

基于改进 YOLO-V4 网络的浅海生物检测模型

毛国君^{1,2}, 翁伟栋¹, 朱晋德¹, 张媛³, 吴富村⁴, 毛玉泽^{5*}

(1. 福建工程学院计算机科学与数学学院, 福州 350118; 2. 福建省大数据挖掘与应用重点实验室, 福州 350118; 3. 獐子岛集团股份有限公司, 大连 116001; 4. 中国科学院海洋研究所实验海洋生物学重点实验室, 青岛 266071; 5. 中国水产科学研究院黄海水产研究所, 青岛 266071)

摘要: 海洋生物智能检测是海洋牧场战略的一部分, 而利用水下机器人在复杂的海洋环境中快速、准确地检测海洋生物是关键问题。由于海底环境复杂、亮度分布不均匀、海洋生物与其生存环境的区分性差、生物被遮蔽或半遮蔽等原因, 准确识别海洋生物是一个巨大的挑战。随着卷积神经网络的发展, 基于深度学习的目标检测算法成为主流, 出现了如 EfficientDet、RetinaNet 和 YOLO-V4 等典型算法。这些基于深度学习的算法都不是完全尽善尽美的, 不能完全满足海洋生物识别的需求。在探测精度、运算速度、密集目标探测效果等方面都有提高的空间。该研究建立了一个海洋生物数据集, 采集了原始图片 1 810 张, 数据增强后得到 7 240 张图片, 它们被分成训练集 (80%) 和测试集 (20%)。其次, 通过引入跨阶段局部网络的概念, 构建了嵌连接 EC (Embedded Connection) 部件, 并将其嵌入到 YOLO-V4 网络的末端, 得到改进的 YOLO-V4 网络。最后, 该研究提出了基于改进 YOLO-V4 网络的海洋生物检测模型 MOD (Marine Organism Detection)。试验结果表明, MOD 模型的 mAP₅₀、mAP₇₅ (交并比阈值为 0.5、0.75 的精度均值) 分别为 0.969 和 0.734, 计算量为 35.328BFLOPs (十亿浮点运算数), 检测帧速为 139 ms (具有图形加速器 GeForce GTX1650 上)。与原始 YOLO-V4 模型相比, MOD 模型的 mAP₅₀ 和 mAP₇₅ 提高了 0.9 和 4.8 个百分点, 而计算量仅提高 0.2%。此外, 对比两种模型的准确率-召回率曲线, MOD 模型的精确度与召回率的平衡点更接近 (1, 1), 因此 MOD 模型能学习精度和效率的平衡性更好。该研究直接面向浅海生物的目标检测问题, 所提供的方法可以为水下机器人精准执行智能捕捞等任务提供有益参考。

关键词: 模型; 深度学习; 目标检测; YOLO-V4; 跨阶段局部网络; 嵌连接

doi: 10.11975/j.issn.1002-6819.2021.12.018

中图分类号: TP391.4

文献标志码: A

文章编号: 1002-6819(2021)-12-0152-07

毛国君, 翁伟栋, 朱晋德, 等. 基于改进 YOLO-V4 网络的浅海生物检测模型[J]. 农业工程学报, 2021, 37(12): 152-158.

doi: 10.11975/j.issn.1002-6819.2021.12.018 <http://www.tcsae.org>

Mao Guojun, Weng Weidong, Zhu Jinde, et al. Model for marine organism detection in shallow sea using the improved YOLO-V4 network[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(12): 152-158. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2021.12.018 <http://www.tcsae.org>

0 引言

海洋经济已成为国民经济新的增长点, 中国已经开始实施海洋牧场工程, 并将其作为新兴战略产业来大力扶持^[1]。事实上, 随着中国国民生活水平的不断提高, 鲍鱼、海参、海胆等珍贵海产品的市场需求与日俱增, 极大促进了浅海养殖为代表的海洋渔业的蓬勃发展。人工潜水监控及捕捞的传统方法不仅耗时、人工成本高, 而且人身伤害大、危险系数高。近年来, 随着水下机器人等技术的日臻成熟和逐步应用, 海洋养殖的信息化和智能化建设迎来新的发展契机。

为实现水下机器人的成功捕捞, 研发海洋生物的自动化目标识别技术尤为重要, 而支撑这一工作的核心技术

是目标检测 (Object Detection) 模型和算法。由于水下获取视频及图像的特殊性, 海洋目标检测在准确率和实时性等方面都有特殊的需求, 所以面向于海洋环境研究海洋生物等目标的高性能检测模型已经成为一个重要的研究方向。本质上说, 目标检测是计算机视觉的一项任务, 旨在解决计算机视觉应用中最基本的问题: 即检测物体的类别和位置, 两个问题的解决程度和效率反映了不同目标检测算法的质量^[2]。

传统的目标检测方法一般由图像特征提取和特征学习两个相对独立的过程来完成。从算法角度, 传统的目标检测方法依赖于图像特征提取算子来预先将图像集抽取成多维特征值集。多维特征值集一旦建立, 图像的学习就可以利用多维数据学习算法来完成知识模式的挖掘。由于传统的图像特征提取算子基本上是基于主观判断的, 而且与知识模式的学习算法分离, 存在提取特征难度大和泛化能力弱的致命缺点, 因此目标的检测精度和效率都很难取得预想的效果^[3]。

相比于传统方法, 基于卷积神经网络 (Convolutional Neural Networks, CNN) 的深度学习目标检测算法逐渐成为新的研究重点^[4]。卷积神经网络是基于深度学习目

收稿日期: 2021-02-09 修订日期: 2021-06-02

基金项目: 国家重点研发项目 (2019YFD0900800/05); 国家自然科学基金项目 (61773415)

作者简介: 毛国君, 博士, 教授, 博士生导师, 研究方向为人工智能和大数据挖掘。Email: 19662092@jtu.edu.cn

*通信作者: 毛玉泽, 博士, 研究员, 博士生导师, 研究方向为健康海水养殖和多营养层次综合养殖。Email: maoyuze@163.com

标检测算法的基础,基于 CNN 的深度学习通过增加 CNN 的卷积层量来提高学习质量。一般而言,基于深度学习的目标检测方法的学习效果要比人工的特征提取和检测方法要好^[5]。从技术角度上说,CNN 使用反向传播算法进行反馈学习,通过自动化的特征提取和多层卷积学习,最大限度地减少人为干预,提升了建模能力^[6]。此外,最新的 CNN 网络利用数据增强(Data enhance)技术来增加学习样本的质量及多样性,有利于 CNN 检测精度提高^[7]。基于深度学习的目标检测算法已经在许多情况得到应用,如已经被用于苦瓜^[8]、西红柿^[9]、荔枝^[10]、生猪脸部^[11]等对象的检测。

基于深度学习的目标检测方法主要有两大类:第一类是基于区域建议的二阶段(Two-stage)目标检测算法,如 R-CNN、Fast R-CNN、Faster R-CNN;第二类是基于回归的一阶段(One-stage)目标检测算法,如 YOLO^[12]、RetinaNet^[13]、EfficientDet^[14]。一阶段检测算法中应用最广的是 YOLO 系列算法。2016 年 5 月,Redmon 等^[12]提出首个基于回归的目标检测方法 YOLO-V1,算法一经提出就受到了研究者的广泛关注。随着 YOLO 系列算法的发展,在 2020 年已经更新到了第四代。YOLO-V4^[15]有着良好的性能,通过标准数据集上的验证,YOLO-V4 是目前最好的目标检测算法之一,检测速度快是 YOLO-V4 的显著特点。

近年来,国内外研究人员已经将 YOLO 网络模型尝试性地应用到海洋生物的检测中。如:Wang 等^[16]在对水下不同类型的鱼类检测时,将迁移学习方法引入到 YOLO 模型中;Xia 等^[17]使用 YOLO-V2 算法,对浅海养殖的海参完成识别,论证了训练样本和优化检测模型对提高准确率的重要性;朱世伟等^[18]为应对高密度水下目标检测任务,提出了基于类加权 YOLO 网络的水下目标检测方法,提升了网络学习的适应性;赵德安等^[19]利用数据增强 Retinex 算法对数据进行预处理,增加了学习样本的数据量和多样性,提升了 YOLO 网络对水下河蟹的识别能力。李庆忠等^[20]设计精简 YOLO 网络,再通过迁移学习的方法训练网络,在对海底鱼类等小目标检测上进行了尝试。

YOLO 算法在某些方面已经得到应用,但是直接应用到浅海养殖生物识别仍然存在问题。试验中表现出的主要问题有:1)由于光线昏暗而且分布不均匀,导致采集的数据清晰度不高,检测精度达不到期望;2)有些海洋生物与生活的环境区分度不够,极易导致漏检的情况;3)生物体的遮挡、半隐藏等现象经常出现,对模型的泛化能力要求更高。因此,本文针对这些浅海海洋生物检测的突出问题,设计了嵌连接 EC (Embedded connection) 部件,并把它成功嵌入到 YOLO-V4 的颈网末端,该部件有望增强网络的学习和泛化能力。

1 海洋生物检测模型

本节主要目的是介绍本文的海洋生物检测模型 MOD (Marine Organism Detection)。为此,从如下 3 个方面依次进行叙述:1)分析原始 YOLO-V4 的结构和工作原理,

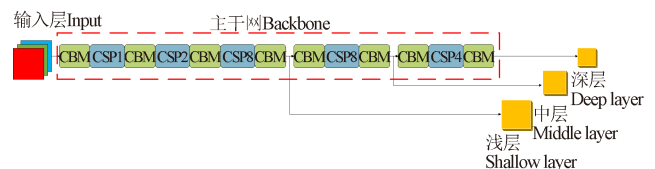
为改进 YOLO-V4 网络提供基本支撑;2)直接针对浅海生物检测的需求,设计嵌入式 EC 部件,得到改进的 YOLO-V4 网络结构;3)基于改进 YOLO-V4 网络,提出海洋生物检测模型 MOD。

1.1 YOLO-V4 网络分析

YOLO 系列算法是目前应用最为广泛的一阶段目标检测算法之一,YOLO-V4 是目前使用最多的版本。YOLO-V4 主要包括输入层、主干网(BackBone)、颈网(Neck)和输出层 4 个基本构件^[21]。输入层接受固定大小的图像,经过主干网进行特征提取后送到颈网进行特征聚合,最后在输出层输出 3 种不同尺度的预测锚框(原文称为 YOLO Head)。

1.1.1 主干网

YOLO-V4 设计的主干网是 CSPDarknet53,它是在 Darknet53 的基础上参考了跨阶段局部网络 CSPNet^[22]提出的。Darknet53 是 YOLO-V3 的主干网,因为它有 53 个卷积层而得名。YOLO-V4 保留了 Darknet53 的框架,但是采用了 CSP 机制,优化了梯度反向传播路径,同时在保证准确率的前提下大大地减少了计算量。图 1 给出了 CSPDarknet53 的基本结构。



注: CBM 为卷积块; CSP 为跨阶段局部网络。

Note: CBM is convolution block; CSP is cross-stage partial network.

图 1 YOLO-V4 主干网结构

Fig.1 Structure of backbone in YOLO-V4

如图 1 所示, CSPDarknet53 有 2 个主要功能单元:

1) CBM (Conv+BN+Mish), 用来控制拼接和采样等工作, 其中 Mish 是 YOLO-V4 使用的新的激励函数(YOLO-V3 使用的是 Leaky_relu); 2) CSP, 一组跨阶段残差单元(如 CSP8 就是 8 个残差单元)。

YOLO-V4 与 YOLO-V3 的主要区别是: YOLO-V4 参考了 CSPNet 机制, 对残差单元的连接结构进行了改进, 加入了一条跨阶段残差的旁路, 形成跨阶段局部连接。这样就有效减少了梯度消失的危险。简单地说, 在 YOLO-V4 的主干网中, 一个残差单元可以形成两条路径: 一条和它的下一个残差单元直接连接; 另一条和 CSP 的末端残差单元连接(旁路)。由于在 YOLO-V4 中增加了旁路, 就可以 CSP 的推理更合理, 而且有效地分割了内部梯度流, 减少了梯度消失的危险, 进而泛化了学习能力。

1.1.2 颈网

颈网是在主干网和预测输出层之间增加的一个特征传递网络, 主要作用是对主干网提取的特征值进行采样聚合, 形成不同尺度下的聚合特征。

YOLO-V4 的颈网采用的是路径聚合网络(Path Aggregation Network, PANet)结构。如图 2 所示, PANet^[23]采用特征金字塔和路径聚合技术, 使得低层信息更容易

传播到高位, 定位更加准确, 同时支持大、中、小三类目标的预测。

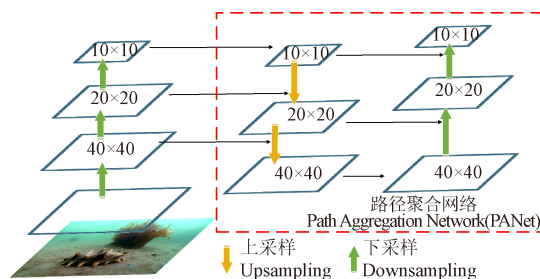


图2 PANet结构示意图

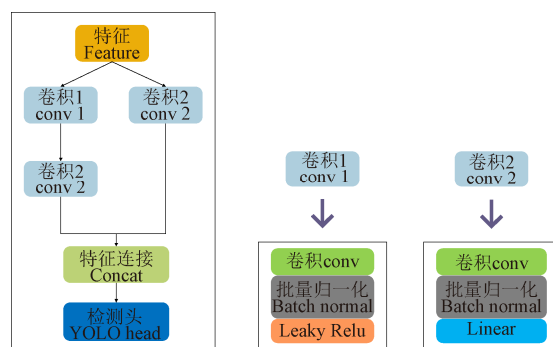
Fig.2 Schematic diagram of PANet structure

1.2 嵌连接 EC 部件设计

设计 EC 部件的目的主要是为了增强特征聚合的适应性和泛化能力。就是说, 面对浅海环境及其生物的特殊性, 需要研究高性能的部件来有效提升特征聚合能力。图3给出了EC部件的结构示意图。如图3a所示, EC部件有两条通路, 一条通过卷积1和卷积2两个单元来完成, 另一条只通过卷积2一个单元来完成。这样, 综合两条通路的结果就可以有效削弱聚合过程中的偏差, 同时卷积1和卷积2使用不同的激励函数也增加EC部件的适应性。如图3b所示, 卷积1是第一类卷积单元, 使用的是 Leaky_ReLU 激励函数, 保留了原来 YOLO-V4 特点。如图3c所示, 卷积2是第二类卷积单元, 使用的是 Linear 激励函数。从预测角度, Linear 函数更快和更容易形成可预测的输出结果。

本文将 EC 部件加入到 YOLO-V4 的颈网末端, 即在

YOLO-V4 颈网的末端增加 EC 部件, 形成了改进 YOLO-V4 网络。



a. EC 部件示意图
a. Schematic diagram of EC components

b. 卷积 1
b. Convolution 1

c. 卷积 2
c. Convolution 2

图3 嵌连接 (EC) 部件

Fig.3 Embedded Connection (EC) components

1.3 模型 MOD 设计

通过在颈网末端加入 EC 部件后, 改进了 YOLO-V4 网络, 作为本文海洋生物检测模型的技术基础。嵌入 EC 部件的 YOLO-V4 网络丰富了颈网的反向传播能力。颈网末端直接连接到 YOLO-V4 的预测端, 增强的反向传播能力有助于提高网络的泛化能力, 减少预测偏差^[24]。同时, 由于 EC 部件被嵌入到整个网络的末端, 使最终的预测加入了浅层的聚合信息, 可以有效地提升预测的精确性。

基于改进 YOLO-V4 网络, 本文设计了海洋生物检测模型 MOD (Marine Organism Detection)。图4给出了 MOD 的详细结构图。

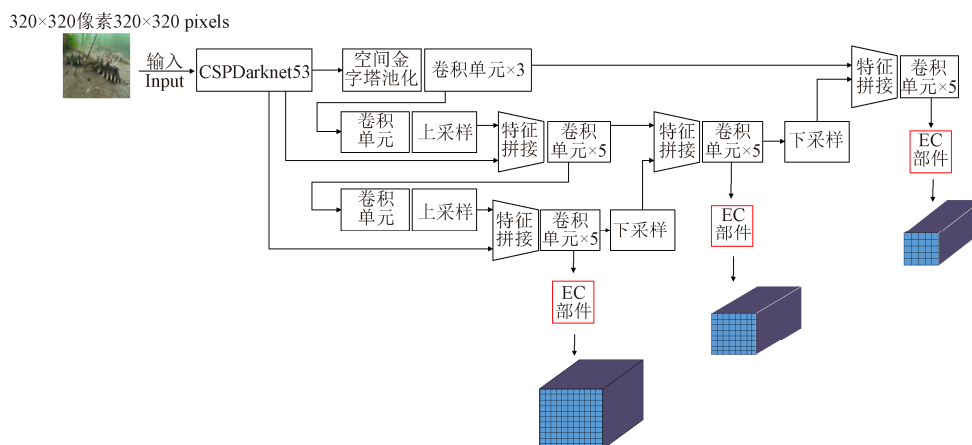


图4 MOD (海洋生物检测) 模型网络结构

Fig.4 Structure of MOD (Marine Organism Detection) model

在 MOD 的结构中, 特征提取阶段是基于 YOLO-V4 的 CSPDarknet53 来完成。主要的变化是在特征聚合阶段, 网络分别在三个尺度上提取特征, 深层信息将经过两次上采样与浅层信息聚合, 聚合后的信息再经过两次下采样分别在像素 40x40、20x20、10x10 尺度下做预测, 它们分别对应着大、中、小 3 种不同尺寸的目标做预测, 深层信息负责检测较大尺寸目标, 浅层信息负责检测较小尺寸目标, 从而提升网络的适应性。在输入信息到检

测头之前, 本文加入了 EC 部件, 这样的设计延缓了梯度消失现象的出现, 有效增强了特征的聚合能力。

2 试验与分析

2.1 试验环境与数据集

本文试验的硬件环境是: 8 G 内存的 Intel® Core™ i7-8700 为中央处理器 (Central Processing Unit), 配有图形加速器 (4 G 的 GeForce GTX 1650), 在 Windows10

计算机系统环境下进行。

使用的数据来自于獐子岛集团的浅海养殖视频数据。通过对视频分帧处理得到原始图片 1 810 张, 利用数据增强技术对原始图片集进行了容量扩充和数据预处理, 因此数据增强后的图片数据量达到 7 420 张, 作为本文训练和测试的数据集。关注的浅海生物种类有 4 种: 鲍鱼、海胆、海参和绿鳍马面鲀。使用 LabelImg 软件对筛选出的图片进行了人工标注, 其中训练集占 80%, 测试集占 20%。值得注意的是, 软件采用的是矩形框来标注检测目标, 而许多浅海生物的外形是极不规则的。以海胆为例: 海胆是呈球状的, 而且周围还布满了伸出的棘刺或管足, 因此在标注时需进行必要的预处理。在图像预处理过程中, 本文重点关注海胆的主体部分, 可以包含少部分棘刺但是不宜过多。此外, 试验中所有信息文件以 XML 格式存储, 包含对应的图片名、目标类别名称、图像尺寸、目标框位置信息等。

2.2 评价指标设定

模型预测时需要交并比 (Intersect Over Union, IOU) 阈值参数进行设定。IOU 阈值的设定与输出预测框有直接关系。一般地, 阈值越大预测准确率越高。本文试验的数据面对复杂海底环境, 检测目标与环境之间的差别较小, 因此增加 IOU 的阈值有利于提高有效目标检测精度。本文试验中使用的评价指标主要包括: mAP_{50} (IOU 阈值为 0.5 的精度均值)、 mAP_{75} (IOU 阈值为 0.75 的精度均值)、帧检测速度 (ms)。 mAP 由召回率和精确率组成的 PR 曲线进一步得到, 是衡量目标检测模型的检测精度的重要指标, 公式 (1) ~ (3) 给出了召回率、精确率和平均精度均值 (mAP) 的计算方法。

$$P = \frac{E}{E + N} \quad (1)$$

$$R = \frac{E}{E + M} \quad (2)$$

$$mAP = \frac{1}{K} \sum_{k=1}^K AP(P, R, k) \quad (3)$$

式中 P 为精确率; R 为召回率; E 为正样本正确划分为正样本的数量; N 为负样本被错误划分为正样本的数量; M 代表正样本被错误划分负样本的数量; AP 代表单类检测目标对应的 PR 曲线与坐标轴围成的积分面积; mAP 代表各个类别检测目标平均检测精度求和的均值; K 代表检测目标类别的数量, 本文试验中 $K=4$ 。

2.3 MOD 模型相比 YOLO-V4 模型有效性

本节试验旨在验证 MOD 模型的有效性, 主要是和原始的 YOLO-V4 模型进行对比试验。试验中 YOLO-V4 和 MOD 模型使用相同的训练集, 设置相同超参数: 输入图像像素尺寸为 320×320 , $batch_size$ 为 16, 初始学习率为 0.001 3, 动量为 0.949, 权重衰减正则系数为 5×10^{-4} , 迭代次数为 200。为了避免训练收敛前学习停滞, 当迭代至 160 时, 降低学习率为初始学习率的 1/10, 当迭代至 180 时降低学习率为初始学习率的 1/100。当用测试集对模型进行测试时, 设置 0.5 和 0.75 两个交并比阈值进行试验。表 1 给出了 YOLO-V4 模型和 MOD 模型迭代次数在 50、

100、150、200 时 mAP_{50} 和 mAP_{75} 的指标。依据表中数据得到, MOD 模型比 YOLO-V4 模型在训练时 mAP 上升得更快, 并且训练结束时 MOD 模型表现更好, mAP_{50} 提高了 0.9 个百分点, mAP_{75} 提高了 4.8 个百分点。表 2 给出两个模型的计算成本及检测速度, MOD 计算成本为 35.328 个 BFLOPs (十亿浮点运算数), 与 YOLO-V4 相比仅提高了 0.2%, mAP_{50} 和 mAP_{75} 均有提高。

表 1 目标检测模型测试结果

Table 1 Test results of target detection model

| 检测网络 Detection network | mAP | 迭代次数 Number of training epochs | | | |
|---------------------------|------------|-----------------------------------|-------|-------|-------|
| | | 50 | 100 | 150 | 200 |
| YOLO-V4 | mAP_{50} | 0.680 | 0.813 | 0.940 | 0.960 |
| MOD | | 0.734 | 0.852 | 0.945 | 0.969 |
| YOLO-V4 | mAP_{75} | 0.058 | 0.216 | 0.364 | 0.686 |
| MOD | | 0.137 | 0.321 | 0.450 | 0.734 |

表 2 模型计算成本及其检测速度

Table 2 Model calculation cost and detection speed

| 检测网络 Detection network | 计算成本 Calculate cost /BFLOPs | 检测帧速 Detection frame rate /ms |
|---------------------------|-----------------------------------|-------------------------------------|
| YOLO-V4 | 35.257 | 139 |
| MOD | 35.328 | 139 |

PR 曲线是一条精确率为纵轴和召回率为横轴的曲线, 观察曲线可以反应出深度学习模型的性能。一般来说, 精确率和召回率是相互制约的度量值。若将所有预测结果都回归, 那么召回率高精确率就会较低; 相反, 若只回归搜索最准确的预测结果, 那么召回率低精确率就会较高, 因此绘制 PR 曲线可以探究模型的综合性能。从图 5 看出, MOD 曲线包裹在 YOLO-V4 曲线外侧, 并且平衡点 (Precision = Recall 时的取值) 位置更接近于坐标 (1, 1), 故 MOD 模型性能更好。

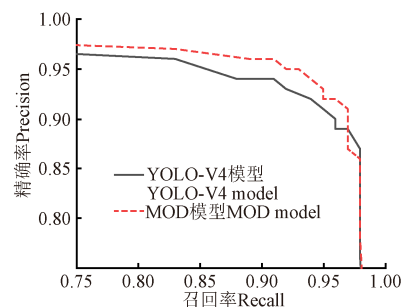


图 5 模型测试集准确率-召回率曲线

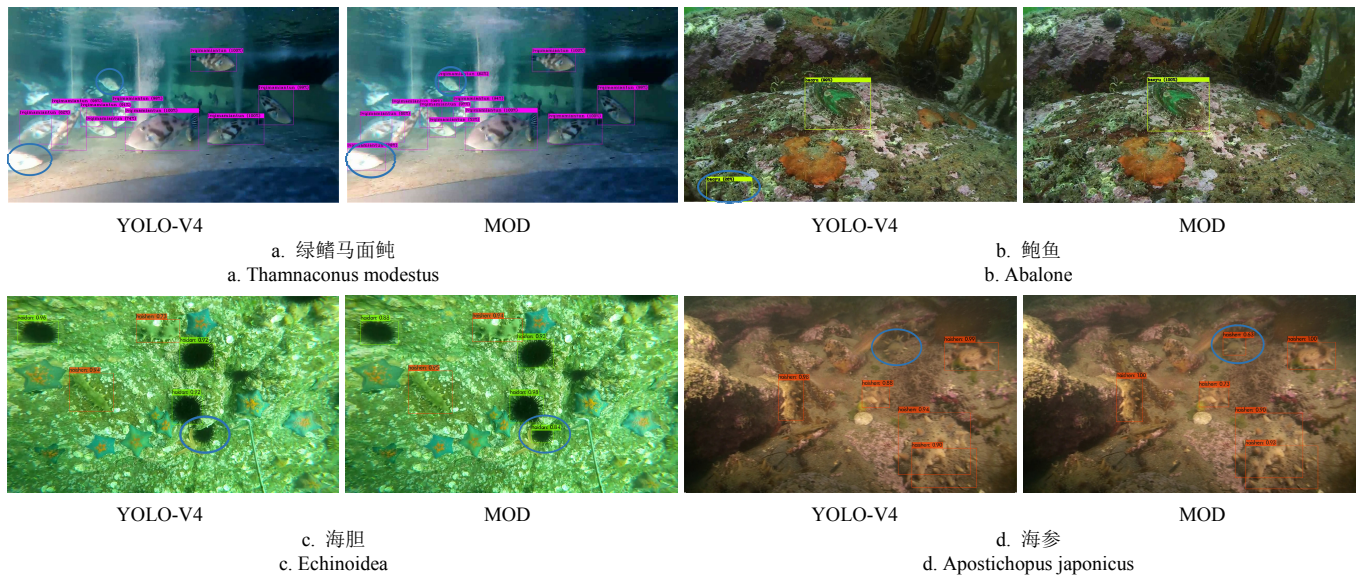
Fig.5 Precision-recall curves of models on test sets

图 6 是使用 YOLO-V4 模型与 MOD 模型分别在检测绿鳍马面鲀、鲍鱼、海胆、海参 4 类浅海生物时的测试图像。从图中可以看出, YOLO-V4 模型在检测鲍鱼的试验中存在误检现象, 检测绿鳍马面鲀、海胆和海参的试验中, 存在漏检现象, 也从一个侧面说明 MOD 模型在检测浅海生物的有效性。

进行测试试验时, 各个类别物种检测准确率结果如表 3 所示。一方面, 在 IOU 阈值为 0.5, 绿鳍马面鲀、海参和海胆的检测中, YOLO-V4 模型和 MOD 模型的检测

基本持平,而在鲍鱼的检测中, mAP_{50} 提高 3.4 个百分点;另一方面,在 IOU 阈值为 0.75, 绿鳍马面鲷、海参、海

胆和鲍鱼的检测中 mAP_{75} 均有提高,分别为 2.5、4.5、3.1、9.2 个百分点。



注: 方框: 模型预测目标; 圆框: 预测差异。
Note: Square frame: Target of model prediction; Round frame: Difference of prediction.

图 6 不同模型检测效果

Fig.6 Detection effect of different model

表 3 模型测试集各个类别准确率结果

| Table 3 Accuracy results of each category in model test set | | | | | |
|---|------------|-------------------------------|------------------------------|------------------|---------------|
| 检测网络 Detection network | mAP | 绿鳍马面鲷 Thamnaconus modestus | 海参 Apostichopus japonicus | 海胆 Echinoidea | 鲍鱼 Abalone |
| YOLO-V4 | mAP_{50} | 0.955 | 0.986 | 0.936 | 0.962 |
| MOD | | 0.956 | 0.990 | 0.933 | 0.996 |
| YOLO-V4 | mAP_{75} | 0.736 | 0.729 | 0.602 | 0.678 |
| MOD | | 0.761 | 0.774 | 0.633 | 0.770 |

2.4 更多模型的对比试验

为进一步验证 MOD 模型的有效性,本文选择了发布于 2018 年的 RetinaNet 算法^[13]和 2020 年的 EfficientDet 算法^[14]作对比试验。这两个算法和 YOLO 是目前最典型的一阶段目标检测算法。表 4 给出了两种模型的主要指标对比。

表 4 不同方法目标检测结果

| Table 4 Target detection results of different methods | | | |
|---|------------|------------|---------------------------------|
| 检测网络 Detection network | mAP_{50} | mAP_{75} | 检测帧速 Detection frame rate/ms |
| RetinaNet | 0.956 | 0.797 | 185 |
| EfficientDet_D0 | 0.648 | 0.524 | 72 |

表 4 (参考表 1) 说明,在被观察的浅海生物数据集下,本文提出的 MOD 模型在 mAP_{50} 指标方面,对比 RetinaNet、EfficientDet_D0 分别高 1.3、32.1 个百分点;在 mAP_{75} 指标方面,对比 RetinaNet 低 6.3 个百分点、EfficientDet_D0 高 21.0 个百分点。在检测速度方面,RetinaNet 模型检测帧速达 185 ms,较 MOD 模型慢 46 ms; EfficientDet 使用 D0 版本,速度较快,检测帧速达 72 ms,较 MOD 模型快 67 ms,但检测准确率较低。综合考虑检测准确率和运行速度,在 4 种模型中 MOD 模型较适合水下机器人智能捕捞任务。

3 结 论

本文针对浅海生物检测的特点,设计了嵌入式 EC 部

件,提出了海洋检测模型 MOD,并建立浅海生物数据集,在浅海生物数据集上对 MOD 模型有效性加以验证。

1) 设计嵌入式 EC 部件,将 EC 部件嵌入网络末端改进了 YOLO-V4,构建了 MOD 模型。本文提出的模型和方法为海洋生物目标检测提供新思路,可以为多类别海洋生物目标识别提供有益参考。

2) 对比原始的 YOLO-V4 模型,MOD 模型检测精度在 mAP_{50} 和 mAP_{75} 指标上分别达到了 0.969 和 0.734,提升了 0.9 和 4.8 个百分点,计算量仅提高 0.2%。特别是检测精度在 mAP_{75} 指标上获得较大提升,说明 MOD 模型比 YOLO-V4 模型对复杂环境的目标检测能力更强。

3) 对比原始的 YOLO-V4 模型,MOD 模型适应性更强。模型在各类别检测精度中均有提高,特别是在 mAP_{75} 指标上最为明显,说明 MOD 模型比 YOLO-V4 模型对各类浅海生物检测适应能力更好。

4) 对比同为第一阶段目标检测算法的 RetinaNet 和 EfficientDet_D0,MOD 模型的检测效果总体优于两种模型。速度和检测精度方面达到了很好的平衡,说明 MOD 模型更适合浅海生物检测任务。

由于条件限制,本文的研究和试验仍有一些需要进一步改进的地方。1) 受网络传输和存储容量限制,本文使用的原始图片数据只有 1 810 张,使用数据增强技术来扩充数据集容量,扩充后的数据图片达到 7 240 张,和大部分同类研究文献的试验数据相当。下一步研究将提升试验环境,采集更多的原始数据,来丰富数据集,提升研究质量;2) YOLO 系列算法目前已经出到第五代,后续将研究 YOLO-V5 算法,进一步在浅海生物数据集上进行针对性改进。

[参 考 文 献]

[1] 阙华勇,陈勇,张秀梅,等. 现代海洋牧场建设的现状与发展对策[J]. 中国工程科学, 2016, 18(3): 79-84.

- Que Huayong, Chen Yong, Zhang Xiumei, et al. Current situation and development countermeasures of modern marine ranch construction[J]. Engineering Science in China, 2016, 18(3): 79-84. (in Chinese with English abstract)
- [2] Fu L, Duan J, Zou X, et al. Fast and accurate detection of banana fruits in complex background orchards[J]. IEEE Access, 2020, 8: 196835-196846.
- [3] 孙晶. 基于深度学习的海底生物目标检测技术研究[D]. 哈尔滨: 哈尔滨工程大学, 2018.
- Sun Jing. Research on Seabed Biological Target Detection Technology Based on Deep Learning[D]. Harbin: Harbin University of Engineering, 2018. (in Chinese with English abstract)
- [4] 赵德安, 吴任迪, 刘晓洋, 等. 基于 YOLO 深度卷积神经网络的复杂背景下机器人采摘苹果定位[J]. 农业工程学报, 2019, 35(3): 164-173.
- Zhao Dean, Wu Rendi, Liu Xiaoyang, et al. Apple positioning based on YOLO deep convolutional neural network for picking robot in complex background[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(3): 164-173. (in Chinese with English abstract)
- [5] 易诗, 李欣荣, 吴志娟, 等. 基于红外热成像与改进 YOLOV3 的夜间野兔监测方法[J]. 农业工程学报, 2019, 35(19): 223-229.
- Yi Shi, Li Xinrong, Wu Zhijuan, et al. Night hare detection method based on infrared thermal imaging and improved YOLOV3[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(19): 223-229. (in Chinese with English abstract)
- [6] Li Y, Wang H, Dang L M, et al. A deep learning-based hybrid framework for object detection and recognition in autonomous driving[J]. IEEE Access, 2020, 8: 194228-194239.
- [7] 薛月菊, 黄宁, 涂淑琴, 等. 未成熟芒果的改进 YOLOv2 识别方法[J]. 农业工程学报, 2018, 34(7): 173-179.
- Xue Yueju, Huang Ning, Tu Shuqin, et al. Immature mango detection based on improved YOLOv2[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2018, 34(7): 173-179. (in Chinese with English abstract)
- [8] 李就好, 林乐坚, 田凯, 等. 改进 Faster R-CNN 的田间苦瓜叶部病害检测[J]. 农业工程学报, 2020, 36(12): 179-185.
- Li Jiuhao, Lin Lejian, Tian Kai, et al. Detection of leaf diseases of balsam pear in the field based on improved Faster R-CNN[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(12): 179-185. (in Chinese with English abstract)
- [9] 赵春江, 文朝武, 林森, 等. 基于级联卷积神经网络的西红柿花期识别检测方法[J]. 农业工程学报, 2020, 36(24): 143-152.
- Zhao Chunjiang, Wen Chaowu, Lin Sen, et al. Tomato florescence recognition and detection method based on cascaded neural network[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(24): 143-152. (in Chinese with English abstract)
- abstract)
- [10] 陈燕, 王佳盛, 曾泽钦, 等. 大视场下荔枝采摘机器人的视觉预定位方法[J]. 农业工程学报, 2019, 35(23): 48-54.
- Chen Yan, Wang Jiasheng, Zeng Zeqin, et al. Vision pre-positioning method for litchi picking robot under large field of view[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(23): 48-54. (in Chinese with English abstract)
- [11] 燕红文, 刘振宇, 崔清亮, 等. 基于特征金字塔注意力与深度卷积网络的多目标生猪检测[J]. 农业工程学报, 2020, 36(11): 193-202.
- Yan Hongwen, Liu Zhenyu, Cui Qingliang, et al. Multi-target detection based on feature pyramid attention and deep convolution network for pigs[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(11): 193-202. (in Chinese with English abstract)
- [12] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016.
- [13] Lin T, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017.
- [14] Tan M, Pang R, Le Q. Efficientdet: Scalable and efficient object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020.
- [15] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv: 2004. 10934, 2020.
- [16] Wang C, Samani H. Object detection using transfer learning for underwater robot[C]//2020 International Conference on Advanced Robotics and Intelligent Systems (ARIS). Taipei, Taiwan: IEEE, 2020.
- [17] Xia C, Fu L, Liu H, et al. In situ sea cucumber detection based on deep learning approach[C]//2018 OCEANS-MTS/ IEEE Kobe Techno-Oceans (OTO). Kobe, Japan: IEEE, 2018.
- [18] 朱世伟, 杭仁龙, 刘青山. 基于类加权 YOLO 网络的水下目标检测[J]. 南京师大学报: 自然科学版, 2020, 43(1): 129-135.
- Zhu Shiwei, Hang Renlong, Liu Qingshan. Underwater object detection based on the class-weighted YOLO net[J]. Journal of Nanjing Normal University: Natural Science Edition, 2020, 43(1): 129-135. (in Chinese with English abstract)
- [19] 赵德安, 刘晓洋, 孙月平, 等. 基于机器视觉的水下河蟹识别方法[J]. 农业机械学报, 2019, 50(3): 151-158.
- Zhao Dean, Liu Xiaoyang, Sun Yueping, et al. Detection of underwater crabs based on machine vision[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(3): 151-158. (in Chinese with English abstract)
- [20] 李庆忠, 李宜兵, 牛炯. 基于改进 YOLO 和迁移学习的水下鱼类目标实时检测[J]. 模式识别与人工智能, 2019, 32(3): 193-203.
- Li Qingzhong, Li Yibing, Niu Jiong. Real-time detection of underwater fish based on improved YOLO and transfer learning[J]. Pattern Recognition and Artificial Intelligence,

- 2019, 32(3): 193-203. (in Chinese with English abstract)
- [21] 燕红文, 刘振宇, 崔清亮, 等. 基于改进 Tiny-YOLO 模型的群养生猪脸部姿态检测[J]. 农业工程学报, 2019, 35(18): 169-179.
- Yan Hongwen, Liu Zhenyu, Cui Qingliang, et al. Detection of facial gestures of group pigs based on improved Tiny-YOLO[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(18): 169-179. (in Chinese with English abstract)
- [22] Wang C, Liao H, Wu Y, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, WA, USA: IEEE, 2020.
- [23] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018.
- [24] He M X, Hao P, Xin Y Z. A robust method for wheatear detection using UAV in natural scenes[J]. IEEE Access, 2020, 8: 189043-189053.

Model for marine organism detection in shallow sea using the improved YOLO-V4 network

Mao Guojun^{1,2}, Weng Weidong¹, Zhu Jinde¹, Zhang Yuan³, Wu Fucun⁴, Mao Yuze^{5*}

(1. School of Computer and Mathematics, Fujian University of Technology, Fuzhou 350118, China; 2. Fujian Provincial Key Laboratory of Big Data Mining and Applications, Fuzhou 350118, China; 3. Zonco Group Co., Ltd., Dalian 116001, China; 4. Key Laboratory of Experimental Marine Biology, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China; 5. Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao 266071, China)

Abstract: Intelligent detection of marine organisms is a significant step of marine ranching strategy. An underwater robot is highly demanding to rapidly and accurately monitor marine organisms in the complex ocean environment. However, there is a relatively low distinction between marine organisms and their living environment, some of which are covered or semi-hidden, due mainly to the low contrast of seabed environment, and uneven distribution of brightness. Therefore, it is a big challenge to accurately identify the specific marine life in the undersea environment. Many target (object) detections using deep learning have emerged, such as EfficientDet, RetinaNet, and YOLO-V4, with the development of convolutional neural networks (CNN) in recent years. Nevertheless, the current network cannot fully meet the specific requirements of marine biological recognition. It is also necessary to improve the detection accuracy, operation speed, and detection efficiency of dense targets. In this study, an improved target (object) detection network using YOLO-V4 was designed to realize the rapid and accurate identification of marine organisms in an aquaculture environment of a shallow sea. A marine biological dataset was firstly established with 7 240 images, which were generated from 1 810 original images after data enhancement. Training (80%) and test datasets (20%) were divided. Data enhancement (suitable for the small data sample learning) effectively enriched the background and elements of the original images, thereby producing much more learning samples than before. As such, an effective expansion of the sample was achieved in the same learning effect as the large sample. Secondly, the Cross-Stage Partial network (CSP) was successfully introduced, while the Embedded Connection (EC) component was designed to detect marine organisms. An improved YOLO-V4 network model was constructed, when the EC was embedded into the end of the YOLO-V4 network. The improved YOLO-V4 network with an EC can be expected to make the gradient flow propagate on different learning paths, while effectively delay the occurrence of gradient disappearance, aiming to improve the detection accuracy and cost-saving calculation. Finally, Marine Organism Detection (MOD) was presented using the improved YOLO-V4 network to achieve a better performance in the complex seabed environments. The experimental results showed that the mAP₅₀ and mAP₇₅ of the MOD model were 0.969 and 0.734, respectively, while the computational complexity was 35.328 billion floating-point operations (BFLOPs), and the detection frame rate was 139 ms on the computer system with a graphics accelerator GeForce GTX 1650. The mAP₅₀ and mAP₇₅ from the MOD increased by 0.9 percent points and 4.8 percent points, respectively, while the amount of computation only increased by 0.2%, compared with the original YOLO-V4 model. Especially, the evaluating indicators in the MOD model improved in all studied categories, where mAP₇₅ presented the most obvious. In addition, the precision and recall values of balance points in the MOD model were closer to (1, 1) in most cases. It can also be reasonable that the learning performance was better in the MOD than the original YOLO-V4 model, compare with the PR curves. Consequently, the finding can provide promising insightful ideas and useful references for the rapid and accurate detection of the marine organisms in an underwater robot of intelligent fishing.

Keywords: models; deep learning; object detection; YOLO-V4; cross-stage partial network; embedded connection