

工程测绘大数据背景下智能数据分析模型的构建与应用

赵飞¹, 吴旭涛²

(1. 西安坐标原点地理空间研究院有限公司, 陕西 西安 710000;

2. 陕西北斗星空规划设计有限公司, 陕西 西安 710000)

摘要 针对工程测绘向“数据驱动”转型中传统人工处理效率低、融合度差、精度有限的瓶颈, 本文构建“数据治理—算法优化—场景适配”智能分析技术体系。数据层通过“空天地”三级采集与“清洗—融合—增强”预处理提升数据质量; 算法层采用“大模型+小模型”架构, 结合 XGBoost、CNN 等算法及知识蒸馏优化策略; 应用层适配输电工程、耕地监测等场景。研究结果显示, 输电工程设计效率提升 70%, 耕地监测产量估算误差降至 10.02%, 为测绘数字化转型提供技术支持。

关键词 工程测绘大数据; 智能数据分析模型; 多源数据融合; 深度学习算法; 工程场景适配

中图分类号: TP31

文献标志码: A

DOI: 10.3969/j.issn.2097-3365.2025.34.013

0 引言

随着 Sentinel 系列卫星、无人机 LiDAR、地面激光扫描等设备的普及, 测绘数据已形成“空天地”一体化采集网络, 单项目数据量从 GB 级跃升至 TB 级, 呈现出多源异构、高维稀疏、实时动态的典型特征^[1]。传统基于人工解译的处理模式面临三重瓶颈: 数据处理周期长(单幅高分辨率影像解译需 72 小时以上)、多源数据融合度低(SAR 与光学影像协同利用率不足 30%)、动态监测精度有限(工程形变预警滞后超 24 小时), 难以满足智能建造对实时性、精准性的需求^[2], 这些问题需通过技术创新、体系构建与生态培育协同破解。本研究为工程测绘数字化转型提供核心技术支撑, 助力新质生产力在基础设施领域落地。

1 工程测绘大数据智能分析技术发展现状

工程测绘作为基础设施建设的核心支撑技术, 正经历从“经验驱动”向“数据驱动”的范式转型^[3]。工程测绘行业正处于“数据爆炸”与“智能转型”的关键交汇期, 传统人工主导的处理模式已难以适配 TB 级多源异构数据的应用需求, 构建精准、高效的智能数据分析模型成为行业突破瓶颈的核心路径。

当前国际研究聚焦两大方向: 一是多源数据融合算法, 如基于极端梯度提升(XGBoost)的遥感特征融合模型, 在冬小麦产量估算中实现 $R^2=0.654$ 、 $nRMSE=10.02\%$ 的精度; 二是行业大模型应用, 如四川测绘地

理信息局基于 DeepSeek 构建的“测智慧”平台, 已实现质检误差率降低 60%、成本核算准确率达 98% 的突破。国内研究则侧重工程场景落地, 如无人机测绘与 CNN 结合的地形提取技术、LiDAR 点云智能分类系统等, 但仍存在模型通用性不足、专业知识融入不深等问题^[4]。

构建适配工程测绘场景的智能数据分析模型, 对行业发展具有双重价值: 在技术层面, 通过融合深度学习与测绘专业知识, 可将数据处理效率提升 3~5 倍, 特征提取精度突破 95% 阈值; 在工程层面, 可实现输电线路选线、耕地动态监测等场景的全流程智能化, 如国网嘉兴供电公司应用智能模型后, 施工图设计周期缩短 70%, 关键信息识别准确率达 96% 以上^[5]。

2 工程测绘大数据智能分析核心技术体系

2.1 数据层技术: 多源异构数据治理

1. 数据采集体系。构建“卫星—无人机—地面”三级采集网络: 光学卫星(Sentinel-2)提供 10 m 级多光谱数据, SAR 卫星(Sentinel-1)实现全天候监测, 无人机 LiDAR 获取亚米级点云, 地面传感器采集厘米级形变数据。以输电工程为例, 该体系可实现地形、植被、建筑等 12 类要素的同步采集, 数据覆盖率提升至 99.2%。

2. 数据预处理流水线。采用“清洗—融合—增强”三步法: 通过孤立森林算法剔除点云异常值(误差小于 0.1 m), 基于注意力机制融合 SAR 后向散射系数与植

被指数 (NDMI、NDVI_{re3} 等核心特征保留率达 98%)，利用生成对抗网络 (GAN) 扩充样本集 (数据量提升 4 倍)。预处理后的数据完整性从 65% 提升至 92%，为模型训练奠定基础。

2.2 算法层技术：智能模型选型与优化

1. 特征提取算法。针对不同数据类型定制方案：遥感影像采用 CNN-LSTM 混合网络，提取光谱、纹理、时序三重特征；点云数据通过 PointNet++ 实现三维特征编码；矢量数据采用图神经网络 (GNN) 挖掘空间拓扑关系。在唐山市冬小麦监测中，红边植被指数 (NDVI_{re2/3}) 与水汽指数 (NDMI) 的特征重要性占比超 60%。

2. 核心预测模型。建立“大模型+小模型”协同架构：以 DeepSeek 等基础大模型为基座，注入 600 余份测绘规范、1 700 万知识点构建专业知识库；针对具体场景部署轻量化模型，如输电线路选线采用强化学习算法 (成本降低 15%)，灾害监测采用 XGBoost 模型 (预警响应时间缩短至 2 小时)。模型对比实验显示，XGBoost 在测绘回归任务中精度显著优于随机森林 (RMSE 降低 0.014 t·hm⁻²) 与支持向量机 (R² 提升 0.115)。

3. 模型优化策略。采用“知识蒸馏+迁移学习”双路径：将大模型知识蒸馏至边缘设备模型 (参数压缩 75%，推理速度提升 3 倍)；通过预训练—微调模式适配新场景，如从耕地监测迁移至建筑形变监测时，仅需 5% 新数据即可达到 90% 精度。

2.3 应用层技术：工程场景适配

开发“数据—模型—服务”一体化接口：通过 X inference 框架调度多模型资源，RAGFlow 框架实现知识精准检索，Dify 框架构建自动化 workflow。该接口已适配六大工程场景：输电工程勘察、耕地动态监测、国土空间规划、生态屏障监管、防灾减灾预警、民生服务响应，形成全链条智能化解决方案。

3 智能数据分析模型构建流程与实例

3.1 标准化构建流程

1. 需求解析与指标定义。明确工程核心指标：在输电线路工程中，定义路径规划成本降低率 (≥ 10%)、地物识别准确率 (≥ 95%)、报告生成效率提升率 (≥ 60%) 等关键指标；在耕地监测中，设定“非农化”识别精度 (≥ 92%)、产量估算误差 (≤ 12%) 等目标。

2. 数据准备与特征工程。以唐山市冬小麦估产模型为例：采集 2024 年 Sentinel-1/2 影像 (覆盖 1.3 万 km²) 与 120 个样点产量数据，构建 24 个特征变量 (含 12 种植被指数、2 种极化系数及 4 种组合参数)，

通过互信息熵筛选出 NDMI、NDVI_{re3} 等 8 个核心特征 (累积重要性占比 98.2%)。

3. 模型训练与验证。采用“五折交叉验证+实时质控”模式：将数据集按 7:3 划分训练集与测试集，XGBoost 模型经 100 轮迭代后，在测试集上实现 R²=0.654、RMSE=0.499 t·hm⁻² 的性能；通过混淆矩阵分析，耕地“非农化”识别的 F1 值达 0.93。

4. 部署与迭代优化。采用“云端+边缘”混合部署：云端部署大模型处理全局数据，边缘设备部署轻量化模型实现实时分析 (如无人机端部署 CNN 模型，影像解译延迟 < 10 秒)。建立模型迭代机制，每季度融入新增工程数据 (≥ 5 000 条) 与专家规则 (≥ 20 条)，确保模型适配性。

3.2 输电工程智能勘察模型实例

1. 模型架构。融合轻量化 CNN 与强化学习的双模块模型：CNN 模块采用 U-Net 架构，实现地形、植被、建筑的语义分割 (mIoU=0.91)；强化学习模块以成本、环保、安全为约束条件，生成最优路径方案 (搜索效率提升 10 倍)。

2. 工程应用效果。在 110 千伏余新线改迁工程中：数据采集阶段，亚米级精度提取地形信息 (误差 ≤ 0.3 m)；路径规划阶段，自动平衡拆迁成本与生态保护，生成 3 套优化方案；成果输出阶段，一键生成施工图与报告 (耗时从 5 天缩短至 1.5 天)。综合效率提升 70%，人工干预减少 85%。

3.3 耕地动态监测模型实例

1. 模型创新点。引入知识图谱增强的 XGBoost 模型：构建包含 2 500 个节点的耕地要素知识图谱，将“土壤类型—植被指数—产量”关联规则融入模型训练，解决传统模型解释性不足问题。

2. 应用成效。在四川省耕地监测中，模型实现三大突破：(1)“非农化”行为识别耗时从 48 小时降至 3 小时；(2)产量估算误差从 15% 降至 10.02%；(3)生成的产量空间分布图与地面真值吻合度达 91% (7.00 ~ 8.00 t·hm⁻² 区间占比 40.75%)。

4 行业应用价值与发展挑战

4.1 多维应用价值

1. 效率提升。智能模型以自动化替代人工，实现全链条效率跃升。遥感影像处理中，传统人工解译 1 m 分辨率影像需 72 小时，智能模型借 CNN-LSTM 网络将周期从 15 天缩至 3 天，某省遥感数据中心应用后年度处理量从 500 幅提至 2 000 幅。工程设计领域，110 千伏余新线改迁工程用“LiDAR 点云+强化学习”，设计

周期从 30 天缩至 9 天。质检环节,某测绘院应用智能模型后效率提 4 倍,10 人/天项目现仅需 2 人/天,漏检率从 8% 降至 1.2%。

2. 精度突破。智能模型融合多源数据与深度学习,突破精度瓶颈。地物识别中,传统人工准确率 75%~85%,智能模型借注意力机制与知识图谱超 96%,某城市建成区测绘中,建筑、道路、绿地识别准确率达 97.3%、98.1%、95.8%,远超传统的 82%、80%、78%。形变监测领域,传统 GPS 精度 ± 2 毫米,智能模型结合激光扫描与 SAR 干涉数据达 ± 0.3 毫米,某地铁工程借此规避施工风险。成本核算上,传统误差超 5%,智能模型融合 1 700 万条数据后误差低于 2%,某输电工程材料成本核算偏差仅 1.2%。

3. 管理升级。智能模型推动管理从“事后核查”转向“主动防控”。防灾减灾中,模型提前 72 小时圈定灾害隐患,某省 2024 年汛期前识别 12 处滑坡风险区,转移 200 余人,减损超 500 万元。耕地保护领域,“空天地”巡护网络将违规发现时间从 15 天缩至 2 天,某县 2024 年发现 32 起违规占耕行为,挽回 120 亩耕地,管控效率提 87%。

4.2 现存挑战

1. 技术瓶颈。多源数据融合与边缘算力成瓶颈。数据融合上,SAR 受大气散射、地表粗糙度影响有斑点噪声,光学影像受云层制约,二者融合需复杂校正,主流算法在平原协同利用率仅 45%,山区降至 35% 以下。边缘设备方面,无人机处理 10 cm 分辨率 LiDAR 点云时,XGBoost 模型推理延迟 65 秒,远超实时监测 10 秒阈值。

2. 数据壁垒。数据标准不统一与安全矛盾突出。各部门格式碎片化,电力用 DWG、国土用 Shapefile、住建用 GeoJSON,转换易失拓扑关系,某跨区域输电工程数据适配耗时 15 天,成本占预算 22%,超行业均值。数据安全上,某省“智慧测绘”平台仅 28% 部门开放核心数据,覆盖率不足 60%;企业拒共享训练数据,致模型样本不足。

3. 人才缺口。“测绘+AI”复合型人才匮乏。高校测绘专业 AI 课程占比不足 10%,某头部企业 2024 年招聘中,200 份简历仅 6 人兼具测绘背景与编程能力,占比 3%,远低于 15% 需求。

5 未来发展方向

一是技术创新。研发跨模态融合大模型,突破 SAR—光学数据协同解译瓶颈,目标将融合精度提升至 85%;探索存算一体芯片在边缘设备的应用,实现复杂模型推理延迟 < 5 秒。二是体系构建。建立全国统一

的工程测绘数据标准体系,制定 12 类核心数据的接口规范;构建“数据交易所+区块链”的共享机制,跨部门数据流通率目标达 70%。三是生态培育。打造“高校—企业—协会”协同培养体系,开设“测绘 AI”交叉学科;开发沉浸式培训系统,从业人员技能更新周期缩短至 6 个月。未来,随着跨模态融合大模型研发、全国统一数据标准建立及“测绘 AI”人才培养体系完善,智能模型将从单一场景赋能向全行业体系重构升级,不仅为工程测绘提供更高效的技术工具,更将推动基础设施建设向“精准化规划、智能化建造、动态化监管”转型,为数字经济与新质生产力发展注入测绘领域的核心动能。

6 结束语

本文系统构建了“数据治理—算法优化—场景适配”三级技术体系,实现测绘专业知识与人工智能技术的深度融合,为行业数字化转型提供可参考的技术方案。在技术实践层面,数据层通过“空天地”一体化采集网络与“清洗—融合—增强”预处理流水线,将数据完整性从 65% 提升至 92%,为模型训练奠定高质量数据基础;算法层创新采用“大模型+小模型”协同架构,结合 XGBoost、CNN 等算法与知识蒸馏、迁移学习优化策略,既保证核心任务(如耕地产量估算)的精度($R^2=0.654$ 、 $nRMSE=10.02\%$),又解决边缘设备部署难题(参数压缩 75%、推理速度提升 3 倍);应用层通过场景化适配,在输电工程、耕地监测等领域验证了模型价值——输电线路设计周期缩短 70%、地物识别准确率超 96%,耕地“非农化”识别耗时从 48 小时降至 3 小时,充分证明智能模型对工程效率与精度的双重提升作用。

参考文献:

- [1] 崔志伟,张博,韩宗文,等.数字化测绘技术在水利工程测量中的应用研究[J].水上安全,2025(06):148-150.
- [2] 王振春.测绘发展现状的工程测量测绘技术应用研究[J].中国设备工程,2024(18):241-243.
- [3] 秦秋金.数字化测绘技术在煤矿测量中的应用研究[J].内蒙古煤炭经济,2025(03):174-176.
- [4] 赵辉煌,黄志标,谢斌.机载雷达测绘技术下大坝沉降量的精准估计[J].北京测绘,2025,39(02):183-188.
- [5] 公艳茹,黄林林,韩晓亮.新城建测绘技术体系构建与地理信息数据管理研究[J].新城建科技,2025,34(02):86-88.