

引用格式:

戚学通, 袁红春. 基于自注意力机制和改进 YOLOv5s 的小目标生物检测[J]. 湖南农业大学学报(自然科学版), 2024, 50(3): 108–114.

QI X T, YUAN H C. Small target biological detection based on self-attention mechanism and improved YOLOv5s[J]. Journal of Hunan Agricultural University(Natural Sciences), 2024, 50(3): 108–114.

投稿网址: <http://xb.hunau.edu.cn>



基于自注意力机制和改进 YOLOv5s 的小目标生物检测

戚学通, 袁红春*

(上海海洋大学信息学院, 上海 201306)

摘要: 为了快速准确地检测出小目标生物(海参、扇贝、海星和海胆)在复杂水下环境的位置及所属种类, 提出一种基于改进 YOLOv5s 的小目标生物检测算法。在特征提取阶段, 引入基于多头自注意力设计的自注意力残差模块, 强化网络全局建模能力的同时, 强化目标特征信息; 在特征融合阶段, 将特征融合网络调整为添加横向连接的双向特征金字塔结构, 增强网络融合不同阶段特征信息的能力; 在检测阶段, 舍弃大目标检测尺度并添加小目标的检测尺度, 提升小目标生物的检测精度; 最后, 引入 α -CIoU 损失函数作为模型边界框回归损失函数, 提高边界框回归精度, 进而提高算法检测准确率。定性试验中, 几乎所有肉眼可见的水产品目标都被改进模型检出, 并正确标记, 体现了改进算法的有效性。 α 值选取试验中, α 值为 2.0 时效果最佳, 平均精度均值(mAP)均优于其他值的, 达到 0.857, 较 α 值为 1.0 时的提升了 0.016。消融试验中, 添加任一优化方法均会提升改进模型的检测精度, 最终改进模型的 mAP 达 0.873, 较原模型的提升了 0.032, 模型参数量减少了 26.8%, 仅有 5 M。对比试验中, 改进模型的 mAP 较 Faster RCNN、YOLOv3、YOLOv4、YOLOv5s、YOLOvX、SSD、NAS-FCOS、改进 YOLOv5 等的提升了 0.020 以上; 改进模型在本地服务器的检测速度达 139 帧/s, 较 YOLOv5s 的提升了 14 帧/s, 略逊于以检测速度著称的 SSD 模型的。可见, 改进模型能满足轻量化和实时性要求。改进模型也成功部署到安卓移动设备中。

关键词: 小目标生物检测; YOLOv5s; 自注意力机制; α -CIoU 损失; 双向特征金字塔; 移动设备部署

中图分类号: S951.2; TP391.413

文献标志码: A

文章编号: 1007-1032(2024)03-0108-07

Small target biological detection based on self-attention mechanism and improved YOLOv5s

QI Xuetong, YUAN Hongchun*

(School of Information, Shanghai Ocean University, Shanghai 201306, China)

Abstract: In order to quickly and accurately detect the location and species of small target organisms(sea cucumbers, scallops, starfish and sea urchins) in complex underwater environments, a small target organism detection algorithm based on improved YOLOv5s was designed in this study. In the feature extraction stage, a self-attention residual module based on multi-head self-attention design was introduced to enhance the global modeling ability of the network while enhancing the target feature information; in the feature fusion stage, the feature fusion network was adjusted to a bidirectional feature pyramid structure with lateral connections to enhance the network's ability to fuse feature information at different stages; in the detection stage, the large target detection scale was discarded and the small target detection scale was added to improve the detection accuracy of small target organisms; finally, the α -CIoU loss function was introduced as the model bounding box regression loss function to improve the bounding box regression accuracy, thereby improving the algorithm detection accuracy. In the qualitative test, almost all aquatic product targets visible to the naked eye were detected and correctly marked by the improved model, which reflects the effectiveness of the improved

收稿日期: 2023-02-22

修回日期: 2024-06-15

基金项目: 国家自然科学基金项目(41776142)

作者简介: 戚学通(1997—), 男, 山东青岛人, 硕士研究生, 主要从事计算机视觉研究, 2448325838@qq.com; *通信作者, 袁红春, 博士, 教授, 主要从事人工智能应用研究, hcyuan@shou.edu.cn

algorithm. In the α value selection test, the best effect was achieved when the α value was 2.0, and the mean average precision(mAP) was better than other values, reaching 0.857, which was 0.016 higher than that when the α value was 1.0. In the ablation experiment, adding any optimization method increased the detection accuracy of the improved model. The mAP of the improved model finally reached 0.873, which was 0.032 higher than that of the original model, and the number of model parameters was reduced by 26.8%, only 5 M. In the comparative experiment, the mAP of the improved model was improved by more than 0.020 compared with Faster RCNN, YOLOv3, YOLOv4, YOLOv5s, YOLOvX, SSD, NAS-FCOS, and improved YOLOv5; and the detection speed of the improved model on the local server reached 139 frames/s, which was 14 frames/s higher than that of the YOLOv5s, slightly lower than that of the SSD model known for its detection speed. It could be concluded that the improved model meets the requirements of lightweight and real-time performance. The improved model was also successfully deployed on Android mobile devices.

Keywords: small target biological detection; YOLOv5s; self-attention mechanism; α -CIoU loss; bidirectional feature pyramid; mobile device deployment

海水养殖业发展迅速, 为人类提供了丰富多彩的水产品, 如海参、海胆、海星、扇贝等。早期渔民对海水养殖生物的捕捞方式主要采用人工捕捞, 这种方式需要消耗巨大的人力和物力资源^[1]。随着深度学习和计算机视觉的快速发展, 人们将水下目标检测模型嵌入到水下机器人进行水下抓取任务, 取得了不错的效果^[2]。

现阶段比较流行的目标检测算法主要有 2 类。一类是 Two-Stage 检测算法, 此类算法的特点是需要先通过专门模块生成候选框(proposal), 再对候选框进行分类以及边界框调整, 代表算法主要有 RCNN^[3-6]系列。HUANG 等^[7]运用改进的 Faster R-CNN 算法检测海胆、扇贝等生物, 该算法检测效果得到了提升, 但整体准确率偏低。SONG 等^[8]利用 MSRCR 算法对 90 种水下生物数据进行增强, 再使用 Mask R-CNN 算法进行检测, 虽然精度得到了提升, 但检测速度仅有 0.69 帧/s, 不能满足实际应用需求。另一类是 One-Stage 目标检测算法, 该类算法将物体位置和目标类别通过一个神经网络模型直接输出检测结果, 提高了检测速度, 代表算法有 YOLO 系列^[9-13]、SSD(single shot multibox detector)^[14]等。洪亮等^[15]利用图像增强技术和改进的 YOLOv3 算法对海珍品进行识别, 准确率提升 2.4%, 但模型内存占 146.7 M。林森等^[16]使用融合注意力机制的轻量化 YOLOv5 算法检测 3 种水下珍品, 检测精度得到了提升, 但算法未有效解决漏检现象。

对此, 笔者基于目前成熟度和通用性都较强

的 YOLOv5 算法进行研究, 提出改进的小目标生物检测算法: 首先, 利用 BoT(Bottleneck Transformer)结构^[17]聚合全局特征的优势, 将自注意力机制引入主干网络, 构成 CNN+Transformer 的混合网络, 使模型可以更好地处理局部信息和全局信息^[18]; 然后, 添加模型输入层到输出层的横向连接, 对不同层提取到的特征进行简单快速融合, 提升模型在复杂水下环境中的特征融合能力, 不再使用原有网络的大目标检测尺度, 增加一个新的小目标检测尺度, 以适应体积更小的水下小目标生物; 最后, 将网络定位损失由 CIoU 损失替换为 α -CIoU 损失, 提高模型对水下小目标生物的定位能力。

1 改进 YOLOv5s 的小目标生物检测算法

1.1 改进 YOLOv5s 算法概述

尽管 YOLOv5s 在目标检测领域表现良好, 但在复杂的水下环境中检测效果较差, 尤其检测小目标生物时常出现误检和漏检现象。为提高检测精度, 提出检测效果更好且更加轻量的水下小目标生物检测模型方案。主要改进点为: 将主干网络最后一个特征提取模块中 Bottleneck 结构替换为更加轻量的 BoT 结构; 引入输入层到输出层的横向连接, 以增强不同区域的特征融合; 将大目标检测头替换为小目标检测头, 提高小目标检测精度; 引入 α -CIoU 损失作为模型的定位损失, 提高边界框回归精度, 减少漏检。模型整体结构如图 1 所示。

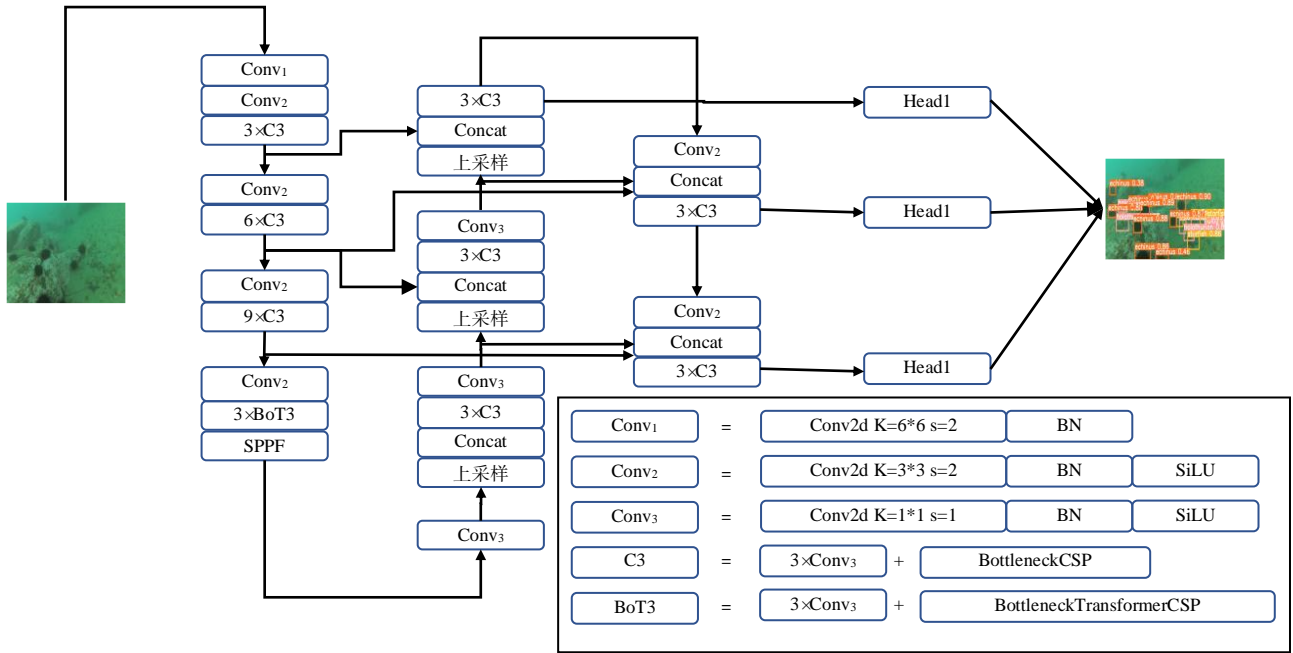


图 1 改进 YOLOv5s 网络结构

Fig.1 Improved YOLOv5s network structure

1.2 骨干网络的改进

考虑到 YOLOv5s 主干网络最后一个阶段 C3 层包含图像强语义信息,减少该层复杂环境对被检测目标的干扰,可以减少算法参数并提升检测效果。借鉴文献[17]中 BoTNet(Bottleneck Transformer Network)的结构,将最后一个阶段 C3 层残差模块 Bottleneck 替换为 BoT 结构,通过引入自注意力机制对卷积捕获的特征图中包含的信息进行处理和聚合,如图 2 所示。BoT 结构用 MHSA(Multi-Head Self-Attention)层替换 3×3 空间卷积, MHSA 是由多个如图 3 所示的自注意力层并联而成。

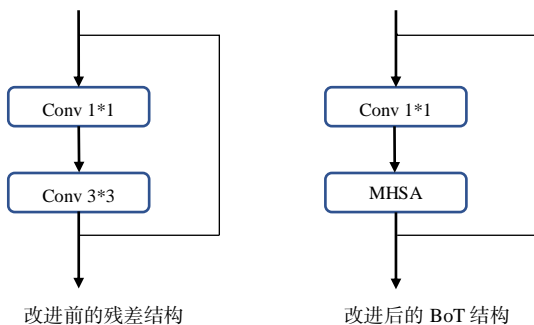


图 2 Bottleneck Transformer 结构

Fig.2 Bottleneck Transformer structure

1.3 损失函数的改进

根据 CIoU 损失公式^[19]和 α -IoU 损失公式^[20]推广得出 α -CIoU 损失公式。

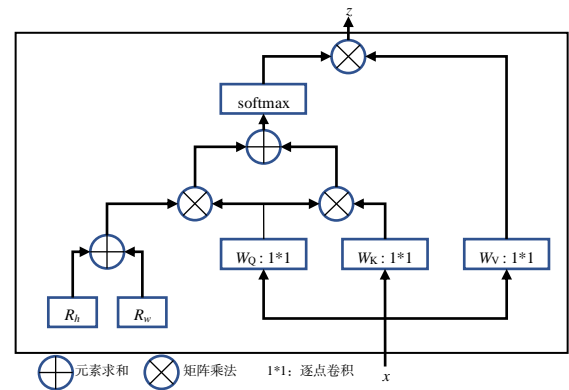


图 3 自注意力层结构

Fig.3 Self-attention layer structure

$$L_{\alpha\text{-CIoU}} = 1 - \text{IoU}^\alpha + \frac{\rho^{2\alpha}(b, b^{\text{gt}})}{c^{2\alpha}} + (\beta v)^\alpha \quad (1)$$

式中: $L_{\alpha\text{-CIoU}}$ 表示 α -CIoU 损失; α 为调节参数; $\rho(b, b^{\text{gt}})$ 表示预测框和目标框中心点之间的欧式距离; c 是包含预测框和目标框的最小矩形框的对角线长度; β 和 v 为预测框相关的调节参数。选用 $L_{\alpha\text{-IoU}}$ 作为边界框回归损失,相比于 L_{CIoU} ,使用合适的 α 值时 $L_{\alpha\text{-IoU}}$ 会增加高 IoU 目标的损失和梯度,进而提高 bbox 的回归精度,可以更好地定位目标生物。

1.4 检测尺度的改进

针对 YOLOv5s 原模型出现小目标漏检的现象,在 YOLOv5s 原有 3 个检测尺度的基础上,舍弃 20×20 的检测层,增加 1 个 160×160 的检测层,具体为在特征融合网络部分增加上采样层和卷积层,

使用主干网络 160×160 特征层与新添加的上采样层相融合, 形成新的三尺度检测(图 1), 以提高模型对水下小目标的检测能力。

1.5 特征金字塔的改进

YOLOv5s 的 neck 部分采用 FPN+PANet 结构, 有一条自顶向下的通道和一条自底向上的通道进行多尺度特征融合, 但最深层的特征层无法很好地吸收骨干网络提取的浅层特征。为此, 在 PANet 的基础上运用双向融合思想, 不仅构造由上而下、由底向上的双向通道, 且在同一尺度的特征间添加横向连接, 分别将 80×80 和 40×40 大小不同位置特征层在 channel 维度上拼接, 构建模型融合多尺度检测后的双向金字塔结构, 如图 4 所示。

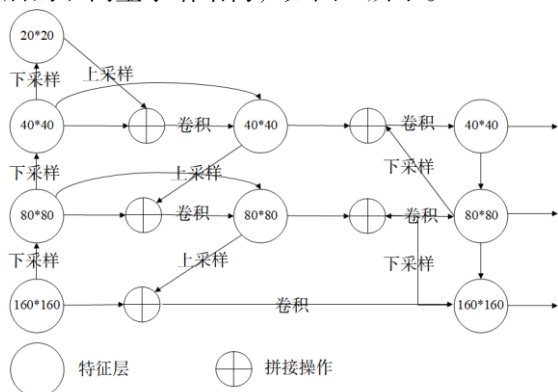


图 4 添加横向连接的双向特征金字塔结构

Fig.4 Adding a bidirectional feature pyramid structure with lateral connections

2 试验准备

2.1 试验环境与相关参数

处理器为 AMD Ryzen 9 5950X 16Core Processor; 显卡为 NVIDIA GEFORCE RTX 3090; 基于 Linux 4.18.0、Python3.8 和 Pytorch 1.10 搭建深度学习框架; GPU 加速软件为 CUDA11.3 和 CUDNN7.6.5。进行的轮数 epochs 大小为 300, batchsize 大小设置为 32, 初始学习率为 0.01, 周期学习率为 0.1, 采

用的优化器为随机梯度下降算法(SGD)。

2.2 试验数据集

选用来自大连理工大学与獐子岛集团水下机器人联合实验室采集的水下图像 RUIE 作为试验数据集。数据集包含扇贝(scallop)、海参(holothurian)、海胆(echinus)、海星(starfish)等 4 类水下生物, 共包含 4 757 张图片, 部分图像存在标记错误或未标记现象, 处理后的数据集含有 3 765 张图片。将数据集拷贝为 2 份: 一份转化为 YOLO 格式的数据集, 以便 YOLOv5s 模型训练测试; 一份保留原先 XML 格式, 以便对比试验模型训练测试。划分数据集为训练集 3 024 张图片, 验证集 378 张图片, 测试集 363 张图片, 共有 25 075 个标签。此外, 为体现算法鲁棒性, 图片未进行清晰处理, 均采用原始图像。

2.3 评价指标

为验证水产品目标检测模型的检测性能, 从速度和精度两方面对模型进行评价。速度方面, 选用每秒检测帧数(FPS)作为评价指标; 精度方面, 选用精确率(P)、召回率(R)、平均精度(AP)和平均精度均值(mAP)作为评价指标。

3 验证试验及其结果

3.1 定性试验

为直观展示算法的性能, 随机从数据集抽取 1 张图片, 分别用原 YOLOv5s 模型和改进模型在同一平台对该图像进行检测, 检测结果如图 5 所示。对比图 5-b 和图 5-c, 原模型未检测出图像左上角海胆目标, 而改进模型可以检测出该目标, 表明改进算法有效减少了小目标漏检现象, 鲁棒性强。对比图 5-a 和图 5-c 可知, 几乎所有肉眼可见的水产品目标都被改进模型检出, 并正确标记, 体现了改进算法的有效性。



a 原图; b YOLOv5s 检测结果; c 改进模型检测结果。

图 5 YOLOv5s 算法与改进算法的检测结果

Fig.5 YOLOv5s algorithm and improved algorithm detection results

3.2 定量试验

通过训练改进模型,得到训练权重文件,使用训练权重对测试集进行检测,得到改进模型的 $P-R$ 曲线如图 6 所示。从图 6 可知,橙色曲线代表的海参与坐标轴围成的面积最小,其 mAP 值也最小,为 0.798;扇贝和海星的检测精度较高, mAP 值分别为 0.872 和 0.886;海胆的检测精度最高, mAP 值达 0.936;整个模型的 mAP 值为 0.873。导致未完全识别的原因是复杂的水下环境干扰,后续将考虑对图像进行清晰化处理,以提升检测识别率。不同种类识别率不一致的原因是数据集各类别数量不一致,数量较少的类别未正确识别会导致准确率降低幅度较大,后续将继续扩充数据集,以提升数据丰富度。图 7 为模型改进前后训练过程的精度变化。由图 7 可知,改进算法训练过程的 mAP 值明显优于原模型,体现了改进算法的优越性。

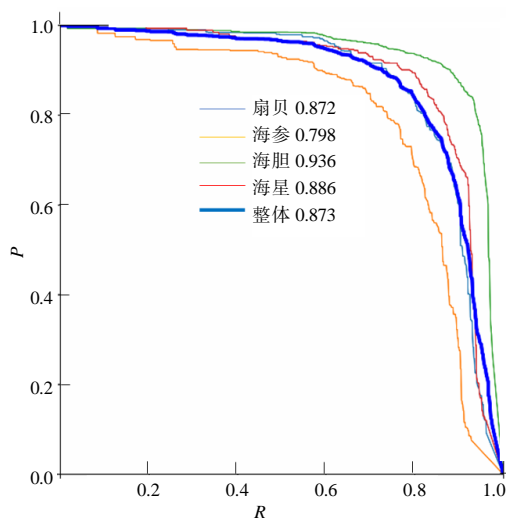
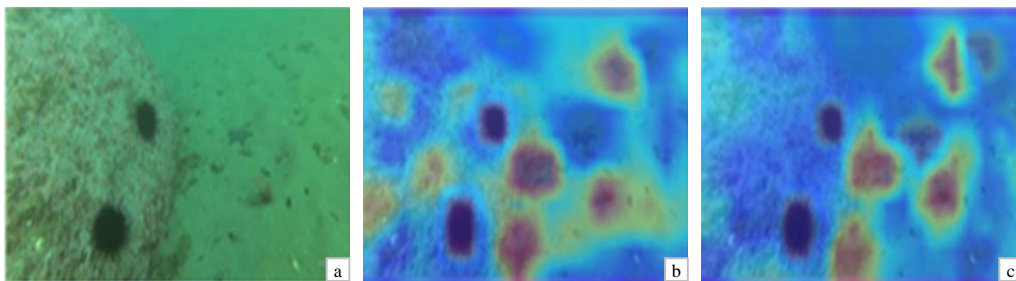


图 6 改进算法的 $P-R$ 曲线

Fig.6 Improved algorithm $P-R$ curve



a 原图; b 未添加 BoT 的图; c 添加 BoT 的图。

图 8 自注意力机制可视化试验结果

Fig.8 Self-attention mechanism visualization experiment

可见,添加 BoT 模块的热力图区域比未添加的更聚焦于目标生物,且未添加 BoT 的算法存在热力图聚焦于背景环境以及目标物体未被聚焦的现象,引入

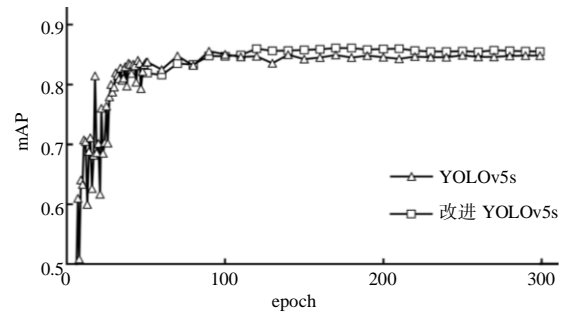


图 7 模型改进前后训练过程的平均精度均值

Fig.7 Average accuracy of the training process before and after model improved

3.3 α 值选取试验

α -CIoU 损失通过调节超参数 α 来满足不同水平的 bbox 回归精度,适当选取 α 值会提高检测器的效果,同时也不会增加训练和推理时间。当 α 值小于 1 时,该损失会降低模型性能,降低了高 IoU 目标的权重,最终产生更多定位较差的目标;而当 α 值大于 1 时,该损失会增加高 IoU 目标的损失和梯度,进而可以获得更准确的边界框回归和目标检测结果。故分别选用 α 值为 1.0、1.5、2.0、2.5、3.0 的 α -CIoU 作为 YOLOv5s 的回归损失,并用这些模型进行检测,对应的 mAP 值分别为 0.841、0.849、0.857、0.852、0.851,说明 α -CIoU 损失对模型的性能提升效果较佳。当 α 值取 2.0 时,模型的 mAP 值最大,故选取 α 值为 2.0 时的 α -CIoU 损失作为整个网络的定位损失。

3.4 自注意力可视化试验

为进一步证明 BoT 模块的有效性,对算法特征的热力图进行了可视化,可直观展示图像中哪部分区域对目标分类贡献程度大,如图 8 所示。从图 8

自注意力机制能够帮助模型强化图像的主体目标,弱化背景,表明自注意力机制的有效性。

3.5 消融试验

改进算法同时对 YOLOv5s 模型进行了损失函数、特征提取、特征融合和检测尺度的改进。为评估不同改进对模型性能的优化程度，设置 6 组消融试验：序号 0 组采用未进行任何改进的原 YOLOv5s 模型；第 1~4 组分别采用单一改进的模型；第 5 组采用全部改进的模型(表 1)。从表 1 可以看出，采用

单一改进都会提升模型的检测精度；同时采用 4 种改进时，算法的检测精度最高，且参数量最少，其 mAP 比原模型的提升了 0.032,参数量比原模型的减少了 26.8%。可见，改进的水产品检测模型在保持轻量化的同时取得了性能提升，适合应用于水下设备对海洋生物的检测。

表 1 消融试验结果

Table 1 Ablation experiments							
序号	YOLOv5s	α -CIoU	BoT	新检测尺度	双向金字塔	mAP	参数量
0	√					0.841	7 020 913
1	√	√				0.857	7 020 913
2	√		√			0.854	6 694 001
3	√			√		0.857	5 379 569
4	√				√	0.850	7 020 913
5	√	√	√	√	√	0.873	5 138 636

3.6 对比试验

为客观验证算法的性能，利用相同的试验环境和同一数据集，与目前在水下目标检测领域应用较广泛的算法(Faster RCNN、YOLOv3、

YOLOv4、YOLOv5s、YOLOvX、SSD、NAS-FCOS^[21]、改进 YOLOv5)进行对比。从表 2 可知，笔者提出的改进 YOLOv5s 算法在检测扇贝、海参、海胆、海星等 4 种生物的 AP 均高于其他算法

表 2 对比试验结果

Table 2 Comparative experiments							
序号	模型	AP				mAP	FPS/(帧 s ⁻¹)
		扇贝	海参	海胆	海星		
0	Faster RCNN	0.185	0.057	0.698	0.295	0.309	30
1	YOLOv3	0.831	0.743	0.922	0.863	0.840	125
2	YOLOv4	0.849	0.732	0.908	0.874	0.841	49
3	YOLOv5s	0.837	0.730	0.924	0.873	0.841	125
4	YOLOvX	0.833	0.735	0.910	0.877	0.839	73
5	SSD	0.524	0.455	0.834	0.734	0.637	148
6	NAS-FCOS ^[21]	0.662	0.701	0.816	0.769	0.737	81
7	改进 YOLOv5 ^[16]	0.827	0.763	0.916	0.883	0.847	70
8	改进 YOLOv5s	0.872	0.798	0.936	0.886	0.873	139

的，mAP 值达 0.873，比其他 YOLO 模型均提升 0.020 以上；速度方面，较原 YOLOv5s 的速度提升了 14 帧/s，略逊于以检测速度著称的 SSD 模型。综合来看，改进 YOLOv5s 算法更好地实现了精度和速度的平衡，更具有使用价值。

3.7 移动设备部署试验

为使改进模型更好应用于真实场景，对 YOLOv5s 算法和改进算法进行了安卓部署。将训练好的模型通过 NCNN 移植到 Android 平台并进行开发，将开发好的项目移植到手机端进行测试。试验采用 Android13 设备系统，CPU 型号为天玑 920，GPU 型号为 Mali-G68 MC4,检测效果如图 9 所示。在安卓系统中，运用 CPU 和 GPU 的改进算法的推理时间分别为 0.112、0.083 s，YOLOv5s 算法的分

别为 0.153、0.101 s，改进算法的检测速度明显优于原模型的。



图 9 移动设备检测效果

Fig.9 Mobile device detection effect

4 结论

本研究中,笔者提出一种基于 YOLOv5s 的小目标生物检测算法,对 YOLOv5s 进行了 4 个方面的改进:回归框损失采用对轻量级模型更加友好且精度更高的 α -CIoU 损失;特征提取网络引入自注意力机制;检测颈部分采用添加横向连接双向特征金字塔结构及预测部分使用新三尺度检测。在水下生物检测试验中,改进算法在保持较高检测准确率的同时保证了轻量化特性,提高了检测速度,并成功部署到安卓系统中。未来可尝试引入图像预处理技术,对采集到的水下图像进行清晰化处理,再进行检测,并将算法应用到更高算力的设备中,应用于实践。

参考文献:

- [1] RICK T C, ERLANDSON J M, VELLANOWETH R L. Paleocoastal marine fishing on the Pacific Coast of the Americas: perspectives from Daisy Cave, California[J]. *American Antiquity*, 2001, 66(4): 595–613.
- [2] AGUIRRE-CASTRO O A, INZUNZA-GONZÁLEZ E, GARCÍA-GUERRERO E E, et al. Design and construction of an ROV for underwater exploration[J]. *Sensors*, 2019, 19(24): 5387.
- [3] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE, 2014: 580–587.
- [4] GIRSHICK R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision(ICCV). Santiago, Chile: IEEE, 2015: 1440–1448.
- [5] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137–1149.
- [6] HE K M, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision(ICCV). Venice, Italy: IEEE, 2017: 2961–2969.
- [7] HUANG H, ZHOU H, YANG X, et al. Faster R-CNN for marine organisms detection and recognition using data augmentation[J]. *Neurocomputing*, 2019, 337: 372–384.
- [8] SONG S J, ZHU J X, LI X H, et al. Integrate MSRCR and mask R-CNN to recognize underwater creatures on small sample datasets[J]. *IEEE Access*, 2020, 8: 172848–172858.
- [9] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Las Vegas, USA: IEEE, 2016: 779–788.
- [10] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Honolulu, USA: IEEE, 2017: 6517–6525.
- [11] REDMON J, FARHADI A. Yolov3: an incremental improvement[EB/OL]. 2018: 1804.02767.(2018–04–08). <https://arxiv.org/abs/1804.02767>.
- [12] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. 2020:2004.10934.(2020–04–23).<http://arxiv.org/abs/2004.10934>.
- [13] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE, 2023: 7464–7475.
- [14] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]//Computer Vision: ECCV 2016: 14th European Conference. Amsterdam, Netherlands: Springer, 2016: 21–37.
- [15] 洪亮, 王芳, 蔡克卫, 等. 面向海洋牧场智能化建设的海珍品实时检测方法[J]. *农业工程学报*, 2021, 37(9): 304–311.
- [16] 林森, 刘美怡, 陶志勇. 采用注意力机制与改进 YOLOv5 的水下珍品检测[J]. *农业工程学报*, 2021, 37(18): 307–314.
- [17] SRINIVAS A, LIN T Y, PARMAR N, et al. Bottleneck transformers for visual recognition[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Nashville, USA: IEEE, 2021: 16519–16529.
- [18] 姜绍忠, 姚克明, 陈磊, 等. 基于 CNN 与 Transformer 混合模型的口罩人脸识别方法[J]. *传感器与微系统*, 2023, 42(1): 144–148.
- [19] ZHENG Z H, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(7): 12993–13000.
- [20] HE J B, ERFANI S, MA X J, et al. Alpha-IoU: a family of power intersection over union losses for bounding box regression[EB/OL]. 2021:2110.13675.(2022–01–22). <http://arxiv.org/abs/2110.13675v2>.
- [21] WANG N, GAO Y, CHEN H, et al. NAS-FCOS: fast neural architecture search for object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Seattle, USA: IEEE, 2020: 11943–11951.

责任编辑: 邹慧玲
英文编辑: 柳 正