

研究论文

视觉大模型驱动高标准农田数字化遥感应用

沈靖^{1,2}, 刘泽^{1,3*}, 何亚文⁴

摘要 提出了一种融合提示工程与遥感特征知识驱动的高标准农田多要素自动提取流程框架,结合视觉大模型的通用分割能力,以及开源数据与遥感特征驱动的专用算法,实现了对农田地块、田间道路、农田防护林及灌溉与排水等关键要素的高效自动识别,以山东省寿光市高标准农田建设区为例,基于高分辨率国产卫星遥感影像开展了农田多要素遥感提取技术验证。实验结果表明,所提方法可在农田项目区实现批量自动处理,显著降低对大量高质量训练样本的依赖和人工处理工作量,并在不同时期、不同分辨率及不同地区的影像上均表现出良好的泛化能力,显著提升了数据生产效率和实际应用潜力。

关键词 高标准农田;农田多要素;遥感应用;视觉大模型;提示工程

农业可持续发展与粮食安全是联合国 2030 年可持续发展目标的重要组成部分^[1]。近年来,中国高度重视耕地保护与农田建设工作^[2-3],通过水利设施、机耕道路、输配电系统、防护林带等基础设施的统筹建设,改善耕地质量、提升地力水平和粮食生产能力,从而实现农田的稳定高产和可持续利用^[4-5]。在全球粮食安全形势日益严峻的背景下,高标准农田建设已成为保障国家粮食安全的核心战略举措。根据规划^[6],到 2030 年,全国将建成 12 亿亩(1 亩=0.0667 公顷)高标准农田,稳定支撑 1.2 万亿斤(0.6 万亿 kg)以上粮食产能。

尽管国家持续加大投入,推动高标准农田建设取

得积极成效,但部分地区仍存在重建设、轻管护等问题,表现为管护体系不健全、技术手段滞后等^[7],导致“断头渠”“花花路”“阴阳田”等不合理现象频发^[8],甚至出现耕地被占用、撂荒等情况,严重影响高标准农田的实际效益^[9]。在此背景下,提升建设与管理的信息化、智能化水平成为亟需解决的关键问题,推动农业数字化转型、构建高标准农田监测与管护体系势在必行。近年来,农业农村部推动农田上图、入库、可视、可查,中国已累计建成高标准农田超 10 亿亩^[10],初步实现数字化成果的空间化管理。高标准农田是一个涵盖“田、水、路、林、电、技、管”等软硬件一体化要素的综合农田体系,其遥感监测与数字化表征技术是实现精准管护与科学决策的核心支撑手段^[11-12]。鉴于当前遥感手段的可及性与可操作性,本研究聚焦于利用光学遥感影像实现农田地块、田间道路、农田防护林、灌溉与排水 4 类关键要素的提取与监测。

在这些要素的遥感监测领域,已有大量基于深度学习的研究取得显著进展,展现出良好的精度与应用

1. 中国科学院地理科学与资源研究所地理信息科学与技术全国重点实验室,北京 100101

2. 中国科学院大学,北京 100049

3. 中国地质大学(北京)人工智能学院,北京 100083

4. 中国石油大学(华东)海洋与空间信息学院,青岛 266580

收稿日期: 2025-05-12; 修回日期: 2025-07-07

基金项目: 国家重点研发计划项目(2022YFB3903501); 青岛市科技惠民示范专项项目(25-1-5-xdny-12-nsh)

作者简介: 沈靖,博士研究生,研究方向为遥感大数据,电子邮箱: shenjing6429@igsrr.ac.cn; 刘泽(通信作者),博士研究生,研究方向为遥感地质学,电子邮箱: 3004220010@email.cugb.edu.cn

引用格式: 沈靖, 刘泽, 何亚文. 视觉大模型驱动高标准农田数字化遥感应用[J]. 科技导报, 2025, 43(18): 86-98; doi:10.3981/j.issn.1000-7857.2025.05.00056

前景。例如,在农田地块识别方面,深度学习技术已广泛应用于区域与中小尺度影像处理,常采用分割网络(segmentation network, SegNet)、金字塔场景解析网络(pyramid scene parsing network, PSPNet)、长短期记忆网络(long short-term memory, LSTM)等网络结构^[13-15],在高分辨率影像上实现 80% 以上的分割精度;在道路提取方面,基于全卷积网络(fully convolutional network, FCN)、编码器-解码器结构网络(UNet)等架构^[16-18]对道路边界或中心线进行提取,通过形态学运算、矢量优化等后处理手段增强连通性;防护林提取则融合 RapidEye、PlanetScope 等亚米级卫星影像,结合 UNet 与残差网络(residual network, ResNet)结构识别树冠中心,实现对印度地区 20 余年期间农田防护林分布及变化的动态追踪^[19];灌溉与排水设施方面,研究普遍采用区域卷积神经网络(region-based convolutional neural network, RCNN)类目标检测框架,实现对大坝、桑基鱼塘、灌区等目标的精准识别^[20-21]。上述成果证明深度学习方法在农田单要素遥感提取中具有良好的适应性与有效性。

然而,目前研究仍存在 2 项突出问题:其一,现有方法多数聚焦单一要素,缺乏面向多要素的系统化遥感提取框架;其二,深度学习模型高度依赖大量高质量的标注样本,且在面对不同区域或时相的遥感影像时,模型易发生性能骤降。这种迁移能力差的问题主要源于数据在时间或空间上具有不同的数据分布,使得传统监督学习方法在实际应用中往往依赖大量专业知识与人工调整,制约其推广与应用。近年来,视觉基础模型(visual foundation models, VFM)^[22-23]为遥感智能化带来新突破。以分割一切模型(segment anything model, SAM)^[24]为代表的新一代大模型,具备零样本目标分割能力,可通过点、框或掩码等多种提示方式引导模型自动完成目标识别。SAM 基于大规模自然图像的预训练,在多任务场景中展现出优异的迁移能力,为遥感影像的智能解析提供了新范式。然而, SAM 在遥感领域的直接应用仍存挑战,需结合领域先验知识与工程化提示设计,引导其高效识别目标对象,从而在降低标注成本的同时实现准确监测。

基于此,本文提出一种融合开源数据、遥感特征与建设规则作为先验知识来源,结合提示工程引导 VFM 的高标准农田遥感数字化监测方法。以山东省

寿光市总面积约 2.3 万亩的高标准农田建设项目区为应用案例,基于 0.5 m 分辨率国产遥感影像,在无需人工标注条件下,实现农田地块、田间道路、农田防护林及灌溉与排水的自动化快速识别,并在多时相、多分辨率、多地区的遥感影像数据上进行应用验证。结果表明,该方法流程极大减少对大量高质量训练样本的依赖及人工处理,提升数据生产效率和成果标准化水平,能有效满足数字化管理需求,展现出良好工程实用价值与推广潜力。

1 高标准农田要素

1.1 高标准农田建设要素体系

高标准农田是指土地平整、集中连片、设施完善、节水高效、农电配套、宜机作业、土壤肥沃、生态友好、抗灾能力强,与现代农业生产和经营方式相适应的旱涝保收、高产稳产的耕地^[6]。国家标准^[11]中,高标准农田建设包含 8 要素:田块整治(田)、土壤改良(土)、灌溉与排水(水)、田间道路(路)、农田防护与生态环境保护(林)、农田输配电(电)、科技服务(技)、管护利用(管),形成软硬件一体化农田体系(图 1)。

《中华人民共和国土地管理法》(2019 年修订)^[25]将土地分为农用地、建设用地和未利用地,其中农用地包括耕地、林地、草地等;《土地利用现状分类》(GB/T 21010—2017)将耕地细分为水田、水浇地和旱地^[26-27]。在概念层面,农田与耕地在广义上同义;狭义上,高标准农田与永久基本农田均属优质耕地,一般农田中的耕地后备资源则为潜在可转化耕地的非耕地资源。按保护等级自高到低可排列为:高标准农田、永久基本农田和农田;按农田建设设施完善度由粗至细、包含关系自大到小可排列为:农用地、农田、耕地、永久基本农田和高标准农田(图 2)。

1.2 农田多要素遥感监测体系

遥感技术具备大范围、快速、多维度观测优势,可显著减少人工调查。但因分辨率和感知能力所限,对小型设施和被遮挡目标监测仍有挑战。本研究聚焦光学遥感影像可识别的 4 要素:农田地块、田间道路、防护林、灌溉与排水,构建多要素监测与分类体系(表 1)。

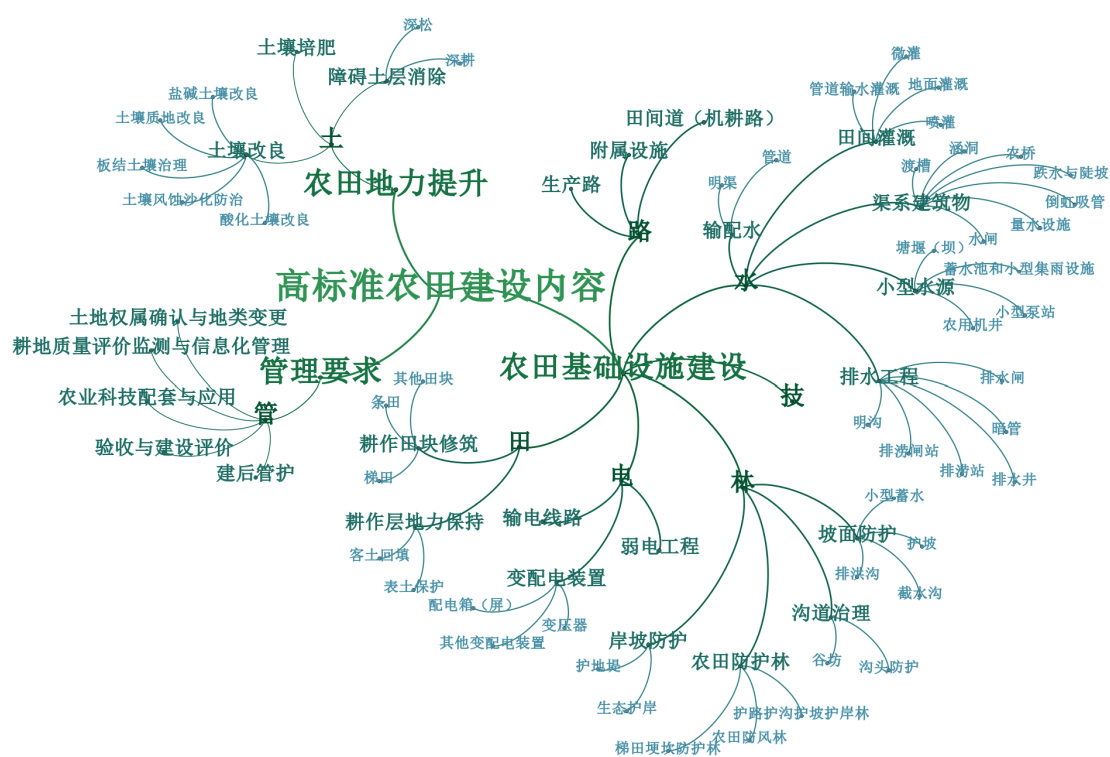


图 1 高标准农田建设 8 个要素内容示意

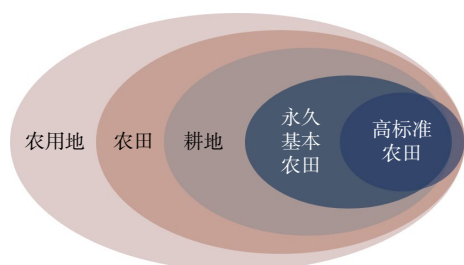


图 2 农田相关概念及其关系

2 视觉大模型算法

本研究以 SAM 为算法核心, 构建了兼具通用性与专用化能力的高标准农田多要素遥感监测流程框架(图 3), 以满足实际应用中的批量、自动化处理。

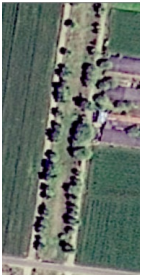
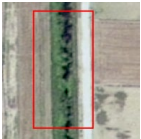
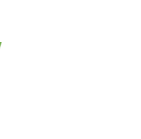
2.1 通用 SAM 算法

SAM 视觉基础模型采用基于提示工程的零样本

表 1 高标准农田多要素遥感监测与分类体系

一级类	二级类	描述	空间尺寸建设标准	遥感特征及时空要求	典型示例
农田地块	条田	地形相对较缓地区, 依据灌排水方向设计的几何形状为长方形或近似长方形的水平田块。水田区域条田可细分为格田	平原区条田长度为200~1000 m (南方宜为 100~600 m); 宽度为 50~300 m; 水田区格田长度为 30~120 m; 宽度为20~40 m	米级分辨率下, 宽度不小于3 m, 面积不小于3 m×3 m	
	梯田	地面坡度相对较陡地区, 依据地形和等高线进行的阶梯状田块	梯田田面长边宜平行等高线布置, 长度为100~200 m。为便于机械作业和田间管理, 水平梯田宽度大于10 m, 坡式梯田宽度大于4 m	米级或亚米级分辨率下, 宽度不小于3 m	
田间道路	田间道(机耕路)	连接田块与村庄、田块之间, 供农田耕作、农用物资和农产品运输通行的道路	宽度为3~6 m	米级分辨率下, 宽度不小于3 m	
	生产道	项目区内连接田块与田间道(机耕路)、田块之间, 供小型农机行走和人员通行的道路	宽度不大于3 m	米级或亚米级分辨率下, 宽度3 m左右	

表 1 高标准农田多要素遥感监测与分类体系 (续)

一级类	二级类	描述	空间尺寸建设标准	遥感特征及时空要求	典型示例
农田防护林	农田防风林	田块周围营造的以防治风沙或台风灾害、改善农作物生长条件为主要目的的人工林	无具体面积建设标准, 农田防护面积比例不小于80%	米级或亚米级分辨率下, 宽度不小于3 m	
	护路护沟 护坡护岸林	田间道路、排水沟、渠道两侧营造的以防止水土流失、保护岸坡安全、提供休憩庇荫场所为主要目的的人工林			
	梯田埂坎 防护林	梯田埂坎处营造的以防止水土流失、保护梯田埂坎安全为主要目的的人工林			
灌溉与排水	蓄水设施	小型塘堰(坝)、蓄水池和小 型集雨设施,用于拦截和集蓄 当地地表径流的挡水建筑物、 泄水建筑物及取水建筑物,可 缓解旱灾时水资源短缺问题, 防范涝灾以及缓解排水压力	无	米级或亚米级分 辨率下, 面积不小于5 m× 5 m	
	沟渠	用于地上输水或排水	无	米级或亚米级分 辨率下, 宽度不小于3 m	
	农桥	用于田间道路跨越洼地、渠 道、排水沟等障碍物	宽度为4~8 m	米级或亚米级分 辨率下, 宽度为3~8 m	
	控水设施	用于田间灌溉、防洪和排涝 的灵活控制, 包含小型泵站、农 用机井、水闸、排涝站、排涝闸 站等	无, 需根据总体布置计算大小	亚米级分辨率	

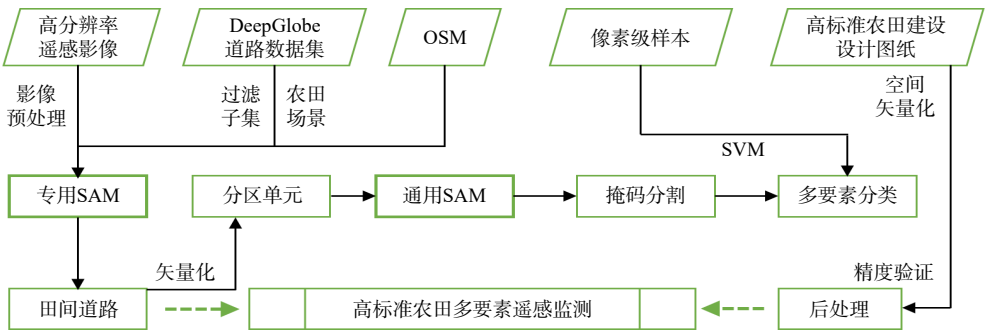


图 3 兼具通用性与专用化 SAM 能力的高标准农田多要素遥感监测流程框架

(zero-shot)学习策略,通过用户可输入点、框或掩码等人机交互的方式实现在完全未知图像或新任务上的目标分割,并适配各类下游任务。实验表明,SAM 在多个任务中的零样本分割性能优于部分全监督方法^[24,28],并广泛应用于医学、视频、机器人及遥感等领域^[29-30]。

SAM 模型由图像编码器(image encoder)、提示编码器(prompt encoder)和掩码解码器(mask decoder)组成。图像编码器采用基于视觉转换器(vision trans-

former, ViT)架构的掩码自编码器(masked autoencoders, MAE)进行预训练,支持高分辨率图像处理;提示编码器处理稀疏提示(点、框)和密集提示(掩码),实现多模态提示信息的高效整合;掩码解码器结合双向自注意力和交叉注意力机制,通过多层感知机(multi-layer perceptron, MLP)输出前景概率。通用SAM算法是指模型在10亿级分割数据集SA-1B(segment anything 1 billion)上训练,具备跨领域零样

本分割能力,可直接应用于从自然场景到遥感、医学影像、工业视觉检测等多种实际场景中。如图4所示,SAM通过构建规则格网采样点作为提示输入,生成整幅影像的分割掩码,但这种基于点提示存在随机性,提示点并不总是目标的正样本,因而在特定任务中需要输入额外提示,或对模型进行专用化改进与后处理,具体内容详见第2.2节。

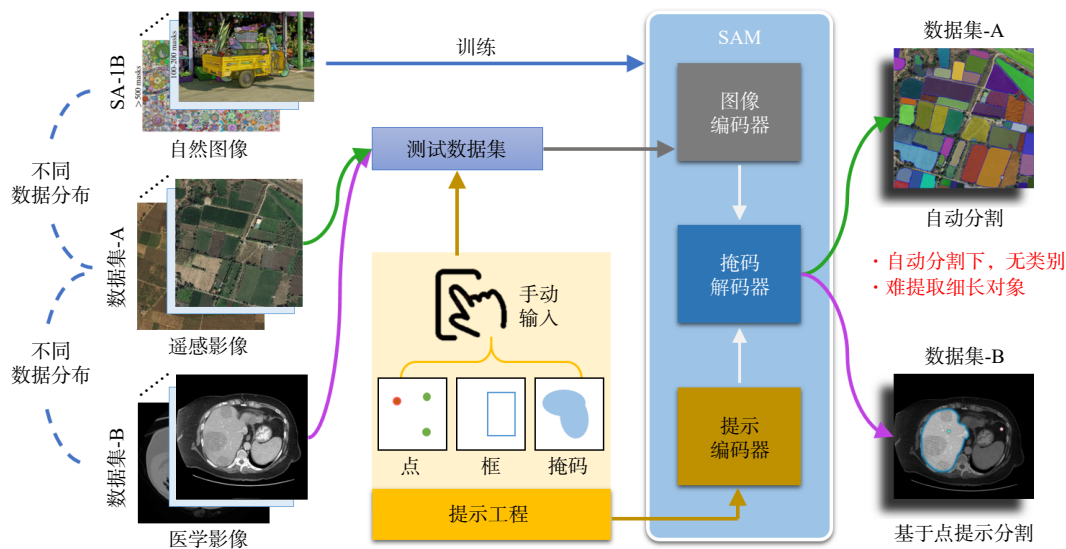


图4 基于SAM的通用算法

2.2 基于SAM的专用算法

在医疗、遥感等复杂场景下,SAM需依赖额外专家先验知识的手动提示输入才能获得理想分割结果^[31-33]。遥感图像本身质量(如分辨率、传感器类型、时相等因素)及人为输入提示工程质量(如类型、位置和数量)差异,导致分割结果的不稳定性。因此,额外的人工输入精准的提示工程,在实际应用中将耗费巨大的人力和时间成本。

其次,针对难以识别的特定对象,如农田道路遥感提取任务中,模型很难准确地描绘出农田场景中细长且分布稀疏的道路。如图5所示,将通用SAM算法应用于0.1 m及1 m分辨率的遥感影像时,能获取独立、面状且边界清晰的地物掩码(如田块、建筑物、树木等),但无法捕获线状形态的地物,如细长的道路、长河流和稀疏分布的林带。即使结构化的混凝土道路在遥感影像上表现出鲜明特征,也无法识别。对此,本研究对SAM算法本身或使用策略进行专用化

改进,以弥补模型在特定对象识别或某些方面能力表现的不足(图6)。

1) 专用化算法改进。

针对田间道路提取的特定任务,采用少量开源道路数据集微调SAM,同时从开放街图(OpenStreet-Map, OSM)道路中心线矢量数据中深度挖掘先验知识,无需额外人工手动输入提示工程,自动化生成适宜的提示输入,以增强SAM模型识别细长道路对象的能力(图7)。该框架主体由微调SAM模型、基于OSM的提示生成器及集成后处理3个部分组成。

在微调SAM模型过程中,选择参数量较小(93.7 M)的ViT-B初始化SAM,以平衡分割性能和计算效率。由于道路在农田与城市场景下的显著差异^[34],本研究从DeepGlobe遥感影像挑战数据集(DeepGlobe Satellite Image Understanding Challenge Dataset, DeepGlobe)^[35]中过滤出138张农田场景子集微调模型,使



图 5 基于 SAM 的不同分辨率遥感图像(0.1 m 和 1 m 分辨率)自动分割结果示意

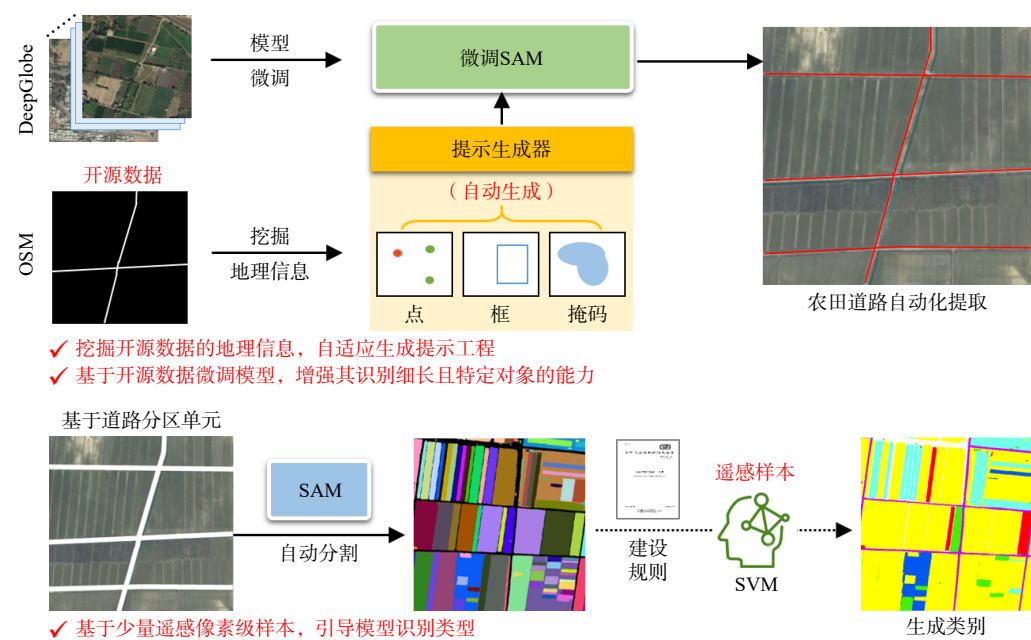


图 6 SAM 专用算法和使用策略

其适应田间道路特征。微调策略选择冻结提示编码器整个模块,仅更新图像编码器和掩码解码器。网络优化选择了 AdamW 优化器,初始学习率为 1×10^{-4} ,权重衰减率为 1×10^{-2} ,批次大小设置为 2,未使用数据增强处理,共迭代训练 200 轮。对于这个二元分类任务,采用结合交叉熵损失 L_{BCE} 、Dice 损失 L_{Dice} 的非加权和损失函数 L ,可表示为:

$$L_{\text{BCE}} = -\frac{1}{N} \sum_{i=1}^N [g_i \log \hat{y}_i + (1 - g_i) \log (1 - \hat{y}_i)] \quad (1)$$

$$L_{\text{Dice}} = 1 - \frac{2 \sum_{i=1}^N g_i \hat{y}_i}{\sum_{i=1}^N (g_i)^2 + \sum_{i=1}^N (\hat{y}_i)^2} \quad (2)$$

$$L = L_{\text{BCE}} + L_{\text{Dice}} \quad (3)$$

式中, g_i 和 $\hat{y}_i$ 分别为像素 i 的真实标签和预测标签, N 为

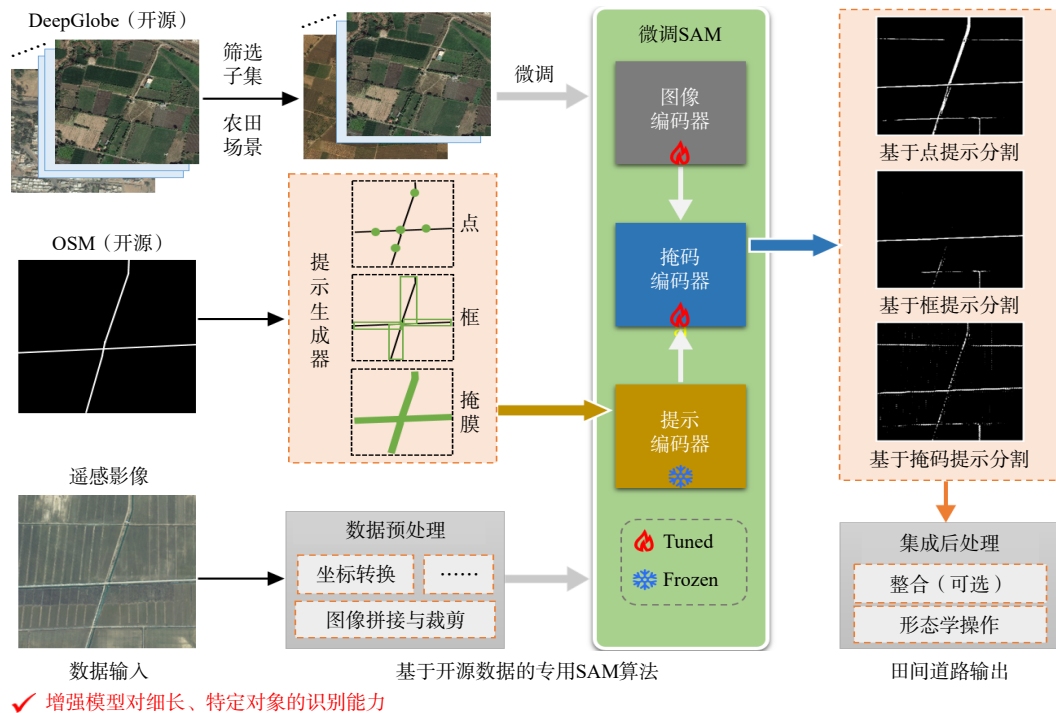


图7 基于开源数据驱动田间道路提取的专用SAM模型结构

输入图像的像素总量。

在提示生成器中，针对点提示，由于OSM数据质量不一致，本研究仅在道路中心线上采样正向点，并排除背景负点。采样前将道路矢量数据栅格化为中心线二值图，再采用 100×100 像素网格采样，保证点沿道路均匀分布并减少数量冗余。针对框提示，将道路矢量数据转换为GeoJSON文件，利用道路起点 (x_s, y_s) 、终点 (x_e, y_e) 的地理坐标生成框提示的左上角和右下角点图像坐标。为避免贯穿整幅图像的道路带来冗余信息，固定短边长（如 $\Delta_x < \Delta_y$ 时，在 x 轴方向固定边长 $l = 200 \text{ px}$ ），根据长宽比 (Δ_x / Δ_y) 将该道路切分为多个小框，以生成若干更精准的边界框。掩码提示则通过形态学膨胀将OSM栅格化的道路中心线扩展为 7 px 宽度的二值栅格图。

在集成后处理中，采用结果级融合策略整合三类提示的分割结果。通过置信度加权融合结果，降低点提示分割结果的噪声、提升框提示对离散点的识别能力。随后使用Candy算子提取骨干道路，并结合形态学膨胀填补缺失。具体操作中，采用 2×10 和 10×2 非对称卷积核迭代了2次，以兼顾水平与垂直方向特征提取，同时减少批量计算的复杂性。

2) 专用化使用策略。

道路是典型的人为线性设施，成为一种天然边界，为农田分区提供可靠的硬约束条件，具有以下优势：(1) 有效区分农田与周边非农用地功能区域；(2) 形态和光谱特征稳定性强，不易受物候变化干扰；(3) 提供多尺度硬边界，更贴合真实农田管理单元区块，避免固定尺寸裁剪造成地物切割与类别混叠，提高农田多要素分类精度和空间一致性。

基于此，本研究提出通用分割模型的专用化策略：首先利用道路网络划分出贴合实际管理单元的农田区域，再借助SAM的零样本分割能力对高分辨率遥感影像生成多要素候选掩码，无需人工提示输入，显著提升自动化与处理效率。为提高模型的分割性能，选择使用参数量约312.3 M的ViT-L作为预训练模型。随后结合各农田要素在光谱、纹理和几何上的差异（表1），选取少量像元构建像素级样本集作为后验知识，通过支持向量机(support vector machine, SVM)对候选掩码分类，实现农田多要素的精确识别。

3 实验结果与分析

3.1 实验研究区

本研究选取山东省寿光市化龙镇2021年建设的

高标准农田项目区作为实验区,总面积约 1533 公顷(图 8)。化龙镇位于寿光市西部,辖 53 个行政村,总耕地约 5700 公顷,是典型的黄淮海平原农业区,农田集中连片、设施完善。建设内容涵盖土壤改良、灌溉与排水、田间道路、农田林网、输配电工程等多方面。

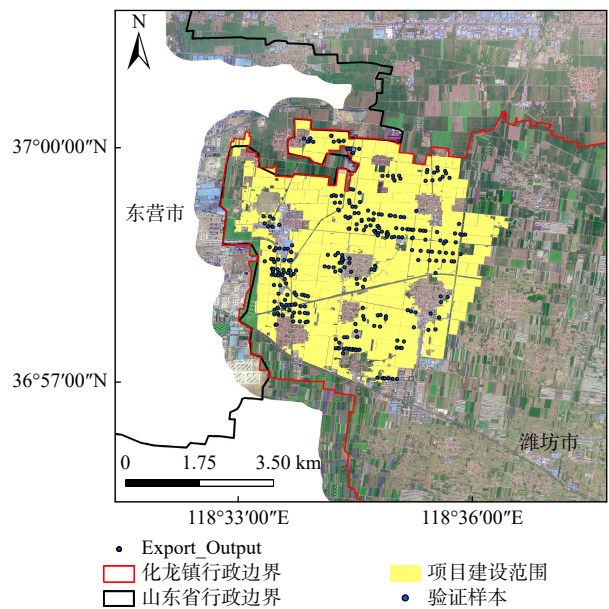


图 8 研究区域概况

3.2 数据集准备

本研究融合了多源遥感与地理信息数据。

1) 遥感影像: ① SuperView 高景卫星影像(2021 年 11 月、2022 年 2 月、2022 年 6 月),空间分辨率为 0.5 m,包含 R、G、B 三波段,用于山东省寿光市的农田多要素细粒度目标分割。② 其他区域影像,包含 1 景覆盖北京大兴区,分辨率为 0.8 m,具备 R、G、B 三波段;1 景覆盖江苏南通市,分辨率为 2 m,由天地图平台下载的 R、G、B 三波段。所有遥感影像经过正射校正、图像配准、波段融合、裁剪与拼接等预处理流程,保证多源影像空间一致性与质量统一性。

2) 辅助矢量: ① 项目边界矢量来自 2021 年政府规划的高标准农田建设范围。② 真值数据基于项目设计图纸,涵盖田间道路、农田防护林、灌排、输配电等建设工程设计,经 Google Earth 核查补充,形成 2 万余像素验证样本(图 8)。③ 开源数据包含 DeepGlobe 道路数据集(0.5 m 分辨率,1024×1024 px 图像大小,筛选 138 张农田场景影像用于 SAM 微调)和 OSM 道路矢量数据(含 27 类道路,经 WGS84 坐标转换与影像配准,可用于生成提示信息辅助 SAM 分割(图 9)。

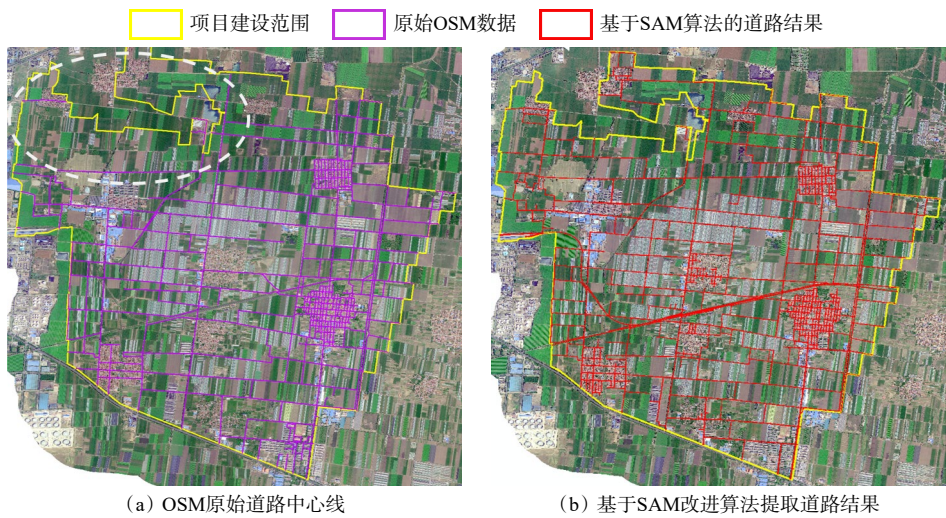


图 9 OSM 原始道路中心线与基于 SAM 改进算法提取道路结果对比

3.3 提取结果

本研究融合 SAM 与监督分类方法,提取了农田地块、道路、防护林、灌排及大棚等关键地物,主要结果如下。

1) 专用 SAM 提取田间道路。利用农田场景道路数据集微调 SAM,并结合 OSM 生成的点、框与掩

码提示,成功提取了实验区主要道路(图 9)。通过中值滤波、形态学膨胀运算与骨架提取等操作提升道路连贯性与边界清晰度。与原始 OSM 相比,结果更完整精准,见图 9 白色虚线圈。

2) 通用 SAM 自动分割。将田间道路结果矢量化,用于重新划分研究区,避免道路两侧地物被切割,

保证分区单元与实际农田管理单元一致,从而减少特征混叠并提升分类精度(图 10)。SAM 在无需额外训练下实现高质量自动分割,为后续分类提供候选掩码。

3) 农田多要素分类。基于 SAM 分割结果,利用少量人工标注的像素样本,训练 SVM 识别农田地块、农田防护林、农田道路等要素。最终整合多类掩码,形成覆盖全区的高标准农田分布图(图 11),为区域农业资源管理与数字化建设提供数据支持。

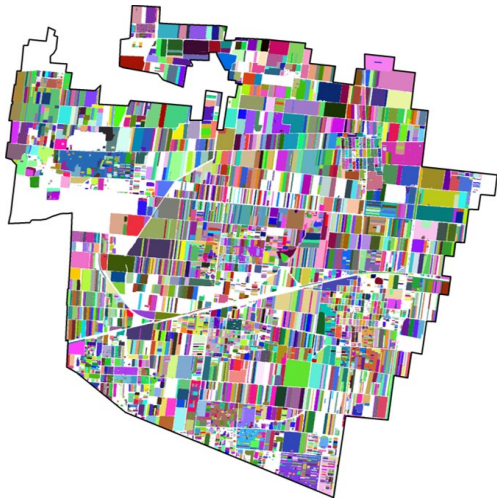


图 10 基于 SAM 的自动分割结果图(目标随机配色)

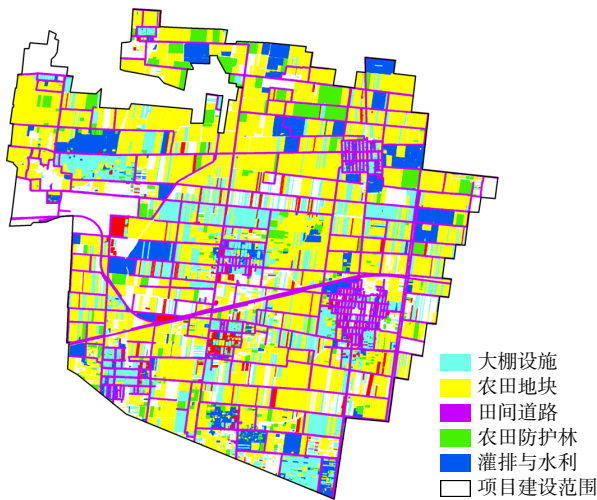


图 11 高标准农田多要素遥感提取结果

3.4 精度验证与分析

本文采用精度(Precision)、召回率(Recall)、F1 分数(F1-score)和 IoU 指标对各类要素的提取结果进行定量评价。

1) 消融实验。基于 DeepGlobe 数据集,开展了包含微调策略、提示工程及模型专用化方案的消融实验,用于分析专用化 SAM 模型各个组件在田间道路

提取任务中的有效性。如表 2 所示,仅使用原始 SAM 算法时各项指标几乎为 0,表明其在窄幅线状道路提取中完全失效,需结合任务专用化与合理提示设计。在提示工程中,点提示与框提示均显著提升性能,而掩码提示表现相对较差。加入集成后处理后,专用化 SAM 模型取得最优表现,F1-score 为 0.71, IoU 为 0.84,证明专用化优化能有效提升线状目标的检测精度。

表 2 专用化 SAM 算法对不同模块进行消融实验评价

方法	Precision	Recall	F1-score	IoU
专用化SAM	0.87	0.77	0.71	0.84
SAM (点)	0.80	0.62	0.68	0.72
SAM (框)	0.85	0.48	0.62	0.64
SAM (掩码)	0.64	0.70	0.67	0.58
原始SAM(有提示工程)	0.05	0.08	0.09	0.04
原始SAM(无提示工程)	0	0.05	0.01	0

2) 对比方法。本节旨在评估所提出方法与其他道路提取方法在遥感影像田间道路提取任务中的有效性和先进性。对比结果如表 3 所示,专用化 SAM 模型在 4 项指标上均显著优于其他方法,F1-score 和 IoU 分别达到 0.71 和 0.84;传统语义分割模型 D-LinkNet 分别仅为 0.49 和 0.35,其他方法几乎失效。结果表明,专用化 SAM 在有限监督条件下亦能实现农田道路的精准提取,适用于样本稀缺场景。

表 3 从 DeepGlobe 道路数据集中过滤出少量农田场景子集的精度评价

方法	Precision	Recall	F1-score	IoU
专用化SAM	0.87	0.77	0.71	0.84
SegFormer ^[36]	0.52	0.16	0.12	0.06
SIINet ^[37]	0.06	0.01	0.01	0
D-LinkNet ^[38]	0.54	0.56	0.49	0.35
DeepLabV3+ ^[39]	0.05	0.24	0.06	0.03

3) 泛化性验证。在实际应用中,高分辨率遥感影像的获取往往成本较高,且在部分地区或特定时间段可能存在数据缺失的问题。为模拟多样化应用场景,设计了基于不同时相、分辨率及区域的泛化性实验,并以高标准农田建设规划设计图与 Google Earth 历史高分辨率影像为真值。经过修正后提取农田地块、田间道路、农田防护林及灌溉排水设施等要素验证样

本。实验采用固定尺寸切片(1024×1024 px)输入模型,结果如图 12 所示。该方法在 3 组不同时相、3 种不同空间分辨率及 3 个不同区域的测试中均表现优于基于 OSM 的伪标签,表现出良好泛化能力。高分辨率影像下的边界刻画清晰,低分辨率则易受噪声干扰,边界信息逐渐模糊。时相变化会导致作物长势及

地表覆盖类型发生变化,甚至可能出现大棚设施新建等地物变化,但本方法仍保持了较高的稳定性和一致性。然而,对于被树木遮挡的区域,仍存在漏分现象;对于大棚或未耕作地块与田间道路等光谱特征高度相似的地物,易出现错分。

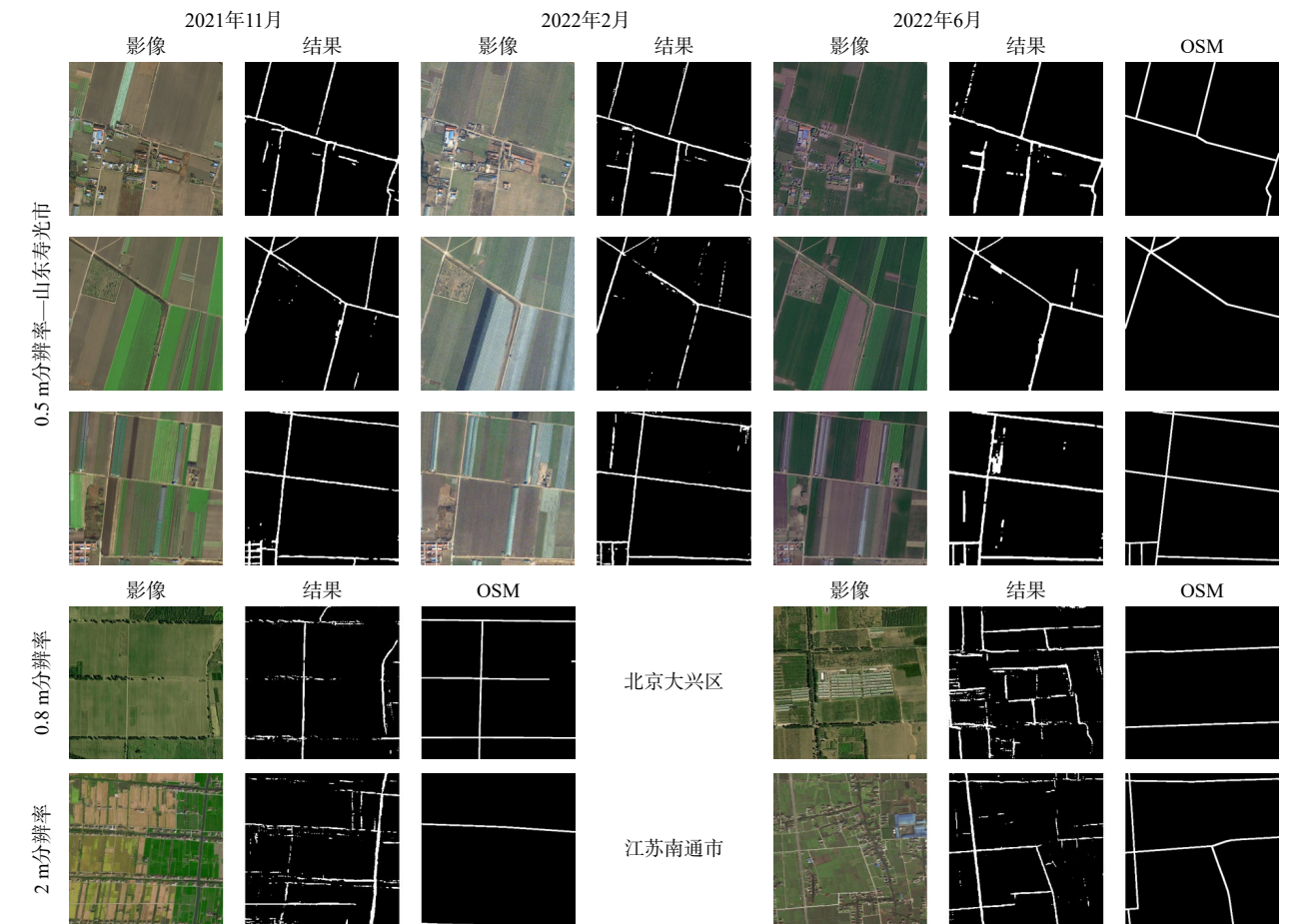


图 12 专用化 SAM 算法在不同时间、不同空间分辨率、不同地区影像的田间道路提取结果

如表 4 所示,本文选取了同一研究区域 3 个不同时相的遥感影像开展实验,进一步验证整体方法流程的有效性及其推广潜力。结果显示,农田地块和田间道路的提取精度较高(F1-score 均超过 0.83),表明视觉大模型在主体农田设施识别任务中表现出较强的鲁棒性。但灌溉排水设施的提取精度波动相对较大(F1-score 为 0.52~0.74),可能是季节气候等因素影响水位变化,进而导致识别结果存在差异。农田防护林的提取性能相对较低(F1-score 介于 0.39~0.52),这主要是由于其在影像中呈现出破碎化、形态不规则且边界模糊等特点,而 SAM 算法更适用于边界清晰的目标,如农田地块和道路。因此,该方法在农田防护林

表 4 高标准农田多要素遥感提取结果精度评价

时相	类型	Precision	Recall	F1-score	IoU
2021年11月	农田地块	0.90	0.89	0.92	0.89
	田间道路	0.92	0.91	0.90	0.86
	农田防护林	0.57	0.46	0.49	0.26
	灌溉与排水	0.80	0.78	0.70	0.65
2022年2月	农田地块	0.87	0.85	0.86	0.83
	田间道路	0.85	0.79	0.83	0.82
	农田防护林	0.36	0.42	0.39	0.30
	灌溉与排水	0.73	0.63	0.52	0.48
2022年6月	农田地块	0.93	0.95	0.93	0.89
	田间道路	0.87	0.90	0.88	0.82
	农田防护林	0.61	0.45	0.52	0.38
	灌溉与排水	0.79	0.75	0.74	0.66

提取方面仍存在不足。针对该问题,未来可结合树木密度估计方法对区域内树木分布进行拟合与筛选,从而有效区分并提取农田防护林,无须依赖树冠形态和考虑季节性变化影响。

4 结果与讨论

本文提出的融合大规模视觉模型与遥感辅助的高标准农田多要素识别流程框架,在典型高标准农田建设区取得了较好的应用成效,并在不同时相、不同分辨率、不同地区的遥感影像上进行了泛化验证。实验结果表明,该方法在农田地块、田间道路等边界清晰、形态规则的目标提取中表现出较高的精度和稳定性。借助提示机制与 SAM 的自动分割能力,无需大量高质量样本即可实现高效分割,兼顾了分割精度与处理效率。然而, SAM 在遥感图像上的表现机制也暴露出一定局限性。具体而言,对于边界模糊、纹理连续变化的对象,如农田防护林,模型识别精度明显下降。这主要是防护林由多个树冠组成,其在遥感影像中呈现冠层叠加、边界不连续、形态破碎的特征,导致 SAM 难以准确分割整条防护林区域,仅能捕捉到部分树冠或分散区域,从而造成漏分与错分。

未来针对防护林的识别问题,可在现有方法基础上结合树木密度估计技术,先估算区域内树木分布密度,进而筛选出分布连片、排列规则的区域作为农田防护林。此方法不依赖树冠具体形态,也无需考虑树冠边界随季节变化的干扰,能够有效提升防护林的整体提取能力和应用稳定性。诸如此类,对于边界模糊、形态复杂或不连续的目标,后续研究需聚焦于目标特征表达与方法融合改进,以提升模型在复杂农田设施识别中的适用性和鲁棒性。

5 结论

1) 改进 SAM 算法能够实现田间道路特定对象提取的专用化任务,基于开源道路数据集微调模型,利用 OSM 矢量数据自动生成提示信息,经过集成与后处理优化能够有效实现了田间道路网络的连续性与结构完整性识别,显著提升了线性要素的自动提取能力。

2) 基于高标准农田多要素遥感监测与分类体系,提出一种融合 SAM 视觉大模型自动化、精准快速的农田多要素遥感提取流程框架,以山东省寿光市典型高标准农田建设项目区为实验区,实现农田地块、田间道路、农田防护林、灌溉与排水等典型要素的识别。

3) 本研究对实际应用具有较强的推广潜力,在不同时相、不同空间分辨率以及不同地区的遥感影像上表现出一定的泛化能力,可有效减少人工标注工作量,提升遥感影像的信息化处理效率,为高标准农田建设效果的动态评估与农业数字化管理提供了可靠的数据支撑。

4) 未来研究中,可进一步提升模型在复杂农田设施识别中的适用性和鲁棒性,改进对边界模糊、形态复杂或不连续的目标的识别精度,并推动其在农业遥感智能识别中的广泛落地应用。

致谢:感谢“坤元”地理科学大模型团队对本论文资料整理工作。

参考文献(References)

- [1] 周琳, 钟钰. 中国粮食安全治理的特征及路径[J]. 科技导报, 2024, 42(16): 6–13.
- [2] 郭珍. 中国耕地保护制度的演进及其实施绩效评价[J]. 南通大学学报(社会科学版), 2018, 34(2): 67–73.
- [3] 魏秀菊, 朱明, 廖艳. 推动农业工程发展 保障我国粮食安全[J]. 科技导报, 2023, 41(20): 5–19.
- [4] 宋文. 基于耕地综合质量的高标准农田建设与监测研究[D]. 北京: 中国地质大学(北京), 2018.
- [5] 王怡婷, 文姝蓉, 果文帅, 等. 中国粮食增产的贡献分解与路径分析[J]. 科技导报, 2024, 42(16): 26–39.
- [6] 国家发展和改革委员会. 全国高标准农田建设规划(2021—2030年)[EB/OL]. (2021-11-02) [2024-09-04]. https://www.ndrc.gov.cn/fggz/fzzlgh/gjjzxgh/202111/t20211102_1302810.html.
- [7] 陈正, 刘瀛弢, 贺德俊, 等. 中国高标准农田建设现状与发展趋势[J]. 农业工程学报, 2023, 39(18): 234–241.
- [8] 孔德明. 断头渠、花花路、阴阳田……高标准农田建设遭遇“打折扣”[EB/OL]. (2020-08-19) [2024-09-04]. http://www.banyuetan.org/jrt/detail/20200819/1000200033134991597755569931596142_1.html.
- [9] 孙春蕾, 杨红, 韩栋, 等. 全国高标准农田建设情况与发展策略[J]. 中国农业科技导报, 2022, 24(7): 15–22.

- [10] 国务院办公厅. 关于切实加强高标准农田建设提升国家粮食安全保障能力的意见[EB/OL]. (2019-11-21)[2024-09-04]. https://www.gov.cn/xinwen/2019-11/21/content_5454258.htm.
- [11] 国家市场监督管理总局, 国家标准化管理委员会. 高标准农田建设 通则: GB/T 30600—2022[S]. 北京: 中国标准出版社, 2022.
- [12] 周成虎. 地理学赋能网络空间认知[J]. 科技导报, 2023, 41(13): 1.
- [13] Xu J F, Zhu Y, Zhong R H, et al. DeepCropMapping: A multi-temporal deep learning approach with improved spatial generalizability for dynamic corn and soybean mapping[J]. *Remote Sensing of Environment*, 2020, 247: 111946.
- [14] Waldner F, Diakogiannis F I. Deep learning on edge: Extracting field boundaries from satellite images with a convolutional neural network[J]. *Remote Sensing of Environment*, 2020, 245: 111741.
- [15] Yang J W, Zhang J S, Zhu S, et al. Cropland mapping in fragmented agricultural landscape using modified pyramid scene parsing network[C]//Proceedings of IGARSS 2019 – 2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2019: 9831–9834.
- [16] Zhu Q Q, Zhang Y N, Wang L Z, et al. A Global Context-aware and Batch-independent Network for road extraction from VHR satellite imagery[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2021, 175: 353–365.
- [17] 王斌, 陈占龙, 吴亮, 等. 兼顾连通性的U-Net网络高分辨率遥感影像道路提取[J]. 遥感学报, 2020, 24(12): 1488–1499.
- [18] 赫晓慧, 陈明扬, 李盼乐, 等. 结合DCNN与短距条件随机场的遥感影像道路提取[J]. 武汉大学学报(信息科学版), 2024, 49(3): 333–342.
- [19] Brandt M, Gominski D, Reiner F, et al. Severe decline in large farmland trees in India over the past decade[J]. *Nature Sustainability*, 2024, 7(7): 860–868.
- [20] Ma Y T, Zhou Z, She X X, et al. Identifying dike-pond system using an improved cascade R-CNN model and high-resolution satellite images[J]. *Remote Sensing*, 2022, 14(3): 717.
- [21] Fang W Z, Sun Y Y, Ji R, et al. Recognizing global dams from high-resolution remotely sensed images using convolutional neural networks[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021, 14: 6363–6371.
- [22] Liu F, Chen D L, Guan Z, et al. RemoteCLIP: A vision language foundation model for remote sensing[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 1–16.
- [23] Hu Y, Yuan J L, Wen C C, et al. RSGPT: A remote sensing vision language model and benchmark[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2025, 224: 272–286.
- [24] Kirillov A, Mintun E, Ravi N, et al. Segment anything[C]//Proceedings of IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2023: 4015–4026.
- [25] 国务院. 中华人民共和国基本农田保护条例[M]. 北京: 法律出版社, 2011.
- [26] 张燕. 我国基本农田保护制度研究[D]. 咸阳: 西北农林科技大学, 2013.
- [27] 国家质量监督检验检疫总局, 中国国家标准化管理委员会. 土地利用现状分类: GB/T 21010—2017[S]. 北京: 中国标准出版社, 2017.
- [28] Ahmadi M, Lonbar A G, Naeini H K, et al. Application of segment anything model for civil infrastructure defect assessment[J]. 2023, arXiv preprint arXiv: 2304.12600.
- [29] Sultan R I, Li C, Zhu H, et al. GeoSAM: Fine-tuning SAM with sparse and dense visual prompting for automated segmentation of mobility infrastructure[J]. 2023, arXiv preprint arXiv:2311.11319.
- [30] Wang D, Zhang J, Du B, et al. Samrs: Scaling-up remote sensing segmentation dataset with segment anything model[J]. *Advances in Neural Information Processing Systems*, 2023, 36: 8815–8827.
- [31] Ji W, Li J J, Bi Q, et al. Segment anything is not always perfect: An investigation of SAM on different real-world applications[J]. *Machine Intelligence Research*, 2024, 21(4): 617–630.
- [32] Chen T R, Zhu L Y, Ding C T, et al. SAM-adapter: Adapting segment anything in underperformed scenes[C]//Proceedings of IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). IEEE, 2023: 3367–3375.
- [33] Ma J, He Y T, Li F F, et al. Segment anything in medical images[J]. *Nature Communications*, 2024, 15(1): 654.
- [34] Shen J, He Y W, Peng J, et al. An open-source data-driven automatic road extraction framework for diverse farmland application scenarios[J]. *Computers and Electronics in Agriculture*, 2025, 235: 110330.
- [35] Demir I, Koperski K, Lindenbaum D, et al. DeepGlobe 2018: A challenge to parse the Earth through satellite images[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2018: 172–181.
- [36] Xie E Z, Wang W H, Yu Z D, et al. SegFormer: Simple and

- efficient design for semantic segmentation with transformers[J]. *Advances in Neural Information Processing Systems*, 2021, 34: 12077–12090.
- [37] Tao C, Qi J, Li Y S, et al. Spatial information inference net: Road extraction using road-specific contextual information [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2019, 158: 155–166.
- [38] Zhou L C, Zhang C, Wu M. D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction[C]//*Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2018: 182–186.
- [39] Chen L C, Zhu Y K, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//*Computer Vision-ECCV 2018*. Cham: Springer International Publishing, 2018: 833–851.

Remote sensing applications for well-facilitated farmland digitalization driven by vision foundation models

SHEN Jing^{1,2}, LIU Ze^{1,3*}, HE Yawen⁴

1. State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China
2. University of Chinese Academy of Sciences, Beijing 100049, China
3. School of Artificial Intelligence, China University of Geosciences (Beijing), Beijing 100083, China
4. College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao 266580, China

Abstract The construction of well-facilitated farmland in China has imposed higher requirements on the digital and refined management of farmland, posing new challenges for the automatic extraction of multi-element information from remote sensing data. This study proposes a framework for the automatic extraction of multiple key farmland elements by integrating prompt engineering with remote sensing feature knowledge. Leveraging the general segmentation capabilities of the vision foundation model, combined with open-source data and feature-driven specialized algorithms, the proposed approach enables efficient automatic identification of critical farmland elements, including plots, field roads, shelterbelts, and irrigation and drainage facilities. Taking the well-facilitated farmland construction area in Shouguang City, Shandong Province as a case study, we conducted technical validation using high-resolution domestic satellite remote sensing imagery. Experimental results demonstrate that the proposed method can achieve batch automatic processing within farmland project areas, significantly reducing dependence on large amounts of high-quality training samples and manual workload. Moreover, it shows good generalization capabilities across images with different acquisition times, spatial resolutions, and regions, thus greatly improving data production efficiency and practical application potential. This research provides a novel approach for the intelligent interpretation of digitalized farmland areas and offers strong technical support for the application of remote sensing in post-construction supervision of agricultural engineering projects.

Keywords well-facilitated farmland; multi-element farmland features; remote sensing application; vision foundation model; prompt engineering ●



(责任编辑 赵庆圆)