

基于改进 Faster R - CNN 的海参目标检测算法

熊海涛¹ 林琪² 宣魁¹ 葛凤丽³ 李娟¹

(1. 青岛农业大学机电工程学院, 青岛 266109;

2. 福建省海洋生物增殖与高值化利用重点实验室, 厦门 361013;

3. 青岛农业大学经济管理学院, 青岛 266109)

摘要: 海参目标检测是实现海参自动化捕捞的前提。为了解决复杂海底环境下背景和目标颜色相近以及遮挡导致的目标漏检问题, 本文在 Faster R - CNN 框架下, 提出了 Swin - RCNN 目标检测算法。该算法的骨干网络采用 Swin Transformer, 同时在结构上融入了多尺度特征提取层和实例分割功能, 提高了算法的自适应特征融合能力, 从而提高了模型在复杂环境下对不同尺寸海参的识别能力。实验结果表明: 本文方法对海参检测的平均精度均值 (mAP) 达到 94.47%, 与 Faster R - CNN、SSD、YOLO v5、YOLO v4、YOLO v3 相比分别提高 4.49、4.56、4.46、11.78、22.07 个百分点。

关键词: 海参; 目标检测; 目标识别; Faster R - CNN; Swin Transformer; 深度学习

中图分类号: TP393 **文献标识码:** A **文章编号:** 1000-1298(2022)S2-0204-06

Sea Cucumber Object Detection Algorithm Based on Improved Faster R - CNN

XIONG Haitao¹ LIN Qi² XUAN Kui¹ GE Fengli³ LI Juan¹

(1. College of Electrical and Mechanical Engineering, Qingdao Agricultural University, Qingdao 266109, China

2. Fujian Key Laboratory of Marine Aquaculture and High-value Utilization, Xiamen 361013, China

3. College of Economics and Management, Qingdao Agricultural University, Qingdao 266109, China)

Abstract: Sea cucumber object detection is the premise of realizing automatic fishing of sea cucumber. To solve the problem of missed object detection caused by occlusion and the color similarity between object and background in the complex seabed environment, Swin RCNN object detection algorithm was proposed under the framework of Faster R - CNN. The backbone network of the algorithm adopted the Swin Transformer, and the multi-dimensional feature extraction layer was integrated into the structure, which improved the adaptive feature fusion ability of the algorithm and improved the object recognition ability of the model for the different sizes of objects under occlusion in complex environments. The actual experimental results showed that the mean average precision achieved 94.47% for the detection of sea cucumbers by the proposed approach, which was increased by 4.49 percentage points, 4.56 percentage points, 4.46 percentage points, 11.78 percentage points, and 22.07 percentage points compared with Faster R - CNN, SSD, YOLO v5, YOLO v4, and YOLO v3, respectively. The research result had certain reference significance for object detection in other complex environments. Therefore, the study of sea cucumber object detection algorithm in complex seabed environment had important theoretical and application value, and also had guiding significance for intelligent identification of other marine products.

Key words: sea cucumber; object detection; object recognition; Faster R - CNN; Swin Transformer; deep learning

收稿日期: 2022-06-10 修回日期: 2022-08-15

基金项目: 国家自然科学基金项目(32073029)、山东省自然科学基金重点项目(ZR2020KC027)、山东省研究生教育质量提升计划项目(SDYJG19134)、国家留学基金项目(201908370048)和福建省海洋生物增殖与高值化利用重点实验室开放课题(2021fjscq08)

作者简介: 熊海涛(1995—),男,硕士生,主要从事机器视觉、人工智能和智能监控研究,E-mail: 2115819631@qq.com

通信作者: 李娟(1969—),女,教授,博士生导师,主要从事机器视觉和人工智能研究,E-mail: lijuan291@sina.com

0 引言

海参收获目前依靠人工潜水进行捕捞,存在人工成本高、效率低和高危险性等问题。研发海参智能捕捞是解决这一问题的有效途径和未来的发展趋势,而海参的准确检测是实现海参智能化捕捞的前提。为此,进行真实海底环境下海参目标检测算法研究不仅具有重要的理论意义,而且具有实际的应用价值。

随着人工智能的发展^[1-2],尤其是深度学习和卷积网络在计算机视觉领域的重大突破^[3-4],出现了许多里程碑式的深度学习目标检测框架^[5]。这些目标检测框架根据检测思想的不同主要分为两类^[6-7]:一类是两阶段检测框架^[8],如 RCNN 系列算法;另一类是一阶段检测框架,如 SSD^[9],YOLO^[10]系列算法。然而,无论哪种框架下的目标识别方法的研究,绝大多数是关于陆地上目标的识别算法^[11],对水下的目标检测研究相对较少,尤其对海参的目标检测算法的研究更少。PENG 等^[12]提出了一种用于水下图像中海参检测的快捷特征金字塔网络,通过引入多种快捷连接方式,提高了特征提取和多尺度特征融合的能力,mAP 达到了 91.5%,高于基准特征金字塔网络。QIU 等^[13]使用轻量化的

SSD(Single shot multi box detector)检测海参,剔除了冗余的滤波器,mAP 为 89.54%。ZENG 等^[14]通过 Faster R-CNN 模型融合对抗遮挡网络(Adversarial occlusion network, AON)实现对海参的有效识别,平均精度仅为 65.6%。LI 等^[15]提出了一种基于 Faster R-CNN 的海参识别和行为分析方法,在准确检测海参的同时绘制其行为轨迹并进行分析。

由于海底复杂多变且海参具有体色随环境变化等特点,从而导致已有算法在识别真实复杂海底环境下海参时存在识别准确率相对不高及泛化能力差的问题^[16]。为此,本文提出一种新的目标检测模型 Swin-RCNN,以解决复杂海底环境下的目标检测所存在的问题。

1 实验材料

1.1 图像采集

本文数据集采集于山东省莱州市海洋牧场(37.42°N, 120.01°E),采集地点位置及部分数据集图像如图 1 所示。数据集由潜水员利用水下相机(GoPro HERO7)拍摄的视频按帧截取获得,并以 JPG 格式存储,其中图像分辨率为 1 920 像素 × 1 440 像素。删除在视频转图像过程中存在的图像质量不佳以及图像中没有海参目标的图像,有效数

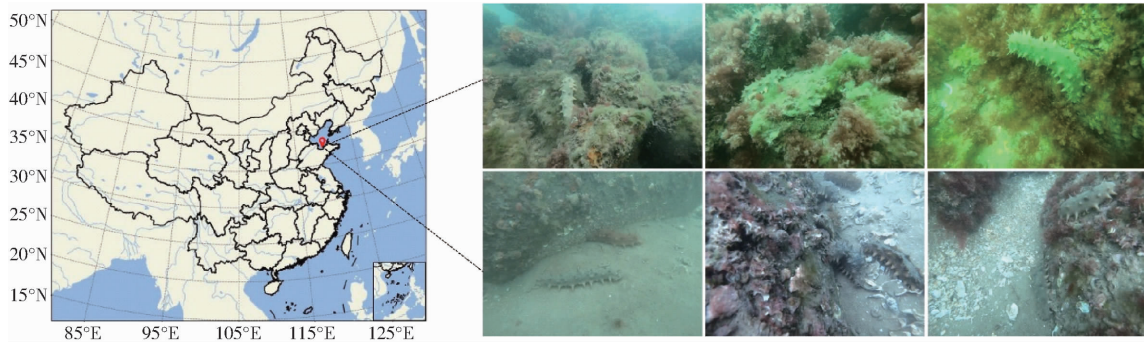


图 1 海参图像的采集地点和示例

Fig. 1 Acquisition site and examples of sea cucumber images

据集共包括 1 500 幅图像。

1.2 数据预处理

1.2.1 数据增强

为了提高模型的泛化能力以及防止模型在训练

过程中出现过拟合,本文采用数据增强的方法对数据集进行扩充。通过对比度增强、翻转 180°、颜色随机、旋转 30°的方法将数据集扩充到 2 000 幅,数据增强结果如图 2 所示。



图 2 数据增强结果示例

Fig. 2 Example of data enhancement results

1.2.2 实验对象标记

海参目标的标记使用 Labelme 标注完成,标记

过程中通过创建方框和坐标轴来人工标记海参,图 3 给出了标注的海参示例,标记完成的海参图像

文件存储类型为 JSON 格式。带注释的样本与每个图像中标记的锚框一一对应,每幅图像可以包含多个锚框,锚框的多少取决于图像中海参的数量。为了满足实验所需的要求,利用 Python 程序把标记好的图像和标签转换成 COCO2017 格式,生成 instances_train2017.json 和 instances_val2017.json 文件的同时将图像以比例 7:3 随机划分为训练集和验证集。



图3 Labelme 标记的海参示例

Fig. 3 Example of labeled sea cucumber by Labelme

1.3 实验设备

本实验在配备 Intel(R) Core(TM) i7-6800K CPU @ 3.40 GHz, RAM 16 GB 和 NVIDIA GeForce RTX 3060 12 GB 的实验平台上进行。软件环境配置为 CUDA11.1.0、CUDNN11.1 和 Python3.8.8。训练过程中采用 StepLR 机制对学习率进行更新,优化

器采用 AdamW,输入图像尺寸为 448 像素 × 448 像素,初始学习率为 0.0001,权重衰减为 0.05,批处理大小为 2,加载数据的线程数目为 2,迭代次数为 200。

2 Swin-RCNN 目标检测算法

Faster R-CNN^[17] 是一种两阶段目标检测模型,通过加入了区域候选网络(Region proposal network, RPN)^[18],弥补了 R-CNN 和 Fast R-CNN 选择性搜索找出所有的候选框而制约算法性能的问题。VGGNet、ResNet 等作为 Faster R-CNN 常用的主干网络,存在获得的特征图分辨率较低,不利于小目标及多尺度物体检测的问题。此外,Faster R-CNN 多使用 CNN 和上采样作为特征提取方法,卷积核大小及池化方法的选择会严重影响检测的精度及效率。Swin Transformer 作为主干网络不仅可以提高模型的多尺度特征提取能力,还尽可能多的保留原始像素信息,同时也减少了计算量。为此,本文采用 Faster R-CNN 架构,提出用 Swin Transformer 作为模型的骨干网络,使用多层卷积逐步提取图像的语义信息,生成不同尺寸的特征图,以提高模型的效率和对不同尺寸海参目标识别的适应性。所提出的 Swin-RCNN 网络结构如图 4 所示。

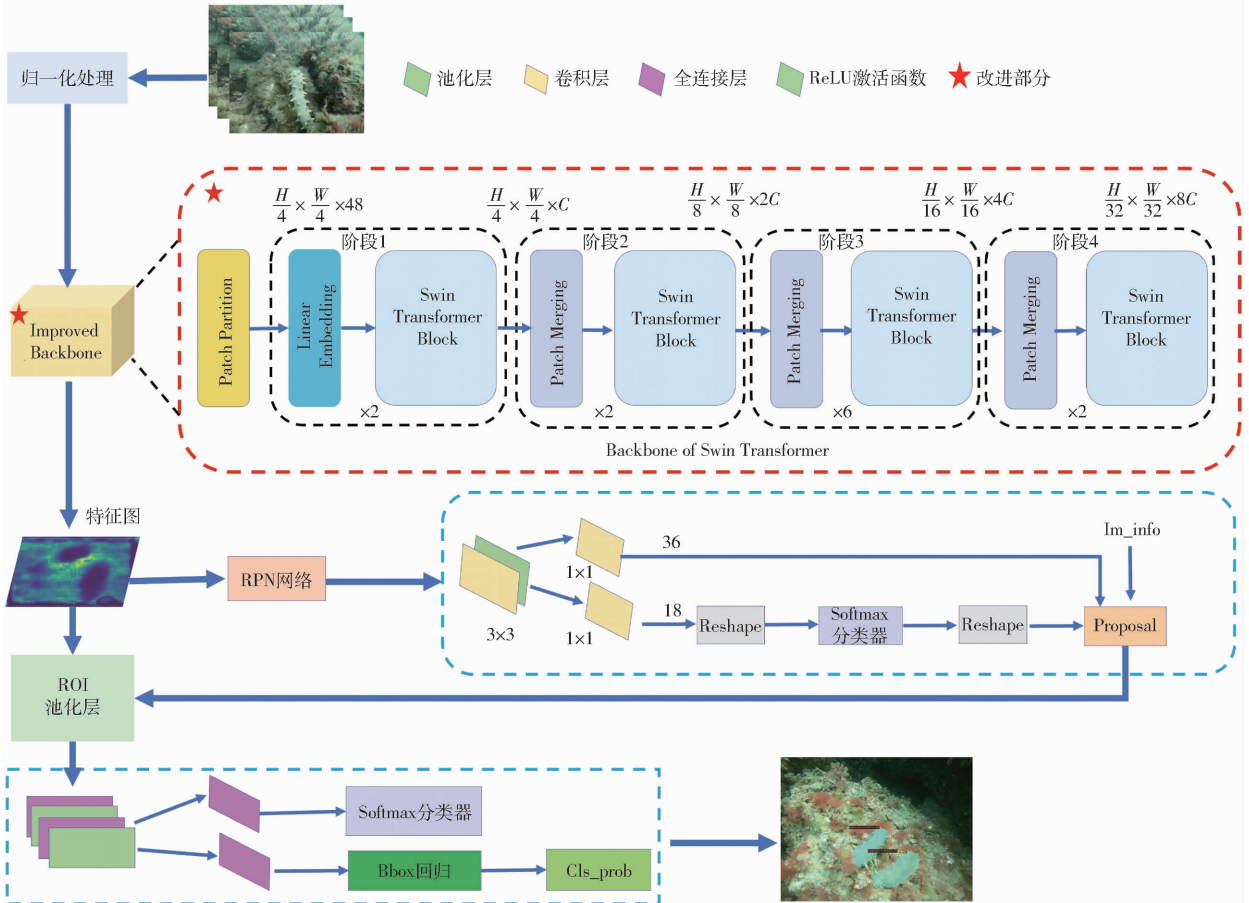


图4 Swin-RCNN 网络结构

Fig. 4 Swin-RCNN network structure diagram

在图 4 中, Swin Transformer 主干网络采取层次化的设计^[19], 一共包含 4 个阶段, 每个阶段都会缩小输入特征图的分辨率, 像 CNN 一样逐层扩大感受野。Swin Transformer 将输入的图像进行 Patch Embedding^[20], 从而将图像裁剪成多个图块并嵌入。在每个阶段里, 由 Patch Merging 和多个 Block 组成。其中 Patch Merging 模块的作用是在每个阶段开始降低图像分辨率。Block 主要由 LayerNorm、Multilayer Perceptron、Window Attention 和 Shifted Window Attention 组成^[21]。在骨干网络 Swin Transformer 结构中, 将阶段 4 输出的特征图进行上采样, 使之与阶段 3 输出的特征图尺寸相同, 进行特征拼接后输入到 Swin_res 模块中, 从而加深网络结构, 提高网络的泛化能力。同时, 由于阶段 3 位于较低层, 提取到的信息大多是细节信息, 阶段 4 提取到的则是抽象语义信息。

在 Swin Transformer 的基础上, 本文所提出的 Swin-RCNN 算法的工作流程如下:

(1) Swin-RCNN 首先从 VOC 或 COCO 数据集中获取输入图像, 并经过 Swin Transformer 提取图像特征。Swin Transformer 通过移动窗口学习目标特征, 故和基于全局的多头自注意力 (Multi-heads self-attention, MSA) 相比, 在窗口内计算自注意力使得窗口多头自注意力 (Window multi-heads self-attention, W-MSA) 能有效减少计算复杂度。Swin Transformer 通过滑动窗口获取图块以及自注意力在窗口内计算的设计使其具有了高效获取特征的能力。另外, Swin Transformer 通过移动窗口操作可以使相邻的两个窗口之间进行交互, 使上下层之间建立了跨窗口连接, 从而达到了全局建模的能力。Swin Transformer 的这些结构使得模型的计算复杂度随着图像尺寸线性增长而不是平方级增长, 能够在特别大的分辨率上进行模型训练, 解决了图像尺度不一和图像的分辨率较大时的目标识别问题。MSA 和 W-MSA 的计算复杂度分别为

$$\Omega_{MSA} = 4h\omega^2c + 2(h\omega)^2c \quad (1)$$

$$\Omega_{W-MSA} = 4h\omega c^2 + 2M^2h\omega c \quad (2)$$

式中 Ω_{MSA} ——MSA 计算复杂度

Ω_{W-MSA} ——W-MSA 计算复杂度

h ——特征图高度

ω ——特征图宽度

c ——特征图深度

M ——每个窗口尺寸

可以看出 Swin Transformer 具有更小的复杂度。

(2) Swin-RCNN 利用提取到的图像特征, 经

过 RPN 网络获取一定数量的感兴趣区域 (ROI)。然后将感兴趣区域和图像特征输入到 RoI 头部, 对这些感兴趣区域进行分类, 判断其属于什么类别, 同时对位置进行微调, 形成较精确的候选框, 从而提高了算法的检测速度。RPN 操作的总体流程如图 5 所示。

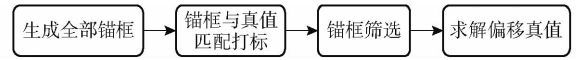


图 5 RPN 操作的总体流程图

Fig. 5 Overall flowchart of RPN operation

在训练过程中, RPN 网络的损失函数为

$$L(p_i, t_i) = \frac{1}{N_{\text{cls}}} \sum_i L_{\text{cls}}(p_i, p_i^*) + \lambda \frac{1}{N_{\text{reg}}} \sum_i p_i^* L_{\text{reg}}(t_i, t_i^*) \quad (3)$$

其中 $L_{\text{reg}}(t_i, t_i^*) = \sum_i \Phi_L(t_i - t_i^*)$

$$L_{\text{cls}}(p_i, p_i^*) = -\lg(p_i p_i^* + (1 - p_i^*)(1 - p_i))$$

$$\Phi_{L_1}(t_i, t_i^*) = \begin{cases} 0.5(t_i, t_i^*)^2 & (|t_i, t_i^*| < 1) \\ |t_i, t_i^*| - 0.5 & (|t_i, t_i^*| \geq 1) \end{cases}$$

式中 L ——RPN 网络损失值

L_{cls} ——分类层损失值

L_{reg} ——回归层损失值

i ——锚点索引

t_i ——预测边界框坐标向量

t_i^* ——真实边界框坐标向量

N_{cls} ——分类样本数

N_{reg} ——回归样本数

p_i ——目标预测概率

λ ——权重参数

Φ_{L_1} ——平滑函数

p_i^* ——锚点判别值, 当锚框为正样本时, $p_i^* = 1$, 否则 $p_i^* = 0$

(3) 感兴趣区域池化层收集输入的特征图和候选框, 综合这些信息提取固定大小的区域特征图, 送入后续的全连接层判定目标类别。最后, 利用固定大小的区域特征图计算候选框类别, 同时再次通过边界框回归获得检验框目标的最终位置。

3 实验与结果分析

3.1 模型评价指标

本文采用损失函数值和平均精度均值作为目标检测模型性能的评价指标。损失函数曲线可以直观地反映迭代次数和模型收敛性的关系。平均精度均值常被用于评价目标检测模型的检测效果, 该值越大, 平均检测精度越高, 表示模型的性能越好。

3.2 实验结果

为了验证所提出模型 Swin-RCNN 的效果, 本文将 Swin-RCNN 和原始的 Faster R-CNN 进行了比较实验, 实验结果如图 6 所示。从图 6a 可看出, Swin-RCNN 经过大约 100 次迭代后, 平均精度均值达到 94% 左右, 并且逐渐稳定在 94.97%, 比 Faster R-CNN 提高 4.49 个百分点。从图 6b 可看出, Swin-RCNN 的损失值曲线逐渐收敛, 损失值越来越小; 当模型迭代大约 30 次时, 损失值逐渐趋于稳定, 模型基本收敛; 与 Faster R-CNN 模型相比, Swin-RCNN 收敛速度更快。

为了更加直观地观察模型的海参检测效果, 本文将未参与模型训练的海参图像分别输入 Faster R-CNN 和 Swin-RCNN 模型进行检测, 结果如图 7 所示。由图 7 可看出, Swin-RCNN 模型能够更加准确地检测出复杂环境下的海参目标, 降低了复杂环境下海参漏检和误检的概率, 模型鲁棒性和泛化能力更强。另外, 本文在 Swin-RCNN 模型中加入了实例分割^[22]的功能, 在正确检测图像中目标的同时还能精确分割每个实例, 有助于解决复杂海底环境下严重遮挡导致小物体丢失全局信息的问题。

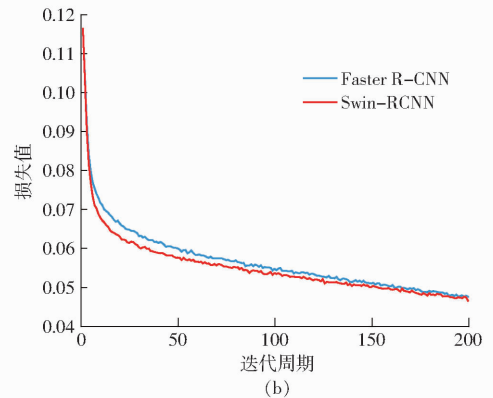
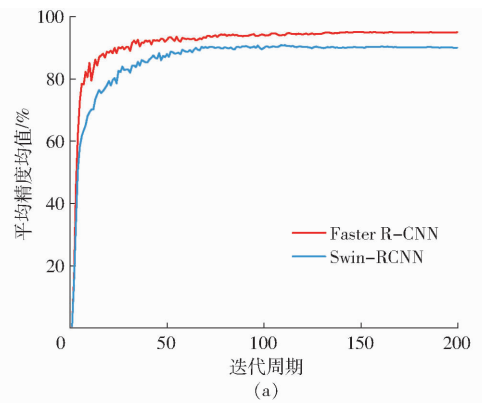


图 6 模型训练对比结果

Fig. 6 Comparative results of model training

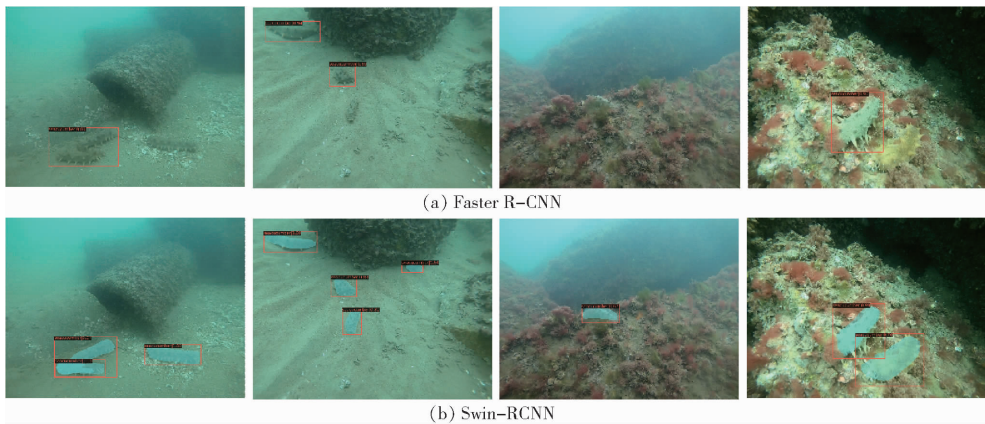


图 7 模型检测结果对比

Fig. 7 Comparative results of model detection

将 Swin-RCNN 与目前常用的目标检测模型进行了对比实验。在使用同一数据集的情况下, 对比结果如表 1 所示。由表 1 可看出, Swin-RCNN 的平均精度均值与 Faster R-RNN、SSD、YOLO v5、

YOLO v4、YOLO v3 相比, 分别提高 4.49、4.56、4.46、11.78、22.07 个百分点。

4 结束语

为了解决复杂海底环境下海参识别难的问题, 本文提出了一种新的目标检测模型 Swin-RCNN, 实现了复杂环境下不同大小海参目标的检测。通过与常用目标检测算法的对比, 验证了本文方法在复杂环境下具有较好的鲁棒性和泛化能力。本研究为复杂海底环境下生物的识别和检测提供了新的思路, 对水下智能化捕捞装置研发有一定的理论和实际意义。

表 1 与其它模型对比结果

Tab. 1 Comparison of results with other models

模型	平均精度均值/%	内存占用量/MB
SSD	89.91	90.60
YOLO v5	90.01	14.40
YOLO v4	82.69	244.00
YOLO v3	72.40	234.00
Faster R-CNN	89.98	108.00
Swin-RCNN	94.47	512.00

参 考 文 献

- [1] ZENG N, QIU H, WANG Z, et al. A new switching-delayed-PSO-based optimized SVM algorithm for diagnosis of Alzheimer's disease[J]. *Neurocomputing*, 2018, 320(3):195-202.
- [2] 赵博, 宋正河, 毛文华, 等. 基于 PSO 与 K-均值算法的农业超绿图像分割方法[J]. *农业机械学报*, 2009, 40(8): 166-169. ZHAO Bo, SONG Zhenghe, MAO Wenhua, et al. Agriculture Extra-green image segmentation based on PSO and K-means clustering[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2009, 40(8): 166-169. (in Chinese)
- [3] 赵博, 毛恩荣, 毛文华, 等. 农业车辆杂草环境下视觉导航路径识别方法[J]. *农业机械学报*, 2009, 40(增刊): 183-186. ZHAO Bo, MAO Enrong, MAO Wenhua, et al. Path recognition for vision navigation system of agricultural vehicle in weed environment[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2009, 40(Supp.): 183-186. (in Chinese)
- [4] 戚超, 左毅, 陈哲琪, 等. 基于改进 VGG16 的大米加工精度分级方法研究[J]. *农业机械学报*, 2021, 52(5): 301-307. QI Chao, ZUO Yi, CHEN Zheqi, et al. Rice processing accuracy classification method based on improved VGG16 convolution neural network[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021, 52(5): 301-307. (in Chinese)
- [5] DENG L, DU H, HAN Z. A carrot sorting system using machine vision technique[J]. *Applied Engineering in Agriculture*, 2017, 33: 149-156.
- [6] XU W, ZHAO L, LI J, et al. Detection and classification of tea buds based on deep learning[J]. *Computers and Electronics in Agriculture*, 2022, 192: 106547.
- [7] KOIRALA A, WALSH K B, WANG Z L, et al. Deep learning-method overview and review of use for fruit detection and yield estimation[J]. *Computers and Electronics in Agriculture*, 2019, 162: 219-234.
- [8] 赵世达, 王树才, 郝广钊, 等. 基于单阶段目标检测算法的羊肉多分体实时分类检测[J]. *农业机械学报*, 2022, 53(3): 400-411. ZHAO Shida, WANG Shucai, HAO Guangzhao, et al. Mutton multiparitime real-time classification and detection based on single-stage object detection algorithm[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022, 53(3): 400-411. (in Chinese)
- [9] SREELATHA R, LAKSHMI R R. Deep learning-based detection system for heavy-construction vehicles and urban traffic monitoring[J]. *International Journal of Advanced Computer Science and Applications*, 2022, 13(3): 355-361.
- [10] QI J, LIU X, LIU K, et al. An improved YOLOv5 model based on visual attention mechanism: application to recognition of tomato virus disease[J]. *Computers and Electronics in Agriculture*, 2022, 194: 106780.
- [11] 刘玲玲, 赵博, 张小超, 等. 谷物品质近红外光谱分析算法的研究[J]. *农机化研究*, 2012, 34(11): 14-19. LIU Lingling, ZHAO Bo, ZHANG Xiaochao, et al. Study on algorithm for grain quality analysis by near infrared spectroscopy [J]. *Journal of Agricultural Mechanization Research*, 2012, 34(11): 14-19. (in Chinese)
- [12] PENG F, MIAO Z, LI F, et al. S-FPN: a shortcut feature pyramid network for sea cucumber detection in underwater images [J]. *Expert Systems with Applications*, 2021, 182: 115306.
- [13] QIU Z, YAO Y, ZHONG M. Underwater sea cucumbers detection based on pruned SSD[C]//2019 IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference, 2019:738-742.
- [14] ZENG L, SUN B, ZHU D. Underwater target detection based on Faster R-CNN and adversarial occlusion network[J]. *Engineering Applications of Artificial Intelligence*, 2021, 100: 104190.
- [15] LI J, XU C, JIANG L, et al. Detection and analysis of behavior trajectory for sea cucumbers based on deep learning[J]. *IEEE Access*, 2020, 8: 18832-18840.
- [16] SHI Z, GUAN C, LI Q, et al. Detecting marine organisms via joint attention-relation learning for marine video surveillance [J]. *IEEE Journal of Oceanic Engineering*, 2022,47: 959-974.
- [17] 彭明霞, 夏俊芳, 彭辉. 融合 FPN 的 Faster R-CNN 复杂背景下棉田杂草高效识别方法[J]. *农业工程学报*, 2019, 35(20): 202-209. PENG Mingxia, XIA Junfang, PENG Hui. Efficient recognition of cotton and weed in field based on Faster R-CNN by integrating FPN[J]. *Transactions of the CSAE*, 2019, 35(20): 202-209. (in Chinese)
- [18] RAO Y, CHENG Y, XUE J, et al. FPSiamRPN: feature pyramid siamese network with region proposal network for target tracking[J]. *IEEE Access*, 2020, 8: 176158-176169.
- [19] CHEN Y, SHU H C, XU W J, et al. Transformer text recognition with deep learning algorithm[J]. *Computer Communications*, 2021, 178: 153-160.
- [20] SHEU M H, JHANG Y S, MORSALIN S M S, et al. UAV object tracking application based on patch color group feature on embedded system[J]. *Electronics*, 2021, 10(15): 1864.
- [21] INGRASSIA S, MORLINI I. Neural network modeling for small datasets[J]. *Technometrics*, 2005, 47: 297-311.
- [22] XU Y W, SABER E, TEKALP A M. Dynamic learning from multiple examples for semantic object segmentation and search [J]. *Computer Vision and Image Understanding*, 2004, 95(3): 334-353.