引文格式：刘爱霞,张奇伟,郭进,等. 基于视觉提示词的耕地实例化分割方法[J]. 地理空间信息, 2024, 22(10): 19-21. doi:10. 3969/j.issn. 1672-4623.2024.10.005

基于视觉提示词的耕地实例化分割方法

刘爱霞[1]，张奇伟[2*]，郭　进[1]，罗　亮[1]
（1. 北京航天世景信息技术有限公司，北京 100089；2. 山东省国土测绘院，山东 济南 250100）

摘　要：耕地是国家粮食安全的根本保证，基于遥感影像的耕地提取为耕地保护，遏制耕地“非农化”，防止“非粮化”提供基本技术支撑。为实现基于高分遥感影像的耕地精细化提取，基于目标检测技术和神经网络大模型（segment anything model，SAM），实现简单标注样本下的耕地实例化分割，提出了一种基于视觉提示词的耕地实例化分割方法（instance segmentation method of farmland based on visual prompts，ISFVP）。在提出的方法中，首先利用耕地目标检测网络检测出耕地的大概范围（以矩形框表示），然后将耕地所处的范围作为提示词输入到网络大模型SAM中，实现耕地范围的精细化提取。实验证明，该方法可从高分遥感影像中自动化检测出耕地，为耕地的精细化提取提供了一种可行的技术方案。

关键词：耕地；遥感影像；目标检测；神经网络大模型
中图分类号：P237　　**文献标志码：**B　　**文章编号：**1672-4623（2024）10-0019-03

Instance Segmentation Method of Cultivated Land Based on Visual Prompts

LIU Aixia[1], ZHANG Qiwei[2], GUO Jin[1], LUO Liang[1]
(1. Beijing Space View Technology Co., Ltd., Beijing 100089, China;
2. Shandong Provincial Institute of Land Surveying and Mapping, Jinan 250100, China)

Abstract: Cultivated land is the fundamental guarantee of national food security. Cultivated land extraction based on remote sensing images provides basic technical support for the protection of cultivated land, to stop any attempt to use it for purposes other than agriculture and specifically grain production. In order to realize the refined extraction of cultivated land based on high-resolution remote sensing images, we used the target detection technology and the large neural network segment anything model (SAM) to realize the instance segmentation of cultivated land under simple labeled samples, and proposed an instance segmentation method of cultivated land based on visual prompts. In the proposed method, we used the cultivated land target detection network to detect the approximate range of cultivated land (represented by a rectangular box) at first. And then, we input the range of cultivated land into the large network model SAM as a prompt word to realize the refined extraction of cultivated land range. Experimental result proves that this method can automatically detect cultivated land from high-resolution remote sensing images, and provide a feasible technical solution for the refined extraction of cultivated land.

Key words: cultivated land, remote sensing image, object detection, neural network big model

耕地作为我国最宝贵资源的战略定位，决定了我们必须实行最严格的耕地保护制度。近年来，党中央、国务院对耕地保护工作给予了前所未有的重视，坚守18亿亩耕地红线在当前和今后相当长时期内是国土资源管理的重中之重。然而，随着形势不断变化，我国耕地保护面临更大的压力和挑战。遥感影像具有覆盖范围大、获取成本低、及时准确的特点，因此基于遥感影像的耕地提取技术在耕地保护监测中发挥着重要作用，获得广泛研究。

当前基于遥感影像的耕地提取方法可以分为面向对象的方法[1-3]和基于深度学习的方法[4-6]。基于面向对象的方法首先对遥感影像进行分割，得到空间上相连，光谱、纹理等特征一致的对象，然后从对象中提取特征进行对象识别分类[7-13]。

虽然基于深度语义分割的耕地提取方法和基于深度实例分割的耕地提取方法获得了长足的发展，并展现出强大的耕地提取能力。然而作为数据驱动的方法，这2种方法都需要标注大量的样本，耕地语义分割样本和实例分割样本需要作业员将耕地的边界精确地描绘出来，费时费力，成本较高。相比于深度语义分割技术和深度实例分割技术，深度目标检测样本仅需要标注矩形框（仅需点击两次鼠标），相比于耕地语义分割样本和实例分割样本的“描边式”标注，标注成本大大降低，样本更加简单。然而深度目标检测技术用于耕地提取仅能获得耕地矩形框，无法获得耕

收稿日期：2023-07-27。
第一作者简介：刘爱霞（1983—），本科，工程师，主要研究方向为摄影测量与遥感，E-mail : 273373771@qq.com。
通信作者：张奇伟（1988—），工程师，主要研究方向为摄影测量与遥感，E-mail : 229330480@qq.com。

地的边界，不能实现耕地的精细化提取。为使用耕地矩形框这种简单样本实现耕地精细化提取，本文基于目标检测技术和神经网络大模型SAM[14]，提出一种基于视觉提示词的耕地实例化分割方法（instance segmentation method of farmland based on visual prompts，ISFVP）。

1 基于视觉提示词的耕地实例化分割

本文提出的方法流程图如图1所示，主要包括3个部分，即影像预处理、耕地目标提取、耕地实例化分割。对于给定高分影像，首先对其进行归一化等预处理，然后输入深度目标检测网络进行耕地检测，得到耕地矩形框（矩形框内包含耕地，矩形框即表示耕地的大概范围），最后将矩形框作为提示词输入到网络大模型SAM中，实现耕地范围的精细化提取。

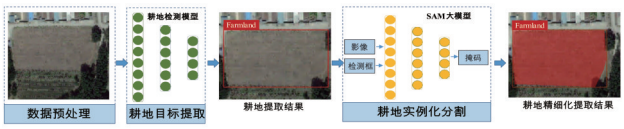


图1 基于视觉提示词的耕地实例化分割

1.1 数据预处理

在预处理阶段，本文对输入数据进行归一化处理，计算公式如下：

$$I=\frac{I-\text{mean}}{\delta} \tag{1}$$

式中，I 为遥感影像；mean 和 δ 分别为影像均值和方差。在训练耕地检测模型时，本文在预处理阶段还需对影像进行随机翻转、截取、拉伸、光谱扰动等操作。

1.2 耕地目标提取

经过预处理阶段之后，将归一化影像输入到耕地检测网络，进行耕地检测。对于耕地检测模型，考虑到检测速度和检测精度2个因素，本文选用PPYOLOE[15]模型。

为了检测出多种尺度的耕地，PPYOLOE模型提取出3个尺度的特征图，然后在3个尺度的特征图上分别提取 20×20 、40×40 、80×80 个检测框（即一共8 400个检测框），最后经过非极大值抑制（Non-Maximum Suppression，NMS）输出最终的耕地目标检测框。PPYOLOE主要包含3个模块，即基础特征提取模块BackBone，多尺度特征融合模块Neck和目标回归模块Head（包含类别逻辑回归和位置回归）。对于输入影像I，PPYOLOE分别将图像分割成 40×40 、80×80 个不重叠的图像块，获得每个图像块的中心 $S=\{(x_i,y_i)|i=1,2,3,\ldots,8\,400\}$。然后将影像I输入到PPYOLOE模型中，首先经过基础特征提取模块BackBone输出3个尺度的特征 $f_b=\{f_{b1},f_{b2},f_{b3}\}$，其中 $f_{b1}\in R^{256\times80\times80}$，$f_{b2}\in R^{512\times40\times40}$，$f_{b3}\in R^{1\,024\times20\times20}$：

$$f_b=\text{BackBone}(I) \tag{2}$$

式中，BackBone(·) 为基础特征提取函数。在 f_b 的基础上，PPYOLOE中的Neck模块采用PAN[16]结构融合多尺度特征信息，提高特征表达能力，并最终输出3个不同尺度的特征图 $f_n=\{f_{n1},f_{n2},f_{n3}\}$，其中 $f_{n1}\in R^{192\times80\times80}$，$f_{n2}\in R^{384\times40\times40}$，$f_{n3}\in R^{768\times20\times20}$，可表示为公式（3）：

$$f_n=\text{Neck}(f_b) \tag{3}$$

最后将多尺度融合特征 f_n 的3个特征图输入Head模块，输出3个尺度特征图上预测的耕地目标，包含耕地矩形框的类别概率 $P=\{f_{p1},f_{p2},f_{p3}\}$，其中 $f_{p1}\in R^{1\times80\times80}$，$f_{p2}\in R^{1\times40\times40}$，$f_{p3}\in R^{1\times20\times20}$，相对于图像块中心S的偏移值 $Dp=\{f_{dp1},f_{dp2},f_{dp3}\}$，其中 $f_{dp1}\in R^{4\times80\times80}$，$f_{dp2}\in R^{4\times40\times40}$，$f_{dp3}\in R^{4\times20\times20}$，可表示为：

$$P,Dp=\text{Head}(f_n) \tag{4}$$

在训练阶段，从预测的8 400个框中筛选出正样本参与耕地类别回归和检测框回归，剩余的为负样本仅参与类别回归。在测试推断阶段，首先基于8 400个图像块的中心S及其对应的偏移值 Dp 和类别概率 P，得到预测的耕地目标检测框，然后对8 400框进行NMS，删除高度重叠冗余的耕地检测框，得到最终的耕地检测框FarmObj:

$$P=\text{Transform}(P) \tag{5}$$

$$Dp=\text{Transform}(Dp) \tag{6}$$

$$T_{\text{Farm_top}}=S-Dp[:,0{:}2] \tag{7}$$

$$T_{\text{Farm_bottom}}=S+Dp[:,2{:}4] \tag{8}$$

$$T_{\text{Farm_obj}}=[T_{\text{Farm_top}},T_{\text{Farm_bottom}}] \tag{9}$$

$$\text{Farm}_{\text{obj}}=\text{NMS}(T_{\text{Farm_obj}},P) \tag{10}$$

式中，Transform(·) 为将耕地矩形框的类别概率 P 和相对于图像块中心S的偏移值 Dp 进行堆叠变换，变换后 $P\in R^{8\,400\times1}$，$Dp\in R^{8\,400\times4}$，Dp 矩阵的前两维（即 $Dp(:,0{:}2)$）为矩形框左上角相对于图像块中心S的位移，后两维（即 $Dp(:,2{:}4)$）表示矩形框右下角相对S的位移。$T_{\text{Farm_top}}$ 和 $T_{\text{Farm_top}}$ 分别为耕地目标矩形框的左上角坐标和右下角坐标，$T_{\text{Farm_obj}}\in R^{8\,400\times4}$ 为预测的8 400个耕地目标矩形框，Farm_{Obj} 为经过NMS滤后的耕地目标矩形框。

1.3　耕地实例化分割

在得到耕地的目标矩形框后，将影像和矩形框FarmObj输入SAM模型，获得矩形框内耕地像素掩码，实现耕地的精细化分割（也即实例化分割）。SAM模型是META公司于2023年发布的多模态大模型，对于输入影像，支持点、矩形框、文本等多种模态的提示词指定需要分割的内容，输出分割掩码。在提出的方法中，本文将检测出来的矩形框作为提示词，用于提取出矩形框内耕地掩码。对于输入的遥感影像，SAM模型首先基于Transformer ViT模型[17]获得图像的特征图：

$$F_{\text{img}} = \text{Im g}E(I) \tag{11}$$

式中，$\text{Im g}E(\cdot)$ 为SAM模型中的图像特征编码；F_{img} 为编码后的遥感图像特征。然后SAM模型对输入的目标检测框进行特征编码：

$$F_{\text{obj}} = \text{obj}E(\text{Farm}_{\text{obj}}) \tag{12}$$

式中，$\text{objE}(\cdot)$ 为矩形框编码模块；F_{obj} 为编码后的特征，最后SAM模型基于 F_{img} 和 F_{obj} 获得耕地掩码：

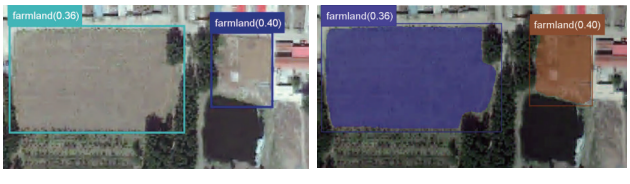
$$\text{Mask} = DE(F_{\text{img}}, F_{\text{obj}}) \tag{13}$$

式中，$DE(\cdot)$ 为特征解码函数；Mask为获得耕地掩码，取值为0或者1，对于矩形框内，属于耕地的像素，取值为1，否则取值为0。本文通过将卫星影像和检测得到的矩形框输入SAM获得最终的耕地掩码，实现耕地的精细化提取。

2　实验结果分析

为验证算法有效性，本文以天津某地为例，标注了3 000张耕地检测数据用于耕地目标检测模型PPYOLOE训练，标注了879张测试数据用于测试。经过测试，本文提出的方法耕地的检测精度 AP 为79.34%，可将大多数耕地正确检测出来。

图2给出了PPYOLOE在测试数据上耕地检测可视化结果。从图2a中可以看出，本文采用的PPYOLOE方法可以将耕地准确检测出来。图2b给出了在耕地目标检测的基础上进行耕地精细化提取的结果，不同地块的耕地用不同的颜色表示，从图2b可以看出，基于PPYOLOE提出的耕地检测框，SAM模型可以进一步地将耕地从耕地检测框中提取出来，具有实用价值。



a 测试样本耕地检测结果　　b 测试样本耕地实例化提取结果

图2　PPYOLOE在测试数据上耕地检测可视化结果

3　结　语

为实现耕地的精细化提取，显著降低样本标注成本，本文联合目标检测和网络大模型SAM，提出一种基于视觉提示词的耕地实例化分割方法。在提出的方法中，本文首先基于目标检测方法对耕地进行粗粒度的耕地目标检测，得到耕地所在的矩形范围，然后将矩形范围作为提示词输入SAM模型中，进行耕地的细粒度提取，获得耕地的精细范围，最终进行耕地的精细化提取。本文提出的方法仅仅利用矩形框标注样本实现耕地的实例化分割，大大降低了对精细化样本标注的依赖，具有实用价值。

参考文献

[1] 牟昱璇,邬明权,牛铮,等. 南方地区复杂条件下的耕地面积遥感提取方法[J]. 遥感技术与应用,2020,35(5):1 127-1 135

[2] 欧阳涛,刘振华,等. 基于ICA面向对象的耕地信息提取[J]. 江苏农业科学,2016,44(7):372-376

[3] 李敏,崔世勇,李成名,等. 面向对象的高分辨率遥感影像信息提取:以耕地提取为例[J]. 遥感信息,2008(6):63-66

[4] 吕林涛,姚建华,黄涛,等. 基于深度学习的遥感影像耕地提取[J]. 数字技术与应用,2023,41(2):98-100

[5] 程锐,魏妍冰,陆苗,等. 基于集成深度学习模型的耕地地块提取[J]. 中国农业资源与区划,2022,43(7):273-281

[6] 王志敏. 基于卷积神经网络的高分辨率遥感影像耕地识别与提取方法研究[D]. 昆明:云南师范大学,2022

[7] 张紫萍,程小凯. 基于分水岭和多尺度分割方法的耕地信息提取[J]. 地理空间信息,2012,10(3):51-53

[8] 徐世武,杨双,孙飞,等. 面向对象的高分辨率影像单值分类耕地提取方法研究[J]. 测绘通报,2014(10):78-81

[9] 张海天,高懋芳,任超. 基于改进Unet++的丘陵地区耕地地块深度分割与提取[J]. 航天返回与遥感,2022,43(4):36-45

[10] Lu R, Wang N, Zhang Y, et al. Extraction of Agricultural Fields Via Dasfnet with Dual Attention Mechanism and Multi-scale Feature Fusion in South Xinjiang, China[J]. Remote Sensing, 2022, 14(9):2 253

[11] Xia L, Luo J, Sun Y, et al. Deep Extraction of Cropland Parcels from Very High-resolution Remotely Sensed Imagery[C]//2018 7th. IEEE International Conference on Agro-geoinFormatics (Agro-geoinformatics), 2018:1-5

[12] Waldner F, Diakogiannis F I. Deep Learning on Edge: Extracting Field Boundaries from Satellite Images with a Convolutional Neural Network[J]. Remote Sensing of Environ-

（下转第41页）