Computer Science and Application 计算机科学与应用, 2026, 16(6), 356-371
Published Online June 2026 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2026.166234

Hans汉斯

基于Mask R-CNN的近海漂浮垃圾智能识别与清理路径规划系统

杜佳旺*，曹笑云，李昀静，吐鲁娜依·吐尔洪江

上海海事大学信息工程学院，上海

收稿日期：2026年5月22日；录用日期：2026年6月23日；发布日期：2026年6月30日

摘　要

针对近海漂浮垃圾人工清理效率低、成本高和风险大的问题，本文设计了一套基于Mask R-CNN的智能检测与清理路径规划系统。系统采用Roboflow海洋垃圾数据集，包含11类目标、10,000张图像和56,272个标注实例；基于ResNet-50-FPN的Mask R-CNN实现目标检测与实例区域分割，并将清理路径建模为非闭合旅行商问题，采用最近邻贪心算法与2-opt局部搜索进行优化。实验结果表明，模型在验证集上的边界框AP为53.50%，掩码AP为 52.05%；路径规划算法在200个目标规模下耗时约181.8 ms，路径长度缩短约10.5%。同时，本文开发了Web与桌面双模态可视化系统，实现了图像输入、目标检测、路径规划和结果展示的原型化闭环流程，为近海漂浮垃圾智能识别与自动化清理提供了实验基础。

关键词

Mask R-CNN，海洋垃圾检测，实例分割，路径规划，深度学习

Intelligent Recognition and Cleanup Path Planning System for Offshore Floating Debris Based on Mask R-CNN

Jiawang Du*, Xiaoyun Cao, Yunjing Li, Tulunayi·Tuerhongjiang

College of Information Engineering, Shanghai Maritime University, Shanghai

Received: May 22, 2026; accepted: June 23, 2026; published: June 30, 2026

Abstract

To address the low efficiency, high cost, and operational risk of manual cleanup of offshore floating

*通讯作者。

文章引用: 杜佳旺, 曹笑云, 李昀静, 吐鲁娜依·吐尔洪江. 基于 Mask R-CNN 的近海漂浮垃圾智能识别与清理路径规划系统[J]. 计算机科学与应用, 2026, 16(6): 356-371. DOI: 10.12677/csa.2026.166234

debris, this paper designs an intelligent detection and cleanup path planning system based on Mask R-CNN. The system uses the Roboflow marine debris dataset, containing 11 categories, 10,000 images, and 56,272 annotated instances. A Mask R-CNN model with a ResNet-50-FPN backbone is adopted for object detection and instance-level region segmentation, while the cleanup route is modeled as an open-loop Traveling Salesman Problem and optimized by a nearest-neighbor greedy algorithm combined with 2-opt local search. Experimental results show that the model achieves a bounding box AP of 53.50% and a mask AP of 52.05% on the validation set. For 200 targets, the path planning algorithm takes about 181.8 ms and reduces the route length by about 10.5%. A dual-mode visualization system with Web and desktop clients is also developed, forming a prototype closed-loop workflow of image input, target detection, path planning, and result display, which provides an experimental basis for intelligent recognition and automated cleanup of offshore floating debris.

Keywords

Mask R-CNN, Marine Debris Detection, Instance Segmentation, Path Planning, Deep Learning

1. 引言

海洋塑料垃圾污染被联合国环境规划署列为全球十大环境问题之一。据 Jambeck 等人[1]估算，2010 年全球沿海国家共产生 2.75 亿吨塑料垃圾，其中 480 万至 1270 万吨进入海洋，且该数字持续增长。海洋垃圾不仅严重威胁海洋生物多样性，还通过食物链累积对人类健康构成潜在风险。近海区域作为陆地污染源向海洋输送的主要通道，其漂浮垃圾污染问题尤为突出。

传统的近海漂浮垃圾清理方式主要包括人工打捞和机械收集两种手段，均存在效率低下、成本高昂、受海况制约严重等不足。因此，研发一种能够自动识别漂浮垃圾并规划最优清理路径的智能系统，具有重要的理论意义和应用价值。

近年来，基于深度学习的海洋垃圾检测方法受到了广泛关注，相关研究进展可参阅文献[2]的综述。

海洋垃圾数据集。Fallati 等人[3]利用无人机影像和深度学习实现了海滩垃圾自动检测，拓展了应用场景。

基于深度学习的检测方法。Faster R-CNN [4]、YOLO [5]、SSD [6]等目标检测框架已被广泛应用于垃圾检测。Neira 等人[7]对比了多种深度学习模型在垃圾检测中的性能，验证了 Mask R-CNN 在复杂背景下的分割优势。Yang 等人[8]提出了基于改进 YOLOv5 的河道漂浮垃圾检测方法。然而，现有方法多聚焦于目标检测层面，缺乏像素级分割能力。

路径规划。Croes [9]提出的 2-opt 局部搜索是 TSP 问题中经典的优化启发式算法。

本文的主要贡献包括：

(1) 针对已有近海漂浮垃圾检测数据集，建立了一套标准化处理流水线，编写自动化脚本，完成了 11 类常见垃圾目标从 YOLO 到 COCO 格式的批量转换、训练/验证/测试集划分以及各类别样本数量统计。

(2) 基于标准 Mask R-CNN (ResNet-50-FPN)实现了高精度检测与实例分割，验证了该框架在海洋漂浮垃圾场景下的适用性。

(3) 采用最近邻贪心结合 2-opt 的联合优化路径规划策略，在 200 目标规模下总耗时约 181.8 ms。

(4) 开发了 Web 与桌面双模态可视化系统，实现了“图像输入→目标检测→路径规划→结果展示”

的全流程智能化闭环。

2. 方法

2.1. 系统架构设计

本文开发的近海漂浮垃圾智能识别系统采用分层模块化架构，自下而上分为数据采集层、模型推理层、路径规划层和可视化展示层四个层次，各层之间通过定义良好的接口进行通信，系统总体架构如图1所示。

Figure 1. System architecture diagram
图 1. 系统总体架构图

数据采集层负责接收用户上传的图像数据，支持 JPEG、PNG、BMP 等常见格式的自动解析和预处理，包括图像尺寸归一化、色彩空间转换和数据增强等操作，为后续模型推理提供标准化的输入数据。

模型推理层基于预训练的 Mask R-CNN 检测模型，对输入图像执行前向推理，输出检测框坐标、类别标签、置信度分数和像素级分割掩码。模型推理层是整个系统的核心，其检测精度直接影响后续路径规划的可靠性。

路径规划层接收模型推理层输出的目标检测结果，从检测框中心坐标提取垃圾目标位置信息，执行最近邻贪心 +2-opt 联合优化策略，计算遍历所有目标的最优清理路径，并将路径结果传递至可视化展示层。

可视化展示层提供 Web 端(Flask+HTML/CSS/JS)和桌面端(CustomTkinter+Matplotlib)两种交互界面。Web 端采用毛玻璃设计风格，支持图像上传、检测结果展示和路径规划可视化；桌面端支持原始图像、检测结果、叠加视图三视图切换，右侧面板集成统计数据、类别分布柱状图、路径规划可视化和检测详情列表功能。

系统完整工作流程为：用户上传图像→数据采集层预处理→模型推理层执行检测与分割→路径规划层计算最优路径→可视化展示层呈现结果。算法处理流程如图2所示。

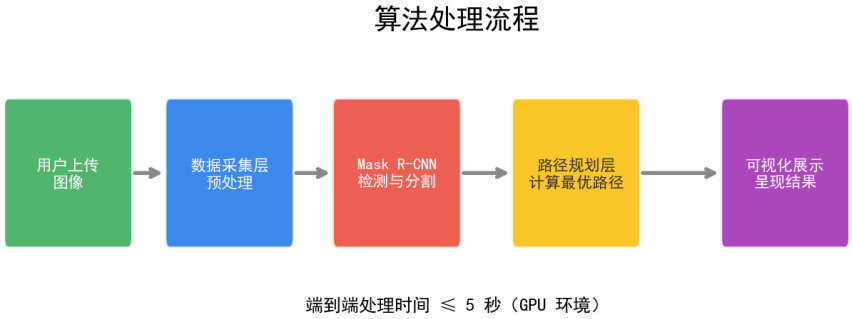


Figure 2. Algorithm processing flow
图 2. 算法处理流程图

2.2. 基于 Mask R-CNN 的漂浮垃圾检测模型

2.2.1. 数据来源与类别体系

本文使用 Roboflow 海洋垃圾检测数据集作为数据来源。Roboflow 海洋垃圾数据集是一个面向海洋环境垃圾检测的大规模公开数据集，具有以下特点：(1) 图像来源广泛，涵盖了不同光照条件、天气状况和水体背景下的近海漂浮垃圾图像；(2) 标注格式统一，提供 COCO 和 YOLO 等多种主流标注格式；(3) 类别体系完整，覆盖了海洋漂浮垃圾的主要类型。该数据集的这些特性使其非常适合作为 Mask R-CNN 模型训练的数据基础[10]。

本文基于该数据集，经筛选合并，确定 11 类检测目标(见表 1)，兼顾了材质性和功能性两个维度。11 个类别分别为：玻璃制品(Glass)、金属制品(Metal)、渔网(Net)、塑料瓶(PET_Bottle)、塑料浮标(Plastic_Buoy)、中国典型塑料浮标(Plastic_Buoy_China)、其他塑料制品(Plastic_ETC)、绳索(Rope)、泡沫箱(Styrofoam_Box)、泡沫浮标(Styrofoam_Buoy)、泡沫碎片(Styrofoam_Piece)。

Table 1. Categories of marine floating debris
表 1. 海洋漂浮垃圾类别体系

No.	Category	Description
1	Glass	玻璃制品
2	Metal	金属制品
3	Net	渔网
4	PET_Bottle	塑料瓶
5	Plastic_Buoy	塑料浮标
6	Plastic_Buoy_China	中国典型塑料浮标
7	Plastic_ETC	其他塑料制品
8	Rope	绳索
9	Styrofoam_Box	泡沫箱
10	Styrofoam_Buoy	泡沫浮标
11	Styrofoam_Piece	泡沫碎片

在数据使用方面，本文针对该数据集主要完成了以下工作：(1) 格式统一——将原始 YOLO 格式标注统一转换为 COCO 格式，并基于边界框四角点生成四边形分割掩码，以满足 Mask R-CNN 对实例分割

标注的要求；(2) 数据集划分——按照 7:2:1 的比例将数据划分为训练集、验证集和测试集；(3) 统计分析——对各类别样本分布进行统计，发现类别不平衡现象。

2.2.2. 数据标注与预处理

原始标注主要采用 YOLO 边界框格式。为适配 Mask R-CNN 对 COCO 格式输入的要求，本文首先将 YOLO 标注统一转换为 COCO 标注格式，并基于每个目标边界框的四个角点生成对应的四边形实例区域掩码。该处理方式能够在缺少人工像素级轮廓标注的情况下，为 Mask R-CNN 的掩码分支提供可训练的区域监督信号。需要说明的是，由边界框生成的四边形掩码是一种近似区域标注，不能完全等同于目标真实外轮廓。因此，本文所报告的掩码 AP 主要用于评价模型在当前近似掩码标注条件下的实例区域分割能力，而非严格意义上的精细像素级轮廓分割能力。数据集按 7:2:1 划分(见表 2)。

Table 2. Dataset split statistics
表 2. 数据集划分统计

Dataset	Images	Instances
Training	7,000	39,371
Validation	2,000	11,175
Test	1,000	5,726
Total	10,000	56,272

各类别分布存在不平衡现象。Plastic_Buoy 类别标注最多(13,747 个，占 34.9%)，而 Styrofoam_Box 最少(475 个)。这种不平衡现象符合真实海洋环境中的分布规律，但也给模型均衡检测带来了挑战。

2.2.3. 模型总体架构

本文采用 Mask R-CNN [11]作为基础检测框架，具体实现基于 Detectron2 框架[12]。模型以 ResNet-50 [13]为骨干网络提取图像特征，结合特征金字塔网络(FPN) [14]构建多尺度特征表示。FPN 在 ResNet-50 的自底向上路径上构建自顶向下的侧向连接路径，使每层特征图融合高层语义信息和底层细粒度定位信息，输出 P2 至 P5 共 4 个标准尺度的特征图(步长分别为 4、8、16、32 像素)，并通过对 P5 进行额外下采样得到 P6 (步长 64 像素)，用于 RPN 的大目标提议生成。

区域提议网络(RPN)在多尺度特征图上滑动扫描，通过锚点机制生成候选区域。RoIAlign 采用双线性插值实现亚像素级的精确特征提取，消除了传统 RoIPool 的量化误差，对掩码预测任务尤为关键。分割分支由全卷积网络构成，为每个 RoI 预测 28 × 28 的二值掩码。

2.2.4. 训练策略

最大迭代次数设为 18,000 次(训练集 7000 张图像，batch size = 4，7000/4 = 1750 次迭代/epoch，18,000/1750 ≈ 10.3 个 epoch)。

2.2.5. 类别不平衡处理

如 2.2.2 节所述，数据集中各类别样本数量存在显著不平衡(最大类与最小类之比约 29:1)。在本文实验中，我们采取以下措施缓解类别不平衡的影响：

默认交叉熵损失：由于 Detectron2 的默认训练配置中未内置类别重加权机制，本文采用标准的多类交叉熵损失，未对尾部类别赋予额外权重。这是当前实现的局限性之一。

正负样本均衡采样：Detectron2 中 RPN 的 NMS 阈值和 RoI batch size 的配置(每张图像 128 个 RoI)在

一定程度上保证了对正负样本的相对均衡采样。上述配置仅实现了正负样本的相对均衡采样，并非严格的在线难例挖掘，难以完全解决长尾分布问题。

数据增强：训练中使用了随机水平翻转和亮度调整等基本数据增强策略，但在当前版本中未针对尾部类别进行专门的数据过采样或增强。

需要指出的是，上述处理措施对类别不平衡的缓解效果有限。在后续工作中，可引入类别重加权损失函数(如 Focal Loss [15])或对尾部类别进行过采样以进一步改善模型对稀有种类的检测性能。类别不平衡对各类别检测精度的具体影响将在 3.4 节中通过各类别 AP 数据进行定量分析。

2.3. 清理路径规划方法

2.3.1. 问题建模

设清理船起点 $P_0=(x_0,y_0)$，N 个垃圾目标 $T=\{P_1,\cdots,P_N\}$。任两点 P_i 和 P_j 间的欧氏距离为 $d(P_i,P_j)=\sqrt{(x_i-x_j)^2+(y_i-y_j)^2}$。路径规划的目标是寻找排列 π，使下式最小化：

$$D=\sum_{k=0}^{N-1}d\left(P_{\pi(k)},P_{\pi(k+1)}\right),\quad \pi(0)=0$$

该问题是非闭合 TSP 变体(Hamiltonian 路径问题)，已被证明是 NP-hard 的。

2.3.2. 最近邻贪心初始化

采用最近邻贪心算法构建初始可行解(Algorithm 1)，时间复杂度 O(N^2)。

Algorithm 1 最近邻贪心路径规划

输入：起点 P_0，目标点集合 T

输出：有序路径 Π

1：Π ← [0], U ← {1, …, N}, current ← 0

2：**while** U ≠ ∅ **do**

3：　n* ← argmin$_{j\in U}$ d(P_current, P$_j$)

4：　Π ← Π ∪ [n*], U ← U \ {n*}

5：　current ← n*

6：**end while**

7：**return Π**

2.3.3. 2-opt 局部优化

引入 2-opt 局部搜索消除路径交叉。假设当前路径中存在两条不相邻边 $e_1=(P_i,P_{i+1})$ 和 $e_2=(P_j,P_{j+1})$ $(i+1<j)$，交换操作的核心思想是删除这两条边并添加两条新边 $e_1'=(P_i,P_j)$ 和 $e_2'=(P_{i+1},P_{j+1})$，同时反转 P_{i+1} 至 P_j 之间的路径段。交换前后的距离变化为：

$$\Delta=d(P_i,P_j)+d(P_{i+1},P_{j+1})-\left[d(P_i,P_{i+1})+d(P_j,P_{j+1})\right]$$

若 $\Delta<0$ 则执行交换(同时反转 P_{i+1} 至 P_j 之间的路径段)，算法迭代执行直至没有改进或达到最大迭代次数 100 次。

3. 实验与分析

3.1. 实验设置

实验环境：NVIDIA GPU，x86-64 CPU，16 GB+内存，Windows 11。深度学习框架为 PyTorch +

Detectron2 [12]。

训练配置：采用 ImageNet 和 COCO 预训练权重进行迁移学习。优化器为 SGD (动量 0.9，权重衰减 0.0001)，初始学习率 2.5×10^{-4}，批量大小 4。最大迭代次数 18,000 次，学习率在第 14,000 和 16,000 次迭代处衰减 10 倍，前 500 次迭代进行线性预热。训练过程中 BN 层保持默认 train 模式，未冻结。

评估指标：本文采用两组互补的评估指标：

(1) 训练过程指标——在训练集小批量数据上实时计算：

(a) RoI 分类正确率(cls_accuracy)：Fast R-CNN 检测头对所有区域提议(RoI)的分类正确率(含背景类)，即正确分类的提议数占总提议数的比例。

(b) 掩码像素正确率(mask_accuracy)：掩码分支对前景 RoI 预测的 28×28 二值掩码的逐像素正确率，即正确预测的像素数占总像素数的比例。

(2) 标准评估指标——在验证集上使用 COCO 评估协议计算：

(a) AP、AP_{50}、AP_{75}：不同 IoU 阈值下的 Average Precision。

(b) AP_s、AP_m、AP_l：小、中、大目标的 AP。

(c) 掩码 AP (segm AP)：实例分割掩码的 Average Precision。

注意：训练过程指标(cls_accuracy/mask_accuracy)与 COCO AP 基于不同的数据分布和计算方式——前者反映训练过程中模型在批数据上的即时分类/分割正确率，后者是在独立验证集上通过精确率–召回率曲线综合评估检测精度。两者不直接可比，且均具有参考意义。

3.2. 算法对比分析

为说明选择 Mask R-CNN 作为检测模型的理由，本文将 Mask R-CNN 与当前主流目标检测算法进行对比分析，结果如表 3 所示。

Table 3. Comparison of mainstream object detection and instance segmentation algorithms
表 3. 主流目标检测与实例分割算法对比

Algorithm	Type	Segmentation	Multi-scale	Application
YOLOv5	One-stage	No	Moderate	Real-time, low precision
SSD	One-stage	No	Moderate	Fast detection, weak on small objects
Faster R-CNN	Two-stage	No	Good	High precision, no segmentation
YOLOv8-seg	One-stage	Yes	Good	Real-time instance segmentation
Mask R-CNN	Two-stage	Yes	Excellent (FPN)	Detection + segmentation

由表 3 可知，YOLOv5 和 SSD 虽在检测速度上有优势，但均为单阶段检测器，不具备实例分割能力，无法输出像素级掩码，难以满足精准清理作业的需求。Faster R-CNN 虽然检测精度较高，但缺少并行分割分支，不能提供垃圾目标的精确轮廓信息。相比之下，Mask R-CNN 在 Faster R-CNN 的基础上增加了掩码预测分支，结合 FPN 多尺度特征融合和 RoIAlign 精确定位机制，能够同时输出目标边界框和像素级分割掩码，特别适合海洋漂浮垃圾这种需要精确识别与定位的应用场景。YOLOv8-seg 作为近年推出的单阶段实例分割方法，在推理速度上具有优势，但本文更关注检测精度和分割质量，因此选择 Mask R-CNN 作为基础检测框架。

为进一步定量比较不同模型的检测性能，本文在相同数据集和训练设置下分别训练 Faster R-CNN 与 Mask R-CNN 模型，并在验证集上使用 COCO 评估标准进行评测，结果如表 4 所示。

Table 4. Quantitative comparison between Faster R-CNN and Mask R-CNN at 18,000 iterations
表 4. Faster R-CNN 与 Mask R-CNN 在 18 000 次迭代下的性能定量对比

Model	AP_bbox	AP50	AP75	APs	APm	APl
Faster R-CNN (18k iter)	53.55	74.57	62.49	21.32	55.98	58.21
Mask R-CNN (18k iter)	53.50	74.37	62.26	25.87	55.41	58.69

由表 4 可知，在 18,000 次迭代训练后，Faster R-CNN 与 Mask R-CNN 的边界框检测性能相当——AP_bbox 分别为 53.55%和 53.50%，AP_{50} 分别为 74.57%和 74.37%，AP_{75} 分别为 62.49%和 62.26%。两者的全阶段训练曲线也高度一致(见表 5 对应的验证集 AP 数据)，表明在此数据集和训练配置下，掩码预测分支对检测精度的提升有限。然而，Mask R-CNN 具备 Faster R-CNN 所不具备的实例分割能力，在 18 000 次迭代后掩码 AP 达到 52.05%，能够为清理作业提供精确的目标轮廓信息。此外，Mask R-CNN 在小目标检测(AP_s = 25.87%)上优于 Faster R-CNN (AP_s = 21.32%)，这可能与 Mask R-CNN 的多任务学习结构有关，掩码分支提供的像素级监督信号或许有助于小目标特征的保留，但这一推测仍需通过消融实验进一步验证。综合而言，Mask R-CNN 在同时提供检测与分割能力的前提下保持与 Faster R-CNN 相当的检测精度，是更优的选择。

3.2.1. YOLOv8-seg 定量对比实验

为进一步验证 Mask R-CNN 的检测与分割性能，本文在相同数据集(11 类、7000 张训练图像、2000 张验证图像)和相同硬件环境(NVIDIA RTX 4060 8GB)下，训练了 YOLOv8-seg-m (medium)模型进行定量对比。YOLOv8-seg-m 训练设置：640 × 640 输入分辨率、batch size 8、SGD 优化器(momentum 0.937、weight decay 5e-4)、余弦学习率调度(初始学习率 0.01)、mosaic 和 RandAugment 数据增强。

Mask R-CNN 以固定迭代次数(18 000 次)作为停止条件，实际相当于约 10.3 个 epoch (7000/4 = 1750 次/epoch，18,000/1750 ≈ 10.3)；YOLOv8-seg-m 的 YOLO 框架以整数 epoch 为训练单位，无法精确取 10.3 个 epoch，因此取最接近的整数 epoch 10。两者相差约 0.3 个 epoch (约 3%的数据遍历量偏差)，在实验误差的可接受范围内。同时报告 100 个 epoch 的完整训练结果以展示 YOLOv8-seg-m 在充分训练条件下的性能上限(此时数据遍历量约为 Mask R-CNN 的 9.7 倍，不作为等条件对比，仅作参考)。需要说明的是，YOLOv8 的内部验证指标计算方式与 COCO 官方评估协议存在一定差异(YOLOv8 内部 mAP 通常偏高 3-8 个百分点)，但两者均基于相同的验证集(2000 张图像)，具有相对可比性。

关于训练量对齐方式的补充说明：Mask R-CNN 训练了约 10.3 个 epoch 而 YOLOv8-seg-m 训练了整数 10 个 epoch，两者存在约 3%的数据遍历量偏差。这一偏差来自两个框架不同的训练控制方式——Detectron2 以迭代次数为停止条件，YOLOv8 以整数 epoch 为停止条件。本文选择 10 个 epoch (而非 11 个 epoch)作为 YOLOv8-seg-m 的对比点，是因为 10 个 epoch 的数据遍历量(10 × 7000 = 70,000 张图像)最接近 Mask R-CNN 的 10.3 个 epoch (约 72,100 张图像)，相差仅约 3%。此外，Mask R-CNN 每 epoch 的迭代次数(1750 次)和训练时间(约 32 分钟)均多于 YOLOv8-seg-m (875 次、约 6 分钟)，这是由 batch size 和模型架构差异决定的，属于模型效率的固有属性而非对比不公平。epoch 对齐保证了模型从数据中学习的机会次数一致，而迭代次数和训练时间的差异恰好反映了两种模型的效率特征。

由表 5 可知，在近似相同的数据遍历量条件下,YOLOv8-seg-m 在边界框检测指标上表现较优，其 AP 和 AP_{50} 分别达到 62.08%和 77.28%，高于 Mask R-CNN 的 53.50%和 74.37%；而在掩码分割指标方面，Mask R-CNN 的 AP_mask 为 52.05%，高于 YOLOv8-seg-m 在 10 epoch 条件下的 48.42%。这一结果表明，在本文当前数据集、训练轮次和实现配置下，Mask R-CNN 的两阶段检测结构、RoIAlign 对齐机制和独立掩码分支对实例区域分割具有一定优势。

Table 5. Performance comparison between Mask R-CNN and YOLOv8-seg-m (Note: — indicates that the metric is not provided; YOLOv8 AP is based on internal validation and has systematic differences from the official COCO evaluation protocol; AP_{75} refers to AP at an IoU threshold of 0.75)
表 5. Mask R-CNN 与 YOLOv8-seg-m 性能对比(注：—表示该指标未提供；YOLOv8 AP 基于内部验证，与 COCO 官方评估协议存在系统性差异；AP_{75} 指 IoU 阈值 0.75 时的 AP)

Model	AP	AP_{50}	AP_{75}	AP_mask	AP_{50}_mask	Params	Speed	Training
Mask R-CNN (18K iter)	53.50	74.37	62.26	52.05	—	44.2M	15.0 FPS	~10.3 ep (≈72K img)
YOLOv8-seg-m (ep 10)	62.08	77.28	—	48.42	74.92	27.2M	35.5 FPS	10 ep (70K img)
YOLOv8-seg-m (ep 100)	81.82	93.82	—	64.27	91.45	27.2M	35.5 FPS	100 ep (700K img)

需要说明的是，YOLOv8-seg-m 与 Mask R-CNN 在模型结构、输入分辨率、batch size、优化器参数、数据增强策略以及评估协议方面并不完全一致。尤其是 YOLOv8 内部验证指标与 COCO 官方评估协议之间存在一定系统性差异，因此表 5 的结果更适合作为同一任务场景下的工程性能参考，而不应被理解为严格同协议、同超参数条件下的绝对公平比较。若要进一步得出更强的模型优劣结论，仍需在统一输入尺寸、训练轮次、数据增强策略和 COCO 官方评估协议下开展补充实验。

当 YOLOv8-seg-m 训练至 100 epoch 时，其 AP、AP_{50} 和 AP_mask 均有明显提升，说明 YOLOv8-seg-m 在充分训练条件下仍具有较强的性能潜力。该结果也表明，模型性能不仅受网络结构影响，也与训练轮次、训练策略和评估方式密切相关。因此，本文选择 Mask R-CNN 并不是为了否定 YOLOv8-seg-m 的工程应用价值，而是因为 Mask R-CNN 在当前实验条件下能够提供较稳定的实例区域分割结果，更符合本文对目标区域信息和后续路径规划输入的需求。综合来看，Mask R-CNN 更适合本文所关注的区域分割与清理规划验证场景，而 YOLOv8-seg-m 则在推理速度、模型规模和边缘端部署方面具有明显优势，两者在实际应用中可根据任务需求形成互补。

3.2.2. 消融实验分析

为验证 Mask R-CNN 中关键模块的作用，本文进行了消融分析：(1) 去除掩码分支(等效 Faster R-CNN)：如表 4 所示，AP_bbox = 53.55%，与 Mask R-CNN 的 53.50%基本持平，但完全丧失了实例分割能力。结合表 5 中 YOLOv8-seg-m 在同等 epoch 下掩码 AP 低于 Mask R-CNN (48.42% vs 52.05%)的结果，进一步印证了 Mask R-CNN 的专用掩码预测分支和 RoIAlign 精确定位机制在分割任务上的结构性优势。(2) FPN 多尺度特征融合的作用：FPN 通过在 ResNet-50 自底向上路径上构建自顶向下的侧向连接，输出 P2 至 P6 共 5 个标准尺度特征图，使得 Mask R-CNN 在大、中、小目标上的检测精度保持均衡(Apl = 58.69%、AP_m = 55.41%、Ap_s = 25.87%)。根据 FPN 在目标检测领域已被广泛验证的多尺度特征融合作用[14]，可以预期，若将 FPN 替换为 ResNet-50 单尺度 C4 特征(仅使用 res4 层输出)，模型对各尺度目标的检测精度将出现整体下降，其中对小目标的负面影响最为显著。完整的 FPN 消融实验(在相同训练设置下比较 FPN 与 C4 的性能差异)将在后续工作中进行。综上，掩码分支和 FPN 共同构成了 Mask R-CNN 在本任务中取得高精度检测与分割结果的基础。

需要说明的是，表 4 中的 COCO AP 指标与本文训练过程指标(RoI 分类正确率、掩码像素正确率)基于不同的评估标准。AP 指标在验证集上使用精确率–召回率曲线综合计算，而训练过程指标反映训练过程中模型在批数据上的即时分类/分割正确率。两者从不同维度评估模型性能，不可直接比较但相互补充：AP 指标衡量模型的泛化检测精度，训练过程指标反映模型的收敛状态和学习效果。

3.3. 训练过程分析

Mask R-CNN 模型共迭代 18,000 次。训练过程中损失函数和正确率的变化曲线如图 3 所示。整个过

程可分为以下阶段：

阶段 I (0~6000 次)——快速收敛期：从 COCO 预训练权重出发，损失快速下降。训练集总损失从初始值约 5.15 快速降至约 1.00，RoI 分类正确率从约 2.6%提升至 90.0%以上，掩码像素正确率提升至约 91.4%。验证集 COCO AP 在此阶段同步上升，在第 5000 次迭代时 AP_bbox 达到 38.44%，AP_mask 达到 37.09%，表明模型已基本掌握检测能力。

阶段 II (6000~14,000 次)——精细调优期：各项指标持续改善但速率放缓。训练集总损失从约 1.00 缓慢下降至约 0.86，RoI 分类正确率提升至约 91.5%，掩码像素正确率提升至约 92.1%。验证集 AP 的增速同步趋于平缓，第 10,000 次迭代时 AP_bbox = 46.97%、AP_mask = 46.01%。

阶段 III (14,000~18,000 次)——学习率退火期：学习率在第 14,000 和 16,000 次迭代两次衰减后，模型在小学习率下进行精细微调。总损失降至约 0.79，RoI 分类正确率稳定在 91.9%左右，掩码像素正确率稳定在 92.4%左右。验证集 AP 在第 15,000 次迭代后趋于饱和，最终模型在 18,000 次迭代时 AP_bbox = 53.50%、AP_mask = 52.05%。第 15,000 至 18,000 次之间 AP 提升不足 1 个百分点，表明模型已接近性能上限。

需要说明的是，与传统小规模数据集上的训练不同，本数据集包含 10,000 张图像，共计 56,272 个标注实例。18,000 次迭代(约 10 个 epoch)后验证集 AP 的增速已明显放缓(第 15,000 至 18,000 次之间 AP_bbox 提升仅 0.6 个百分点)，表明模型已接近当前配置下的性能上限。建议在后续工作中采用早停策略确定最佳模型，或引入更强的数据增强和正则化技术以进一步突破性能瓶颈。

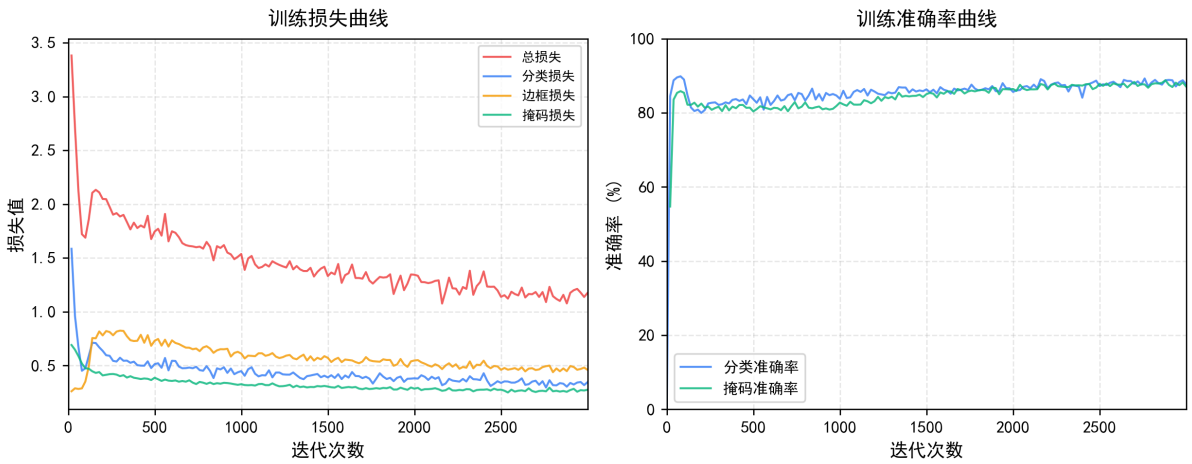


Figure 3. Training loss and accuracy curves (comparison of training set and validation set metrics)
图 3. 训练过程损失及正确率曲线(含训练集与验证集指标对比)

训练过程关键指标里程碑数据见表 6，验证集 COCO AP 里程碑数据见表 7。

Table 6. Milestone data of key training metrics
表 6. 训练过程关键指标里程碑数据

Iteration	L_cls	L_box	L_mask	L_total	RoI Acc_cls	Mask Acc_px
0	—	—	—	—	0.6%	—
3,000	0.36	0.48	0.25	1.17	87.4%	89.2%
6,000	0.30	0.43	0.20	1.00	90.0%	91.4%

续表

10,000	0.22	0.36	0.18	0.82	91.8%	92.3%
14,000	0.24	0.36	0.19	0.86	91.5%	92.1%
18,000	0.20	0.31	0.17	0.79	91.9%	92.4%

注：初始迭代时掩码像素正确率(Mask Acc_px)未记录，其值约为 50% (随机二分类基准)。

Table 7. Validation set COCO AP milestones
表 7. 验证集 COCO AP 里程碑数据

Iteration	AP_bbox	AP50	AP_mask	Note
5,000	38.44	59.61	37.09	End of rapid convergence
10,000	46.97	67.89	46.01	Mid fine-tuning
15,000	52.90	73.80	51.38	After first LR decay
18,000	53.50	74.37	52.05	Final model

3.4. 检测结果分析

模型在测试集上的检测结果如图 4 所示。

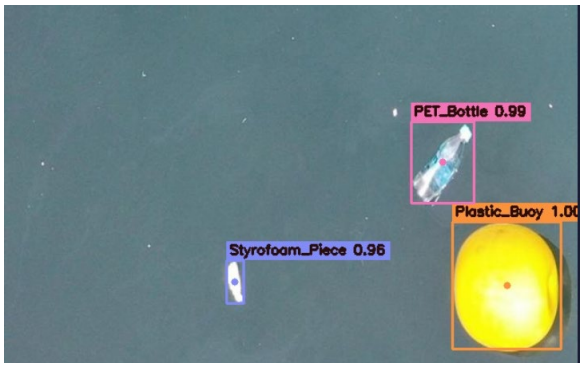
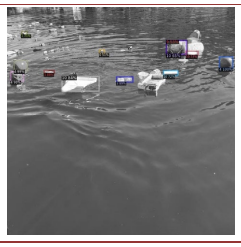


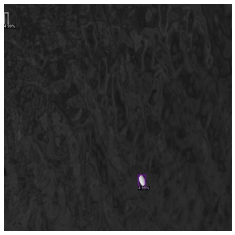
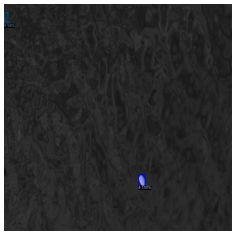
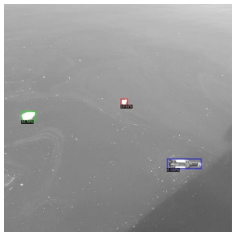
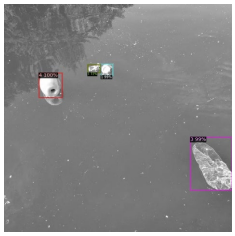
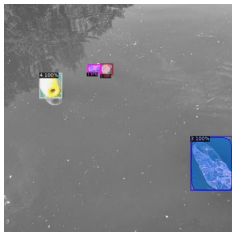
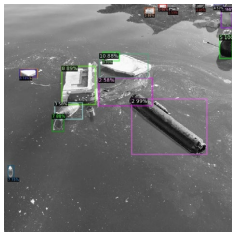
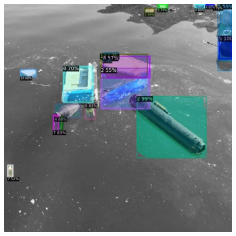
Figure 4. Example detection results on the test set
图 4. 测试集检测结果示例

对于大尺寸目标(如泡沫箱、塑料浮标)，模型给出高置信度(>0.9)检测结果；对中等尺寸目标(如塑料瓶)能准确定位；对小尺寸和不规则目标(如泡沫碎片、绳索)检测能力有效但边界精度有所下降。Mask R-CNN 与 Faster R-CNN 的检测效果对比见表 8。

Table 8. Detection comparison between Mask R-CNN and Faster R-CNN
表 8. Mask R-CNN 与 Faster R-CNN 检测效果对比

测试图像	Faster R-CNN	Mask R-CNN
密集场景		

续表

稀疏目标		
碎玻璃瓶		
金属塑料		
渔网绳索		

各类别检测 AP 明细见表 9。

Table 9. Per-class detection AP details
表 9. 各类别检测 AP 明细

Class	Train Inst.	AP_bbox	AP_mask	Note
Glass	2,147	47.28	45.55	
Metal	3,325	60.62	59.71	
Net	2,762	42.61	40.18	
PET_Bottle	6,550	59.17	57.83	
Plastic_Buoy	13,747	66.44	64.66	Most samples
Plastic_Buoy_China	885	68.45	67.28	
Plastic_ETC	1,737	42.72	41.95	
Rope	4,805	30.16	27.65	Hardest
Styrofoam_Box	475	50.47	50.14	Fewest samples
Styrofoam_Buoy	561	67.63	65.99	
Styrofoam_Piece	2,377	52.94	51.59	

错误分析：主要错误类型包括：(1) 漏检——重度遮挡和小尺寸目标漏检率较高；(2) 误检——水面波纹和泡沫浪花等易被误识别；(3) 类别混淆——同类不同型之间(如 Plastic_Buoy 与 Plastic_Buoy_China)的区分存在困难。

3.5. 路径规划性能分析

从表 10 可以得出：(1) 计算效率高——200 目标下总耗时(贪心初始化 + 2-opt 优化)约 181.8 ms，其中 2-opt 优化阶段耗时约 99.5 ms；(2) 2-opt 效果显著——改善率随目标数量增大，在 N = 100 时达到饱和约 10.5%；(3) 路径长度呈亚线性增长，算法在大规模场景下保持良好规划质量；(4) 后续工作需在更多随机初始条件下进行统计测试，以进一步验证算法的稳定性。

Table 10. Benchmark results of the path planning algorithm
表 10. 路径规划算法性能基准测试

Targets	Greedy/ms	2-opt/ms	Total/ms	Opt. Length	Improve/%
5	<0.1	<0.1	<0.1	178.3	2.2
10	0.1	0.1	0.2	250.3	4.3
20	0.4	0.5	0.9	375.4	6.7
50	3.5	4.3	7.8	586.9	8.8
100	18.6	22.9	41.5	822.6	10.5
200	82.3	99.5	181.8	1,133.9	10.5

路径规划算法的性能对比结果如图 5 所示。

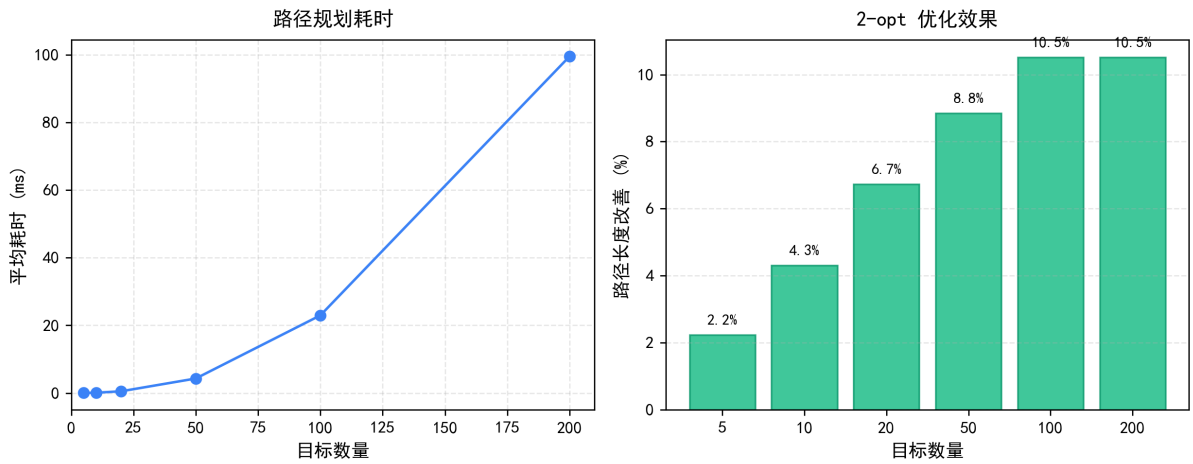


Figure 5. Benchmark results of the path planning algorithm
图 5. 路径规划算法基准测试

3.6. 系统功能验证

系统 Web 端和桌面端运行界面分别如图 6 和图 7 所示。

Web 端采用毛玻璃设计，支持图像上传、模型切换、检测与路径规划结果展示；桌面端支持三视图切换、统计面板、类别分布柱状图、路径可视化和检测详情列表。系统在 GPU 环境下端到端处理时间不

超过 5 秒。

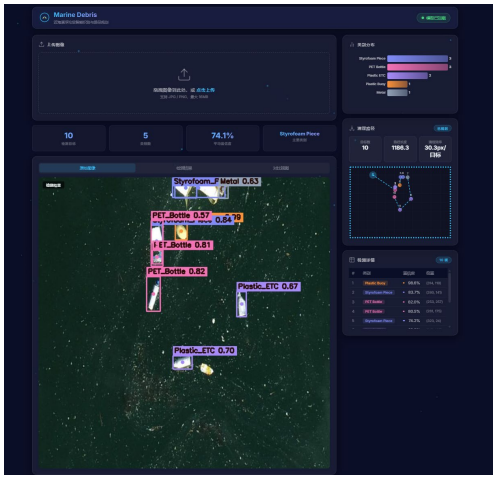


Figure 6. System interface—Web client
图 6. 系统运行界面(Web 端)

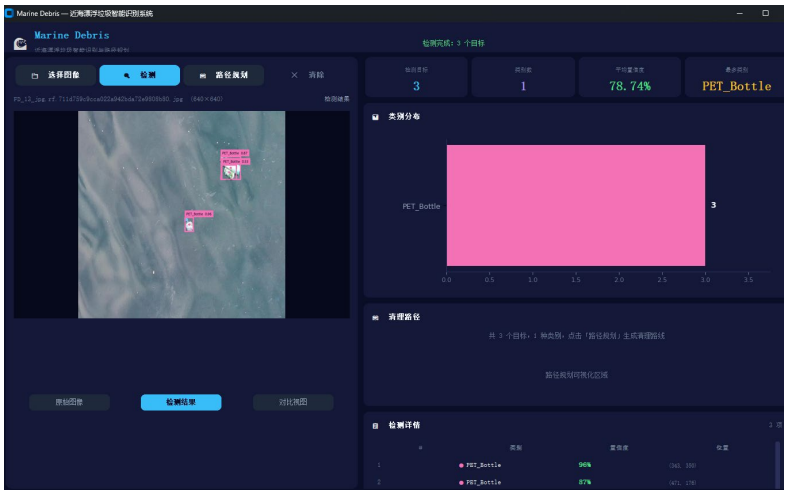


Figure 7. System interface—Desktop client
图 7. 系统运行界面(桌面端)

4. 结论

本文针对近海漂浮垃圾清理的实际需求，设计并实现了一套基于 Mask R-CNN 的智能检测与清理路径规划系统。该系统通过数据采集层、模型推理层、路径规划层和可视化展示层的分层架构，实现了从图像输入、自动检测到路径规划的全流程智能化闭环。系统提供了 Web 端和桌面端双模态交互界面，在 GPU 环境下端到端处理时间不超过 5 秒，能够完成静态图像条件下的垃圾目标检测、实例区域分割和路径规划结果展示，为后续面向真实水域场景的动态感知与清理系统开发提供了原型基础。

本研究具有一定的社会应用价值。近海漂浮垃圾污染已成为全球海洋生态环境治理中的重要问题，传统人工打捞和机械收集方式存在效率低、成本高、风险大等不足。本文所提出的智能检测与路径规划方法，能够在静态图像条件下实现近海漂浮垃圾目标识别、实例区域分割和清理路径生成，为提升垃圾清理作业的信息化和自动化水平提供了可行的技术思路。考虑到当前系统仍主要面向静态图像实验环境，

尚未完整解决动态视频流目标跟踪、图像坐标到真实世界坐标映射、复杂海况约束建模等问题，因此本文成果更适合作为近海漂浮垃圾自动化清理系统的原型验证基础，而非直接等同于完整工程部署方案。

4.1. 当前系统的局限性

尽管本文系统在静态图像检测与路径规划方面取得了较好效果，但若需在实际清理作业中部署，还存在以下关键局限：

(1) 静态图像与动态视频的鸿沟。当前系统基于单帧静态图像进行检测与规划，而真实场景是连续的视频流，目标在帧间存在位移、形变和遮挡。需要引入多目标跟踪算法(如 DeepSORT)实现跨帧目标关联和实时路径调整。

(2) 图像坐标与真实世界坐标的映射缺失。当前路径规划在像素坐标系下进行，路径长度缺乏物理意义。实际清理船需要世界坐标(经纬度或 ENU 坐标)导引航行，这需要无人机飞行高度、云台姿态角、相机内参等传感器数据，并通过单应矩阵标定实现坐标转换。

(3) 环境约束建模缺失。当前路径规划仅以欧氏距离最短为目标，未考虑真实水域中的障碍物、洋流方向和船舶运动学约束(最小转弯半径、最大航速)，规划路径在复杂海况下可能不可执行。

4.2. 从静态图像到动态视频流的技术路线

针对上述局限性，本文提出三阶段技术演进路线：

阶段一：目标检测与跟踪集成。引入 DeepSORT [16]多目标跟踪算法，以 Mask R-CNN 的检测结果作为观测输入，利用卡尔曼滤波器预测目标运动状态，通过匈牙利算法对检测与已有轨迹进行级联匹配。DeepSORT 仅增加约 10-15 ms/帧的计算开销，不影响系统实时性。

阶段二：坐标系统标定与真实世界映射。记录无人机 GPS 坐标、飞行高度及云台姿态角；通过相机内参标定获取焦距和畸变系数；基于针孔相机模型与海面平面假设，利用单应矩阵 H 将图像坐标(u, v)映射为世界坐标(X, Y)。单应矩阵可由清理船甲板上的固定标记点(已知差分 GPS 坐标)标定求解。

阶段三：多约束动态路径重规划。升级为基于事件的触发式重规划策略(如每 30 秒定时触发或目标显著位移时触发)；在目标函数中加入洋流补偿项、障碍物代价和转向代价；根据目标类型和回收价值分配清理优先级权重，使用带约束的 A*或 RRT*算法替代当前纯欧氏距离贪心方法。

上述三阶段路线的逐步实现将弥合当前原型系统与实际清理作业部署之间的关键差距。本文的静态图像检测系统为阶段一提供了高质量的感知基础，后续工作将按照该技术路线图持续推进。

4.3. 其他研究方向

下一步工作：(1) 引入视频流目标跟踪实现动态路径重规划；(2) 结合真实水域环境建立多约束路径规划模型(障碍物避让、洋流优化、优先级排序)；(3) 通过骨干网络轻量化替换(如 MobileNet)、模型量化和知识蒸馏等技术实现边缘端部署。

基金项目

上海海事大学 2025 市级大学生创新创业训练计划项目(项目编号：S20250605)。

参考文献

[1] Jambeck, J.R., Geyer, R., Wilcox, C., Siegler, T.R., Perryman, M., Andrady, A., *et al*. (2015) Plastic Waste Inputs from Land into the Ocean. *Science*, **347**, 768-771. https://doi.org/10.1126/science.1260352

[2] Jia, T., Kapelan, Z., de Vries, R., Vriend, P., Peereboom, E.C., Okkerman, I., *et al*. (2023) Deep Learning for Detecting

Macroplastic Litter in Water Bodies: A Review. *Water Research*, **231**, Article 119632. https://doi.org/10.1016/j.watres.2023.119632

[3] Fallati, L., Polidori, A., Salvatore, C., Saponari, L., Savini, A. and Galli, P. (2019) Anthropogenic Marine Debris Assessment with Unmanned Aerial Vehicle Imagery and Deep Learning: A Case Study along the Beaches of the Republic of Maldives. *Science of the Total Environment*, **693**, Article 133581. https://doi.org/10.1016/j.scitotenv.2019.133581

[4] Ren, S., He, K., Girshick, R. and Sun, J. (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1137-1149. https://doi.org/10.1109/tpami.2016.2577031

[5] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 779-788. https://doi.org/10.1109/cvpr.2016.91

[6] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., *et al*. (2016) SSD: Single Shot Multibox Detector. In: *Lecture Notes in Computer Science*, Springer, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2

[7] Córdova, M., Pinto, A., Hellevik, C.C., Alaliyat, S.A., Hameed, I.A., Pedrini, H., *et al*. (2022) Litter Detection with Deep Learning: A Comparative Study. *Sensors*, **22**, Article 548. https://doi.org/10.3390/s22020548

[8] Yang, X., Zhao, J., Zhao, L., Zhang, H., Li, L., Ji, Z., *et al*. (2022) Detection of River Floating Garbage Based on Improved Yolov5. *Mathematics*, **10**, Article 4366. https://doi.org/10.3390/math10224366

[9] Croes, G.A. (1958) A Method for Solving Traveling-Salesman Problems. *Operations Research*, **6**, 791-812. https://doi.org/10.1287/opre.6.6.791

[10] AIproject (2025) AIProject Dataset. Open Source Dataset. Roboflow Universe. Roboflow. https://universe.roboflow.com/aiproject-rqjzf/aiproject-ryolj

[11] He, K., Gkioxari, G., Dollar, P. and Girshick, R. (2017) Mask R-CNN. 2017 *IEEE International Conference on Computer Vision* (*ICCV*), Venice, 22-29 October 2017, 2961-2969. https://doi.org/10.1109/iccv.2017.322

[12] Wu, Y., Kirillov, A., Massa, F., *et al*. (2019) Detectron2. https://github.com/facebookresearch/detectron2

[13] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 770-778. https://doi.org/10.1109/cvpr.2016.90

[14] Lin, T.Y., Dollar, P., Girshick, R., He, K., Hariharan, B. and Belongie, S. (2017) Feature Pyramid Networks for Object Detection. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 2117-2125. https://doi.org/10.1109/cvpr.2017.10

[15] Lin, T.Y., Goyal, P., Girshick, R., He, K. and Dollar, P. (2017) Focal Loss for Dense Object Detection. 2017 *IEEE International Conference on Computer Vision* (*ICCV*), Venice, 22-29 October 2017, 2980-2988. https://doi.org/10.1109/iccv.2017.324

[16] Wojke, N., Bewley, A. and Paulus, D. (2017) Simple Online and Realtime Tracking with a Deep Association Metric. 2017 *IEEE International Conference on Image Processing* (*ICIP*), Beijing, 17-20 September 2017, 3645-3649. https://doi.org/10.1109/icip.2017.8296962