doi:10.6041/j.issn.1000-1298.2025.04.028

基于无人机遥感和弱监督学习的城乡结合部固体废弃物识别

冯权泷[1]　张鑫虹[1]　师宏大[1,2]　牛博文[1]　陈泊安[1,3]　高秉博[1]

(1.中国农业大学土地科学与技术学院,北京 100083;2.北京宏福国际农业科技有限公司,北京 102601;
3.武汉大学遥感信息工程学院,武汉 430079)

摘要:针对当前固体废弃物在遥感影像中标注困难、特征复杂、边界提取精度低等问题,提出了一种基于弱监督学习的两阶段方法:第 1 阶段采用图像级标签数据集,在 5 种网络模型中进行对比实验,最后选择 Swin Transformer 作为特征学习模型。然后,采用梯度加权类激活映射图进行特征区域可视化,以得到热力图。随后融合自适应阈值法和色差法作用于热力图,获取固体废弃物粗轮廓。第 2 阶段采用 DeepSnake 模型进行优化以得到精细化轮廓。利用无人机多光谱遥感影像数据,对河北省廊坊市开发区内 6 个典型城乡结合地区进行实验。第 1 阶段,对 5 种网络模型进行测试,Swin Transformer 优势显著,精确率 93.8%、召回率 95.0%、F1 分数 94.4%,同时通过注意力区域可视化对比也显示其识别效果最好;自适应阈值和色差法融合法的粗轮廓提取在二值化对比实验下显示出优势。第 2 阶段,精轮廓提取定量分析采用了 COCO 数据集的评价指标平均精度(Average precision,AP)进行评价,在 IOU 为 0.5 时,AP 为 91.3%;IOU 为 0.75 时,AP 为 77.5%。同时,在 1、2 阶段轮廓提取的定性比较下,显示出 DeepSnake 的优化作用。实验结果表明:本研究能够利用图像级标签数据集精确识别提取固体废弃物,具有显著的精度优势,可为我国城乡生态环境治理等提供可行方法。

关键词:固体废弃物;无人机遥感;场景识别;特征提取;弱监督学习;城乡结合部

中图分类号:P283.8;TP75　　**文献标识码**:A　　**文章编号**:1000-1298(2025)04-0303-10　　**OSID**:

Solid Waste Identification in Urban-rural Fringe Areas Based on UAV Remote Sensing and Weakly Supervised Learning

FENG Quanlong[1]　ZHANG Xinhong[1]　SHI Hongda[1,2]　NIU Bowen[1]　CHEN Boan[1,3]　GAO Bingbo[1]

(1. *College of Land Science and Technology*, *China Agricultural University*, *Beijing* 100083, *China*
2. *Beijing Hongfu International Agricultural Technology Co., Ltd.*, *Beijing* 102601, *China*
3. *School of Remote Sensing and Information Engineering*, *Wuhan University*, *Wuhan* 430079, *China*)

Abstract: Aiming to address the challenges of difficult annotation, complex features, and low boundary extraction precision for solid waste in remote sensing imagery, a two-stage method was proposed based on weakly supervised learning: in the first stage, an image-level labeled dataset was utilized to conduct comparative experiments among five network models, ultimately selecting the Swin Transformer as the feature learning model. Subsequently, the gradient-weighted class activation mapping was employed for feature region visualization to obtain heatmaps. These heatmaps were further processed by using a combination of adaptive thresholding and color difference methods to obtain a rough outline of the solid waste. In the second stage, the DeepSnake model was employed for optimization to achieve refined contours. This study utilized unmanned aerial vehicle (UAV) multispectral remote sensing image data to conduct experiments in six typical urban-rural interface areas within the Langfang Development Zone, Hebei Province. The results of the experiments were as follows: in the first stage, testing of the five network models revealed a pronounced advantage for the Swin Transformer in feature extraction quantitative analysis, with a precision of 93.8%, recall of 95.0%, and F1 score of 94.4%. Visualization of attention regions also indicated that it had the best recognition effect. The coarse outline extraction by using the combination method of adaptive thresholding and color difference demonstrated

收稿日期:2024-03-04　修回日期:2024-10-20

基金项目:国家重点研发计划项目(2022YFB3903504)

作者简介:冯权泷(1988—),男,副教授,博士生导师,主要从事机器、深度学习及在农业遥感中的应用研究,E-mail:fengql@cau.edu.cn

superiority in the binary comparison experiment. In the second stage, quantitative analysis of fine contour extraction evaluated by using the average precision (AP) metric from the COCO dataset, yielded an AP value of 91.3% at IOU 0.5 and 77.5% at IOU 0.75; moreover, qualitative comparison of contour extraction between the first and the second stages highlighted the optimization effect of DeepSnake. The results demonstrated that this study can accurately identify and extract solid waste by using an image-level labeled dataset, offering pronounced accuracy advantages and providing a viable method for the ecological environment management of urban and rural areas in China.

Key words: solid wastes; UAV remote sensing; scene recognition; feature extraction; weakly supervised learning; urban-rural fringe areas

0 引言

固体废弃物是人类在生产、消费和生活中产生的固态或半固态物质,具有不规则堆积、流动性差等特性[1]。随着我国社会的迅速发展和人民生活水平的提高,生活区域已经无法容纳居民生产活动产生的固体废弃物,导致城市周边和农村废弃地带固体废弃物随意堆积的现象频发。大部分固体废弃物不能降解,影响土壤质地,并散发恶臭、滋生病菌,威胁人类健康和环境。因此,开展固体废弃物的精确提取具有必要性,明确的废弃物空间分布和覆盖面积,对于垃圾污染定量评估的准确性[2]、规划垃圾处理和资源回收设施布局[3],以及优化垃圾处理能力等都具有十分重要的意义。

固体废弃物识别领域源于早期的实地调查,后随着遥感技术的发展,逐步转向遥感影像的应用。BAGHERI 等[4]利用目视解译方法对固体废弃物空间分布进行统计,这是基于遥感影像固体废弃物识别的早期尝试之一。但该方法主观性强、效率低,难以满足大规模识别需求,自动检测识别方法[5-6]应运而生。虽然实现了固体废物物的自动检测,但主要依赖传统的机器学习方法,并且需要手工选择图像特征,其泛化能力和鲁棒性可能不足以应对复杂、异质的场景。与上述方法相比,深度学习在计算机视觉任务中表现突出[7-9]。

固体废弃物识别领域已经从目视解译过渡到传统算法和深度学习的应用,并取得了显著进展。但目前仍存在一些问题:①很多研究仍需要大量的标注样本进行训练,以获得准确的特征信息。但固体废弃物在遥感影像中具有边界轮廓不清晰、易混淆和不易标注等特点,导致标注困难、成本高等问题。②当前的研究大多数是直接采用监督式深度学习模型,没有针对固体废弃物在具体场景中的差异进行优化调整。这些差异包括固体废弃物的背景、分布、形状尺寸等,而这些差异特征又会影响废弃物识别精度[10-11]。③当前的网络模型对于固体废弃物这种形状不规则目标进行实例分割的效果不佳,即只能进行粗略定位,而识别目标边界时,并不能准确预测像素对象属于哪一类别,轮廓边界划分不够精确。

针对以上问题,本文提出一种两阶段递进的提取策略。利用不精确监督信息的弱监督学习[12]方法,仅需图像级标签数据便可实现固体废弃物的边界提取,以解决标注困难问题;以 Swin Transformer 作为特征学习模型,其分层式注意力机制可在不同尺度上提取目标的特征,同时其移动窗口的特点增强模型对于图像局部信息的感知能力,从而更好地捕捉固体废弃物的分布、形状尺寸等特征,以应对场景特异性问题;通过 DeepSnake 循环卷积进行轮廓特征学习,回归轮廓点偏移实现轮廓变形,对目标边界分割进行优化。以期为研究固体废弃物分布、面积统计提供可行方法。

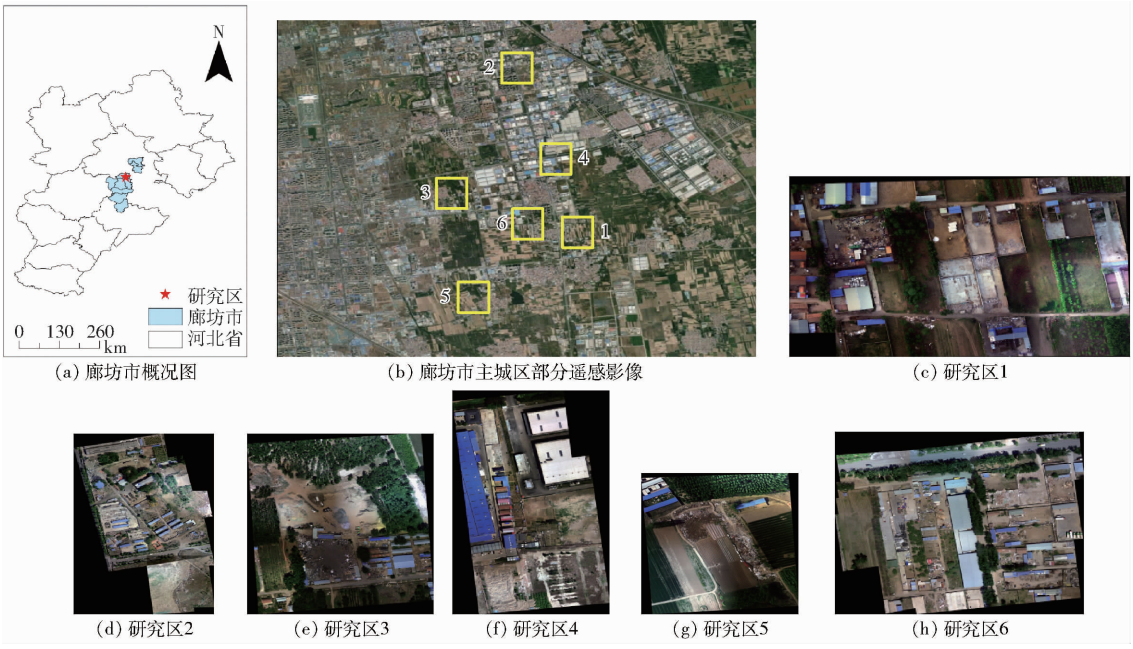
1 研究区概况与数据集构建

1.1 研究区概况

研究区为京津冀战略核心区域——廊坊市开发区,是廊坊市工业经济示范区,是城镇与乡村多融合区域,固体废弃物产生量大、种类多。选择开发区内 6 个具有代表性子区域作为实验对象(图 1)。

1.2 数据源及其预处理

采用大疆精灵 4 RTK 型多光谱无人机,于研究区废弃物上方恒定高度 100 m 处获取多光谱影像数据,该高度下所得遥感影像空间分辨率为 2.74 cm,足以清晰反映固体废弃物的形状和轮廓细节,便于识别提取。相机镜头焦距为 6 mm,最大光圈 f/2.8,测光模式偏中心平均,航向重叠率 85%,旁相重叠率 90%,飞行速度 2 m/s。多光谱相机由 5 个通道组成,分别为蓝光波段(中心波长 450 nm)、绿光波段(中心波长 570 nm)、红光波段(中心波长 675 nm)、红边波段(中心波长 730 nm)和近红外波段(中心波长 850 nm)。实验时间为 2021 年 5 月 1—2 日,为充分利用有限时间,研究选取 6 个在地理位置、区域类型等方面具代表性的实验区进行飞行测试,天气晴朗无云,风力 4 ~ 5 级。共采集 2 730 幅 TIFF 格式的无人机影像。



(a) 廊坊市概况图　(b) 廊坊市主城区部分遥感影像　(c) 研究区1

(d) 研究区2　(e) 研究区3　(f) 研究区4　(g) 研究区5　(h) 研究区6

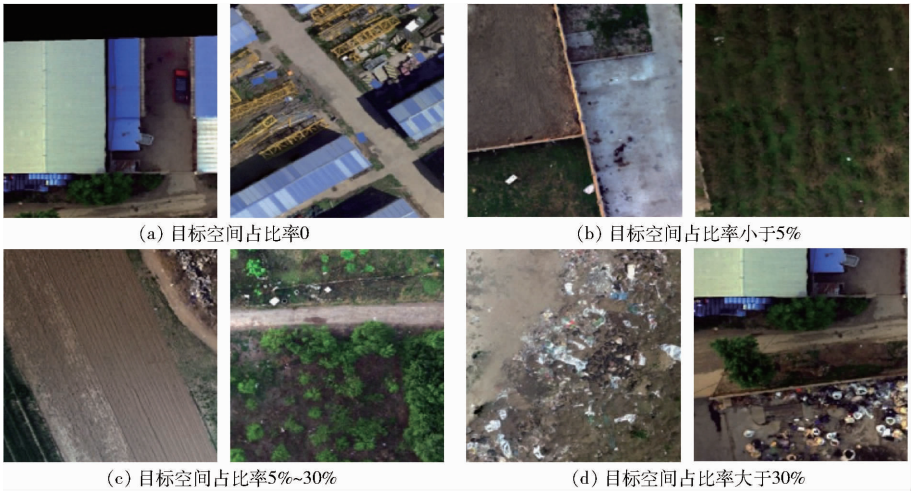
图 1　研究区概况

Fig. 1　Overviews of research area

为消除环境的干扰，使用多光谱相机自带白板进行辐射校正实验，并通过 PhotoScan 对影像进行辐射校正。然后对影像进行真彩色合成，并利用 PhotoScan 拼接生成覆盖整个研究区的遥感影像。在拼接中，会进行方向校正以确保影像的方向对齐（南北正向），而该过程会使最终影像边界处产生未覆盖区域，表现为黑色。最后，将影像裁剪成 512 像素 ×512 像素的小图，该尺寸能够较好涵盖成堆固体废弃物，以满足后续深度学习需要。

1.3　图像级标签数据集构建

基于不精确监督信息的弱监督学习方法进行固体废弃物识别，因此在构建样本数据集时，仅需人工标注每幅影像中是否存在固体废弃物即可，即将样本分为废弃物样本和非废弃物样本两类。在分类时，选择包含农田白膜、彩钢瓦、工厂建筑、耕地、马路、水体、森林植被、山体等场景的样本，其中工地堆放用具、渣土堆不属于固体废弃物。筛选这些具有代表性与多样性的样本，以提升模型在不同场景下的识别能力，避免因样本类别单一导致过拟合问题。此外，在标注样本时，按照影像中目标场景像素不同占比进行标注。为保证模型有效学习到目标特征[13]，通过设置不同占比数值进行测试，以及结合场景的实际情况，最后选择占比超过 30% 的为废弃物样本，占比低于 5% 的为非废弃物样本，占比在此区间的不纳入数据集，由此确立样本构建规则，见图 2。最终，得到 512 像素 ×512 像素的数据样本图像共计 1 081 幅，详情见表 1。本研究所有实验都按照比例 5∶1 随机将样本划分为训练集和测试集，同



(a) 目标空间占比率0　(b) 目标空间占比率小于5%

(c) 目标空间占比率5%~30%　(d) 目标空间占比率大于30%

图 2　固体废弃物识别规则

Fig. 2　Classification standards for solid waste

时随机抽取 30% 的训练集样本作为模型的验证集，防止模型过拟合。

表 1 样本数据分类与分布

Tab. 1 Sample data classification and distribution

幅

类别	研究区 1	研究区 2	研究区 3	研究区 4	研究区 5	研究区 6	总计
固体废弃物	54	174	75	101	80	31	515
非固体废弃物	46	214	96	86	83	41	566
总计	100	388	171	187	163	72	1 081

2 研究方法

2.1 技术路线

本研究搭建了一种基于弱监督学习的遥感影像中固体废弃物的精细化识别技术路线(图 3)，通过图像级标签数据实现像素级的固体废弃物精细化提取。该技术路线分为 2 个阶段：第 1 阶段，首先构建 5 种不同网络模型，分别为 ResNet50[14]、VGG - 19[15]、Efficientnet - b3[16]、VIT[17] 和 Swin Transformer，通过比较分析，最终选择 Swin Transformer 模型来进行特征提取。随后结合梯度加权类激活映射图(Gradient-weighted class activation mapping，Grad - CAM)[18] 与自适应阈值、色差法融合法，得到固体废弃物粗轮廓。第 2 阶段，使用 DeepSnake 模型，进一步细化粗轮廓，最终得到固体废弃物精细轮廓。

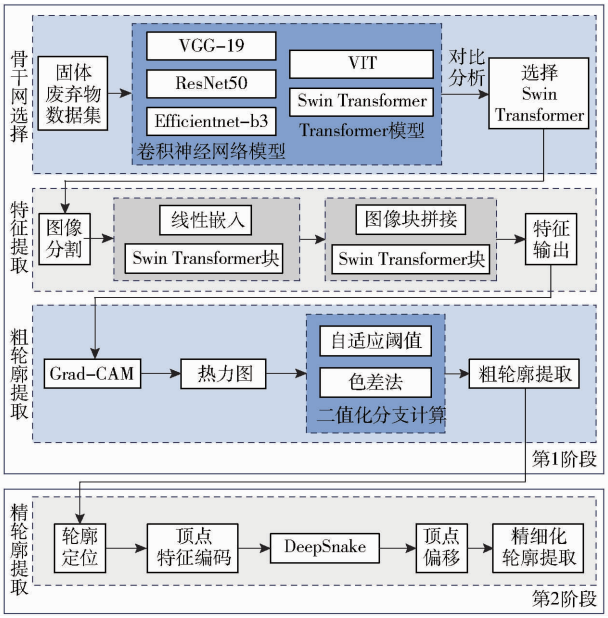


图 3 技术路线

Fig. 3 Technology roadmap

2.2 粗轮廓提取

2.2.1 基于 Swin Transformer 的特征提取

构建 5 种网络模型：以 VGG - 19、ResNet50、Efficientnet - b3 代表前沿的卷积神经网络模型，以 VIT、Swin Transformer 代表前沿的 Transformer 模型，使用图像级标签数据进行对比实验。其中，采用精确率、召回率、F1 分数作为定量评价指标，使用 Grad - CAM 方法的注意力区域可视化作为定性评价指标，通过对比实验，最终选择 Swin Transformer 作为特征提取模型。

Swin Transformer 模型对固体废弃物特征提取的网络结构如图 4 所示，图中，H、W、C 分别表示影像高度、宽度、维度。该模型将输入影像($512 \times 512 \times 3$)分割成互不重叠的尺寸为 4×4 的图像块，即每个图像块保留 $4 \times 4 \times 3$ 的特征维度。在第 1 层中，图像块通过线性嵌入，向量维度 $C = 3$，以符合 Swin Transformer 块的输入要求。接着图像进入 Swin Transformer 块进行特征提取。Swin Transformer 块利用自注意力非重叠窗口(Window multi-head self-attention，W - MSA)和自注意力非重叠移动窗口(Shifted window multi-head self-attention，SW - MSA)来学习图像特征，基于窗口的自注意力提高了计算效率，而移动窗口(向右下移动 1/2 倍窗口)则实现了窗口之间的通信。为获得多尺寸特征信息且提高计算效率，模型采用分层级策略，即通过第 2、3、4 层中的图像块拼接(将特征图中邻近小块合并，并在通道方向拼接，进行层归一化与线性层的全连接)进行降采样，这样即可在网络的不同层次上处理图像特征。最后，第 4 层输出将被用作最终的固体废弃物特征提取结果，完成对输入图像特征提取。

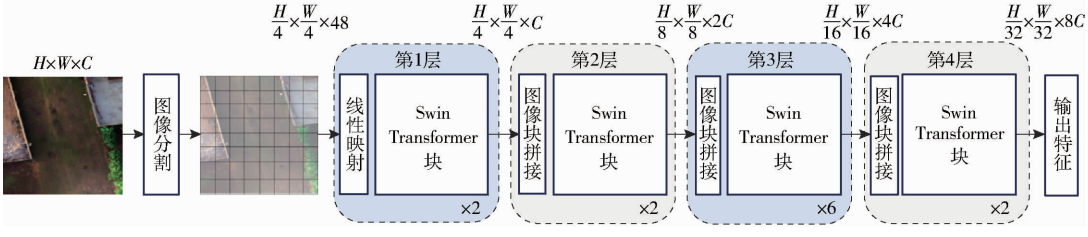


图 4 固体废弃物特征提取网络结构图

Fig. 4 Network architecture diagram for solid waste feature extraction

2.2.2 基于 Grad - CAM 的热力图生成

Grad - CAM 方法是 SELVARAJU 等[18] 提出的基于卷积神经网络模型的可视化方法，它可以将卷积神经网络用于目标识别的分类特征在原图上以热

力图显示，从而反映目标在图像中的位置和区域。本研究中，采用 Grad－CAM 方法突出显示 Swin Transformer 模型在图像级标签数据中进行分类预测的关键区域，生成的热力图即可揭示固体废弃物的位置。最终将热力图与原始影像合并显示。

2.2.3　基于自适应阈值和色差法融合的固体废弃物粗轮廓提取

从热力图到固体废弃物的具体轮廓，需要对热力图进行二值化处理。本研究采用了自适应阈值和色差法融合策略来提取目标轮廓，具体步骤如下：先将热力图进行灰度处理、去噪处理；然后应用自适应阈值法，以最大化前景与背景之间的类间方差为阈值设定原则，对图像进行二值化，得到亮度较高的目标区域；接着，在 HSV 色彩空间中，通过多次实验确定了在 90% 情况下能够准确覆盖目标区域的 HSV 阈值范围，色差法即利用阈值生成掩码层并作用于热力图，保留固体废弃物区域并去除非固体废弃物区域，生成对应的二值化结果；最后，将自适应阈值法和色差法的二值化结果融合重叠，即采用像素级别的“交”操作，只有同时为 255 的像素被判定为固体废弃物，否则为背景，从而获得固体废弃物的粗轮廓。

2.3　精细化轮廓提取

为进一步获取固体废弃物的精细化轮廓，本研究采用基于深度学习的 DeepSnake 模型。该模型通过循环卷积结构处理输入的粗轮廓顶点，并基于学习到的特征获取每个顶点需要调整的偏移量，以尽可能准确地包围固体废弃物真实边界，经过多次迭代更新，最终得到更为精确的轮廓。

DeepSnake 网络由主干网络、融合网络以及预测网络部分构成（图 5）。其中主干网络是由 8 层 CirConv－BN－ReLU 层组成，主要用来完成轮廓上各顶点的特征学习。其中批量归一化（Batch normalization，BN）作用是在梯度下降算法过程中将 Batch 数据进行归一化处理，8 层 CirConv－BN－ReLU 之间都采用残差跳跃连接。

融合网络是由 1×1 卷积和最大池化层来组成，主要用来融合主干网络学习到各个顶点的多尺度特征信息。其中 1×1 卷积和最大池化层可以将轮廓上各顶点的特征值进行融合拼接到一起，随后将融合后的特征进行降维转发到预测网络。

预测网络是由 3 个 1×1 卷积层和 2 个修正线性激活函数（Rectified linear unit，ReLU）组成，其主要用来将预测特征映射成顶点位置的偏移量，最后将轮廓上所有顶点的偏移量形成一个集合输出。

该算法要求输入的是图像的一个初始轮廓，而

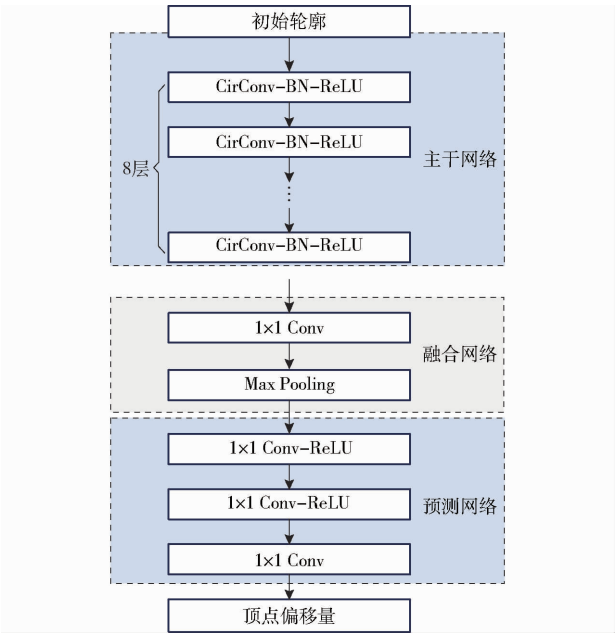


图 5　DeepSnake 网络结构图

Fig. 5　DeepSnake network architecture

在上一阶段已经获取到固体废弃物的粗轮廓，其中轮廓包含每个边线顶点的值，通过在该模型中进行学习，将会输出包含有输入轮廓所有顶点的位置偏移量的集合，最后顶点偏移得到最终的轮廓边界，见图 6。

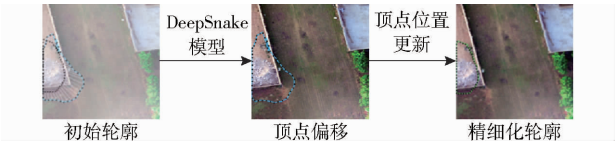


图 6　精细化轮廓提取示意图

Fig. 6　Illustration of refined boundary extraction

3　实验结果与分析

3.1　第 1 阶段实验结果与分析

3.1.1　特征提取模型定量分析

针对 5 个特征提取模型，利用图像级标签数据进行测试评估。对于二分类研究问题，模型的定量评价采用混淆矩阵衍生指标。其中，精确率（Precision）是指模型预测为固体废弃物的结果中，真实固体废弃物占比；召回率（Recall）是指真实固体废弃物样本中，被模型正确预测为固体废弃物的占比。表 2 为各个模型对固体废弃物预测分类的结果。

从表 2 可以看到，Swin Transformer 预测废弃物优势显著，精确率、召回率、F1 分数都取得了最好的结果。

3.1.2　特征提取模型定性分析

图 7 是用 Grad－CAM 方法对注意力区域进行可视化，对 5 种模型识别结果定性分析结果图。在

表 2 第 1 阶段测试集混淆矩阵评价指标计算结果

Tab. 2 Evaluation metrics calculation results for confusion matrix of the first stage test set %

模型	精确率	召回率	F1 分数
ResNet50	89.2	92.5	90.7
VGG－19	93.3	87.5	90.3
Efficientnet－b3	79.8	88.7	84.0
VIT	90.2	92.5	91.6
Swin Transformer	93.8	95.0	94.4

分布相对集中、废弃物面积大的场景 1 下，Swin Transformer 准确识别出大面积废弃物，而其他模型性能有限，特别是 ResNet50 和 VGG－19；在废弃物分布相对集中但面积较小、有相似非目标物体的场景 2 中，Swin Transformer 不仅对大面积目标进行准确分类，还能较好地识别小面积目标，相较之下，其他模型存在漏分和错分的问题；在废弃物分布不集中、面积较小的场景 3 中，Swin Transformer 是唯一能准确识别出废弃物的模型；在背景环境嘈杂、废弃物分布不规律的场景 4 中，Swin Transformer 仍展现出良好的识别效果，而其他模型则存在准确性较低的问题；面对图像中固体废弃物占比较小的场景 5 时，Swin Transformer 在这种极端的环境下依旧能识别出固体废弃物的范围。

场景 1～5 中，包含固体废弃物在研究区中存在

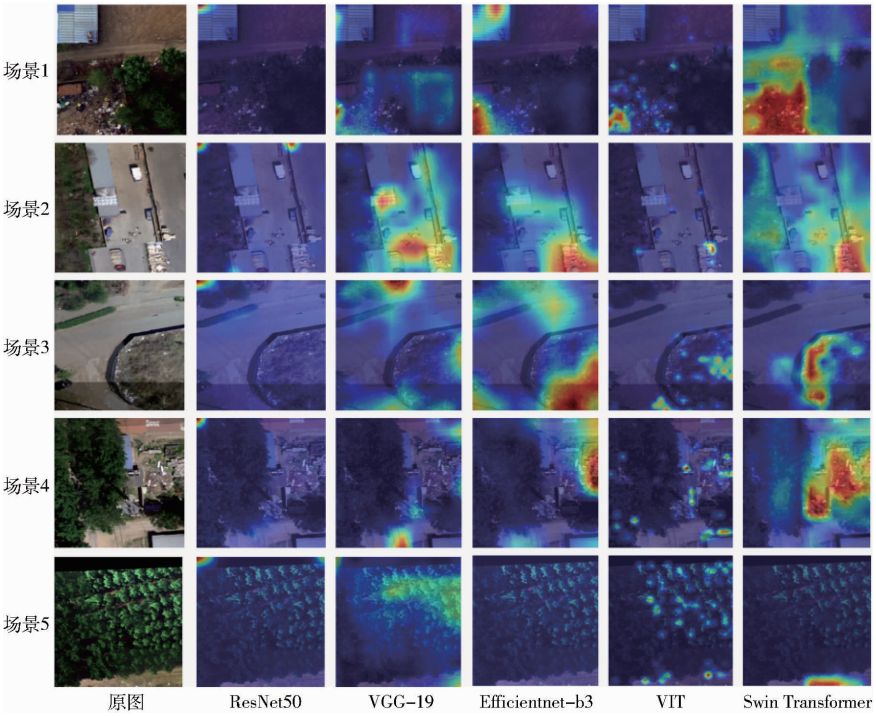


图 7 特征提取模型定性对比示意图

Fig. 7 Qualitative comparative representation of feature extraction models

的各种形态，以上结果可以看出，Swin Transformer 模型在废弃物识别上表现出极好的效果，但是其注意力区域容易将周边小部分非样本环境加入进来，因此需要第 2 阶段实验来进一步提升固体废弃物识别精确度。

3.1.3 二值化对比

Swin Transformer 的热力图及不同二值化方法应用结果显示（图 8），自适应阈值二值化涵盖亮度较高但颜色为蓝色的非注意力区域，而色差法的结果则包含亮度较低但颜色为红色的非目标区域。通过将两者的结果融合重叠，能够弥补各自不足，更精确地划分出固体废弃物区域。

3.2 第 2 阶段实验结果与分析

3.2.1 精轮廓提取模型定量分析

精细化轮廓提取阶段采用 COCO 数据集格式的数据，因此评价指标选择 COCO 数据集评价指标 AP。AP 即在精确率-召回率（Precision-recall，PR）曲线上求 Precision 的平均值，可以用积分的形式表示。但是在具体实验中，一般不直接求解 PR 曲线，而是采用在 PR 曲线上取若干个点来近似计算 AP。

实验采用 IOU 为 0.5 时的 AP_{50} 和 IOU 为 0.75 时的 AP_{75} 作为评价指标。

3.2.2 精轮廓提取模型定性分析

通过将第 1 阶段（Swin Transformer）和第 2 阶段（Swin Transformer + DeepSnake）的结果进行可视化对比，以展示第 2 阶段精轮廓提取优化效果，图 9 为对比结果。通过两阶段结果对比可知：

（1）第 2 阶段中，固体废弃物目标在分类图层中的占比明显大于第 1 阶段，其边缘更接近真实废弃物的边缘。

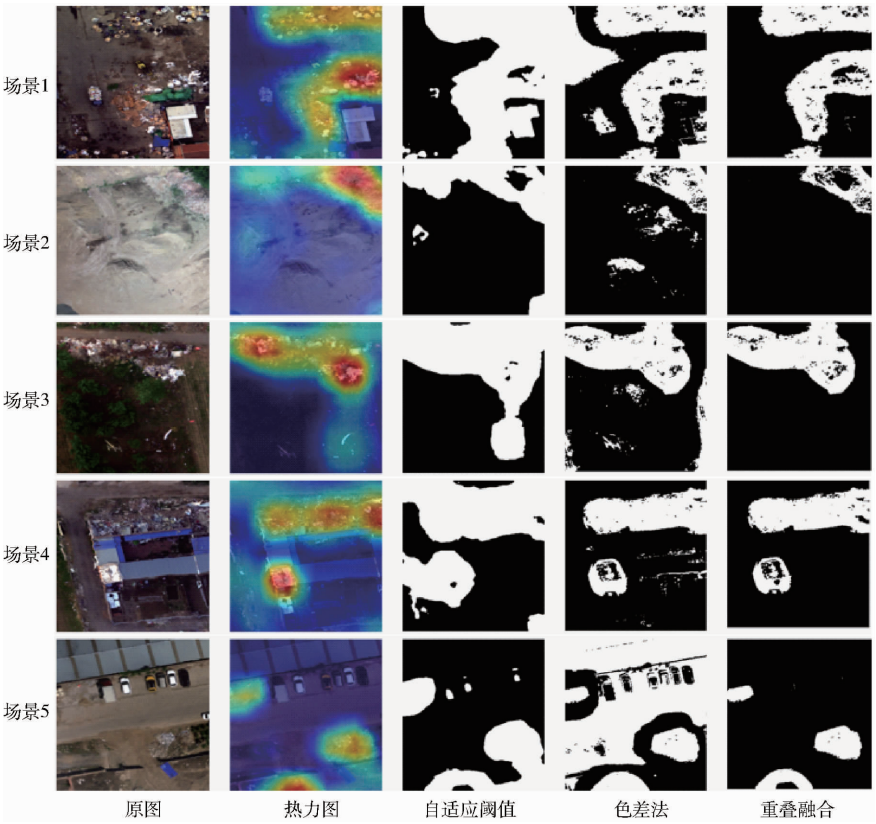


图 8　热力图及不同二值化方法应用结果

Fig. 8　Heatmap and results from different binarization techniques applied

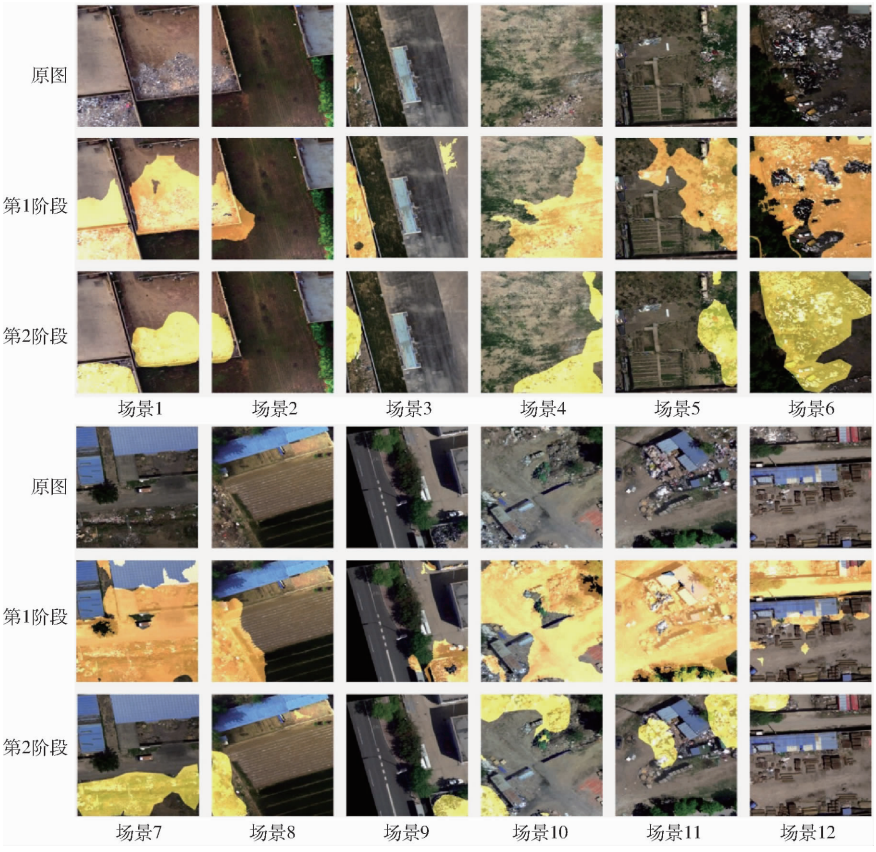


图 9　典型场景固体废弃物识别对比

Fig. 9　Comparative diagram of solid waste recognition in typical scenarios

(2)对于第 1 阶段中连片的识别区域,第 2 阶段能够将这些区域准确分割,从而精确划分固体废弃物区域。这表明第 2 阶段识别结果更为细化,减少了对错误目标的覆盖。

(3)对于第 1 阶段中的错误识别场景,例如树顶、地板、屋顶和荒地等,第 2 阶段进行了排除和过滤操作,减少了对非目标物体的大面积覆盖,使图层更多地包含目标信息。

3.3 最终实验结果与分析

3.3.1 识别结果

采用本文模型对 6 个研究区的全部影像进行识别,得到了最终的固体废弃物识别结果,详见图 10。

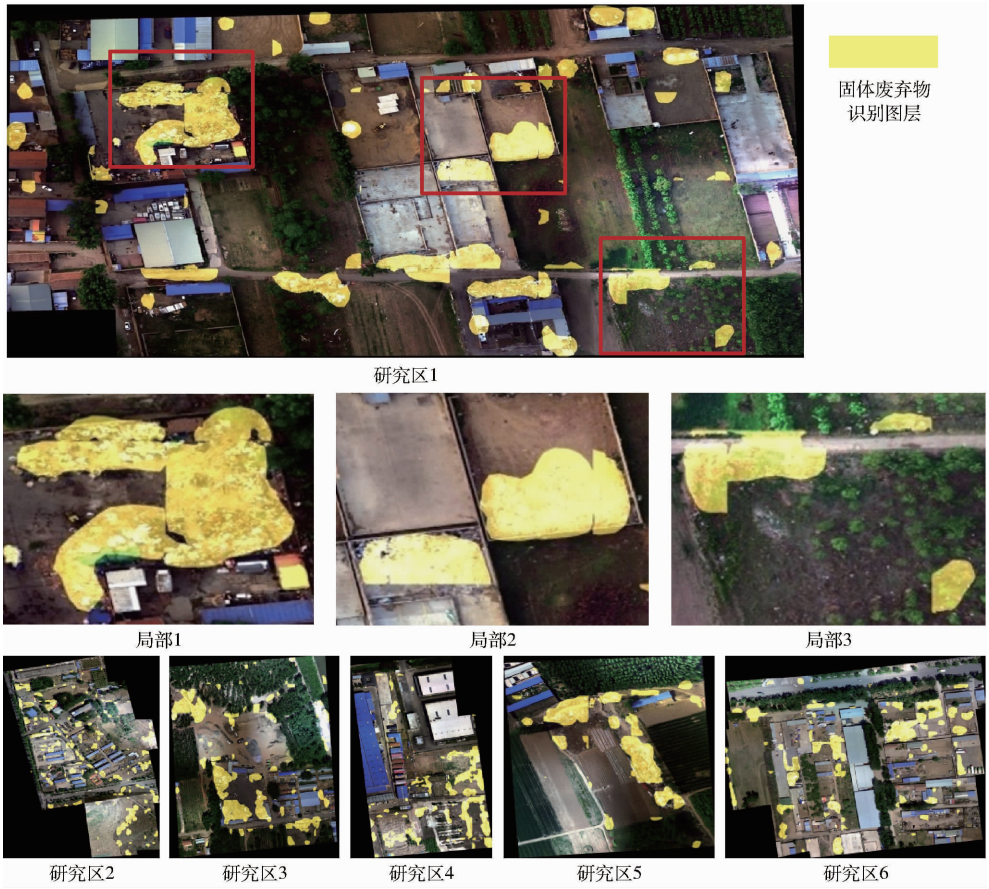


图 10 固体废弃物识别结果

Fig. 10 Solid waste identification results

可以观察到,在这些区域中,固体废弃物基本都已被成功识别并提取,其轮廓清晰可见。这显示出此模型在遥感场景中对固体废弃物识别的良好效果。

3.3.2 错误场景识别分析

图 11 为固体废弃物错误识别案例。经过分析,可以将错误识别原因分为:①特征相似引发的错误。如与废弃物颜色、亮度相似的深色地板和红色屋顶,同废弃物形状相近、具有不规则边缘的人行道破损白漆和树伞阴影,这些相似的视觉特征导致了模型的误判。②特征学习不足导致的错误。如未将耕地中固体废弃物识别出来,这主要是由于这类废弃物的训练样本较少,使模型在训练阶段未能充分学习到明显的识别特征。

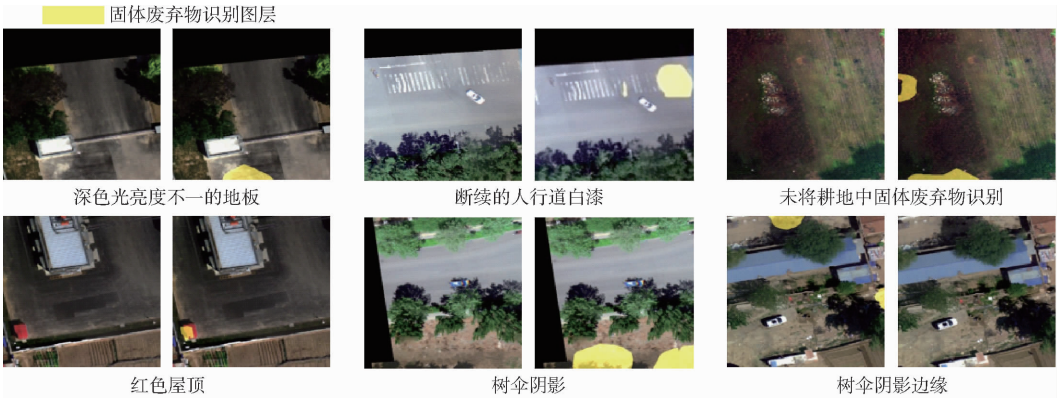


图 11 固体废弃物错误识别案例

Fig. 11 Cases of solid waste misidentification

4　讨论

4.1　识别方法比较

为了进一步验证本文模型对固体废弃物识别性能，将使用 CenterNet + DeepSnake 模型[19]、Dance 模型[20]与本文模型进行比较。所选择的 2 个模型都是实例分割方法，能够对固体废弃物进行识别提取，对比结果如表 3 所示。从表 3 可以看出：本文模型 AP_{50}、AP_{75}均为最高。比 CenterNet + DeepSnake 模型均高出近 50 个百分点，比 Dance 模型均高出近 40 个百分点。这一结果证明了本文模型在性能上具有显著的优势，同时也突出 Swin Transformer 作为特征提取模型的优越性，其特征的学习通过分层注意力机制和移动窗口特质，有效捕捉遥感影像中固体废弃物的分布、形状及尺寸等关键特征，体现了其处理场景异质性问题时的卓越能力。

表 3　不同模型提取固体废弃物平均精度对比

Tab. 3　Comparison of average precision for solid waste extraction using different models　%

模型	AP_{50}	AP_{75}
本文模型	91.3	77.5
CenterNet + DeepSnake	41.6	21.2
Dance	56.5	35.1

此外，与其他固体废弃物识别方法相比，本文模型同样表现出优势。例如 SILVESTRI 等[21]曾利用 IKONOS 卫星数据通过 NDVI（归一化植被指数）识别固体废弃物，但该方法在大面积区域的识别可靠性较低，因 NDVI 易受非固体废弃物因素影响。BAO 等[22]曾基于无人机影像和光谱阈值法识别固体废弃物场，但光谱重叠问题容易导致误分类，尤其在光谱波段较少的情况下。SUN 等[9]的 SRAF－Net 模型尽管在泛化性和效率上表现良好，但只能生成固体废物场址的边界框，无法精细提取边界，且需大量手动标注数据。相比之下，本研究提出的 Swin Transformer + DeepSnake 两阶段方法不依赖光谱信息，并通过弱监督学习减少了手工标注需求，DeepSnake 模型在边界分割上的能力使固体废弃物的边界提取更加精细，展现了方法的优越性。

4.2　不足与展望

由于获取高质量的无人机影像样本耗时费力，难以获得足够多的固体废弃物样本进行训练，因此本研究特意选取 6 个不同特征的区域作为试点研究，这些区域包括了不同规模的工厂，以及居民生活区等多种场景，固体废弃物呈现的形态包括集中型与分散型等不同情况，旨在通过典型场景下样本多样性，弥补数量上的不足，保证模型能够进行更充分的特征学习。

本文方法在所选研究区内固体废弃物的精确识别上已显示良好效果。然而，其在更大范围固体废弃物场景下的普遍适用性仍然需要进一步验证。鉴于无人机影像获取成本高昂，未来研究探索可将本文方法应用于更易于获取且覆盖面更广的卫星遥感影像上，以进一步扩展此方法在固体废弃物监控管理中的应用意义。

5　结论

（1）为解决固体废弃物在遥感影像中标注困难、特征复杂、边界提取精度低等问题，实现其高精度识别提取，提出了一种基于不精确监督信息的弱监督学习方法，进行两阶段递进识别：首先构建了图像级标签数据集，第 1 阶段采用 Swin Transformer 作为特征学习模型，并基于 Grad－CAM 进行特征可视化，再利用自适应阈值与色差法融合进行粗轮廓提取；第 2 阶段采用 DeepSnake 模型，对粗轮廓进一步优化，最终实现了固体废弃物轮廓精细化识别。

（2）利用无人机在廊坊市开发区上空获取 6 个典型城乡结合地区的遥感影像数据进行实验，并与 CenterNet + DeepSnake 和 Dance 等多个模型进行对比分析。实验结果表明，本文构建的方法能够利用图像级标签数据精确提取固体废弃物，与其他分类方法相比具有显著的精度优势，有效证明了方法实用性，对于快速、精确获取固体废弃物分布、面积统计提供了可行方法。

参　考　文　献

[1] GAUTAM S, BREMA J, DHASARATHAN R. Spatio-temporal estimates of solid waste disposal in an urban city of India: a remote sensing and GIS approach[J]. Environmental Technology & Innovation, 2020, 18: 100650.

[2] 谭亚荣. 环境污染核算体系研究[D]. 杨凌：西北农林科技大学，2007.
TAN Yarong. Study on the accounting system of environmental pollution[D]. Yangling: Northwest A&F University, 2007. (in Chinese)

[3] 陶建格. 中国环境固体废弃物污染现状与治理研究[J]. 环境科学与管理，2012，37(11)：1－5，31.
TAO Jian'ge. Current status and management of solid waste pollution in China[J]. Environmental Science and Management, 2012, 37(11): 1－5, 31. (in Chinese)

[4] BAGHERI S, HORDON R M. Hazardous waste site identification using aerial photography: a pilot study in Burlington County, New Jersey, USA[J]. Environmental Management, 1988, 12(1): 119 – 125.

[5] BIOTTO G, SILVESTRI S, GOBBO L, et al. GIS, multi-criteria and multi-factor spatial analysis for the probability assessment of the existence of illegal landfills[J]. International Journal of Geographical Information Science, 2009, 23(10): 1233 – 1244.

[6] 刘亚岚,任玉环,魏成阶,等. 北京 1 号小卫星监测非正规垃圾场的应用研究[J]. 遥感学报,2009,13(2):320 – 326.
LIU Yalan, REN Yuhuan, WEI Chengjie, et al. Study on monitoring of informal open-air solid waste dumps based on Beijing – 1 images[J]. Journal of Remote Sensing, 2009, 13(2): 320 – 326. (in Chinese)

[7] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Commun. ACM, 2017, 60(6): 84 – 90.

[8] LIN K, ZHOU T, GAO X, et al. Deep convolutional neural networks for construction and demolition waste classification: VGGNet structures, cyclical learning rate, and knowledge transfer[J]. Journal of Environmental Management, 2022, 318: 115501.

[9] SUN X, LIU Y, YAN Z, et al. SRAF – Net: shape robust anchor-free network for garbage dumps in remote sensing imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(7): 6154 – 6168.

[10] ZHANG C L, LI Y, WU J. Weakly supervised foreground learning for weakly supervised localization and detection[J]. Pattern Recognition, 2023, 137: 109279.

[11] NIU B, FENG Q, YANG J, et al. Solid waste mapping based on very high resolution remotesensing imagery and a novel deep learning approach[J]. Geocarto International, 2023, 38(1): 2164361.

[12] WANG X, ZHU Z, YAO C, et al. Relaxed multiple-instance SVM with application to object discovery[C]// Proceedings of the IEEE International Conference on Computer Vision. IEEE, 2015: 1224 – 1232.

[13] CHEN J, HE F, ZHANG Y, et al. SPMF – Net: weakly supervised building segmentation by combining superpixel pooling and multi-scale feature fusion[J]. Remote Sensing, 2020, 12(6):1049.

[14] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016:770 – 778.

[15] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]// arXiv Preprint, arXiv:1409.1556, 2014.

[16] TAN M, LE Q. Efficientnet: rethinking model scaling for convolutional neural networks[C]// Proceedings of the 36th International Conference on Machine Learning. PMLR, 2019: 6105 – 6114.

[17] ALI A M, BENJDIRA B, KOUBAA A, et al. Vision transformers in image restoration: a survey[J]. Sensors, 2023, 23(5): 2385.

[18] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad – CAM: visual explanations from deep networks via gradient-based localization[J]. International Journal of Computer Vision, 2020, 128(2): 336 – 359.

[19] PENG S, JIANG W, PI H, et al. Deep snake for real-time instance segmentation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 8533 – 8542.

[20] LIU Z, LIEW J H, CHEN X, et al. DANCE: a deep attentive contour model for efficient instance segmentation[C]// Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). IEEE, 2021: 345 – 354.

[21] SILVESTRI S, OMRI M. A method for the remote sensing identification of uncontrolled landfills: formulation and validation[J]. International Journal of Remote Sensing, 2008, 29(4): 975 – 989.

[22] BAO Z, SHA J, LI X, et al. Monitoring of beach litter by automatic interpretation of unmanned aerial vehicle images using the segmentation threshold method[J]. Marine Pollution Bulletin, 2018, 137: 388 – 398.