

基于改进 YOLO v4 的单环刺螠洞口识别方法

冯娟^{1,2} 梁翔宇¹ 曾立华^{3,4} 宋小鹿¹ 周玺兴⁵

(1. 河北农业大学信息科学与技术学院, 保定 071001; 2. 河北省农业大数据重点实验室, 保定 071001;
3. 华南师范大学生命科学学院, 广州 510631; 4. 广东省水产健康安全养殖重点实验室, 广州 510631;
5. 河北农业大学机电工程学院, 保定 071001)

摘要: 针对养殖池塘内单环刺螠自动采收和产量预测应用需求,提出一种基于深度学习的单环刺螠洞口识别方法,以适用于自动采收船的嵌入式设备。该方法通过将 YOLO v4 的主干网络 CSPDarkNet53 替换为轻量级网络 Mobilenet v2,降低网络参数量,提升检测速度,并在此基础上使用深度可分离卷积块代替原网络中 Neck 和 Detection Head 部分的普通卷积块,进一步降低模型参数量;选取带色彩恢复的多尺度视网膜 (Multi-scale retinex with color restoration,MSRCR) 增强算法进行图像增强;利用 K-means++ 算法对数据集进行重新聚类,对获得的新锚点框尺寸进行线性缩放优化,以提高目标检测效果。在嵌入式设备 Jetson AGX Xavier 上部署训练好的模型,对水下单环刺螠洞口检测的平均精度均值 (Mean average precision,mAP) 可达 92.26%,检测速度为 36 f/s,模型内存占用量仅为 22.2 MB。实验结果表明,该方法实现了检测速度和精度的平衡,可满足实际应用场景下模型部署在单环刺螠采收船嵌入式设备的需求。

关键词: 单环刺螠洞口; 目标检测; 图像增强; 锚点框优化; YOLO v4

中图分类号: TP391.4; S979 文献标识码: A 文章编号: 1000-1298(2023)02-0265-10 OSID: 

Urechis unicinctus Burrows Recognition Method Based on Improved YOLO v4

FENG Juan^{1,2} LIANG Xiangyu¹ ZENG Lihua^{3,4} SONG Xiaolu¹ ZHOU Xixing⁵

(1. College of Information Science and Technology, Hebei Agricultural University, Baoding 071001, China
2. Hebei Key Laboratory of Agricultural Big Data, Baoding 071001, China
3. College of Life Science, South China Normal University, Guangzhou 510631, China
4. Key Laboratory for Healthy and Safe Aquaculture, Guangzhou 510631, China
5. College of Mechanical and Electrical Engineering, Hebei Agricultural University, Baoding 071001, China)

Abstract: In order to realize the real time detection of *Urechis unicinctus* burrows in the actual aquaculture pond scene, and provide support for the automatic harvesting and yield prediction of *Urechis unicinctus*, a deep learning based identification method of *Urechis unicinctus* burrows was proposed. In view of the limited storage space of the embedded equipment of harvesting vessel and high real time requirements for target detection, the YOLO v4 model had a large number of parameters and a slow detection speed. By replacing the backbone network CSPDarkNet53 of YOLO v4 with a lightweight Mobilenet v2 to reduce the amount of network model parameters and improve the detection speed. On this basis, depthwise separable convolution blocks were used instead of the normal convolution blocks in the Neck and Detection Head parts of the original network to further reduce the number of model parameters. For the poor quality of underwater images, the multi-scale retinex with color restoration (MSRCR) algorithm was selected for image enhancement. Finally, for the original anchor box obtained by clustering the COCO dataset, which was not suitable for small target recognition, the K-means++ algorithm was used to recluster the dataset and optimize the linear scaling of the obtained new anchor box size to obtain the most suitable anchor box for the dataset in order to improve the target detection effect. To simulate the automatic capture scene of *Urechis unicinctus*, a set of image acquisition equipment with an unmanned

收稿日期: 2022-03-17 修回日期: 2022-05-25
基金项目: 河北省重点研发计划项目(20327217D)
作者简介: 冯娟(1979—),女,副教授,主要从事农业信息化研究,E-mail:zlh52103@gmail.com
通信作者: 曾立华(1981—),男,副教授,主要从事智慧渔业研究,E-mail:zenglh@cau.edu.cn

ship as the main body was built, and an image data set was established through the collected video to conduct experiments. The trained model deployed on the embedded device Jetson AGX Xavier can detect mean average precision (mAP) of underwater *Urechis unicinctus* burrows up to 92.26% with detection speed of 36 f/s and model size of only 22.2 MB. Experiments showed that the method achieved a better balance of detection speed and accuracy and can meet the demand of practical application scenarios where the model was deployed in the embedded equipment of the *Urechis unicinctus* harvesting vessel. It provided a reference for the subsequent automatic harvesting of *Urechis unicinctus* and yield prediction in breeding ponds.

Key words: *Urechis unicinctus* burrows; target detection; image enhancement; anchor optimization; YOLO v4

0 引言

单环刺螠是中国北方沿海常见的一种海洋底栖无脊椎动物,俗名海肠,体壁富含多种人体需要的生物活性肽,肉质鲜美,具有很高的营养价值及经济价值^[1]。单环刺螠有潜沙穴居习性,在海底营造非标准的U型洞穴生活,洞口向上,身体能潜到40 cm深度,洞口直径约3 cm^[2]。目前我国仍采用手工挖掘成体的方式,其缺陷较多,容易造成单环刺螠死亡或残损,且捕捞效率低、成本高,影响池塘内刺参的生态混养等^[3]。本文对单环刺螠洞口进行识别研究,是实现自动化捕捞、养殖池塘内产量预测、确定最佳单环刺螠投放比例的必要准备,目前尚未见相关报道。

水下图像质量退化是准确识别单环刺螠洞口的主要难点。此前,国内外学者对水下目标识别进行了广泛的研究。QIAO等^[4]利用对比度受限的自适应直方图均衡化(Contrast limited adaptive histogram equalization, CLAHE)算法增加海参刺和海参体的对比度,利用主动轮廓法对海参进行分割。吴健辉等^[5]提出一种基于图像处理技术的鱼种类自动识别方法,识别4种主养鱼类,对团头鲂的识别精度为100%,对鲫、鲤、草鱼的识别精度为96%。在实际应用中,由于水下图像信息的复杂性,传统目标检测算法只能适用于某种特定环境,无法应用在不同场景,算法检测的精确率和速度都有待提高^[6]。深度学习给水下目标识别带来了新的发展方向,在农业^[7-10]、工业^[11-13]等多个领域得到了广泛应用。DAI等^[14]使用带有内部特征和核校准模块的残差网络进行水生生物识别,在自建数据集上识别Top-1准确率提高至95.7%。张明华等^[15]通过改进Cascade R-CNN和图像增强进行夜晚鱼类的检测。张佳林等^[16]利用机器视觉进行鱼类摄食分类,综合准确率达到了89%。

上述研究对象多为水中生物,对于水下洞口识别暂无涉及,且图像采集多在水质清的环境下进行。

但养殖塘塘底受光线较弱和存在杂质等因素的影响,目标检测效果并不理想。当前基于深度学习的目标检测算法主要分为两类,基于Two-stage的目标检测算法,基于One-stage的SSD和YOLO系列算法。基于Two-stage的检测算法具有高检测精度的特点,缺点是检测速度较慢;基于One-stage的检测算法检测精度相对较低,但检测速度快,能够满足实时检测需求。捕捞船捕捞时模型需部署在船上的嵌入式设备中进行实时检测,模型尺寸与检测速度要求较高。

针对上述问题,选择One-stage系列综合表现最稳定的YOLO v4为基础目标检测模型,替换其主干网络为轻量型网络,在此基础上使用深度可分离卷积块替换原网络中的普通卷积块,降低模型参数量,提高检测速度;采用MSRCR增强算法对数据集进行图像增强,提高图像质量;对锚点框尺寸进行优化,降低预设锚点框尺寸带来的误差,进一步提升检测精度。最后,将研究模型部署在嵌入式设备上,测试模型效果。

1 材料与方法

1.1 图像采集

实验地点位于河北省秦皇岛市附近的一块单环刺螠养殖池塘。模拟实际场景下捕捞船的捕捞,搭建了一套由水下相机、自主游弋式巡航无人船、Jetson AGX Xavier构成的图像采集装置,获取单环刺螠养殖塘底影像。自主游弋式巡航无人船(图1a)(山一科技)配备两个电池供电的无刷电机,由采集人员在池塘内遥控航行。无人船搭载成像设备——水下摄像头,镜头分辨率为352像素×240像素,获取塘底视频保存至嵌入式设备Jetson AGX Xavier上。

1.2 数据集构建与标注

视频获取时间为2020年11月15日,采用垂直摄影的方式,共获取视频21.55 min,利用Python程序每间隔25帧截取一幅图像,人工复检删除质量较

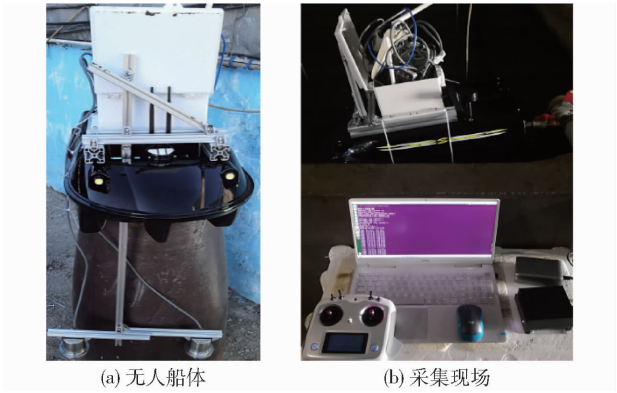


图 1 图像采集装置
Fig. 1 Image acquisition device

低的图像后共保留 782 幅图像。为丰富数据集、提高模型泛化能力,对图像进行数据增强操作,每幅图像进行水平镜像、垂直镜像及水平垂直镜像,将数据集扩增至 3 128 幅。将 3 128 幅图像随机划分为训练集、测试集和验证集,其中(训练集 + 验证集)与测试集图像数量比例为 9:1;训练集与验证集图像比例为 9:1,训练集 2 533 幅,验证集 282 幅,测试集 313 幅。

使用目标检测算法训练前,需要对采集的原始

图像进行人工标注。选择当前常用的 Labellmg 标注工具对水下单环刺螈洞口进行标注,选定图像中仅有单环刺螈洞口一个分类,使洞口中心点位于标注框中心,标注完成后自动生成含有类别和位置信息的 xml 文件。为满足 YOLO v4 神经网络的训练要求,将制作好的数据集保存为 PASCAL VOC 格式,以便用于网络训练与测试,在 3 128 幅图像中共含边框标注样本 30 896 个。

1.3 图像增强

水下图像由于水中介质吸收光能、散射改变光的方向等,图像对比度低、清晰度低、质量退化^[17]。为提升检测精度,对数据集进行图像增强处理。目前主流的水下图像增强方法包括 3 类:①基于空间域的增强方法,包括灰度变化和直方图均衡化等。②基于颜色恒常性的增强方法,包括色域映射和 Retinex 理论等。③基于先验知识的增强方法,包括暗通道先验去雾和基于 R 通道的先验颜色恢复等^[18]。

选取当前主流的 CLAHE 算法^[19]、暗通道先验去雾算法^[20]、MSRCR 算法^[21]对数据集进行图像增强对比实验,其效果见图 2。结果显示,MSRCR 算

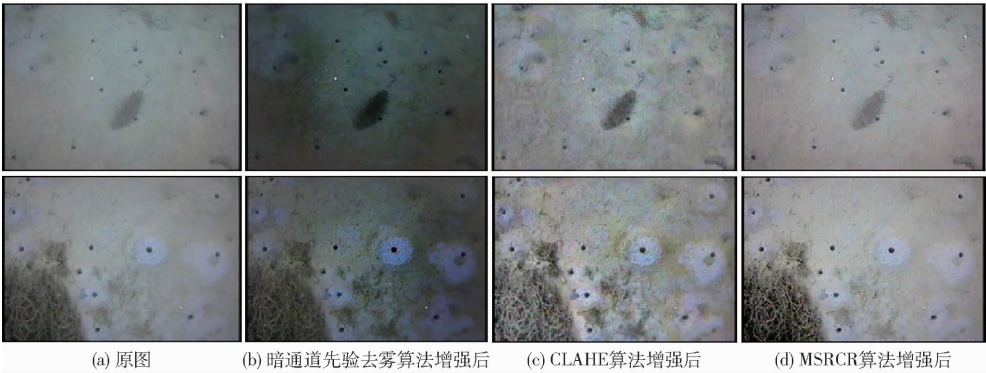


图 2 不同方法增强效果对比
Fig. 2 Comparisons of different image enhancement methods

法增强后的数据集检测精度提升最多。

根据 GIMP 中 MSRCR 的思想,MSRCR 算法有动态对比参数 Dynamic、最大尺度、尺度数 3 个超参数可以对增强后的图像质量进行动态调整。经不同参数组对比实验,如图 3 所示,尺度数与 Dynamic 对增强图像质量的影响最大,针对此数据集,结果显示 Dynamic 取 1 时目视增强效果最好,尺度数取值越大,水下图像增强后的对比度越强,当其取值在 6 ~ 8 之间时能保持图像最适宜的清晰度增强。对于最大尺度,针对此数据集,实验结果显示取值以大于 300 效果最佳,300 ~ 500 目视效果接近。最终采用 Dynamic 为 1,最大尺度为 500,尺度数为 8 作为 MSRCR 选定参数对数据集进行图像增强处理。

2 单环刺螈洞口识别模型优化与部署

2.1 YOLO v4 模型改进

相比于 Two - stage 方法,One - stage 是通过基于回归的方式来提取特征,无需生成大量的候选窗口,直接对输入图像进行目标检测和分类,属于端到端的目标检测。YOLO v4 是 One - stage 系列中最为先进的目标检测方法之一,综合表现稳定。

YOLO v4^[22]由主干网络(Backbone)、颈部网络(Neck)和预测头网络(Detection head)3 部分组成,整体上是基于 YOLO v3 的改进,采用了很多新的目标检测算法思想,如 Mosaic 数据增强、标签平滑、余弦退火算法等。在检测精度方面,由于 YOLO v4 模型预设锚点框是基于 COCO 数据集获取,与单环刺

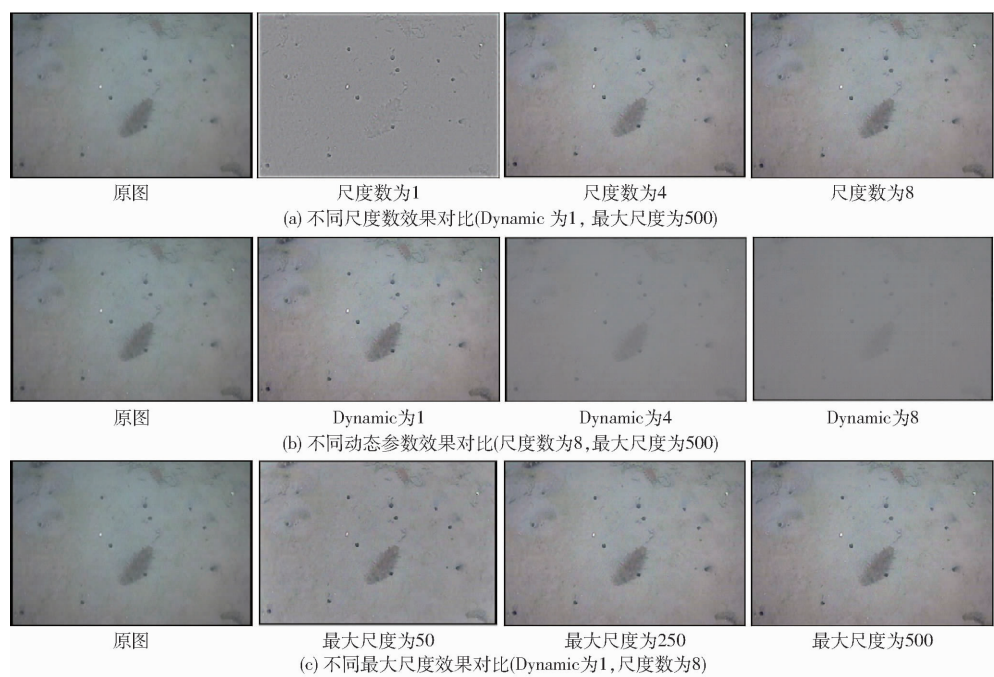


图3 MSRCR 算法不同参数效果对比

Fig. 3 Comparisons of effects of different parameters of MSRCR algorithm

螭洞口目标在尺寸上差距较大,导致检测精度有所下降。在检测速度与模型尺寸方面,自动采捕船嵌入式设备算力及存储空间资源有限,采捕时洞口检测实时性要求较高,而 YOLO v4 模型主干网络 CSPDarkNet53 网络参数量大,特征提取过程中计算量大,计算时间长,导致整体 YOLO v4 模型检测速度与模型尺寸无法满足要求。针对以上问题,在 YOLO v4 模型的基础上进行了针对性改进,使其更适用于采捕船嵌入式设备的部署。

2.2 单环刺螭洞口识别模型优化

2.2.1 主干特征提取网络替换

YOLO v4 原始主干特征提取网络 CSPDarkNet53 模型复杂、检测实时性差、模型尺寸大。为满足模型部署到采捕船的嵌入式设备上实现单环刺螭洞口实时检测的要求,需要将 YOLO v4 主干网络替换为更高效的轻量型网络。

MobileNet v2 是 MobileNet v1 的改进版,在 MobileNet v1 的基础上利用 Inverted residual 策略加入了反向残差模块 (Inverted residual block),利用 Linear bottleneck 策略将最后的 ReLU6 层变成了 Linear 形式,利用 Expansion layer 策略使深度卷积可以在更高维度进行。经一系列改进后的 MobileNet v2 网络模型的参数量与计算量明显降低,检测速度有所提升,使用 Mobilenet v2 替换 YOLO v4 的主干特征提取网络 CSPDarkNet53。Mobilenet v2^[23] 作为轻量型神经网络,使用 3×3 逐通道卷积与 1×1 逐点卷积的结构组合,实现目标的特征提取及特征信息的融合,3×3 的逐通道卷积层参数占比少,网络

层的计算量及参数占比主要集中在 1×1 的逐点卷积。逐通道卷积参数占比的减少使得 Mobilenet v2 能够提高网络训练效率,加快运算速度实现快速检测,解决因硬件设备限制导致的目标检测速度缓慢问题。

2.2.2 深度可分离卷积

Mobilenet v2 等轻量型网络结构中,会有深度可分离卷积 (Depthwise separable convolution, DSC)^[24],由逐通道卷积和逐点卷积两部分结合起来提取特征。相比常规的卷积操作,深度可分离卷积可以减少网络参数量,节约计算成本,提高检测速度。例如输入特征图尺寸为 $H \times W$ 、输入通道数为 M ,卷积核尺寸为 $K \times K$,输出通道数为 N ,输出特征图尺寸为 $D_F \times D_F$ 。普通卷积计算量 N_s 为

$$N_s = K \times K \times M \times N \times D_F \times D_F \tag{1}$$

深度可分离卷积计算量由逐通道卷积与逐点卷积组成。逐通道卷积计算量 N_d 为

$$N_d = D_F \times D_F \times M \times K \times K \tag{2}$$

逐点卷积计算量 N_p 为

$$N_p = D_F \times D_F \times M \times N \tag{3}$$

深度可分离卷积与普通卷积计算量比值为

$$\frac{N_d + N_p}{N_s} = \frac{1}{K^2} + \frac{1}{N} \tag{4}$$

由计算结果可以看出,深度可分离卷积的计算量减少与输出通道数和使用的卷积核尺寸相关,如当卷积核为 3×3 时,深度可分离卷积的计算量约为普通卷积的 1/9,有利于提高模型的推理速度。

2.2.3 反向残差模块

Mobilenet v2 的一个重要改进就是反向残差模块的使用,整个 Mobilenet v2 都由反向残差模块组成。反向残差模块可以分为两部分:左边是主干部分,首先利用 1×1 的普通卷积进行升维,使用批量标准化(Batch normalization,BN)层对特征进行归一化处理,然后利用 3×3 的深度可分离卷积进行特征提取,接上批次 BN 层与 ReLU 6 激活函数,最后再利用 1×1 的普通卷积实现降维,并使用 BN 层进行归一化处理。右边是残差边部分,反向残差模块将输入与左半部分的输出直接融合作为残差边。当步长为 2 时,为使输入与输出的维度匹配,不采用短连接,如图 4 所示。

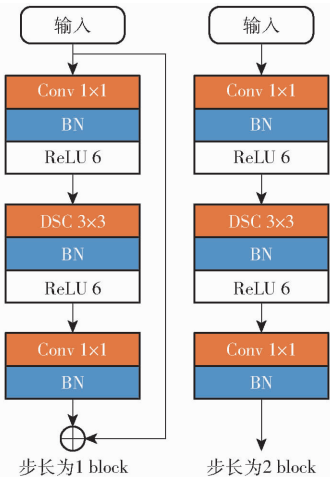


图 4 反向残差模块
Fig. 4 Inverted residual block

2.3 改进 YOLO v4 模型

为简化网络结构,降低模型尺寸,满足模型部署在嵌入式设备上实现单环刺螈洞口实时检测的要

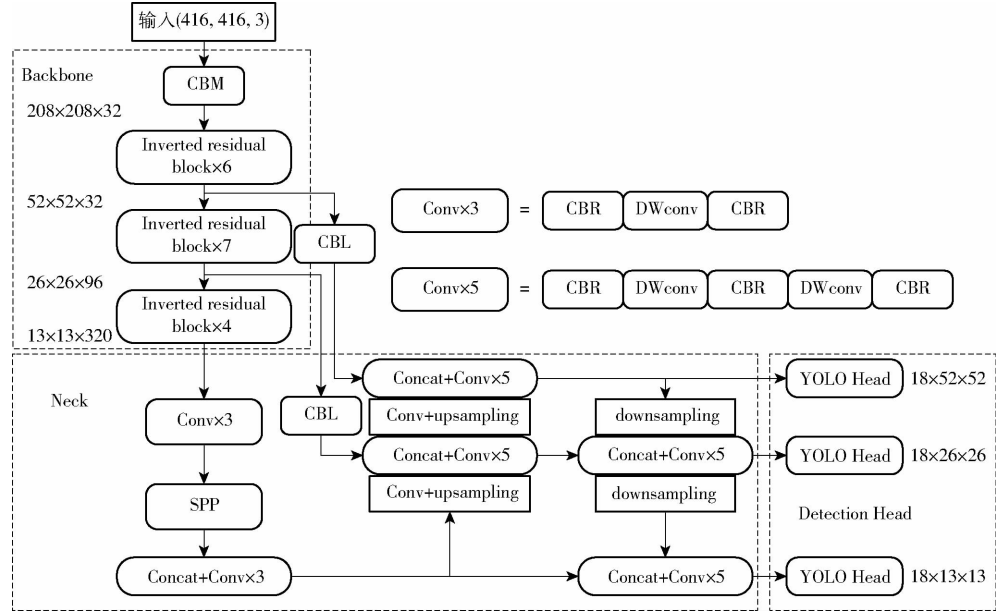


图 5 改进后的 YOLO v4 算法网络结构图
Fig. 5 Improved YOLO v4 network structure

求,利用轻量型网络 Mobilenet v2 替换 YOLO v4 的 CSPDarkNet53 网络,将其输出作为颈部网络部分的输入,Mobilenet v2 主要由 3 个 6 层、7 层、4 层的反向残差模块构成,同时为进一步减少模型参数量,使用深度可分离卷积块替换 YOLO v4 中 Neck 和 Detection Head 部分的 3×3 和 5×5 普通卷积块,改进后的 YOLO v4 网络结构如图 5 所示。其中 DWconv 代表一个 3×3 的深度可分离卷积和一个 1×1 的普通卷积组成的深度可分离卷积块,CBR 由卷积层(Convolutional)、BN 层和激活函数层(ReLU 6)组成。

2.4 锚点框聚类与优化

YOLO v4 网络按大中小比例预设了 9 个共 3 类锚点框,分别用于 3 个不同尺度的 YOLO Head 来预测目标,预设锚点框尺寸是在 COCO 数据集上通过聚类获得。预设的锚点框尺寸在常规场景下的目标检测任务已经具有较好的适用性,但在极端场景小目标识别方面,使用预设的锚点框会增大误差,影响模型的检测精度。为提升单环刺螈洞口检测精度,针对数据集重新进行锚点框聚类。利用 K-means 聚类算法对数据集进行聚类获取锚点框,尺寸为 (5, 6)、(6, 7)、(7, 6)、(7, 7)、(8, 8)、(10, 7)、(8, 9)、(10, 10)、(14, 13),经训练测试发现,检测精度相较预设锚点框有所下降。分析原因是 K-means 对初始聚类中心的选取敏感,不同初始聚类中心聚类的结果差别较大,易造成选取不准确。

针对 K-means 选取初始聚类中心的弊端,利用 K-means ++ 算法解决该弊端。K-means ++ 从输入宽高的集合 N 中随机选择一个作为第一个聚类中

心 O_1 , 然后计算每个点 x 到已有最近聚类中心点的距离 $D(x)$, 同时计算出每个样本被确定为下一个聚类中心的概率 $P(x)$ 为

$$P(x) = \frac{D(x)^2}{\sum_{x \in N} D(x)^2}$$

(5)

其中 $D(x) = 1 - I_{ou}$

(6)

式中 I_{ou} ——聚类中心与每个标注框的交并比

选择 $P(x)$ 最大的样本作为下一个聚类中心, 即距离当前 n 个聚类中心越远的点会有更高的概率被选为第 $n + 1$ 个聚类中心, 重复上述操作直到选定 K 个中心点。最后不断重复计算每个样本到 K 个聚类中心点的距离, 将样本点划分至距离最近的聚类中心的类中并更新聚类中心, 直到锚点框的大小不再发生改变, 得到新的锚点框尺寸为 (7, 12)、(7, 10)、(7, 8)、(8, 12)、(10, 10)、(10, 17)、(11, 13)、(14, 20)、(18, 26)。

经训练测试发现, 经过 K-means++ 算法重新聚类后的锚点框对于目标检测的 mAP 略有提升, 但召回率和调和平均值 (F1 值) 有所下降。分析是由于数据集目标类型单一, 标记框尺寸集中, 导致 K-means++ 聚类的锚框尺寸集中, 无法体现 YOLO v4 模型的多尺度输出优势。针对此问题, 采用对锚框线性缩放扩展的方法, 将锚框尺寸进行缩放拉伸,

计算公式为

$$\begin{cases} X'_1 = \alpha X_1 \\ X'_9 = \beta X_9 \\ X'_i = \frac{X_i - X_1}{X_9 - X_1} (X'_9 - X'_1) + X'_1 \\ Y'_i = X'_i \frac{Y_i}{X_i} \end{cases}$$

(7)

式中 X_i ——第 i 个经聚类获得的锚点框宽
 Y_i ——第 i 个经聚类获得的锚点框高
 X'_i ——经过缩放扩展后的锚点框宽
 Y'_i ——经过缩放扩展后的锚点框高
 α ——锚点框缩放倍数
 β ——锚点框扩展倍数

为确定最佳的比例参数, 将参数 α 固定设置为 0.3, 参数 β 依次设置为 2、3、 $\cdots$ 、10 进行对比实验。如表 1 所示, 当 α 设置为 0.3 时, 随着 β 的变化, mAP 呈现非标准的折线变化, mAP 最高点时 $\alpha = 0.3, \beta = 5$ 。如表 2 所示, 设置参数 β 为 5, α 取不同值进行实验 (α 设为 0.1 时缩放后部分锚点框为 0, 舍弃该参数), 由实验结果可得, 当 $\alpha = 0.7, \beta = 5$ 时能够取得对于单环刺螭洞口的最佳识别精度, 此时锚点框尺寸为 (4, 8)、(4, 5)、(4, 4)、(11, 17)、(27, 27)、(27, 46)、(35, 41)、(58, 83)、(90, 130)。

表 1 $\alpha = 0.3$ 时不同 β 的实验结果

Tab. 1 Experimental results of different β when $\alpha = 0.3$

%

指标	β								
	2	3	4	5	6	7	8	9	10
mAP	89.29	88.81	88.34	90.31	88.57	89.11	90.09	86.44	90.03
F1 值	89.40	90.39	90.34	90.81	89.84	90.11	90.25	87.52	90.32
召回率	88.92	89.65	90.45	90.75	89.59	89.74	89.96	86.81	90.58
精确率	89.88	91.15	90.23	90.86	90.09	90.49	90.54	88.24	90.07

表 2 $\beta = 5$ 时不同 α 的实验结果

Tab. 2 Experimental results of different α when $\beta = 5$

%

指标	α								
	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
mAP	88.91	90.31	89.93	89.49	91.35	91.73	91.29	90.61	
F1 值	90.46	90.81	90.51	89.96	90.19	90.53	90.09	89.97	
召回率	90.54	90.75	90.74	89.41	90.33	89.84	89.95	89.34	
精确率	90.37	90.86	90.28	90.51	90.04	91.23	90.23	90.61	

2.5 模型嵌入式设备部署

利用 Jetson AGX Xavier 进行模型嵌入式设备部署, 将在 PC 端训练的 YOLO v4 和改进模型导入 Jetson AGX Xavier 开发板, 并配置相应的运行环境。在直接运行模型测试时发现改进模型相比于 YOLO v4 检测速度确实有很大提升, 但仍无法满足实时检

测要求。针对该问题, 利用嵌入式平台广泛应用的 TensorRT 进行模型推理加速, TensorRT 是 NVIDIA 高性能深度学习推理优化器, 可为深度学习模型的推理提供低延迟、高吞吐量的加速度, 支持包括 Pytorch 在内的多种深度学习框架部署推理。利用 TensorRT 加速需将模型转换为 TensorRT 支持格式, 转换过程为 Pytorch 格式 (Pth)——开放神经网络交换格式 (Onnx)——TensorRT 格式 (Trt), 转换时模型计算精度为半精度计算 (FP16), 利用 Python API 实现 trt 文件的导入与初始化。模型运行检测图像如图 6 所示。

3 实验结果与分析

3.1 实验平台和参数设置

模型训练和 PC 端测试均在同一台计算机进

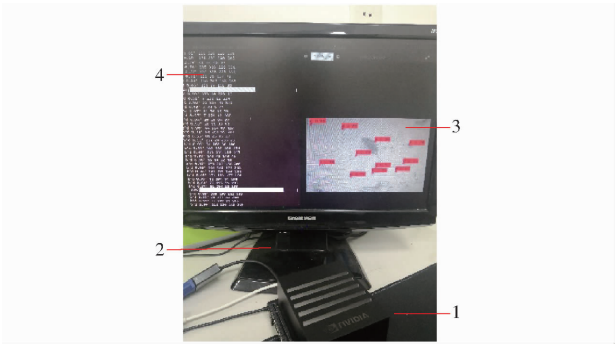


图 6 Jetson AGX Xavier 部署测试

Fig. 6 Experiment diagram on Jetson AGX Xavier

1. Jetson AGX Xavier 2. 外接显示器 3. 检测图像 4. 运行界面

行, CPU 为 Intel Core i5 9300H, GPU 为 NVIDIA GeForce GTX 1650, 使用 Pytorch 1. 6. 0 作为深度学习框架, CUDA 10. 2, CUDNN 7. 5. 6, Windows 10 操作系统, 运行内存 12 GB。

模型的嵌入式端部署测试在 Jetson AGX Xavier 上进行, 搭载 64 个 Tensor Core 的 512 核 NVIDIA Volta GPU, CPU 为 8 核 NVIDIA Carmel ARMv8. 2 64 位 CPU, 内存为 32 GB, 显存为 32 GB。相应配置系统为 ARM 版 Ubuntu 18. 04 操作系统, 模型运行环境的配置为 Python 3. 8、Pytorch 1. 6. 0、CUDA 10. 2、TensorRT 7. 1。

训练参数设置如下: 训练集样本 2 533 幅, 测试集样本 313 幅, 验证集 282 幅, 为防止过拟合, 训练共设置 100 个迭代周期 (Epoch), 前 50 个周期的初始学习率为 0. 001, 样本批量大小 (Batchsize) 为 8, 后 50 个周期学习率设置为 0. 000 1, Batchsize 为 4, 非极大值抑制 (Non-maximum suppression, NMS) 为 0. 3, 置信度阈值为 0. 5, 交并比 (Intersection over union, IOU) 阈值为 0. 5。

3.2 评价指标

利用精确率 (Precision, P)、召回率 (Recall, R)、F1 值、平均精度均值 (mAP)、模型尺寸和检测速度来评估模型的综合性能。

精确率是指在所有检测出的单环刺螈洞口目标中检测正确的概率, 召回率是指所有的正样本中正确识别的概率, F1 值表示精确率与召回率的调和平均数, 平均精度均值表示模型在各个召回率下的精确率平均值。检测目标仅为单环刺螈洞口, 平均精度 (Average precision, AP) 也等于 mAP, 检测精度的各指标优先级由高到低为 mAP、F1 值、 P 、 R 。

3.3 增强算法性能评价

由水下环境复杂、能见度低等因素导致图像质量较低不利于模型对目标的识别, 使用当前主流的 3 种水下图像增强算法: CLAHE 算法、暗通道先验去雾算法、MSRCR 算法对数据集进行图像增强。增

强算法的选择在相关水下目标识别研究中^[25-26]多为研究者目视估计选择, 受研究者个人因素影响较大。为评价算法性能, 利用 3 种算法增强后的图像与原始图像在相同软硬件环境下进行目标检测效果对比实验, 对比实验结果见表 3。

表 3 不同增强算法检测结果对比

Tab. 3 Comparison of detection effects of different enhancement algorithms

增强算法	精确率	召回率	F1 值	mAP
无增强	89. 79	88. 00	88. 89	87. 04
暗通道先验去雾	90. 33	90. 08	90. 21	89. 24
CLAHE	90. 20	87. 67	88. 92	87. 74
MSRCR	90. 44	90. 14	90. 29	89. 30

由表 3 可知, CLAHE 算法增强后的图像 mAP 略有提升, 但召回率有所下降。MSRCR 算法与暗通道先验去雾算法在检测精度方面提升效果接近, 但 MSRCR 算法对单环刺螈洞口目标检测精度提升最高且颜色分布更加平衡, 保持了较好的视觉观感, 最终选用 MSRCR 算法作为模型的图像增强算法。

3.4 锚点框优化效果

利用 YOLO v4 预设锚点框、K-means 聚类得到的锚点框、K-means + + 聚类得到的锚点框以及 K-means + + 聚类后线性缩放优化的锚点框 (K-means + + E) 共 4 种锚点框在相同软硬件环境下进行训练并测试, 对单环刺螈洞口目标的检测性能结果如表 4 所示。

表 4 不同算法获得的锚点框检测效果对比

Tab. 4 Comparison of priori box detection effect obtained by different methods

算法	精确率	召回率	F1 值	mAP
YOLO v4	89. 79	88. 00	88. 89	87. 04
K-means	88. 17	81. 47	84. 69	85. 82
K-means + +	89. 79	83. 18	86. 36	88. 86
K-means + + E	91. 23	89. 84	90. 53	91. 73

由图 7 可以看出, 使用 K-means + + E 锚点框模型的 $P-R$ 曲线下包含的面积明显高于其他 3 组,

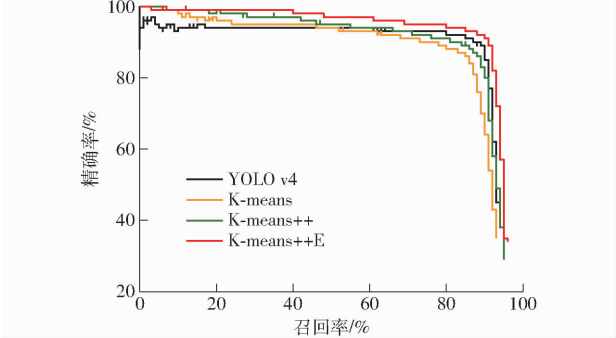


图 7 不同锚点框的检测结果 $P-R$ 曲线

Fig. 7 $P-R$ curves of detection results of different anchors

说明 mAP 更高,更适合作为单环刺螈洞口数据集的锚点框。

3.5 综合改进后模型性能评价

3.5.1 训练损失值对比与最优模型确定

使用相同训练集、验证集和测试集分别在 YOLO v4 与改进模型上进行训练,对比迭代损失值变化曲线,如图 8 所示。

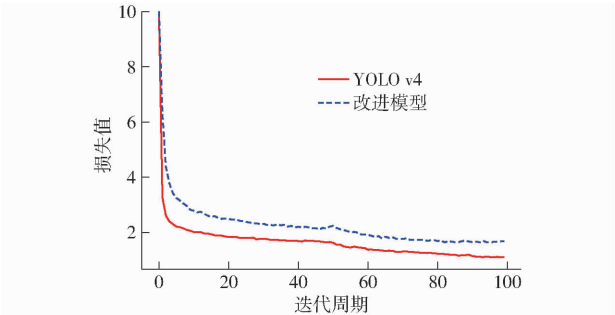


图 8 YOLO v4 和改进模型训练的损失值曲线

Fig.8 Loss curves of YOLO v4 and improved model training

由图 8 可见,改进模型与 YOLO v4 模型的损失值曲线均能够在前 10 个迭代周期中迅速下降,表明轻量化后的改进模型能够快速收敛。从第 11 到第 50 个周期损失值逐步下降,在第 50 个周期由于学习率与 Batchsize 的改变出现振荡。之后直到第 90 个周期仍处于缓慢下降状态,在第 90 个周期之后,改进模型损失值稳定在 1.6 左右,模型拟合结束。为确定实验最优模型,所有对比实验模型在训练稳定后的 5 个最低损失值模型为最优模型前提下,综合考虑 mAP 与其他 3 项精度指标确定最优模型,最

终选用第 98 个迭代周期迭代完成后保存的模型为单环刺螈洞口识别模型,此时模型具有最高的 mAP 与召回率,也具有较高的 F1 值和精确率。

3.5.2 模型轻量化效果

为展示模型轻量化效果,将 YOLO v4 与仅替换主干网络为 Mobilenet v2 的模型以及进一步替换网络内 Neck 和 Detection Head 部分普通卷积块为深度可分离卷积块的改进模型进行参数量和模型尺寸的对比,对比结果如表 5 所示。

表 5 网络结构参数量对比

Tab.5 Comparison of network structure parameters		
网络模型	参数量	模型内存占用量/MB
YOLO v4	63 937 686	244. 0
YOLO v4 – Mobilenet v2	40 390 242	154. 0
改进模型	11 975 749	46. 5

由表 5 可得,改进模型参数量由 63 937 686 降低至 11 975 749,而模型尺寸由 244 MB 降低至 46.5 MB,可见模型轻量化效果明显,有效减小了模型尺寸,更适于部署在资源有限的嵌入式设备。

3.5.3 不同模型性能指标比较

为测试模型效果,在相同软硬件环境下将改进模型与轻量化 YOLO v4 和当前主流的 One – stage 目标检测算法 SSD、YOLO v3、YOLO v4、YOLO v4 – tiny 进行对比实验,对模型进行综合评估。在检测精度方面,由于 mAP 和 F1 值均同时兼顾了检测模型的精确率和召回率,主要采用 mAP 和 F1 值来评价模型的检测精度,评估结果如表 6 所示。

表 6 不同模型综合效果对比

Tab.6 Comparison of comprehensive effects of different models

网络模型	精确率/%	召回率/%	F1 值/%	mAP/%	模型内存占用量/MB	检测速度/(f·s ⁻¹)	每帧图像推理时间/ms
SSD	79.94	15.69	26.23	68.13	90.6	15.23	65.65
YOLO v4 – tiny	75.67	70.45	72.97	68.12	22.4	110.67	9.04
YOLO v3	93.92	86.23	89.91	91.89	235.0	20.75	48.20
YOLO v4	94.31	94.05	94.18	93.49	244.0	16.97	58.90
轻量化 YOLO v4	89.79	88.00	88.89	87.04	46.5	45.15	22.20
改进模型	91.02	90.96	90.99	92.48	46.5	45.31	22.10

由表 6 可知,在检测精度方面,改进模型的平均精度均值达到 92.48%,相比 One – stage 的 SSD、YOLO v4 – tiny、YOLO v3 提高 24.35、24.36、0.59 个百分点,相比 YOLO v4 降低 1.01 个百分点;F1 值达到 90.99%,相比 SSD、YOLO v4 – tiny、YOLO v3 分别提高 64.76、18.02、1.08 个百分点,相比 YOLO v4 降低 3.19 个百分点。在检测速度方面,改进模型为 45.31 f/s,比 SSD、YOLO v3、YOLO v4 提高 30.08、24.56、28.34 f/s。改进模型的内存占用量仅为

46.5 MB,相比SSD、YOLO v3 和 YOLO v4 分别减小 44.1、188.5、197.5 MB,更适用于嵌入式设备的部署。

3.5.4 不同模型检测效果对比

为直观展示模型检测性能,分别采用表 6 中 YOLO v4、轻量化 YOLO v4 与改进模型对不同的单环刺螈洞口图像进行检测。

如图 9 所示,模型轻量化后,目标检测精确率有所下降,在第 1 幅图像中左下角出现错检,在第 2 幅

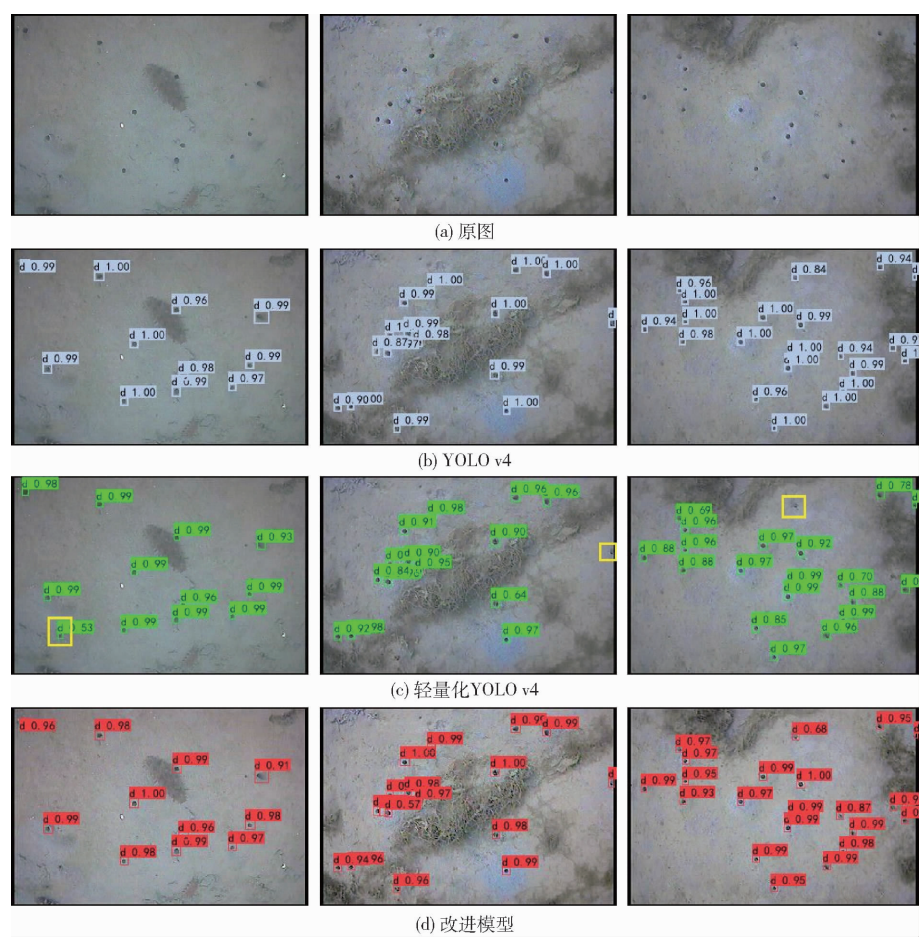


图 9 不同模型单环刺螈洞口识别效果

Fig. 9 Recognition effect of *Urechis unicinctus* burrows with different models

最右侧和第 3 幅最上方出现漏检,但在改进之后模型没有错检且完全检测出所有单环刺螈洞口,模型检测目标的置信度在 0.9 以上的更多,表明改进模型对检测目标的置信度有一定加强作用,检测效果更好,检测能力更强,满足实际应用场景下单环刺螈洞口的实时检测精度要求。

3.5.5 模型嵌入式设备部署测试

为验证模型是否满足部署在嵌入式设备实时检测的要求,在 Jetson AGX Xavier 开发板进行模型部署对比实验。对 YOLO v4、改进模型和经过 TensorRT 加速的 YOLO v4(trt)、改进模型(trt)进行性能评估,实验结果如表 7 所示,其中模型尺寸为模

表 7 模型嵌入式设备部署效果对比

Tab. 7 Comparison of model embedded device deployment effects

网络模型	精确率/%	召回率/%	F1 值/%	mAP/%	模型内存占用 量/MB	检测速度/ (f·s ⁻¹)	单帧图像推理 时间/ms
YOLO v4	94.31	94.05	94.18	93.49	256.3	9.08	110.14
改进模型	91.02	90.96	90.99	92.48	48.8	18.55	53.91
YOLO v4(trt)	94.29	93.83	94.06	93.27	133.3	19.14	52.25
改进模型(trt)	91.00	90.75	90.88	92.26	22.2	36.00	27.78

型文件在 Ubuntu 系统的实际占用空间。

由表 7 可知,未经 TensorRT 加速的 YOLO v4 与改进模型在 Xavier 开发板上由于硬件条件的限制检测速度大幅下降,经 TensorRT 加速的模型由于模型转换导致检测精度略微降低,但模型尺寸与推理时间大大减少。经 TensorRT 加速后的 YOLO v4(trt)相较于 YOLO v4,检测速度提升 2.11 倍,为

19.14 f/s,无法满足单环刺螈洞口实时检测的要求。经 TensorRT 加速后的改进模型(trt)在精确率、召回率、F1 值、mAP 分别下降 0.02、0.21、0.11、0.22 个百分点,但检测速度达到 36.00 f/s,相较于直接推理的 18.55 f/s,加速比约为 1.94,且模型尺寸仅为 22.2 MB。实验结果表明,经过 TensorRT 加速后的改进模型满足部署在嵌入式设备上实时检测单环刺

螭洞口的需求。

3.6 讨论

将 YOLO v4 主干网络替换为轻量型网络会导致 mAP 一定程度的下降,但在经过图像增强和锚点框优化后,在检测精度方面,模型平均精度均值能达到 92.48%,低于 YOLO v4 模型 1.01 个百分点,检测速度提升约 3 倍,且模型尺寸仅为 YOLO v4 模型的 1/5;而在检测速度与模型尺寸方面,YOLO v4-tiny 虽然获得了 110.67 f/s 的最高检测速度与 22.4 MB 的最小模型内存占用量,但是对于小目标或遮挡目标的检测效果远差于 YOLO v4 网络,mAP 仅为 68.12%,无法满足检测精度要求。最后,通过嵌入式端设备检测效果对比实验,经过 TensorRT 加速后的改进模型在检测精度方面 F1 值为 90.88%,mAP 为 92.26%,检测速度为 36.00 f/s,模型满足部署在单环刺螭自动采捕船嵌入式设备上实时检测的检测精度及检测速度要求。

4 结论

(1)以单环刺螭洞口为研究对象,提出了适于移植到采捕船嵌入式设备的、基于深度学习的单环刺螭洞口识别方法。在嵌入式设备上,改进模型的

F1 值为 90.88%,mAP 为 92.26%,满足实际生产需要,可用于水下单环刺螭洞口实时识别,是实现养殖环境中单环刺螭自动采捕和产量预测的必要前提,其研究思路为后续研发相关自动化设备提供了技术支持。

(2)针对水下图像质量退化问题,在网络训练前对图像采用 MSRCR 增强方法,图像增强后 F1 值提升 1.4 个百分点,mAP 提升 2.26 个百分点。

(3)针对 YOLO v4 模型冗余,模型尺寸大,检测速度慢的问题,将 YOLO v4 的特征提取网络由 CSPDarkNet53 替换为 Mobilenet v2,并将模型中 Neck 和 Detection Head 部分的普通卷积块替换为深度可分离卷积块,减小模型尺寸,提升检测速度。嵌入式端模型内存占用量仅为 22.2 MB,检测速度为 36.00 f/s,单幅图像的推理时间为 27.78 ms,满足实时检测的要求。

(4)针对数据集中多为小目标的特点,使用 K-means++ 聚类算法对数据集进行重新聚类,并将聚类获得的锚点框进行线性缩放扩展,以获取更适合数据集的锚点框,改进后的锚点框相比预设锚点框 F1 值提升 1.64 个百分点,mAP 提升 4.69 个百分点。

参 考 文 献

[1] 许星鸿,孟霄,甘宏涛,等.单环刺螭的繁殖生物学[J].水产学报,2020,44(8):1275-1285.
XU Xinghong, MENG Xiao, GAN Hongtao, et al. Reproductive biology of *Urechis unicinctus*[J]. Journal of Fisheries of China, 2020,44(8):1275-1285. (in Chinese)

[2] 曹梅,辛益,陈吉圣,等.单环刺螭特征特性及养殖技术[J].现代农业科技,2021(20):178-180.
CAO Mei, XIN Yi, CHEN Jisheng, et al. Characteristics of *Urechis unicinctus* and breeding techniques[J]. Modern Agricultural Science and Technology, 2021(20):178-180. (in Chinese)

[3] 大连海洋大学.一种海肠采捕船:CN111528192A[P]. 2020-08-14.

[4] QIAO Xi, BAO Jianhua, ZENG Lihua, et al. An automatic active contour method for sea cucumber segmentation in natural underwater environments[J]. Computers and Electronics in Agriculture, 2017,135:134-142.

[5] 吴健辉,张晓飞,杨敏,等.基于鱼体背部弯曲潜能算法的四种主养鱼类识别[J].淡水渔业,2016,46(4):101-106.
WU Jianhui, ZHANG Xiaofei, YANG Min, et al. Identification of four fish species based on bending potential algorithm of fish back contour[J]. Freshwater Fisheries, 2016,46(4):101-106. (in Chinese)

[6] 张弓.基于卷积神经网络的水下目标识别研究[D].镇江:江苏科技大学,2020.
ZHANG Gong. Target recognition based on convolutional neural network[D]. Zhenjiang: Jiangsu University of Science and Technology, 2020. (in Chinese)

[7] 刘芳,刘玉坤,林森,等.基于改进型YOLO的复杂环境下番茄果实快速识别方法[J].农业机械学报,2020,51(6):229-237.
LIU Fang, LIU Yukun, LIN Sen, et al. Fast recognition method for tomatoes under complex environments based on improved YOLO[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020,51(6):229-237. (in Chinese)

[8] 刘天真,滕桂法,苑迎春,等.基于改进YOLO v3的自然场景下冬枣果实识别方法[J].农业机械学报,2021,52(5):17-25.
LIU Tianzhen, TENG Guifa, YUAN Yingchun, et al. Winter jujube fruit recognition method based on improved YOLO v3 under natural scene[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021,52(5):17-25. (in Chinese)

[9] LI Fei, WANG Yun, LI Yiming, et al. Tied bilateral learning for aquaculture image enhancement[J]. Computers and Electronics in Agriculture, 2022,199:107180.

[10] CAO Shuo, ZHAO Dean, LIU Xiaoyang, et al. Real-time robust detector for underwater live crabs based on deep learning[J]. Computers and Electronics in Agriculture, 2020,172:105339.

[11] 余永维,杜柳青,闫哲,等.基于深度学习特征的铸件缺陷射线图像动态检测方法[J].农业机械学报,2016,47(7):407-412.
YU Yongwei, DU Liuqing, YAN Zhe, et al. Dynamic detection of casting defects radiographic image based on deep learning feature[J]. Transactions of the Chinese Society for Agricultural Machinery, 2016,47(7):407-412. (in Chinese)