

基于改进Faster R-CNN模型的水面漂浮物检测方法

黄海源, 赵子豪*, 张海刚, 薛元飞

深圳职业技术学院粤港澳大湾区人工智能应用技术研究院, 广东 深圳

收稿日期: 2021年11月27日; 录用日期: 2021年12月22日; 发布日期: 2021年12月29日

摘要

水面环境治理长期以来都是生态环保的一项重要工作。然而, 人工清理水面垃圾的方式难以满足实际工作需求。近年来, 深度学习驱动的目标检测算法被成功应用在诸多领域。为解决传统检测方法效率低的问题, 本文提出了一种水面漂浮物检测方法。该方法以Faster R-CNN模型为基础, 对其主干网络做了改进。在自建的小型水面漂浮物数据集上进行实验, 模型识别精度达到77.02%, 相较于其它模型有至少2.56%的提升。此外, 大量对比实验表明, 该模型具有良好的检测性能, 基本满足实际需求。

关键词

水面漂浮物, 目标检测, 卷积神经网络, Faster R-CNN

River Drifting Garbage Detection Based on the Improved Faster R-CNN

Haiyuan Huang, Zihao Zhao*, Haigang Zhang, Yuanfei Xue

Institute of Applied Artificial Intelligence of the Guangdong-Hong Kong-Macao Greater Bay Area, Shenzhen Polytechnic, Shenzhen Guangdong

Received: Nov. 27th, 2021; accepted: Dec. 22nd, 2021; published: Dec. 29th, 2021

Abstract

River governance has always been a significant task of ecological environmental protection. However, manual cleaning for river drifting garbage is difficult to meet the actual work needs. In recent years, object detection algorithms have been successfully applied in many fields. To solve the problem of low efficiency of traditional methods, this paper proposes a detection method for river

*通讯作者。

文章引用: 黄海源, 赵子豪, 张海刚, 薛元飞. 基于改进 Faster R-CNN 模型的水面漂浮物检测方法[J]. 计算机科学与应用, 2021, 11(12): 3108-3116. DOI: 10.12677/csa.2021.1112314

drifting garbage. This method is based on the Faster R-CNN model and improves its backbone network. The experiments are performed on the self-built small drifting garbage data set, and the model recognition accuracy reaches 77.02%, which is 2.56% higher than other classical models. Finally, some comparison experiments show that the proposed model has good detection performance and basically meets the actual needs.

Keywords

River Drifting Garbage, Object Detection, Convolutional Neural Network, Faster R-CNN

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

水面环境治理长期以来都是我国国家生态环境部和社会环保组织高度重视的一项环保工作。据《全国水利普查公报》统计,我国河流总长度为 150.85 万公里,湖泊总面积为 7.80 万平方公里。传统的漂浮物清理方式是利用清理船在各段水域巡视,人工搜集水面漂浮物。清理船出巡由专员负责,且船上的环保人员在清理前需要预先定位漂浮物。在漂浮物较多或是面积较大的水域,目前的清理方式仍无法实现机械化与无人化[1]。此外,由于水上作业易受天气影响,清理工作存在较大不确定因素。从质量和效率两方面来看,传统的清理方式难以满足如今环保工作的迫切需求。

随着计算机视觉技术的发展,许多优秀的神经网络模型随之而生。基于卷积神经网络(Convolutional Neural Network, CNN) [2]的目标检测算法在近十年也开始崭露头角。计算机视觉目标检测模型辅以监控系统,在未来将成为实时监测水面环境情况的高效工具。现阶段,目标检测算法主要分为一步检测(one-stage)和两步检测(two-stage)。基于 one-stage 的代表性算法,适合对实时性要求较高的工业环境;基于 two-stage 的算法,适合对准确率要求较高的工业环境。近年来,以上算法所实现的目标检测模型在国内已得到广泛应用。2017 年,曹诗雨等[3]利用 Fast R-CNN 算法实现的车辆目标检测模型,检测效果优良。2019 年,刘等[4]利用改进的区域卷积神经网络 Faster R-CNN 算法识别空中飞行物,模型各类别 P-R 曲线积分值均值(mean Average Precision, mAP)达 75.6%。同年,雷李义[5]利用自制小型水面漂浮物数据集,对比了 Faster R-CNN、R-FCN [6]和单步多尺寸检测器(Single Shot MultiBox Detector, SSD) [7]三种目标检测算法模型,证明了 Faster R-CNN 在检测水面漂浮物时的泛化性能优于其它算法。相较于车辆、飞行物等固定目标,水面漂浮物形态多样,易受水流及光照影响。要对其实现有效检测,仍需对模型做进一步改进。

本文构建了一种改进的 Faster R-CNN 模型。在自建的小型水面漂浮物数据集上,本文进行了模型性能对比实验,改进的模型实现了 77.02%的 mAP,相较于其它经典模型有所提升。最后,本文基于改进模型在多种场景中做了漂浮物检测实验,均取得较好的检测结果。这进一步表明了本文所提出的模型具有较好的检测性能与泛化能力,基本能够满足落地需求。

2. 目标检测相关算法

2.1. SSD

SSD 是刘等人在 2016 年提出的一种 one-stage 目标检测算法。SSD 采用 VGG16 [8]的特征提取结构,

并利用池化层缩放特征图,使得特征图尺寸在训练中有多种变化。这种特殊设计使得 SSD 在检测小目标物时精度较高,也是其性能优于其它算法的重要原因。基于以上结构的 SSD 算法相对容易实现,检测性能良好,检测速度可达到每秒 46 帧(Frames Per Second, FPS)。

2.2. YOLOv3

YOLOv3 是 Redmon [9]基于 YOLOv1 [10]和 YOLOv2 [11]提出的改进版 one-stage 算法。在主干网络(backbone)上,YOLOv3 使用更深层的 Darknet53 网络,并去除原有的池化层,转而使用多个卷积核降维,并利用分批标准化(Batch Normalization, BN) [12]算法使得训练更易收敛。因网络结构加深,虽然 YOLOv3 的推理速度有所下降,但检测精度显著提升,且推理速度依然能满足实时检测的需求。

2.3. YOLOv4

YOLOv4 是 Bochkovskiy [13]于 2020 年发布的 YOLO 系列算法。Bochkovskiy 在 YOLOv3 框架中做了大量改进。首先,使用 Mosaic 数据增强方法处理数据;其次,将主干网络替换为 CSPDarknet53,并在特征提取网络上加入了 SPP [14]等算法;最后,使用完整交并比(Complete Intersection over Union, CIoU)作为回归损失。比起同期的大多数目标检测算法,YOLOv4 的各方面性能都更为优异。

2.4. Faster R-CNN

Faster R-CNN 是 Girshick [15]等基于 R-CNN、Fast R-CNN 改进的 two-stage 目标检测算法。R-CNN [16]通过选择性搜索(Selective Search)生成候选框,并利用支持向量机(Support Vector Machine, SVM)分类特征向量和边框回归(Bounding Box Regression)调整预测边框。因 R-CNN 提取特征较为费时,Girshick 在 2015 年提出 Fast R-CNN 算法[17],对原有网络结构做出调整:直接将图像输入网络生成特征图,利用特征矩阵分类和边界框预测。以上调整使得模型训练速度大幅提高,检测精度亦有明显提升。

2016 年, Girshick 提出了 Faster R-CNN 框架。Girshick 使用区域候选网络(Region Proposal Network, RPN)代替费时低效的 Selective Search 算法生成候选框。Faster R-CNN 将候选框投影到特征图上用于分类和边界框预测,并对 RPN 和 Fast R-CNN 两部分网络进行交替训练。使用 RPN 网络后,相较于最初的 R-CNN, Faster R-CNN 的训练速度从每张图片 47 s 优化到 1 s 以内。因使用 RPN 减少了大量不必要的重复运算, Faster R-CNN 算法在训练效率和检测精度上都有较大提升。

3. 本文所提出的方法

3.1. 模型构建

在河道场景中,水面漂浮物大小、水波以及气候等因素都会影响模型的检测结果,故本方法选用检测精度表现更好的 Faster R-CNN 算法。同时,由于水纹波动及光照影响,水面漂浮物检测相较传统的目标检测任务更为困难。此外,传统的 Faster R-CNN 算法在数据集规模较小时特征提取能力受限,较难满足实际应用需求。

针对上述问题,本方法改进了 Faster R-CNN 原有网络结构,使得改进模型能在同等数据集规模下,提取更多特征用于学习。本方法构建改进的 Faster R-CNN 模型如图 1 所示,首先对训练集图像进行预处理,包括图像翻转、去噪、高斯滤波等;并在 ResNet-50 [18]网络框架中增加 9 个网络模块,每个模块包括 3 个卷积层、3 个 BN 层、1 个 ReLU 函数激活层。以上改进使得模型的特征提取能力明显提升。通过每次训练前的随机图像预处理,原数据集的训练集规模成倍扩大,使得模型可以在更多场景中训练,以增强其泛化性能。改进后的模型解决了特征提取困难的问题,小数据集也能提供模型更多的特征用于学习。

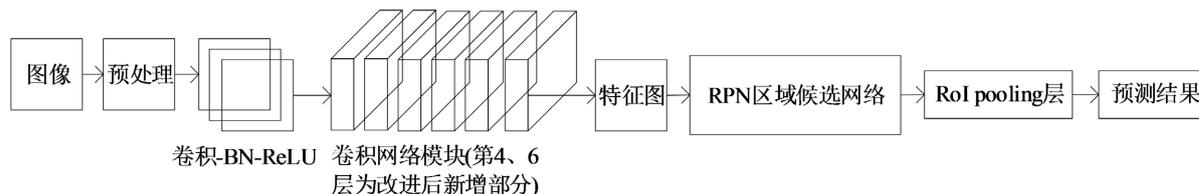


Figure 1. The improved Faster R-CNN model structure

图 1. 改进的 Faster R-CNN 模型结构

3.2. 模型性能评估

3.2.1. 数据集

本文所构建的水面漂浮物数据集由 680 张水面图像构成，涵盖江河、湖泊、海面、水库等多种环境。构建的数据集在保证图像质量的同时尽可能选择多样化的场景数据，以训练出泛化性能更强的模型。图像原始尺寸为 1280 × 960，在预处理时统一缩放为 600 × 800。本文使用 labelImg 软件标注图中漂浮物的边框信息，保存为 XML 格式。本次实验按约为 7:2:1 的比例划分训练集、验证集和测试集，并将图片名称及路径保存在 Main 目录下的 train.txt、val.txt 和 test.txt 文本文件中。构建的数据集目录树如图 2 所示。

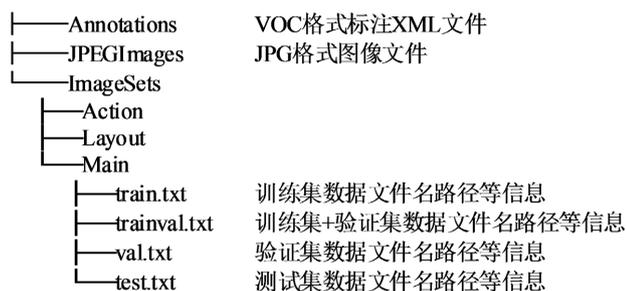


Figure 2. Dataset catalog tree

图 2. 数据集目录树

3.2.2. 模型评估

表 1 为不同参数设置下取得的 mAP 值，在改进模型的调参实验中，精度最优的参数组合为表 1 首行。各模型的训练损失多在迭代次数(epoch)为 50 至 100 间收敛，本文通过模型每次训练反馈的损失和精度调整 epoch，改进模型的 mAP 值在 epoch 为 75 时表现最好；学习率的调整范围在 1e-3 至 1e-2 之间，对学习率的调整主要参考主流 CNN 算法的实验经验，改进模型在学习率为 5e-3 时训练损失收敛稳定，过小则收敛较慢，过大则难以收敛。通过多次实验，得到表 1 中多种参数组合所训练模型的 mAP 值。

Table 1. Model and parameter comparison experimental results

表 1. 模型及参数对比实验结果

	迭代次数	学习率	优化器	mAP
改进 Faster R-CNN	75	5e-3	SGD	77.02%
改进 Faster R-CNN	75	5e-3	Adagrad	64.23%
改进 Faster R-CNN	50	5e-3	SGD	75.42%
改进 Faster R-CNN	75	1e-3	SGD	50.87%
改进 Faster R-CNN	75	1e-2	SGD	71.07%

Continued

Faster R-CNN	75	5e-3	SGD	71.20%
Faster R-CNN	75	1e-2	SGD	69.71%
Faster R-CNN	75	1e-3	SGD	46.44%
Faster R-CNN	50	5e-3	SGD	74.46%
SGD	75	5e-4	Adam	61.91%
YOLOv3	75	1e-3	Adam	63.00%
YOLOv4	100	1e-3	Adam	64.69%

此外，基于最佳参数组合，本文还对比了几种模型的其他性能指标，如精确率、召回率及 FPS。如表 2 所示，在本实验中，SSD 模型在 FPS 值上表现十分优异，64.43 FPS 完全能够满足实时检测的需求，因其网络结构简单，SSD 模型的 mAP 值稍逊其它模型。YOLO 系列的两个模型总体表现均衡，YOLOv4 相比于 YOLOv3，在模型结构上做出了大量改进和优化，使得 mAP 值有近 2 个百分点的提升，相对地导致 FPS 值有所下降。改进 Faster R-CNN 模型的 FPS 值相较于其它模型略显不足；但其 mAP 值高于其它模型，且模型在测试中表现出了较好的泛化性能，整体性能及检测精度优于其它模型，同时也有效地解决了本文提出的相关问题。

Table 2. Comparison of evaluation metrics across models

表 2. 各模型评估指标的比较

	精确率	召回率	FPS	mAP
改进 Faster R-CNN	89.90%	78.07%	6.38	77.02%
Faster R-CNN	91.58%	76.32%	6.52	74.46%
SSD	53.57%	65.79%	64.43	61.91%
YOLOv3	71.43%	61.40%	36.58	63.00%
YOLOv4	65.22%	67.59%	27.66	64.69%

4. 场景对比实验

考虑到水面环境复杂多样，本文基于多种不同河道场景进行对比实验，以检验所提出的模型在数据集规模较小、环境较为复杂的情况下是否具有良好的检测性能。实验选择的变量包括目标物尺寸大小、目标物与拍摄位置距离远近、水波水色及水流速度等，选择标准参照变量的代表性及其是否能体现模型的优缺点。

第一个实验对尺寸较大的物体进行检测。这是因为水面漂浮物大量聚集时，待检测目标体积较大，仅能识别小尺寸物体的模型无法适应所有场景。实验目的在于评估模型对尺寸变量的泛化性。

实验结果见图 3，除 YOLOv3 外，各模型均能实现基本的检测功能。改进的 Faster R-CNN 模型对较大尺寸物体的检测性能卓越，而训练集多为小尺寸物体，可见改进的 Faster R-CNN 模型对物体尺寸有不错的泛化性。再观察其它 3 种模型的实验结果，Faster R-CNN 模型检测性能较好，但受水色影响出现了误检现象；YOLOv3 模型对大尺寸物体的检测精度较高，对中等尺寸物体检测精度偏低；YOLOv4 模型与改进的 Faster R-CNN 模型性能相差甚微，但置信度偏低。

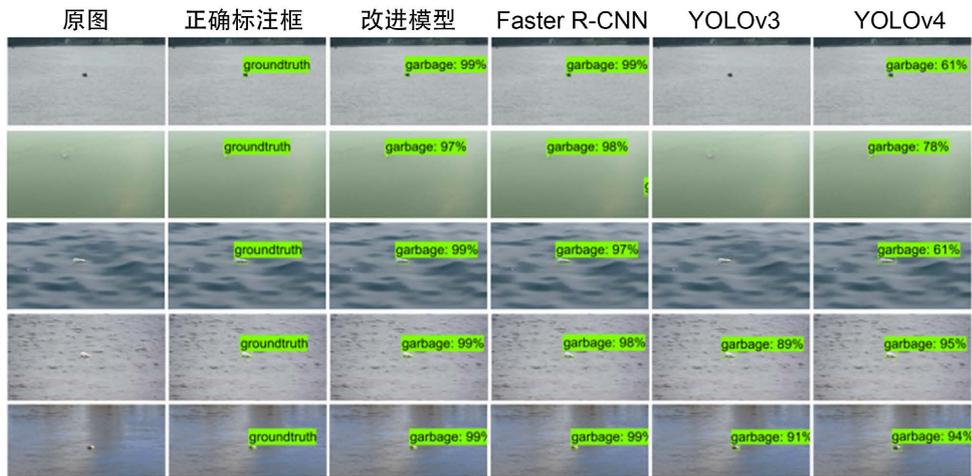


Figure 3. Comparative experimental results of large-size drifting object detection
图 3. 大尺寸漂浮物检测对比实验结果

第二个实验对尺寸较小的物体进行检测。因训练集多为小尺寸物体，故本实验对模型检测小尺寸物体的性能要求较高。实验目的在于评估模型的特征提取能力。

实验结果如图 4 所示，各模型均实现了对较小尺寸物体的基本检测功能。改进的 Faster R-CNN 模型对小尺寸物体的检测性能十分优异，且受物体颜色变化影响小；Faster R-CNN 模型与改进模型的性能差异较小；YOLOv3 与 YOLOv4 同样表现良好，但在不同颜色物体的置信度上低于改进模型。

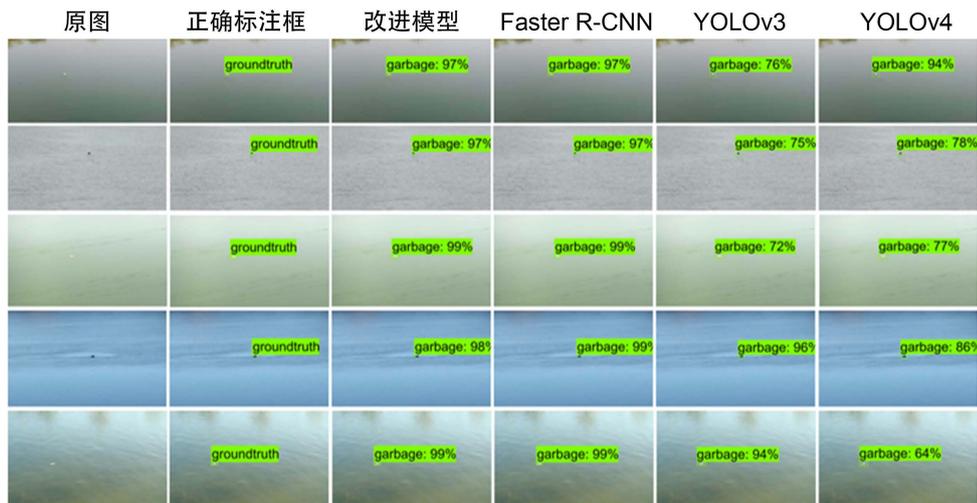


Figure 4. Comparative experimental results of small-size drifting object detection
图 4. 小尺寸漂浮物检测对比实验结果

第三个实验对远距离物体进行检测对比。水域周边通常有定距安装的监控设施，若模型能提前定位检测较远物体，清理工作能更有条不紊地进行，同时能减少意外发生的概率，节省安置监控设施的成本。

实验结果如图 5 所示，除本文提出的改进模型外，其余模型对第一张图片的检测效果较差。改进的 Faster R-CNN 模型检测性能优异，置信度在水色与物体本色比较接近时略受影响；Faster R-CNN 模型检测性能较好，但在水色复杂时出现了漏检现象，泛化性一般；YOLO 系列算法检测性能一般，同样受到了水色影响，且总体置信度偏低。

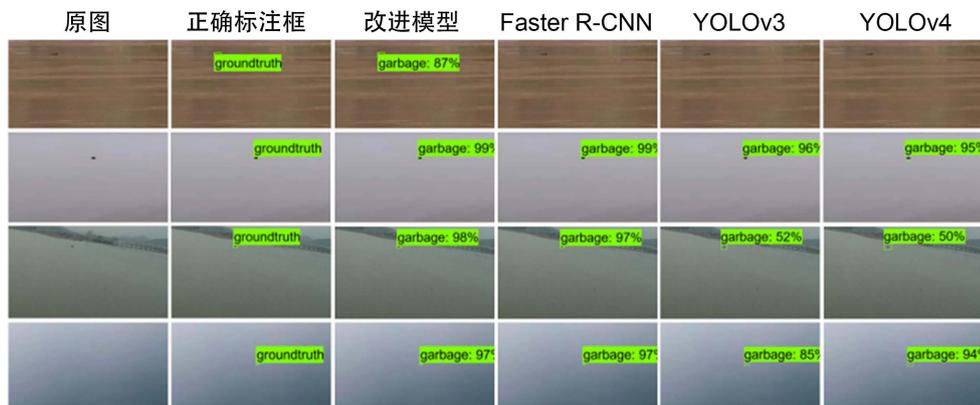


Figure 5. Comparative experimental results of long-range drifting object detection

图 5. 远距离漂浮物检测对比实验结果

第四个实验对近距离物体进行检测，目的在于评估模型的核心检测性能。因近距离物体的特征更容易提取，模型受环境影响的概率较低，相比在其它场景的检测性能应该更强。通常情况下，近距离检测性能优异的模型在多数场景同样表现良好。

实验结果如图 6 所示，多数模型对近距离较大尺寸的物体能有效检测。改进的 Faster R-CNN 模型检测性能最佳，在近距离场景对不同大小的物体都能定位；Faster R-CNN 模型检测性能一般，无法定位近距离小物体；YOLO 系列算法检测性能一般，仅能检测近距离较大物体，对小物体检测精度极低。

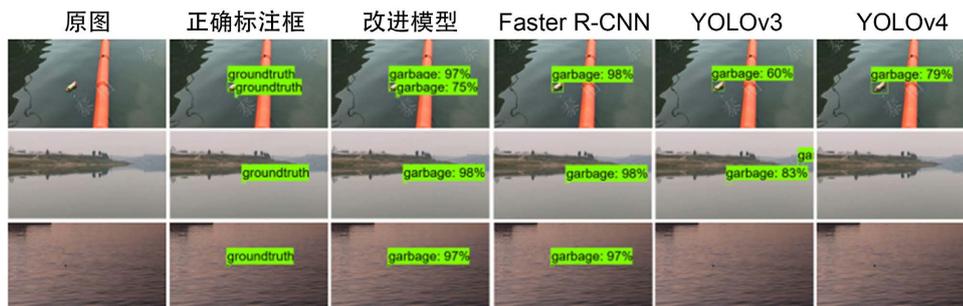


Figure 6. Comparative experimental results of short-range drifting object detection

图 6. 近距离漂浮物检测对比实验结果

第五个实验基于多种水波震荡时的场景进行检测。水波是水域生态最常见的环境变量之一，且对模型性能影响较大。受水波震荡影响越小，模型在实际应用中表现越好，实验目的为考察各模型的泛化性能。

实验结果如图 7 所示，除 YOLO 系列模型外，其它模型基本检测性能良好，略受环境影响。改进的 Faster R-CNN 模型检测性能十分优异，受水波影响极小；Faster R-CNN 模型表现较好，但在面对物体较多的场景，易出现漏检现象；YOLOv3 模型检测性能一般，受水波影响较大，难以定位；YOLOv4 模型表现稍好，但检测性能不稳定，略受水波因素影响。

第六个实验在不同天气下的场景进行目标检测对比。因天气环境影响持续时间较长，而漂浮物的清理周期较短，若模型能在不同天气场景中保持良好的检测性能，意味着其落地可靠性有了基本保障。

实验结果如图 8 所示，各模型均能实现基本检测功能。改进的 Faster R-CNN 模型在雨天、阴天场景的表现依旧乐观，几乎不受雨滴影响；Faster R-CNN 模型同样表现优异；YOLOv3 模型略受阴天影响出现了漏检现象；YOLOv4 模型在雨天、阴天时检测性能较好，置信度较 Faster R-CNN 模型略低。

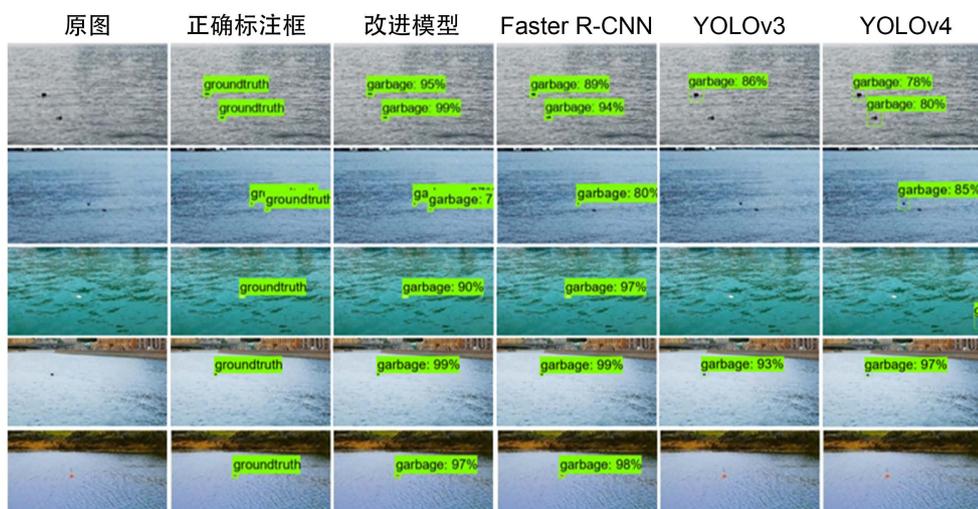


Figure 7. Experimental results of drifting object detection under different water wave scenarios
图 7. 不同水波场景下的漂浮物检测实验结果

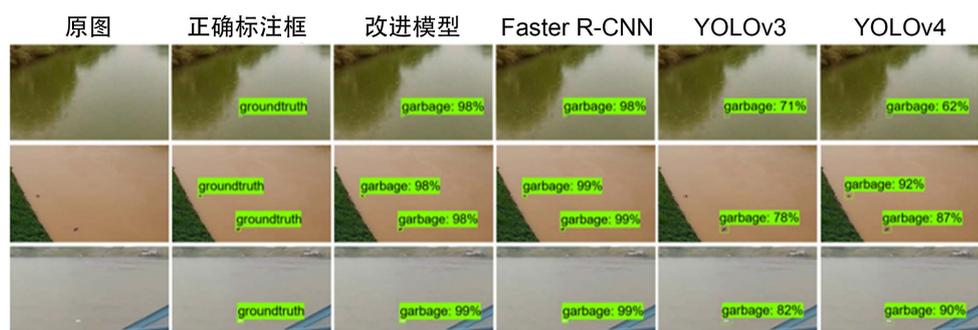


Figure 8. Experimental results of drifting object detection under different weather conditions
图 8. 不同天气下的漂浮物检测实验结果

综合分析以上实验结果，改进的 Faster R-CNN 模型总体表现最优。主干网络使用改进的 ResNet-50 相比于 VGG16 和原始 ResNet-50 有更大量级的训练参数，使得模型具有更好的特征提取和学习能力。本文提出的改进模型对大、小尺寸物体的检测性能极高，相较于 YOLO 系列模型在精度和置信度上有明显优势，相较于 Faster R-CNN 模型的泛化性能更强。而在对远、近距离物体的检测中，改进模型性能优异，相比于其它模型在精度和置信度上的表现都更为稳定。此外，改进模型受水波、水色、天气等环境因素影响较小，相比于其它模型更能适应不断变化的场景。模型在实验中的不足是检测速度偏慢，需要高性能的计算机支持。

5. 结论

针对传统人工清理水面漂浮物方式低效的问题，本文提出了基于改进 Faster R-CNN 的水面漂浮物自动检测方法。本文通过图像预处理、多次冻结训练，有效地改善了传统目标检测模型识别精度低、特征提取难的问题。同时，本文改进模型主干网络的卷积结构，优化了模型泛化性能因数据集规模较小而表现不佳的问题。大量对比实验结果表明，本文改进的目标检测模型各方面性能良好，基本能满足多数河道场景的漂浮物检测需求。针对模型检测速度偏慢的情况，后续工作的方向在于优化模型框架，深入发掘水面漂浮物的更多特征，并有针对性地提出契合的网络结构，减少检测过程的参数计算量。

致 谢

感谢深圳市科技计划资助(RCBS20200714114940262), 广东省普通高校重点领域专项(2020ZDZX3082), 广东省农村科技特派员项目(KPT20200220), 汽车新技术安徽省工程技术研究中心开放基金(QCKJ202103)的资助和支持。

基金项目

深圳市科技计划资助(RCBS20200714114940262), 广东省普通高校重点领域专项(2020ZDZX3082), 广东省农村科技特派员项目(KPT20200220), 汽车新技术安徽省工程技术研究中心开放基金(QCKJ202103)。

参考文献

- [1] 原建洋, 窦岩, 王少文, 刘赛赛. 市面上典型水面垃圾清理船对比研究[J]. 科技创新导报, 2019, 16(7): 67-68.
- [2] Gu, J., Wang, Z., Kuen, J., Ma, L. and Wang, J. (2018) Recent Advances in Convolutional Neural Networks. *Pattern Recognition*, **77**, 354-377. <https://doi.org/10.1016/j.patcog.2017.10.013>
- [3] 曹诗雨, 刘跃虎, 李辛昭. 基于 Fast R-CNN 的车辆目标检测[J]. 中国图像图形学报, 2017, 22(5): 671-677.
- [4] 刘聪聪, 应捷, 杨海马, 刘瑾, 李筠. 基于区域卷积神经网络的空中飞行物识别算法[J]. 传感器与微系统, 2021, 40(1): 110-113+117.
- [5] 雷李义. 基于深度学习的水面漂浮物目标检测及分析[D]: [硕士学位论文]. 南宁: 广西大学, 2019.
- [6] Dai, J., Li, Y., He, K. and Sun, J. (2016) Object Detection via Region-Based Fully Convolutional Networks. *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016*, Barcelona, 5-10 December 2016, 379-387.
- [7] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S.E., Fu, C.Y. and Berg, A.C. (2016) Ssd: Single Shot Multibox Detector. *European Conference on Computer Vision*, Amsterdam, Vol. 1, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2
- [8] Simonyan, K. and Zisserman, A. (2015) Very Deep Convolutional Networks for Large-Scale Image Recognition. *International Conference on Learning Representations*, San Diego, 7-9 May 2015, 1-14.
- [9] Redmon, J. and Farhadi, A. (2018) Yolo3: An Incremental Improvement.
- [10] Redmon, J., Divvala, S.K., Girshick, R.B. and Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 779-788. <https://doi.org/10.1109/CVPR.2016.91>
- [11] Redmon, J. and Farhadi, A. (2017) YOLO9000: Better, Faster, Stronger. *IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 6517-6525. <https://doi.org/10.1109/CVPR.2017.690>
- [12] Ioffe, S. and Szegedy, C. (2015) Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *International Conference on Machine Learning*, Lille, 7-9 July 2015, 448-456.
- [13] Bochkovskiy, A., Wang, C.Y. and Liao, H.Y.M. (2020) Yolov4: Optimal Speed and Accuracy of Object Detection.
- [14] He, K., Zhang, X., Ren, S. and Sun, J. (2015) Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **37**, 1904-1916. <https://doi.org/10.1109/TPAMI.2015.2389824>
- [15] Ren, S., He, K., Girshick, R.B. and Sun, J. (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **39**, 1137-1149. <https://doi.org/10.1109/TPAMI.2016.2577031>
- [16] Girshick, R.B., Donahue, J., Darrell, T. and Malik, J. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 580-587. <https://doi.org/10.1109/CVPR.2014.81>
- [17] Girshick R.B. (2015) Fast R-CNN. *IEEE International Conference on Computer Vision*, Santiago, 7-13 December 2015, 1440-1448. <https://doi.org/10.1109/ICCV.2015.169>
- [18] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 770-778. <https://doi.org/10.1109/CVPR.2016.90>