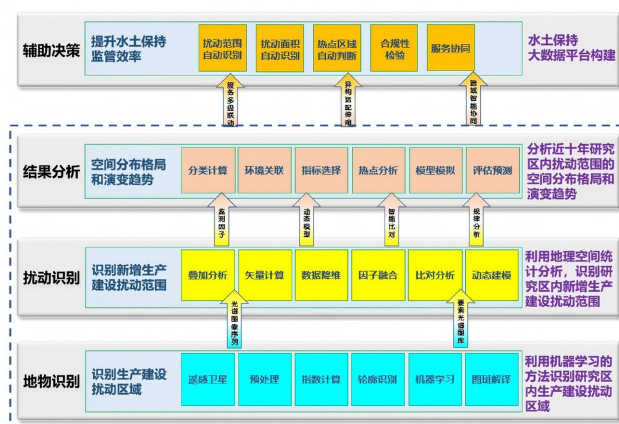


Figure 1. Overall organizational structure diagram
图 1. 总体组织架构图

3. 项目技术路线及研究内容

3.1. 技术路线

围绕项目总体研究目标,通过调用 ENVI/IDL 和 ArcGIS Engine 软件的图像处理功能,基于遥感数据和实地调研数据,实现水土保持生产建设项目的自动化识别功能,所识别的内容包括生产建设项目扰动区域的地理位置、形状轮廓和面积范围。并以浦东新区近十年(2010 年和 2020 年)内扰动范围的空间分布格局和演变趋势为案例进行系统分析研究。技术路线中,输入资料为 Landsat 系列卫星的波段数据,归一化植被指数数据、归一化建筑指数数据、归一化水体指数,纹理特征数据,输出数据为扰动图斑的统计数量、新增扰动图斑统计(包括地理位置、数量、面积和轮廓特征)。

3.2. 研究内容

3.2.1. 识别研究区内生产建设扰动区

整理和收集浦东区的遥感卫星影像;遥感数据进行预处理,包括几何校正、大气校正、辐射校正、拼接和裁剪等[1];计算归一化植被指数、归一化建筑指数及纹理特征,区分硬化地表和自然覆被地[2];基于 Sobel 算子识别生产建设项目区域和非生产建设项目的地物轮廓。利用机器学习,建立多个已知区域的典型地物样本区,提取特征波段、指数和纹理特征,面向对象的识别区域内生产建设扰动区域;利用实地调查和更高分辨率的遥感或航拍数据校验分类精度。

3.2.2. 识别研究区内新增生产建设扰动范围

利用地理空间统计分析,将两个不同时期的识别后的数据进行叠加分析,筛选出新增的生产建设区域,识别其空间地理位置和面积,计算其斑块平均面积、数量和形状特征;将 2020 获取的扰动图斑数据与实际上报的生产建设防治责任范围(水保监管 4.0 系统的矢量图)进行空间比对(对比图斑超过 50 个),将 2010 年至 2020 年新增扰动图斑数据,运用更高精度的遥感和航空影像比对(对比图斑超过 50 个),校验其合规性并进行现场核实,记录方法识别的精度。并选取典型扰动项目 30 个进行了现场核实。对满足防治责任范围矢量化要求的项目进行合规性初步分析,将监管区域扰动图斑矢量图(用 Y 表示,虚线)与防治责任范围矢量图(用 R 表示,实线)进行空间叠加分析,判定生产建设项目扰动合规性[3],如图 2 所示。

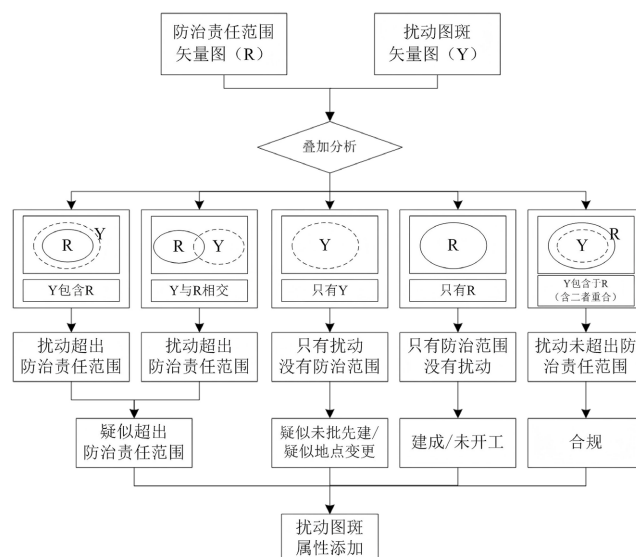


Figure 2. Compliance analysis technical process
图 2. 合规性分析技术流程

3.2.3. 分析研究区内扰动范围的空间分布格局和演变趋势

在获取 2010 年及 2020 年上海市浦东新区内的生产建设扰动区域数据的基础上，分析生产建设扰动区域的数量和空间分布格局和热点区域。热点分析的目标就是识别出具有统计显著性聚类的区域，因为这说明这些事件正在受某些空间过程因子的影响，存在空间上的关联性，本研究冷热点分析包括三部分，具体包括 2010 年上海市浦东新区生产建设项目扰动区冷热点分布，2020 年上海市浦东新区生产建设项目扰动区冷热点分布，以及 2010 年至 2020 年上海市浦东新区新增生产建设项目扰动区冷热点分布。

4. 技术方法流程

4.1. 基于卫星遥感数据，利用机器学习的方法识别研究区内生产建设扰动区域

① 基于 Landsat TM/ETM 数据，以上海市浦东新区行政区域作为研究区，以两个不同时期(2010 年和 2020 年)同一季节少云(云量 < 5%)的高分辨率多波段遥感影像作为基础数据，其中 Landsat TM/ETM 为水土保持生产建设扰动范围识别数据集，实地调查为分类结果的验证数据集。具体数据参见表 1。Landsat 系列数据来源于地理空间数据云，2010 年 Landsat 7 TM，日期 2010 年 5 月 25 日，轨道号分别是，条带号 118，行号 38 和 39；2020 年的 Landsat 7 ETM，日期是 2020 年 4 月 18 日，轨道号分别是，条带号 118，行号 38 和 39。

Table 1. Introduction to Landsat satellite data
表 1. Landsat 卫星数据介绍

TM 传感器	光谱范围(纳米)	分辨率(米)
Band 1	0.45~0.52	30
Band 2	0.52~0.60	30
Band 3	0.63~0.69	30
Band 4	0.76~0.90	30
Band 5	1.55~1.75	30
Band 6	10.40~12.50	120 (30)
Band 7	2.08~2.35	30

② 基于 ENVI 和 ArcGIS 对遥感数据进行预处理，数据预处理的过程包括几何校正(地理定位、几何精校正、图像配准、正射校正等) [4]，几何校正控制误差参数在 1 个象元内，图像融合、图像镶嵌、图像裁剪、去云及阴影处理和大气校正等几个环节[5]。

③ 计算归一化植被指数(NDVI)、归一化建筑指数(NDBI)、归一化水体指数(NDWI)及纹理特征[1]。利用自然间断点分级法(Jenks)，区分硬质化地表和自然覆被地。归一化建筑指数可以较为准确地反映建筑用地信息，数值越大表明建筑用地比例越高，建筑密度越高。归一化植被指数是反映农作物长势和营养信息的重要参数之一[6]。

④ 基于 Sobel 算子识别生产建设扰动区域的轮廓；利用机器学习(支持向量机分类器)，建立多个已知区域的生产建设扰动区域典型地物样本区，提取特征波段、指数和纹理特征，面向对象的识别区域内生产建设扰动区域，Sobel 算子基本计算公式如下：

$$A = \begin{pmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{pmatrix}$$

$$B = \begin{pmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix}$$

4.2. 利用地理空间统计分析，识别研究区内新增生产建设扰动范围

- ① 基于 ArcGIS 软件，利用地理空间统计分析，将两个不同时期(2010 年和 2020 年)研究区范围内的识别后生产建设扰动区域矢量数据进行叠加分析，筛选出 2010 至 2020 年新增的生产建设扰动区域，识别其空间地理位置和面积，地理位置包含扰动地表每个小斑块的经度和纬度信息，面积代表扰动地表的覆盖范围大小。
- ② 利用景观格局指数计算其斑块平均面积、数量和形状特征；将获取的数据与实际建设单位上报的生产建设防治责任范围(即扰动图斑)进行空间比对，校验其合规性。景观指数能定量反映景观的结构组成与空间配置，是浓缩景观格局信息的指标[7]。
- ③ 判定“未批先建、未批先变、未验先投”的生产建设项目并进行现场核实，记录方法识别的精度。

4.3. 基于生产建设扰动区域识别数据，分析近十年研究区内扰动范围的空间分布格局和演变趋势

- ① 在获取 2010 年和 2020 年上海市浦东新区内的生产建设扰动区域数据的基础上，分析生产建设扰动区域 的数量和空间分布格局，统计各街道生产建设扰动区域的面积和数量；
- ② 基于 ArcGIS 软件，利用空间热点分析，识别生产建设扰动范围的空间热点和冷点区域；
- ③ 记录热点区域所处的街道区域及数量，明确浦东新区生产建设活动对自然地表造成的扰动情况，为浦东新区的水土保持工作和水土流失治理提供数据支撑和理论依据。定量探讨新增水土保持生产建设项目空间冷热点分布特征，揭示水土保持生产建设项目聚集程度和变化趋势，并可用于识别“热点区域”以及数据的异质检验。

5. 研究结果

本研究利用地理空间统计分析，将两个不同时期的数据识别后进行叠加分析，筛选出新增的生产建设区域，识别其空间地理位置和面积，计算其斑块平均面积、数量和形状特征。以 2010 年和 2020 年浦东新区各街道范围内新增生产建设扰动图斑为例，2010 年~2020 年，浦东新区新增扰动图斑 1143 个，新增扰动面积 1148.57 km²。

经对 2020 获取的扰动图斑数据与实际上报的生产建设防治责任范围(水保监管 4.0 系统的矢量图)进行空间比对(对比图斑超过 50 个)，具体数据参见表 2。经现场核实，基于机器学习的生产建设扰动范围识别技术总体精度正确率达 86.2%。

Table 2. Detailed information on production and construction disturbance patterns within the streets of Pudong New Area in 2010 and 2020
表 2. 2010 年和 2020 年浦东新区各街道范围内生产建设扰动图斑详细信息

FID	镇名称	斑块边界长度(km)	数量(个)	面积(km ²)
0	北蔡镇	0.29	29	12.22
1	曹路镇	0.45	36	12.52
2	川沙新镇	0.74	53	87.82
3	大团镇	0.54	44	80.95

续表

4	东明路街道	0.12	16	1.35
5	高东镇	0.39	33	24.04
6	高桥镇	0.29	27	11.12
7	高行镇	0.29	19	4.93
8	航头镇	0.56	68	96.86
9	合庆镇	0.33	34	40.18
10	沪东新村街道	0.09	12	7.12
11	花木街道	0.21	24	10.31
12	惠南镇	0.62	36	59.76
13	金桥镇	0.34	32	17.16
14	金杨新村街道	0.12	27	9.46
15	康桥镇	0.50	36	32.74
16	老港镇	0.58	45	49.33
17	陆家嘴街道	0.11	19	1.18
18	南汇新城镇	0.75	95	178.78
19	南码头路街道	0.09	14	7.3
20	泥城镇	0.44	22	11.41
21	浦兴路街道	0.11	13	1.72
22	三林镇	0.5	11	10.51
23	上钢新村街道	0.11	17	5.69
24	书院镇	0.41	49	62.34
25	唐镇	0.34	29	14.34
26	塘桥街道	0.09	13	7.97
27	万祥镇	0.28	39	29.12
28	潍坊新村街道	0.09	10	2.86
29	新场镇	0.56	43	80.48
30	宣桥镇	0.66	58	81.54
31	洋泾街道	0.11	11	3.84
32	张江镇	0.5	36	31.22
33	周家渡街道	0.11	9	0.52
34	周浦镇	0.54	48	42.79
35	祝桥镇	0.91	36	17.09

6. 结论与建议

通过基于机器学习的生产建设扰动范围识别技术研究,以及相关技术在上海市浦东新区水土保持监管工作中的成功应用,证明了该方法的可靠性,未来可推广到上海市其他各区县及类似省份水土保持工作中,可大大减少人工工作量。

但是该方法也存在一定的局限性：一是对影像资料的依赖性较强，采用的高时空分辨率数据成本高、覆盖范围小、重访周期长，难以满足动态监测的要求；二是处理混合像元能力有限，可能存在同物异谱或同谱异物；三是对不同类型建设项目存在识别差异，因水土保持扰动类型的多样性，不同扰动类型的光谱、纹理和形状特征差异巨大，一个模型很难高精度地识别所有类型的扰动。

为更好地推进“基于机器学习的生产建设扰动范围识别技术”从理论研究走向实际业务应用，建议从以下两个方面进行改进：一是推动一体化平台建设，将机器学习模型与 GIS、业务管理系统深度集成；二是加强多学科交叉研究，将遥感识别与水土保持学、土壤学等学科融合，从而更好地推进该技术的发展应用。

参考文献

- [1] 易扬, 胡昕利, 史明昌, 等. 基于 MODIS NDVI 的长江中游区域植被动态及与气候因子的关系[J]. 生态学报, 2021, 41(19): 7796-7807.
- [2] 易扬, 张桂莲, 张浪, 等. 基于高分二号遥感影像的城市植被三维绿量研究[J]. 园林, 2020(11): 2-7.
- [3] 徐坚, 钟秀娟, 王崇任, 丁杨, 高之栋. 生产建设项目水土保持“天地一体化”监管模式研究[J]. 中国水利, 2019(20): 48-50.
- [4] 王爽. 遥感与物联网技术在智慧农业灾情精准分析中的应用研究[J]. 产业与科技论坛, 2023, 22(14): 31-33.
- [5] 韩帅. 呼和浩特城市空间结构演变研究[D]: [硕士学位论文]. 呼和浩特: 内蒙古师范大学, 2020.
- [6] 王志国. 基于多源数据的典型草原植被与土壤相关性研究[D]: [博士学位论文]. 呼和浩特: 内蒙古农业大学, 2014.
- [7] 沈中健, 曾坚. 闽南三市城镇发展与地表温度的空间关系[J]. 地理学报, 2021, 76(3): 566-583.