

采用改进 SSD 网络的海参目标检测算法

张 岚¹, 邢博闻^{1,2*}, 李 彩², 李硕峰³

(1. 上海海洋大学工程学院, 上海 201306; 2. 热带海洋环境国家重点实验室(中国科学院南海海洋研究所), 广州 510308;
3. 哈尔滨工程大学智能科学与工程学院, 哈尔滨 150001)

摘要: 随着海参养殖业快速发展, 利用水下机器人代替人工作业的海参智能捕捞已成为发展趋势。浅海环境复杂, 海参体色与环境区分性差、海参呈现半遮蔽状态等原因, 导致目标识别准确率低下。此外由于景深运动, 远端海参作为小目标常常未被识别成功。为解决上述问题, 该研究提出一种基于改进 SSD 网络的海参目标检测算法。首先通过 RFB (Receptive Field Block) 模块扩大浅层特征感受野, 利用膨胀卷积对特征图进行下采样, 增加海参细节、位置等信息, 并结合注意力机制, 对不同深度特征进行强化, 将计算得出的权重与原特征信息相乘以此获得特征图, 使结果包含最具代表性的特征, 也抑制无关特征。最后实现特征图融合, 进一步提升水下海参的识别精度。以实际拍摄的视频进行测试验证, 在网络结构层面上, 对传统算法进行改进。试验结果表明, 基于改进的 SSD 网络的海参目标检测算法的平均精度均值为 95.63%, 检测帧速为 10.70 帧/s, 相较于传统的 SSD 算法, 在平均精度均值提高 3.85 个百分点的同时检测帧速仅减少 2.8 帧/s。与 Faster R-CNN 算法和 YOLOv4 算法进行对比试验, 该研究算法在平均精度均值指标上, 分别比 YOLOv4、Faster R-CNN 算法提高 4.19 个百分点、1.74 个百分点。在检测速度方面, 该研究算法较 YOLOv4、Faster R-CNN 算法分别低 4.6 帧/s、高 3.95 帧/s, 试验结果表明, 综合考虑准确率与运行速度, 改进后的 SSD 算法较适合进行海参智能捕捞任务。研究结果为海参智能捕捞提供参考。

关键词: 图像识别; 深度学习; 算法; 海参捕捞; SSD 网络

doi: 10.11975/j.issn.1002-6819.2022.08.034

中图分类号: TP391.4

文献标志码: A

文章编号: 1002-6819(2022)-08-0297-07

张岚, 邢博闻, 李彩, 等. 采用改进 SSD 网络的海参目标检测算法[J]. 农业工程学报, 2022, 38(8): 297-303. doi:

10.11975/j.issn.1002-6819.2022.08.034 <http://www.tcsae.org>

Zhang Lan, Xing Bowen, Li Cai, et al. Algorithm for detecting sea cucumbers based on improved SSD[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(8): 297-303. (in Chinese with English abstract)

doi: 10.11975/j.issn.1002-6819.2022.08.034 <http://www.tcsae.org>

0 引言

近年来, 随着全世界海参需求量的增长, 海参养殖业也得到快速发展, 并创造了巨大的经济效益^[1]。海参养殖业的快速发展也暴露出诸多问题, 主要体现在依旧采用传统人工捕捞的作业方式, 存在效率低、风险高等问题。因此采用水下机器人进行智能捕捞已成为行业发展的主要趋势^[2-3], 该技术的核心是完成海参目标的智能检测。

目前, 深度学习算法在目标检测领域中不断发展应用, 逐渐成为实际生产中目标识别的重要手段之一。该技术的优势在于可进行目标特征的自动提取与利用, 并在大量传统识别应用中取得了良好的效果^[4-7]。Wang 等^[8]通过 BP 神经网络进行海参目标识别, 利用 RGB 图像和深度图像作为提高准确性的先验知识, 以此提高识别精度。Guo 等^[9]采用不同配置的深度残差网络进行海参目标

识别, 具有较强的鲁棒性。Yao 等^[10]提出了基于 SSD (Single Shot MultiBox Detector) 和深度可分离卷积的海参检测方法, 利用 MobileNet-SSD 来检测海参, 具有运算量少, 实时性较高的优点。上述学者及其他研究人员在对海参目标识别深入研究后, 均提出了一系列具有实际应用价值的方法。但绝大多数研究只针对静态海参, 未考虑到水下机器人运动中景深变化引起的海参小目标现象, 因此存在漏检等误差情况。

本文针对上述自然水域中存在的海参目标识别问题, 提出了一种改进 SSD 网络的目标检测算法。在传统 SSD 算法基础上, 通过引入 RFB (Receptive Field Block) 模块与注意力机制, 来增强深层特征语义信息, 同时对浅层特征进行下采样以此扩大感受野, 实现水下海参目标的高精度检测, 以期能为海参智能捕捞提供参考。

1 SSD 目标检测算法

传统 SSD 目标检测算法由 Liu 等^[11]提出。此类目标检测算法的核心是利用不同卷积层的特征图对结果进行预测。

1.1 传统 SSD 算法的模型结构

传统 SSD 算法使用 VGG16 作为主干网络, 并将 VGG16 的两个全连接层转换为卷积层, 为后续的多尺度特征提取奠定基础。传统 SSD 算法网络结构将 6 个特征

收稿日期: 2021-12-12 修订日期: 2022-03-30

基金项目: 热带海洋环境国家重点实验室(中国科学院南海海洋研究所)开放课题(LTOZZ1917); 上海市 2022 年度地方院校能力建设项目(22010502200)

作者简介: 张岚, 研究方向为机器视觉、图像识别。

Email: zhang258708921@163.com

*通信作者: 邢博闻, 副教授, 机器人工程本科专业负责人, 研究方向为海洋环境监测、海洋机器人技术。Email: bwxing@shou.edu.cn

层与最后检测分类层相连接, 形成不同维度特征图, 进行回归预测^[12-14]。

1.2 传统 SSD 算法的损失函数

传统 SSD 算法损失函数分为两个部分: 分类损失函数和定位损失函数^[15], 即:

$$L(x, c, l, g) = \frac{1}{N} (L_{conf}(x, c) + \alpha L_{loc}(x, l, g)) \quad (1)$$

式中 L_{conf} 为分类损失函数; L_{loc} 为定位损失函数; N 为样本数量; α 为加权系数; x 为当前搜索预测框类别的匹配信息; c 为标注类别; l 表示搜索预测框边界坐标; g 表示标注边界框坐标。

其中定位损失函数如下式所示, $x_{ij}^k = \{1, 0\}$ 表示第 i 个搜索预测框与第 j 个真实框关于类别 k 是否匹配; l_i^m 为预测框; \hat{g}_j^m 为真实框; N_i 为所匹配的样本数量; Pos 为正样本; Box 为搜索预测框中心点坐标以及预测框宽和高的集合; $smooth_{L1}$ 为 $L1$ 的误差函数。

$$L_{loc}(x, l, g) = \sum_{i \in Pos} \sum_{m \in Box} x_{ij}^k smooth_{L1}(l_i^m - \hat{g}_j^m) \quad (2)$$

分类损失函数如下式所示, \hat{c}_i^p 表示第 i 个搜索预测框中目标为 P 的概率; \hat{c}_i^0 表示第 i 个搜索预测框中未检测出目标的概率; x_{ij}^p 表示为搜索预测框 i 与真实框 j 关于 P 类别匹配情况; Neg 为负样本; Pos 为正样本。

$$L_{conf}(x, c) = - \sum_{i \in Pos} x_{ij}^p \lg(\hat{c}_i^p) - \sum_{i \in Neg} \lg(\hat{c}_i^0) \quad (3)$$

$$\hat{c}_i^p = \exp(c_i^p) / \sum_p \exp(c_i^p) \quad (4)$$

1.3 传统 SSD 算法目标检测性能分析

传统 SSD 算法目标检测算法采用多尺度特征图进行目标预测, 即采用大小不同的特征图进行预测, 并在卷积生成的 6 个特征图上分别生成 6 个长宽比不同的先验框^[16-18]。

如图 1 所示, 较低层级的特征图较大, 但相对的单位感受野比较小, 适合用于检测小目标; 较高层级的特征图较小, 相对的单位感受野较大, 适合检测大目标。因此对于小目标的识别常依赖于底层特征图, 然而底层特征卷积层数较少, 使得特征提取并不充分, 导致语义区分较差。

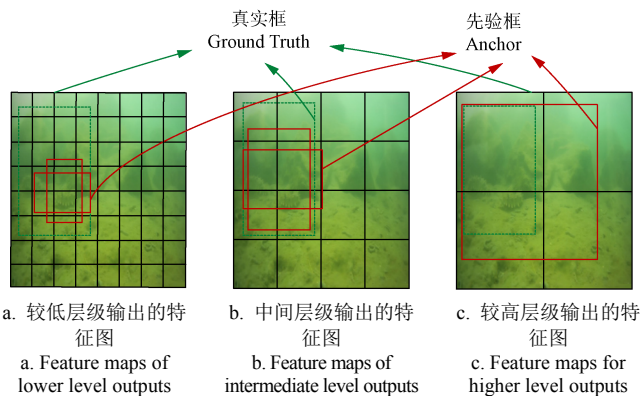


图 1 传统 SSD 算法不同层级输出结果

Fig. 1 Output results from different levels of SSD algorithm

水下机器人识别水下海参时, 由于景深差异, 远端海参呈现为小目标。在使用传统 SSD 目标检测算法识别水下海参时, 由于对海参细节检测能力不足, 无法对水下海参进行精确定位。为弥补该缺点, 需要对传统 SSD 算法进行优化。

2 基于 RFB 和注意力机制的 SSD 算法目标检测

2.1 引入 RFB 模块的 SSD 网络

针对传统 SSD 算法依赖于浅层特征对小目标进行识别所导致的识别率不佳的情况, 本文通过引入 RFB 模块, 在残差网络的基础上加入了膨胀卷积结构, 有效增加目标识别感受野, 进一步提高检测精度^[19-20]。

RFB 模块与残差网络结构相似, 均为多分支卷积模块。由多分支卷积层和膨胀卷积层两部分组成。膨胀卷积是在标准卷积的基础上加入空洞, 以此增加感受野。因此膨胀卷积在标准卷积的基础上增加了一个超参数, 即膨胀因子 K_d ^[21]。膨胀后的卷积核尺寸:

$$K_d = L \cdot (K - 1) + 1 \quad (5)$$

式中 K_d 为经过膨胀后的卷积核尺寸; L 为膨胀因子; K 为原始的卷积核尺寸。

当膨胀因子为 2 时, 在 3×3 卷积核下, 膨胀卷积后卷积核尺寸为 5×5 , 相较于传统卷积拥有更大的感受野, 使每个卷积包含更多语义信息^[22-24]。

图 2 所示为 RFB 模块处理流程。RFB 模块由 3 个分支模块组成, 每个分支的底层分别经过 1×1 、 3×3 、 5×5 不同尺寸卷积核进行处理, 之后分别进行膨胀因子为 1、3、5 的 3×3 卷积核空洞卷积, 从而获得 3 个不同感受野, 最后将所有的特征图进行连接处理^[25]。

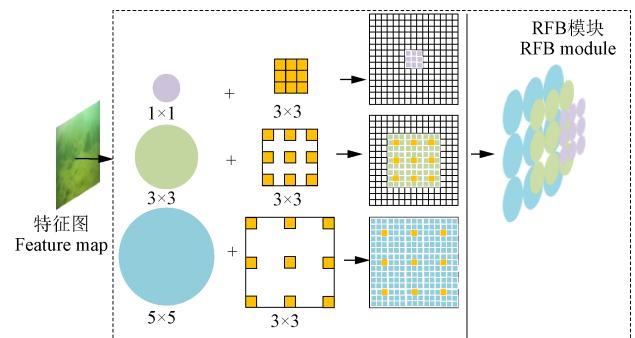


图 2 RFB 模块处理流程图

Fig. 2 RFB module processing flowchart

引入 RFB 模块仅对浅层特征进行处理, 以此获得更加丰富的浅层语义信息。针对深层特征图同样需要进行特征处理的问题, 本文算法进一步引入注意力机制, 对各深度特征图进行不同程度的强化。

2.2 引入注意力机制的 SSD 网络

注意力机制是模拟人类注意力的重要方法。本文在改进的 SSD 网络基础上引入空间注意力机制和通道注意力机制, 对不同深度的特征进行提取强化。将最终得到的多层特征图进行融合处理, 期望取得更优异的检测效果, 进而提升对水下海参的识别精度。

首先对不同层次的特征引入空间注意力机制。当输入一幅多通道特征图后，空间注意力会对该特征图中不同空间作用域关系进行学习训练，对更具有代表性的局部特征赋予更高权重后，生成二维空间权重图 W 。将二维权重图 W 与相对应的位置空间相乘，从而获得具有代表性的特征图。空间注意力的训练机制如下：首先通过全局最大池化和全局平均池化操作，在每一个空间位置得到两个特征代表值，然后通过 3×3 的卷积运算($f^{3 \times 3}$)将这两个特征代表值融合处理，得到空间注意力图，最后利用 sigmoid 激活函数生成 $0 \sim 1$ 的空间注意力权重，计算公式如下：

$$W_s(F) = \text{sigmoid}\{f^{3 \times 3}\{[\text{MaxPool}(F); \text{AvgPool}(F)]\}\} \quad (6)$$

式中 $W_s(F)$ 为经过空间注意力机制处理后的特征图； F 为输入的多通道特征图； $f^{3 \times 3}$ 为 3×3 的卷积运算； AvgPool 为平均池化； MaxPool 为最大池化。

继续引入通道注意力，输入多通道特征图，通道注意力机制首先过滤掉无关的通道特征，利用特征图中每个通道间的关系来学习得到权重数组；最后将其乘以相对应的通道^[26]。本质上通道注意力机制是针对更具代表性的通道进行学习训练；对通道特征进行重新标定；对具有高响应的通道特征给予更高的权重使其变为强特征，以此获得通道注意力特征图^[27]。通道注意力机制计算公式如下所示。

$$W_s'(F) = \text{sigmoid}\{\text{MLP}[\text{AvgPool}(F)] + \text{MLP}[\text{MaxPool}(F)]\} \quad (7)$$

式中 $W_s'(F)$ 为结果特征图； MLP 为多层感知器； AvgPool 为平均池化； MaxPool 为最大池化。

引入注意力机制的 SSD 算法流程如图 3 所示。在进行目标识别任务时，若以同样的关注权重对待每一个特征，不仅会影响识别精度，还会增加网络收敛所需要的时间。因此引入注意力机制后在一定程度上提升了精度均值，进一步提升水下海参的识别精度。

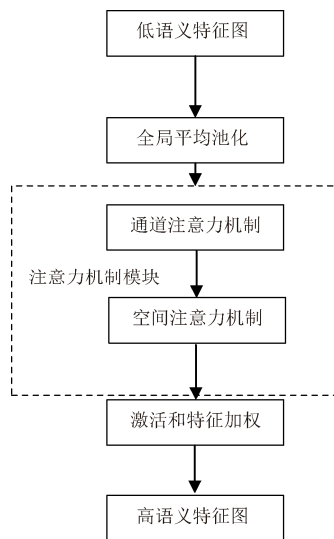


图 3 引入注意力机制的 SSD 算法流程
Fig.3 SSD algorithm flow introducing attention mechanism

2.3 改进后的 SSD 网络结构

改进后的 SSD 网络结构，在加入 RFB 模块的基础上，对已生成的 6 个特征图引入注意力机制。通过 RFB 模块模拟人类视觉的感受野，进一步加强网络的特征提取能力，使浅层特征包含更多的语义信息，提升对海参目标的识别能力^[28]。引入注意力机制，首先对各相关特征通道、特征空间的关系进行建模；之后将建模所得的对应权重与原特征信息相乘以此获得空间特征图与通道特征图。特征图既包含了最具代表性的特征，也抑制无关特征^[29-30]。通过改进的 SSD 网络结构，不仅能够提高对海参目标的识别精度，还能进一步降低漏检率、误检率。

本文首先选择 Conv4_3 的特征和 FC7 层的特征进行特征融合，融合后经过 RFB 模块，获得新的有效特征层 P1，再次进行卷积操作，以获得符合要求的特征图^[30]。接下来对 FC7 层单独进行 RFB 处理，增大感受野，获得特征图 P2，对 P2 进行 stride 为 2 的 RFB 处理，获得 P3，以此获得 P1(38×38)、P2(19×19)、P3(10×10)、P4(5×5)、Conv10_2、Conv11_2 六个不同尺度的有效特征图。传统的 SSD 算法网络结构如图 4 所示。改进后的 SSD 网络结构如图 5 所示。

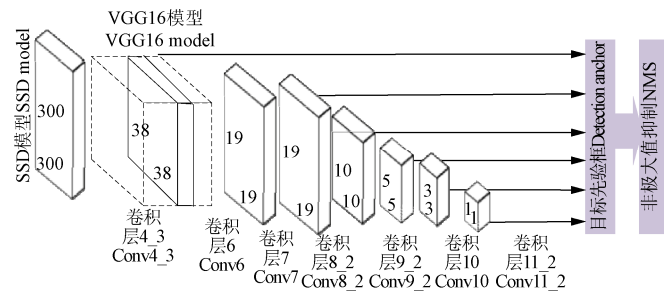
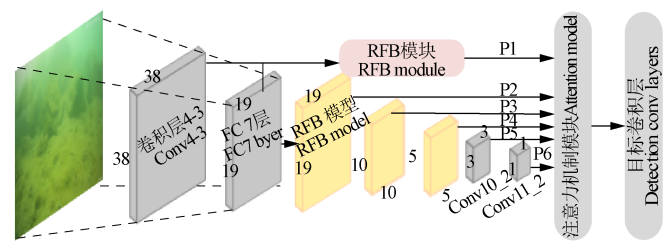


图 4 传统 SSD 算法网络结构图
Fig.4 Network-structure diagram of traditional SSD algorithm



注：P1-P6 为模型预测目标。

Note: P1-P6 are model prediction targets.

图 5 改进后的 SSD 网络结构
Fig.5 Improved SSD network structure

改进后的 SSD 网络结构，采用了膨胀卷积对特征图进行了下采样^[30]，将深层特征映射成浅层输入，进行更深层次的特征提取，使得低层特征图具有更大的感受野，并且具备更多的语义信息，进一步提升对小目标的检测精度。

3 材料与方法

本文以实际拍摄的视频进行了测试验证，针对传统 SSD 算法水下海参识别率低的问题，在网络结构层面上

提出了改进算法。为了验证本文算法的可行性, 本文所采用的计算环境为: CPU 为 Intel(R)Core(TM) i7-9750H, GPU 选用 NVIDIA GeForce RTX 2060, 操作系统选用 Ubuntu20.04, 使用的深度学习框架为 keras 2.1.5。

3.1 试验数据

试验数据来自于大连瓦房店谢屯镇的浅海养殖海参视频。通过对视频进行分帧处理, 得到原始海参图片 1 710 张, 利用数据扩增对图像数据进行翻转、剪切、旋转、缩放等操作, 对原始数据集进行扩充, 得到 4 005 张扩增处理后的图片数据, 以此作为本文的数据集。使用 LabelImg 软件对筛选的图片进行人工标注, 其中 70% 作为训练集, 30% 作为测试集。LabelImg 软件采用矩形框来标注海参目标, 重点关注海参的主体部分, 标签信息以 XML 格式存储, 包含对应的图片名、类别名、图像大小、标注框的位置信息等。

3.2 评价指标设定

在利用模型进行预测前, 首先需要对交并比 (Intersection-over-Union, IoU) 参数进行设置。交并比为目标检测中的一项概念, 是算法产生的候选框与原标记框的交叠率, 即它们的交集与并集的比值。本文主要使用 IOU 阈值为 0.5 的平均精度均值 (mAP₅₀)、帧检测速度 (帧/s)。平均精度均值 (mAP) 是目标检测模型中常用的评价指标, 由召回率 (Recall) 和精确率 (Precision) 组成的 P-R 曲线计算得到^[31]。式 (8)、(9)、(10) 为精确率、召回率、mAP 计算公式。

$$\text{Precision} = \text{TP}/(\text{TP}+\text{FP}) \quad (8)$$

$$\text{Recall} = \text{TP}/(\text{TP}+\text{FN}) \quad (9)$$

$$\text{mAP} = \sum_{i=1}^k \text{AP}_i / k \quad (10)$$

式中 FP 为假正例; FN 为假负例; TP 为真正例; TN 为真负例; AP_i 是指一个类别的平均精度; k 为类别数目。

4 结果与分析

4.1 改进后的 SSD 模型有效性验证

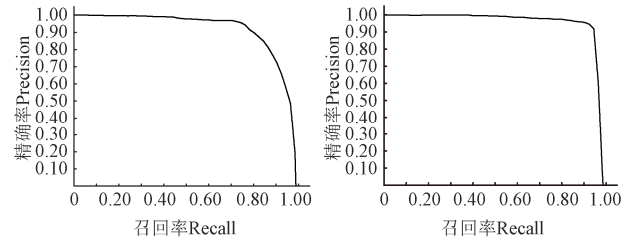
为验证改进后的 SSD 模型的有效性, 本文与传统的 SSD 模型的识别效果进行对比试验。改进的 SSD 模型与传统 SSD 模型使用相同的训练集, 设置如输入图像尺寸为 300×300 像素, batch_size 为 16 等一系列相同的参数。在训练阶段利用迁移学习的思想, 将训练分为冻结阶段和解冻阶段。在冻结阶段, 学习率为 5×10⁻⁴, 冻结模型主干, 对网络进行微调。在解冻阶段, 学习率为 1×10⁻⁴, 解冻模型主干, 调整特征提取网络。用测试集对模型进行测试时, 以通过计算 mAP₅₀ 来进行对比试验。

如表 1 所示, 与传统算法相比, 本文提出的算法在海参数据集上平均精度均值提高 3.85 个百分点的同时检测帧速减少 2.8 帧/s, 改进的 SSD 算法检测帧速基本满足海参识别要求, 实现了对海参的检测精度的有效提高。

图 6 为模型改进前后 P-R 曲线对比图。利用 P-R 曲线反映模型性能。

表 1 算法改进前后测试结果

检测网络 Detection network	平均精度均值 Mean Average Precision mAP ₅₀ /%	迭代次数 The number of iterations	检测帧速 Detect frame rate (帧·s ⁻¹)
本文算法	95.63	200	10.70
传统 SSD 算法	91.78		13.50



a. 传统 SSD 模型
b. 改进 SSD 模型

图 6 模型改进前后 P-R 曲线对比图

Fig.6 Comparison of P-R curves before and after model improvement

如图 6 所示, 改进后的 SSD 模型的 P-R 曲线与坐标轴 X, Y 之间的面积更大, 并且其平衡点更接近于坐标点 (1,1), 说明改进后的 SSD 模型的系统性能较好。

图 7 是利用改进前后的 SSD 模型检测海参的测试图像。设置 IOU 阈值为 0.5, 在对海参的检测中, 如图 7 所示, 图 7a 对比显示传统的 SSD 算法发生误检。图 7b 结果显示, 传统的 SSD 算法针对大小海参目标时置信度仅为 69%, 75%。改进后的 SSD 算法针对大小海参目标时, 置信度分别提高了 6 个百分点、24 个百分点。图 7c 单目标识别对比结果显示, 本文算法较传统的 SSD 算法, 置信度提高 12 个百分点。图 7d 对比显示传统的 SSD 算法发生了漏检。由于数据集为水下机器人采集的视频分帧得到, 图 7c、图 7d 为相邻帧, 对比得到随着物距增加, SSD 算法出现了漏检, 而本文算法仍能识别出小目标海参。根据图 7 结果显示, 传统的 SSD 算法会发生误检、漏检现象, 主要因为在自然环境下, 由于水下机器人运动, 远端海参常常为小目标, 小目标识别更依赖于低层特征图, 而低层特征图进行多次卷积丢失了较多细节信息。本文算法在一定程度上降低了漏检率, 并且相较于传统算法, 本文算法随着物距增加仍能对远端海参小目标进行识别; 并且被传统 SSD 算法误检的部分也能被本文算法正确检测, 在一定程度上降低了误检率。

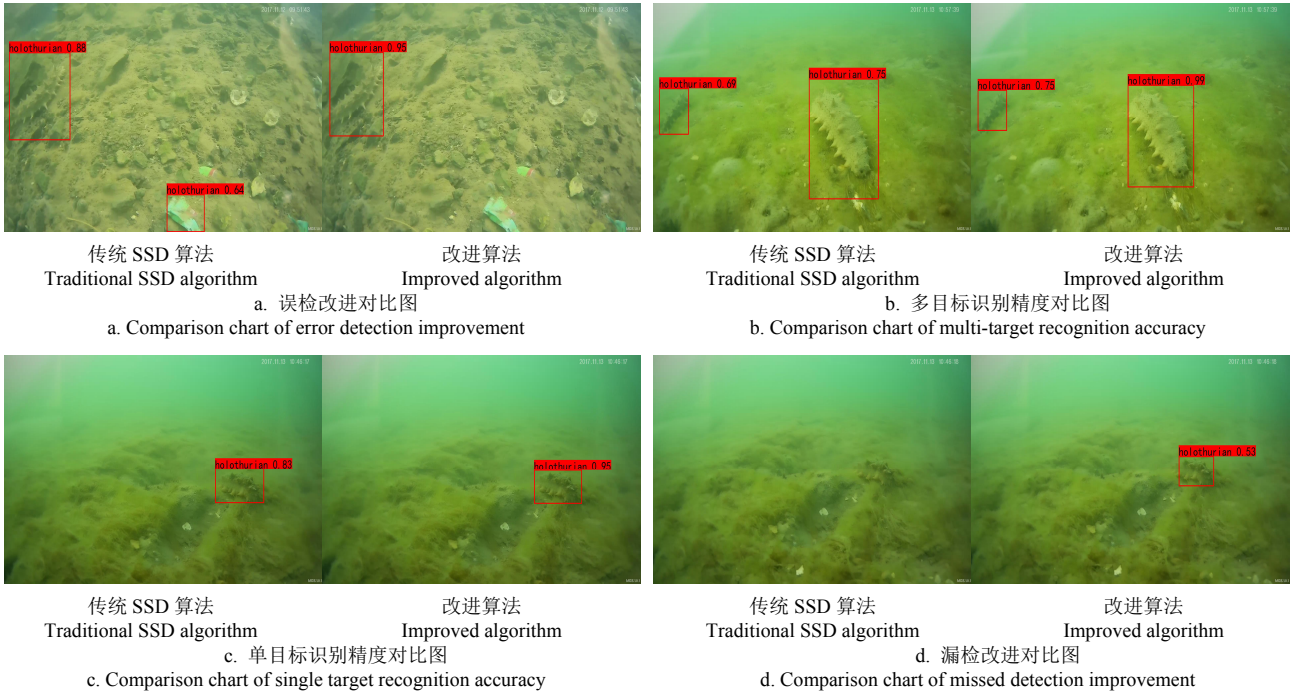
4.2 不同模型对比

为了进一步证明本文算法的有效性, 本文选择 Faster R-CNN 算法和 YOLOv4 算法进行对比试验。Faster R-CNN 算法与 YOLOv4 为目前典型的一阶段、二阶段目标检测算法。图 8 为 4 种算法的 P-R 曲线对比图。

参考图 8、表 2 说明, 在同一个海参数据集下, 本文算法在 P-R 曲线上的表现, 优于 YOLOv4 与 Faster R-CNN, 说明本文算法性能更佳。本文算法在 mAP₅₀ 指标上, 分别比 YOLOv4、Faster R-CNN 算法提高 4.19 个百分点、1.74 个百分点。在检测速度方面, 本文算法较 YOLOv4、Faster R-CNN 算法分别低 4.6 帧/s、高 3.95 帧/s。

综合考虑，目标检测算法目标识别性能，本文算法较合

适进行水下海参识别工作。



注：方框为模型预测目标。
Note: Square frame is a target of the model prediction.

图 7 模型改进前后海参识别对比图

Fig. 7 Comparison of sea cucumber identification before and after model improvement

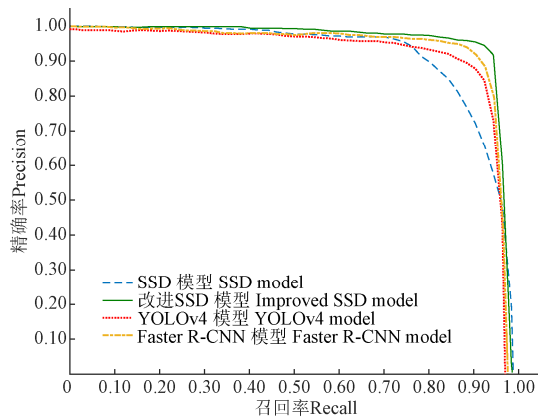


图 8 4 种模型 P-R 曲线对比图

Fig.8 Comparison of P-R curves of four models

表 2 3 种算法测试结果
Table 2 Three algorithms test results

检测网络 Detection network	平均精度均值 Mean Average Precision mAP ₅₀ /%	迭代次数 The number of iterations	检测帧速 Detect frame rate (帧·s ⁻¹)
本文算法	95.63		10.70
YOLOv4	91.44	200	15.30
Faster R-CNN	93.89		6.75

5 结 论

本文针对实际环境下海参目标特点，设计了改进的 SSD 算法，在建立海参数数据集基础上，对改进的 SSD 算法进行了有效性验证。

1) 设计加入 RFB (Receptive Field Block) 模块，增

大浅层特征感受野，并引入注意力机制，对不同深度特征进行强化，得到更加精准的海参细节。使得位于不同景深的远端和近端海参目标均能够成功识别，从而提高动态下海参目标的识别精度。

2) 相较于传统 SSD 模型，本文算法平均精度提升 3.85 个百分点，为 95.63%；检测帧速度仅减少 2.8 帧/s。说明了本文算法较传统的 SSD 算法具有更强的目标检测能力。

3) 对比 YOLOv4 与 Faster R-CNN 算法，本文算法在 mAP₅₀ 指标上分别比 YOLOv4、Faster R-CNN 算法提高 4.19 个百分点、1.74 个百分点。在检测速度方面，本文算法较 YOLOv4、Faster R-CNN 算法分别低 4.6 帧/s、高 3.95 帧/s。综合考虑，本文算法的目标检测能力较好，海参目标识别适用性更高。

下一步的研究，后续将对算法实时性进行优化，并结合水下海参体色随环境变化的特点，对海参图像进行预处理+识别，以实现海参的高效识别。

[参 考 文 献]

[1] 董鹏, 周烽, 赵惊惊, 等. 基于双目视觉的水下海参尺寸自动测量方法[J]. 计算机工程与应用, 2021, 57(8): 271-278.
Dong Peng, Zhou Feng, Zhao Congcong, et al. Automatic measurement of underwater sea cucumber size based on binocular vision[J]. Computer Engineering and Applications, 2021, 57(8): 271-278. (in Chinese with English abstract)

[2] 岳冬冬, 方辉, 樊伟, 等. 中国智能渔业发展现状与技术需求探析[J]. 渔业信息与战略, 2019, 34(2): 79-88.
Yue Dongdong, Fang Hui, Fan Wei, et al. On the status and

- technical demands of intelligent fisheries development in China[J]. *Fishery Information and Strategy*, 2019, 34(2): 79-88. (in Chinese with English abstract)
- [3] 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.
Mao Guojun, Wen 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)
- [4] 李道亮, 包建华. 水产养殖水下作业机器人关键技术研究进展[J]. *农业工程学报*, 2018, 34(16): 1-9.
Li Daoliang, Bao Jianhua. Research progress on key technologies of underwater operation robot for aquaculture[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2018, 34(16): 1-9. (in Chinese with English abstract)
- [5] 高芳芳, 武振超, 索睿, 等. 基于深度学习与目标跟踪的苹果检测与视频计数方法[J]. *农业工程学报*, 2021, 37(21): 217-224.
Gao Fangfang, Wu Zhenchao, Suo Rui, et al. Apple detection and counting using real-time video based on deep learning and object tracking[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2021, 37(21): 217-224. (in Chinese with English abstract)
- [6] 李小占, 马本学, 喻国威, 等. 基于深度学习与图像处理的哈密瓜表面缺陷检测[J]. *农业工程学报*, 2021, 37(1): 223-232.
Li Xiaozhan, Ma Benxue, Yu Guowei, et al. Surface defect detection of Hami melon using deep learning and image processing[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2021, 37(1): 223-232. (in Chinese with English abstract)
- [7] 孙红, 乔金博, 李松, 等. 基于深度学习的玉米拔节期冠层识别[J]. *农业工程学报*, 2021, 37(21): 53-61.
Sun Hong, Qiao Jinbo, Li Song, et al. Recognition of the maize canopy at the jointing stage based on deep learning[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2021, 37(21): 53-61. (in Chinese with English abstract)
- [8] Wang Q, Li S, Qin H, et al. Super-resolution of multi-observed RGB-D images based on nonlocal regression and total variation[J]. *IEEE Transactions on Image Processing*, 2016, 25(3): 1425-1440.
- [9] Guo X, Zhao X, Liu Y, et al. Underwater sea cucumber identification via deep residual networks[J]. *Information Processing in Agriculture*, 2019, 6(3): 307-315
- [10] Yao, Qiu Z, Zhong M. Application of improved MobileNet-SSD on underwater sea cucumber detection robot[C]//2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). IEEE, 2019, 1: 402-407.
- [11] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[C]//European Conference on Computer Vision. Switzerland: Springer, Cham, 2016: 21-37.
- [12] Ning C, Zhou H, Song Y, et al. Inception single shot multibox detector for object detection[C]//2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). IEEE, 2017: 549-554.
- [13] Chen H, Wang Q, Yang G, et al. SSD object detection algorithm with multi-scale convolution feature fusion[J]. *Journal of Frontiers of Computer Science & Technology*, 2019, 13(6): 1049.
- [14] Kumar A, Zhang Z J, Lyu H. Object detection in real time based on improved single shot multi-box detector algorithm[J]. *EURASIP Journal on Wireless Communications and Networking*, 2020(1): 1-18.
- [15] Arora A, Grover A, Chugh R, et al. Real time multi object detection for blind using single shot multibox detector[J]. *Wireless Personal Communications*, 2019, 107(1): 651-661.
- [16] Leng J, Liu Y. An enhanced SSD with feature fusion and visual reasoning for object detection[J]. *Neural Computing and Applications*, 2019, 31(10): 6549-6558.
- [17] Nguyen A Q, Nguyen H T, Tran V C, et al. A visual real-time fire detection using single shot multibox detector for uav-based fire surveillance[C]//2020 IEEE Eighth International Conference on Communications and Electronics (ICCE). IEEE, 2021: 338-343.
- [18] 黄豪杰, 段先华, 黄欣辰. 基于深度学习水果检测的研究与改进[J]. *计算机工程与应用*, 2020, 56(3): 127-133.
Huang Haojie, Duan Xianhua, Huang Xinchen. Research and improvement of fruits detection based on deep learning[J]. *Computer Engineering and Applications*, 2020, 56(3): 127-133. (in Chinese with English abstract)
- [19] Chen J, Liu R, Tong Y, et al. Synthetical application of multi-feature map detection and multi-branch convolution[J]. *EURASIP Journal on Wireless Communications and Networking*, 2019, 2019(1): 1-8.
- [20] Zhao Y, Zhao J, Zhao C, et al. Robust real-time object detection based on deep learning for very high resolution remote sensing images[C]//IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2019: 1314-1317.
- [21] Liu S, Huang D. Receptive field block net for accurate and fast object detection[C]//Proceedings of the European Conference on Computer Vision (ECCV). Switzerland: Springer, Cham. 2018: 385-400.
- [22] Hou S, Li Y, Pan Y, et al. A face detection algorithm based on two information flow block and retinal receptive field block[J]. *IEEE Access*, 2020, 8: 30682-30691.
- [23] Liu X, Zhou S, Kuang H, et al. Local adjustment block net for object detection[C]//2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE). IEEE, 2020: 1655-1658.
- [24] Mao Y, Chen K, Diao W, et al. Beyond single receptive field: A receptive field fusion-and-stratification network for airborne laser scanning point cloud classification[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2022, 188: 45-61.
- [25] Jiao L, Zhang S, Dong S, et al. RFP-Net: Receptive field-based proposal generation network for object detection[J]. *Neurocomputing*, 2020, 405: 138-148.
- [26] 张海涛, 张梦. 引入通道注意力机制的 SSD 目标检测算法[J]. *计算机工程*, 2020, 46(8): 264-270.

- Zhang Haitao, Zhang Meng. SSD target detection algorithm with channel attention mechanism[J]. Computer Engineering, 2020, 46(8): 264-270. (in Chinese with English abstract)
- [27] 鞠默然, 罗江宁, 王仲博, 等. 融合注意力机制的多尺度目标检测算法[J]. 光学学报, 2020, 40(13): 132-140.
- Ju Moran, Luo Jiangning, Wang Zhongbo, et al. Multi-scale target detection algorithm based on attention mechanism[J]. Acta Optica Sinica, 2020, 40(13): 132-140. (in Chinese with English abstract)
- [28] Zheng Y Y, Kong J L, Jin X B, et al. Real-time vegetables recognition system based on deep learning network for agricultural robots[C]//2018 Chinese Automation Congress (CAC). IEEE, 2018: 2223-2228.
- [29] Jiang H, Shi T, Bai Z, et al. Ahcnet: An application of attention mechanism and hybrid connection for liver tumor segmentation in ct volumes[J]. IEEE Access, 2019, 7: 24898-24909.
- [30] Shang T, Dai Q, Zhu S, et al. Perceptual extreme super-resolution network with receptive field block[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. IEEE, 2020: 440-441.
- [31] Li X, Yang Z, Wu H. Face detection based on receptive field enhanced multi-task cascaded convolutional neural networks[J]. IEEE Access, 2020, 8: 174922-174930.

Algorithm for detecting sea cucumbers based on improved SSD

Zhang Lan¹, Xing Bowen^{1,2*}, Li Cai², Li Shuofeng³

(1. College of Engineering Science and Technology, Shanghai Ocean University, Shanghai 201306, China; 2. State Key Laboratory of Tropical Oceanography, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou 510308, China; 3. College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001, China)

Abstract: Intelligent fishing of sea cucumbers has been an ever-increasing trend using underwater robots in recent aquaculture, instead of the conventional manual operations. However, there is a relatively low distinction between the sea cucumbers and the complex living environment, some of which are semi-hidden in the natural ocean. It is easy to induce the low accuracy of an underwater robot in the recognition of the sea cucumber targets. Particularly, the remote sea cucumbers cannot be recognized as the small targets with the depth of field during the movement of an underwater robot in the natural environment. In this study, the object detection algorithm was proposed for the sea cucumbers using improved Single Shot multibox Detector (SSD) network deep learning. Firstly, the shallow-feature receptive field was improved to increase the location information using a receptive field block. The spatial attention and channel attention mechanisms were then combined to strengthen the features of different depths in the network. The original feature information was multiplied to obtain the weight between each feature channel and feature space. As such, the most representative features were achieved in the channel and spatial feature maps without the irrelevant features. Finally, the fusion of the feature map was performed to further improve the precision of sea cucumber recognition. The actual video was taken to verify the model during testing in the experiment. The improved recognition rate of underwater sea cucumber was obtained at the level of network structure. The experimental results show that the Mean Average Precision (mAP₅₀) was 95.63% for the target detection of sea cucumber using the improved SSD network, and the detection frame rate was 10.7 frame/s. Specifically, the mAP₅₀ increased by 3.85 percentage points, while the detection frame rate was only reduced by 2.8 frame/s, compared with the traditional SSD. The precision-recall (P-R) curves were compared before and after the model improvement. There was a larger area between the P-R curve of the improved SSD model and the X and Y coordinate axes, and the balance point was closer to the coordinates (1, 1), indicating the better performance of the improved SSD model. The Faster R-CNN and YOLOv4 were selected to verify the effectiveness of the improved SSD. The mAP₅₀ values of the improved model were 4.19 and 1.74 percentage points higher than those of the YOLOv4 and Faster R-CNN, respectively, indicating the better system performance of the improved model on the P-R curve than those algorithms. The detection speed of the improved model was 4.6 frame/s lower than that of YOLOv4, whereas, that was 3.95 frame/s higher than that of Faster R-CNN. Consequently, the improved SSD was more suitable for the underwater robot of sea cucumber in the intelligent fishing task, considering the target detection accuracy and running speed. The finding can provide a strong reference for the intelligent fishing of sea cucumbers in aquaculture.

Keywords: image recognition; deep learning; algorithms; sea cucumber fishing; single shot multibox detector