

水下模糊目标的自适应检测方法

刘照龙 袁宏武 余倩

Adaptive detection methods for fuzzy underwater objects

LIU Zhaolong, YUAN Hongwu, YU Qian

引用本文:

刘照龙, 袁宏武, 余倩. 水下模糊目标的自适应检测方法[J]. 应用光学, 2025, 46(5): 1024–1033. DOI: 10.5768/JAO202546.0502002

LIU Zhaolong, YUAN Hongwu, YU Qian. Adaptive detection methods for fuzzy underwater objects[J]. Journal of Applied Optics, 2025, 46(5): 1024–1033. DOI: 10.5768/JAO202546.0502002

在线阅读 View online: <https://doi.org/10.5768/JAO202546.0502002>

您可能感兴趣的其他文章

Articles you may be interested in

多维特征点空间的红外弱小目标检测方法

Multi-dimensional feature point space infrared dim target detection method

应用光学. 2020, 41(6): 1268–1276 <https://doi.org/10.5768/JAO202041.0606002>

基于双路多尺度金字塔池化模型的显著目标检测算法

Salient target detection algorithm based on dual-channel multi-scale pyramid pooling model

应用光学. 2021, 42(6): 1056–1061 <https://doi.org/10.5768/JAO202142.0602007>

基于改进YOLOv3的小目标检测算法

Small object detection algorithm based on improved YOLOv3

应用光学. 2024, 45(4): 732–740 <https://doi.org/10.5768/JAO202445.0402002>

基于改进Faster R-CNN的红外目标检测算法

Infrared target detection algorithm based on improved Faster R-CNN

应用光学. 2024, 45(2): 346–353 <https://doi.org/10.5768/JAO202445.0202001>

基于改进SSD的车辆小目标检测方法

Detecting method of small vehicle targets based on improved SSD

应用光学. 2020, 41(1): 150–155 <https://doi.org/10.5768/JAO202041.0103004>

基于改进RoI Transformer的遥感图像多尺度旋转目标检测

Multi-scale oriented object detection based on improved RoI Transformer in remote sensing images

应用光学. 2023, 44(5): 1010–1021 <https://doi.org/10.5768/JAO202344.0502001>



关注微信公众号，获得更多资讯信息

文章编号: 1002-2082 (2025) 05-1024-10

引用格式: 刘照龙, 袁宏武, 余倩. 水下模糊目标的自适应检测方法 [J]. 应用光学, 2025, 46(5): 1024-1033.

LIU Zhaolong, YUAN Hongwu, YU Qian. Adaptive detection methods for fuzzy underwater objects[J]. Journal of Applied Optics, 2025, 46(5): 1024-1033.



在线阅读

水下模糊目标的自适应检测方法

刘照龙¹, 袁宏武^{1,2}, 余倩¹

(1. 安徽建筑大学 电子与信息工程学院, 安徽 合肥 230000; 2. 安徽新华学院 大数据与人工智能学院, 安徽 合肥 230000)

摘 要: 水下环境存在光线衰弱、色彩失真、复杂背景干扰、目标尺度多样以及目标特征模糊等问题。针对目标尺度多样性和特征模糊难定位等问题, 提出了一种基于 Faster R-CNN (faster region-based convolutional neural networks) 改进的水下目标检测算法。首先, 在特征提取中引入可切换的空洞卷积, 解决了特征提取过程中图像全局上下文信息丢失而导致的特征损失问题; 其次, 使用递归特征金字塔使高层特征与底层特征多次交互融合, 增强模型对水下小目标以及复杂形状目标的检测能力; 最后, 引入了一种基于引导锚框的建议网络, 该网络根据图像的语义特征, 动态生成锚点更为稀疏且形状自适应的锚框, 显著地提高了模型对水下目标检测的准确性及定位能力。实验表明: 改进后的算法在水下数据集 DUO 上 mAP (mean average precision) 提高了 5.7%, 并且在通用目标检测数据集 VOC 上也有较好的表现。

关键词: 水下目标检测; 可切换空洞卷积; 递归特征金字塔; 特征模糊; 锚框引导建议网络

中图分类号: TN202; TP391.41

文献标志码: A

DOI: 10.5768/JAO202546.0502002

Adaptive detection methods for fuzzy underwater objects

LIU Zhaolong¹, YUAN Hongwu^{1,2}, YU Qian¹

(1. School of Electronic and Information Engineering, Anhui Jianzhu University, Hefei 230000, China; 2. School of Big Data and Artificial Intelligence, Anhui Xinhua University, Hefei 230000, China)

Abstract: Underwater environments usually suffer from issues such as light attenuation, color distortion, complex background interference, diverse target scales, and blurred target features. This paper addresses challenges related to diverse target scales and difficulties in feature localization by proposing an improved underwater target detection algorithm based on Faster R-CNN ((sg_Faster R-CNN). Firstly, we introduced switchable atrous convolution in feature extraction to address the problem of feature loss due to the absence of global contextual information during feature extraction. Secondly, we used a recursive feature pyramid to enable multiple interactions between high-level and low-level features, enhancing the capability of model to detect small underwater targets and complex-shaped objects. Lastly, we introduced a proposal network based on guided anchor boxes, which could dynamically generate anchors that were sparser and shape-adaptive based on the image's semantic features, significantly improving the accuracy and localization ability of the model for underwater target detection. Experiments demonstrate that the improved algorithm achieves a 5.7% increase in

收稿日期: 2024-04-12; 修回日期: 2024-07-10

基金项目: 安徽省自然科学基金资助项目 (2108085MF230); 安徽省高校自然科学研究重点项目 (2022AH051867); 安徽省高校优秀科研创新团队项目 (2022AH010099)

第一作者: 刘照龙, E-mail: 1716932646@qq.com

通信作者: 袁宏武, E-mail: yuanhongwu@axhu.edu.cn

mAP (mean average precision) on the DUO underwater dataset and also performs well on the general-purpose object detection dataset VOC.

Key words: underwater object detection; switchable atrous convolution; recursive feature pyramid; feature blur; region proposal by guided anchoring

引言

目标检测是计算机视觉领域中重要且具有挑战性的任务之一。当前的目标检测算法按照处理流程分为一阶段和两阶段的算法,这两种检测算法在检测策略、网络结构以及性能方面有着不同的特点。一阶段目标检测算法采用密集策略,在单个网络中直接完成目标检测,无需生成候选框。它将目标检测问题转化为回归问题,通过在图像中密集采样锚框并同时预测其类别和位置,实现了端到端的检测。这样的设计使得算法具有较快的检测速度,但精度较低,适用于实时性要求较高的场景。这类算法的代表如 YOLO(you only look once)、SSD(single shot multibox detector)、RetinaNet 等^[1-3]。两阶段目标检测算法则采用分阶段策略,在第一阶段通过区域建议网络 RPN(region proposal network)生成候选框,然后在第二阶段对这些候选框进行分类和回归。常见的两阶段目标检测算法包括 Faster R-CNN、Cascade R-CNN 等^[4-5],虽然其相比一阶段目标检测算法速度较慢,但在目标定位和检测精度上具有良好的表现。

随着智能捕捞的兴起,利用水下机器人进行智能捕捞已成为发展趋势。目标检测是水下捕捞的关键环节,它能够准确识别和定位水下目标,保证了捕捞的精准性和高效性。水下目标检测作为目标检测的特定应用,受到了各国研究学者的关注。LIN W H 等^[6]提出了 ROIMIX 的数据增强方法,通过模拟重叠、遮挡和模糊的目标,旨在帮助模型隐式学习高密度目标检测能力。WANG Z 等^[7]使用 Poisson GAN 的数据增强方法,来改变目标的数量、位置、大小,解决了水下目标类别不平衡和大质量小目标问题。DG-YOLO^[8]通过使用 WQT 的数据增强方法,将海洋、湖泊、河流等不同水下环境泛化为统一领域,减少不同图像之间的色彩差异,提高了模型的检测效果。SWIPENET^[9]使用高分辨率和语义丰富的特征图,显著地提高了小目标检测的准确性,并且还提出了一种新的样本重加权损失,减少了噪声对目标检测的影响。FAN B 等^[10]提出了一种具有特征增强和锚点细化的水

下检测框架,该框架采用复合连接主干来增强特征表示,引入感受野增强模块来利用多尺度特征,并通过学习偏移量来细化多尺度特征,以更好地与锚对齐,在一定程度上解决了目标定位不准确的问题。YE H C H 等^[11]提出了一种颜色转换模型,将彩色图像转换为相应的灰度图像,解决了水下颜色吸收的问题,从而提高了目标检测性能。ZHAO Z 等^[12]基于水下视频目标检测,利用前景和背景的细微差异提取重要特征,提出了一种增强的路径聚合网络,解决了线性上采样导致的语义信息利用率不足的问题。LIANG X 等^[13]提出了双头结构和位置编码模块,通过在感兴趣区域特征上应用外部注意力模块来提升模型检测能力。SONG P 等^[14]提出了一种基于目标的先验概率对硬例样本重加权的方法,解决了正负样本不平衡的问题。当前水下目标检测的研究方向主要有:1) 数据增强方面,模拟实际的遮挡模糊等情况,增强模型对真实世界复杂情况的适应能力;2) 架构创新方面,通过改进骨干网络、引入多尺度特征融合、精细化锚点设计、利用注意力机制等手段,优化特征提取和目标定位过程,提升对小目标的检测性能;3) 水下环境色彩处理方面,通过色彩转换和色彩补偿来恢复目标的真实特征,提高模型的检测能力;4) 样本选择和重加权方面,调整样本权重或设计新的损失函数,解决正负样本不平衡问题,引导模型从大量背景噪声中识别目标。

本文针对水下目标尺度多样性和特征模糊难定位的问题,提出了一种改进 Faster R-CNN 的水下目标检测算法。本文的主要工作如下:

1) 在 ResNet 特征提取中引入可切换的空洞卷积 SAC(switchable atrous convolution)^[15],在不增加参数数量的前提下,扩大了感受野,在卷积操作时动态地选择合适的空洞率,从而捕获更多的局部细节。

2) 采用递归特征金字塔 RFP(recursive feature pyramid)替代传统的特征金字塔 FPN(feature pyramid network),使图像高层特征与底层特征多次交互融合,增强了模型对水下小目标以及复杂形状

目标的检测能力,提高了水下目标检测的准确性和鲁棒性。

3) 为了应对水下目标因环境因素引起的特征模糊而难以精确定位的问题,引入了一种基于锚框引导的建议网络 GA-RPN(region proposal by guided anchoring)^[16]。该网络利用图像的深层次语义特征来动态生成适应性更强、分布更为稀疏的锚框结构。相较于传统固定形状和密度的锚框设计,GA-RPN 能够在不同的图像区域自适应地调整锚框大小与比例,更精准地匹配潜在目标的锚框,从而显著提高了模型对水下目标检测的准确性及定位能力。

1 改进的 Faster R-CNN 水下目标检测算法

水下目标检测相比于传统的目标检测更具有挑战性。水下光线在传播过程中会发生衰减和散射,导致图像中的目标模糊且色彩失真。水下环境中存在许多障碍物、浮游生物、水草等复杂背景干扰,与目标具有相似的外观特征,容易使得目标与背景混淆。由于水下目标遮挡导致的目标尺度

多样性,水下环境引起的形状及纹理特征的模糊难定位等问题,对通用目标检测模型 Faster R-CNN 提出改进,得到的本文算法模型如图 1 所示,流程图如图 2 所示。

算法由 4 部分组成:特征提取、多尺度的特征融合、生成高质量的建议框和精确定位。图像首先通过嵌入 SAC 卷积的 ResNet 进行特征提取,提取图像中目标的形状、纹理特征以及图像内容的语义特征,把提取后的特征图输入到 RFP 进行循环迭代的特征融合,使带有丰富语义的高层特征经过上采样后与带有丰富空间的底层特征融合,增强了算法对多尺度目标特征细节信息的提取。将融合后每一层的特征图都输入到 GA-RPN 中,GA-RPN 有大量锚框,在每个锚框中产生 2 个形状信息的概率图和 1 个位置信息的概率图,结合形状信息和位置信息的分支会生成许多高质量且稀疏的建议框,另一分支的形状概率图会在原始特征图上做形状的特征适应,之后将建议框和特征图输入到 ROI(region of interest)中做精确目标定位。最后通过全连接层进行目标分类和边界框回归。

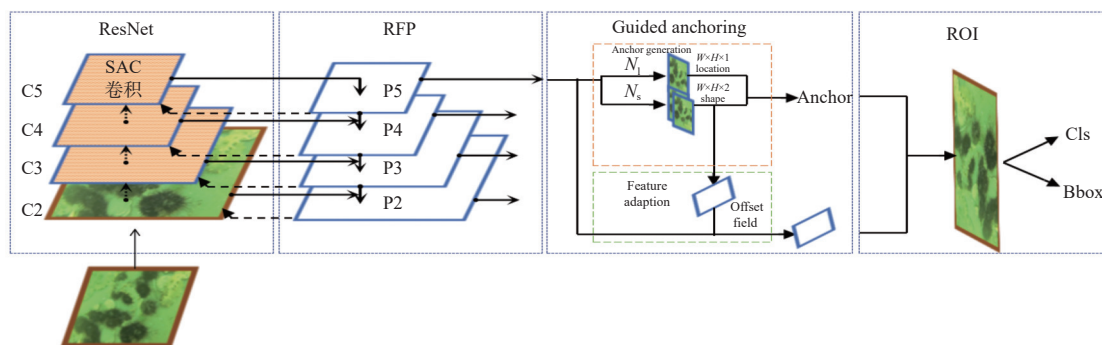


图 1 本文算法 sg_Faster R-CNN 模型图

Fig. 1 Architecture of sg_Faster R-CNN algorithm in this paper

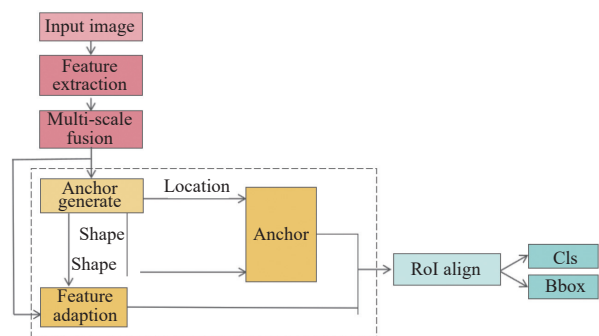


图 2 本文算法 sg_Faster R-CNN 流程图

Fig. 2 Flowchart of sg_Faster R-CNN algorithm in this paper

1.1 可切换的空洞卷积 (SAC)

在传统卷积操作中,卷积核在特征图上以固定步长逐点进行滑动,导致感受野有限,不足以捕捉到目标的全局上下文信息,而随着网络深度的增加,每一层卷积都会因为下采样导致空间分辨率的损失。空洞卷积 AC(atrous convolution)引入了膨胀率,在卷积核的元素之间插入若干个零,会在特征图元素之间产生空洞,能够跨越更大的空间范围去感知重要特征。SAC 基于 AC,并引入了动态调整空洞率机制,其主要思想是允许模型根据输入特征自动选择合适的空洞率,识别相同物体

在不同尺度下的特征, 实现对目标尺度变化的自适应处理。如图 1 所示, 在 ResNet 的 C3、C4、C5 中添加了 SAC 卷积。SAC 结构由 3 部分组成, 分别是 2 个全局上下文模块和 1 个 SAC 卷积, 如图 3 所示。在 ResNet 中将普通的卷积替换为 SAC, 有利于适应不同尺寸的同类目标, 增强模型对目标的理解和感知能力。

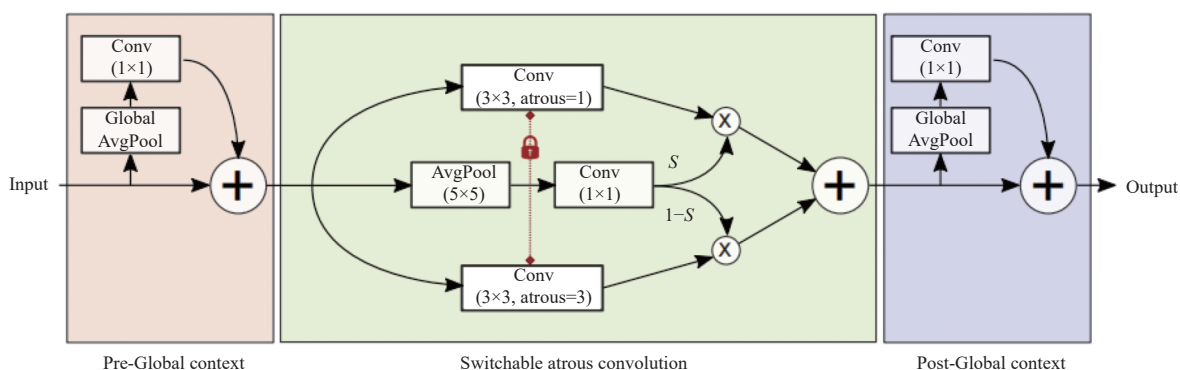


图 3 SAC 内部结构图

Fig. 3 Internal structure diagram of SAC

1.2 递归特征金字塔 (RFP)

在特征提取过程中, 由于深层网络的下采样导致高层特征图的空间分辨率较低, 会丢失小目标的详细位置信息和局部特征, 虽然底层特征具有较高的分辨率, 但语义信息不够丰富。为了有效地解决多尺度目标检测的问题, FPN 在每个层级将自顶向下的上采样特征与同一级别的底层特征进行融合, 确保了每个层次上的特征既包含丰富的语义信息, 又拥有足够的空间精度。但在处理尺度变化较大或遮挡严重的水下目标时, FPN 在提取水下目标检测中多尺度特征能力方面较差, 为此本文提出利用递归特征金字塔 RFP 来解决该问题, RFP 递归式的融合机制有助于模型更好地挖掘和利用多尺度信息, 使图像高层特征与底层特征多次交互融合, 在不同层次特征之间建立更为紧密的关联性, 能够显著提升检测模型对细微结构和全局上下文的捕捉能力, 从而增强对多尺寸水下目标的识别准确性。图 4 和图 5 分别给出了 RFP 结构图和递归展开结构图。

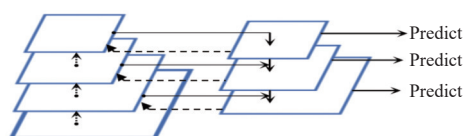


图 4 RFP 结构图

Fig. 4 Structure diagram of RFP

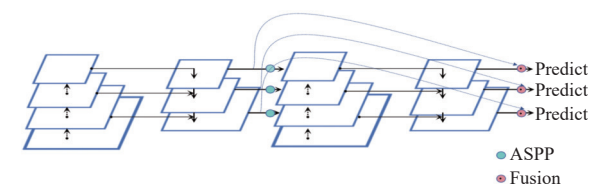


图 5 RFP 递归展开结构图

Fig. 5 Structure diagram of recursive unfolding RFP

RFP 的实现是在 FPN 的基础上添加了反馈连接, 其公式如式(2)~式(3):

$$f_i' = F_i'(f_{i+1}', x_i') \quad (2)$$

$$x_i' = B_i'(x_{i-1}', R_i'(f_i'^{t-1})) \quad (3)$$

式中: f_i' 表示第 i 阶段的输出特征; B_i 表示自底向上的第 i 阶段; F_i 表示自顶向下的第 i 阶段; x_i 表示输入图像的特征图; R_i 表示将特征转换连接回自底向上主干之前的特征转换; t 表示递归多少次。

1.3 引导锚点建议网络 (GA-RPN)

传统的区域建议网络 (RPN) 通常采用预定义的不同尺度和纵横比的锚框覆盖特征图上的每一个像素位置, 大量锚框可能会对背景或其他不感兴趣的区域进行预测, 从而产生大量的无效候选框, 增加了后续处理阶段的计算负担, 并可能导致检测精度下降。引导锚点建议网络 (GA-RPN) 是 RPN 改进的一种策略方法, 它允许根据训练数据自适应地调整锚框的大小、形状以及分布, 能够生

成更具有针对性的锚框布局,使得锚框更多地集中在潜在目标可能出现的区域,减少无用锚框的数量,进而提高检测性能和计算效率。引导锚点建议网络如图6所示,图中对于每个输出特征图,使用锚点生成模块分别预测锚框位置和形状,然后在原特征映射上加入特征自适应模块,使新特征映射能够识别锚点形状。

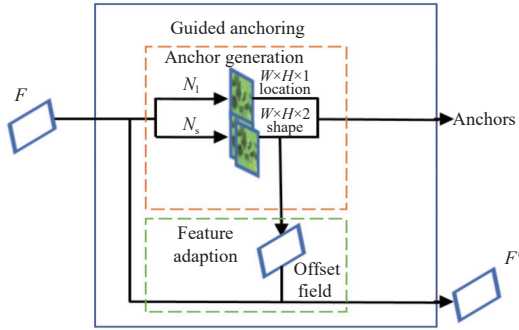


图6 GA-RPN 结构图

Fig. 6 Structure diagram of GA-RPN

1.3.1 锚框的位置预测

锚框位置预测分支(location)会生成一个与输入特征图 F 大小相同的概率图,概率映射 $p(i, j|F)$ 是使用子网络 N_1 预测的。该网络是对特征图进行 1×1 卷积,获得对象得分图,然后通过sigmoid函数将其转换为概率值。根据生成的概率图,通过选择相应概率值高于预定义阈值的位置来确定对象可能存在的活动区域。这个过程可以过滤掉90%的区域,由于不需要考虑那些被排除的区域,将随后的卷积层替换为掩模卷积,以获得更有效的推理。

1.3.2 锚框的形状预测

锚框形状预测分支(shape)会生成两个与输入特征图 F 大小相同的概率图,预测每个位置的最佳形状 (w, h) ,在特征图上预测出的锚框需要经过相应的变换和映射,才能对应到原图上的实际大小,变换公式如式(4)~式(5):

$$w = \sigma \cdot s \cdot e^{dw} \quad (4)$$

$$h = \sigma \cdot s \cdot e^{dh} \quad (5)$$

式中: dw 和 dh 是形状预测分支的输出,然后将其映射到最佳位置 (w, h) ; s 是步距; σ 是一个变量,设置为8。

1.3.3 锚框引导的特征适应

传统的锚框在整个特征图上是均匀的,即在每个位置上具有相同的形状和比例。然而本文锚框的形状因位置而异,大型锚框的功能应该对大区

域的内容进行编码,而小型锚框的功能具有较小的细节编码能力,因此基于底层锚点形状在每个单独位置进行转换特征:

$$f'_i = N_T(f_i, w_i, h_i) \quad (6)$$

式中: f_i 是第 i 个位置的特征; w_i 和 h_i 是对应的锚框的宽和高; N_T 是 3×3 的可变形卷积。从锚框形状预测分支的输出中预测一个偏移域,对带有偏移量的原始特征映射进行可变形卷积,然后进一步执行分类和边界框回归。

1.3.4 Focal loss 损失函数

GA-RPN使用Focal loss损失函数,针对水下目标类别的不平衡,能够避免在训练过程中只关注简单样本,忽略困难样本而造成的检测精度低问题。Focal loss定义如式(7):

$$L_{ol}(p_i) = \begin{cases} -\alpha(1-p_i)^\gamma \log(p_i), & y_i = 1 \\ -(1-\alpha)p_i^\gamma \log(1-p_i), & y_i = 0 \end{cases} \quad (7)$$

式中: p_i 是锚框 i 的输出物体的分数; y_i 是锚框 i 的标签, $y_i=1$ 代表正样本,否则为负样本; α 和 γ 是超参数, α 和 γ 设置为0.25和2.0。

2 实验数据集与评价指标

2.1 数据集

本文在公开的水下图像数据集DUO上进行实验,该数据集用于机器人拾取的水下目标检测,包含了多种水下图像的集合,并且具有合理的注释。它包含7782张图像,其中6671张用于训练,1111张用于测试,种类包括海胆(echinus)、海星(starfish)、海参(holothurian)、扇贝(scallop)。

Pascal VOC是一个通用的目标检测数据集,总共包括20个类别。数据集包括VOC2007和VOC2012。VOC2007数据集有9963张图片,其中5011张图片作为训练集,其余4952张图片作为测试集。VOC2012有11540张图片。本文在VOC2007和VOC2012上训练,在VOC2007测试集评估。

2.2 评价指标

本文使用平均精度(mean average precision, mAP)指标,它由召回率(R)和准确率(P)计算得出,包括AP50(IoU=0.5), AP75(IoU=0.75)和AP。mAP定义为

$$P_{AP} = \sum_{i=1}^n (R_n - R_{n-1}) P_n \quad (8)$$

$$P_{mAP} = \frac{\sum_{i=1}^n P_{AP}}{N} \quad (9)$$

式中: P_{AP} 和 P_{mAP} 值越大,检测精度越高,说明算

法检测越准确; R_n 和 P_n 分别代表第 n 个检测目标的召回率和准确率。

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (10)$$

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \quad (11)$$

式中: N_{TP} 表示准确检测的目标数量; N_{FN} 表示漏检的目标数量; N_{FP} 表示错误检测的目标数量。

2.3 实验环境及训练过程

本实验的实验环境如表 1 所示。

表 1 实验环境表

Table 1 Experimental environment specifications

Environment	Property
System	Ubuntu 22.04
CPU	Intere(R) Core(TM) i9-13900K
GPU	NVIDIA GeForce RTX 4090
CUDA	Version12.2
Compiling Environment	PyTorch 2.0.1+Python 3.8

在 DUO 数据集采用 SGD 优化器, 初始化学习

率为 0.005, 权重衰减为 0.0001, 动量为 0.9。训练轮次 12 轮, 在第 8 轮和第 11 轮将学习率除以 10 以找到最合适参数, 最后一轮达到最好效果。

3 实验结果与分析

3.1 不同模型分析对比

为了验证本文方法的有效性, 在 DUO 数据集上进行测试, 并与经典的目标检测算法检测效果进行对比。本文对比的两阶段目标检测算法有: Faster R-CNN^[4]、Cascade R-CNN^[5]、Detectors^[15]、Dynamic R-CNN^[17]、Libra R-CNN^[18]、SABL(side-aware boundary localization)^[19]。对比的一阶段的目标检测算法有: RetinaNet^[3]、FCOS(fully convolutional one-stage)^[20]、ATSS(adaptive training sample selection)^[21]、AutoAssign^[22]、Deformable DETR(defo-rmable transformers)^[23], 水下目标检测算法有: RoIAttn^[13]、Boosting R-CNN^[14] 等。不同的算法模型的检测结果如表 2 所示。

表 2 不同模型在 DUO 数据集上检测效果

Table 2 Detection performance of different Models on the DUO dataset

Model	AP	AP50	AP75	Echinus	Starfish	Holothurian	Scallop	FDR/%	MDR/%	Param/M	FPS
Faster R-CNN	63.1	83.3	71.0	71.2	73.3	64.0	44.1	39.3	33.4	41.1	63.4
Cascade R-CNN	64.8	83.1	71.9	72.0	74.3	66.3	46.4	39.1	33.8	68.9	51.8
Dynamic R-CNN	64.8	81.8	72.5	72.1	75.8	65.5	45.8	36.7	36.3	41.1	61.6
Libra R-CNN	63.6	82.6	71.8	71.4	73.6	64.9	44.5	47.4	31.7	41.4	58.7
SABL	65.8	83.6	72.8	72.4	75.4	66.6	48.8	38.7	34.3	41.9	54.4
Detectors	67.8	84.9	74.8	74.9	76.0	68.3	52.1	36.3	33.3	123.2	21.2
RetinaNet	62.8	82.0	70.9	70.1	73.6	64.4	42.9	49.5	35.9	36.2	68.7
FCOS	62.9	83.6	70.0	72.2	73.4	62.5	43.3	31.7	40.1	31.8	70.1
ATSS	66.8	84.9	74.5	72.6	76.5	67.7	49.3	32.9	38.5	31.9	65.8
AutoAssign	65.3	84.8	73.0	73.5	74.8	65.4	47.3	32.9	38.6	35.9	65.9
Deformable DETR	63.3	83.9	71.5	68.8	72.8	63.8	47.7	35.4	36.6	39.8	41.6
RoIAttn	64.1	84.2	73.0	70.4	73.4	65.7	46.8	39.7	34.6	47.5	42.2
Boosting R-CNN	67.5	84.9	76.1	73.8	76.4	67.9	51.8	46.3	30.8	45.9	66.2
sg_Faster R-CNN	68.7	86.2	76.4	75.3	77.5	68.6	53.6	31.9	34.3	96.0	19.6

从表 2 中可以看出, 通用目标检测算法和专为水下环境设计的目标检测算法在处理水下图像时, 由于水下环境受到光的散射、吸收、色彩失真以及复杂海底地形的影响, 水下目标检测算法优于通用目标检测算法。在水下目标检测时, 通用的目标检测算法如 Faster R-CNN、Cascade R-CNN、RetinaNet、FCOS 存在检测精度低的问题, 原因是

有水下环境光照散射和吸收、水下目标聚集、形态不规则等因素的影响。而针对类别不平衡提出的 Dynamic R-CNN、Libra R-CNN、ATSS、AutoAssign 等算法在水下目标检测中表现较好, 主要原因在于它们能够有效地处理水下环境中目标分布不均匀的情况, 通过合理的类别权重分配或样本重采样等方法, 提高对少数类别目标的检测能力, 有效

解决了水下目标检测中的类别不平衡问题,从而提升了整体的检测精度。SABL 算法通过引入对目标侧边信息的敏感性,使得模型能够更加关注目标物体的边界细节,从而在水下目标检测中更准确地识别和定位目标。相比之下,Detectors 针对目标检测任务中的尺度变化、遮挡和复杂背景等挑战提出的算法,在水下目标检测中效果显著,主要因为水下环境中目标的尺度常常变化较大,目标出现遮挡和背景复杂干扰等问题。RoIAttn 算法通过对感兴趣的目标区域进行注意力机制的引入,使得模型能够更加聚焦于水下目标区域的关键特征,从而提高了水下目标检测的准确性和鲁棒性。Boosting R-CNN 是一种专注于解决水下目标检测中类别不平衡问题的算法,它采用了重加权的思想,对困难样本增加权重,从而使得模型更加关注那些难以识别的目标。

本文提出的算法在虚检率(FDR)方面表现出色,明显低于其他对比算法,这表明该算法在将非目标物体错误识别为目标物体方面具有显著优势。然而,在漏检率(MDR)方面,与其他模型的表现相差不大,这意味着该算法在识别所有真正目标物体方面并未显著优于其他模型。与原模型 Faster R-CNN 模型相比,本文算法的漏检率略有上升,但虚检率显著下降。这表明我们的改进算法在检测精确度上表现优异,能更有效地避免将非目标物体误识别为目标物体。算法的复杂度用参数量(Param)来展现,实时性采用了每秒帧数(FPS)作为衡量标准。本文提出的算法使用了 SAC 卷积替换普通卷积、递归特征金字塔替代传统特征金字塔,并使用 GA-RPN 提升目标的定位和检测能力。虽然这些改进提升了算法的检测精度和鲁棒性,但也相应地增加了模型的参数量和计算复杂度,从而导致 FPS 较低。实验结果显示,虽然本文算法在参数量和计算复杂度方面略有增加,但在检测精度和处理复杂场景的能力上有显著提升。这表明本文提出的改进算法在某些应用场景中具有明显优势。

本文算法在 Detectors 思想的基础上,通过引入锚框引导建议网络,在水下目标检测中能够有效地生成适应性更强、形状更自适应的锚框,进一步提升了水下目标检测的精度和鲁棒性。在 DUO 数据集上,本文算法相比 Faster R-CNN 算法的 mAP 值提高了 5.6%,在水下目标的类别中也达到了最好的效果,特别是在水下模糊以及小目标

聚集的情况下,例如扇贝(scallop)的检测效果最明显,相比原模型 Faster R-CNN 检测精度提升了 9.5%。和 Detector 算法相比,在海星(starfish)上的检测效果精度增长的最明显,验证了本文的改进方法可以更好地适应目标形状。

3.2 不同模型的检测结果

图 7 展示了不同模型在 DUO 水下数据集的检测结果,黄色框为海胆,黑色框为海星,红色框为海参。可以看到,由于水下环境偏绿,导致和环境相似的海星难以被察觉,第 1 行中,其他模型都未检测出黑色框标注的海星,第 2 行中,红色框标注的海参检测效果较明显,可以看到其他算法均有漏检的海参。图 8 为检测对比图,从图 8(c)和图 8(d)中可以清晰看到,原模型 Faster R-CNN 存在很多漏检,本文模型 sg_Faster R-CNN 可以检测到存在的难以观察到的目标。

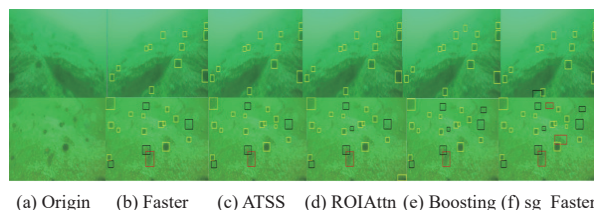


图 7 检测结果图

Fig. 7 Detection results chart

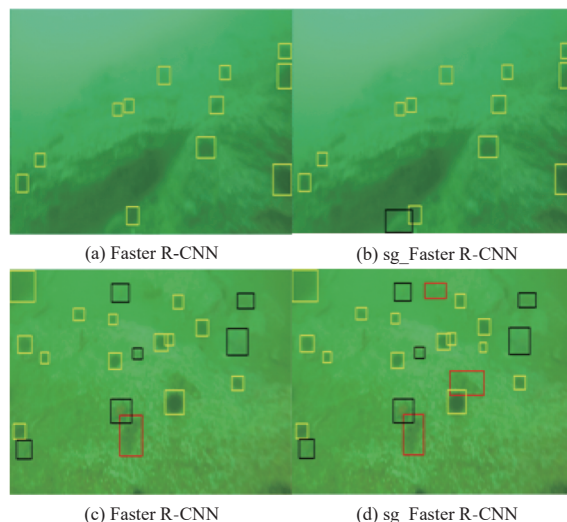


图 8 检测对比图

Fig. 8 Detection comparison chart

3.3 消融实验

3.3.1 SAC 的对比实验分析

如表 3 所示,通过引入 SAC 卷积, Faster R-CNN 模型在 AP、AP50 和 AP75 各项指标上均有所

提升,尤其是在针对扇贝这一具有显著尺度变化特点的目标类别上,其性能改善最为明显:相较于 Faster R-CNN 提高了 5.2% 的平均精度,达到了

49.3%。对比同样用于处理尺度变化问题的 DCN (deformable convnets)^[24] 及其升级版本 DCNV2^[25], SAC 在扇贝类别的准确率提升上也更为明显。

表 3 添加不同卷积在 DUO 数据集上检测效果

Table 3 Detection performance with different convolutional layers on the duo dataset

Model	AP	AP50	AP75	Echinus	Starfish	Holothurian	Scallop
Faster R-CNN	63.1	83.3	71.0	71.2	73.3	64.0	44.1
+SAC	65.7	84.9	74.5	73.7	74.0	66.0	49.3
+DCN	64.6	83.8	72.6	71.8	74.2	66.4	46.0
+DCNV2	64.6	84.1	72.1	72.1	74.1	65.3	46.7

3.3.2 RFP 的对比实验分析

如表 4 所示,在水下复杂环境下的目标检测任务中,采用 RFP 模块改进后的 Faster R-CNN 模型相较于仅使用 FPN 或 PAFPN 表现出更优的性能。

RFP 通过递归式的特征融合机制,有效提升了模型对水下目标多种尺度变化的适应性和对细微结构与全局上下文信息的捕捉能力,从而增强了对不同尺寸水下目标的识别准确性。

表 4 不同颈部在 DUO 数据集上检测效果

Table 4 Detection performance with different neck architectures on the duo dataset

Model	AP	AP50	AP75	Echinus	Starfish	Holothurian	Scallop
Faster/FPN	63.1	83.3	71.0	71.2	73.3	64.0	44.1
+RFP	65.8	83.4	73.1	74.1	74.7	66.5	48.0
+PAFPN	63.4	83.3	71.1	71.1	73.6	64.6	44.2

3.3.3 模型的消融实验

从表 5 可以看出,通过引入 SAC 卷积对全局上下文信息的有效捕获与利用,成功解决了特征提取时空结构信息丢失的问题,使得模型能更好地适应图像中目标物体位置及形态的变化,在扇贝的检测中,效果最为明显。RFP 实现了高层语义特征与底层细节特征之间的多次交互融合,有

利于模型对于尺寸较小且形状复杂的水下目标识别能力的增强。为应对水下环境导致的目标特征模糊问题,通过引入 GA-RPN,根据图像内在语义信息动态生成更稀疏且形状自适应的锚框,从而提升了模型在定位模糊边界及各种尺度水下目标上的准确性。证明了本文所提算法在水下目标检测领域的优越性能。

表 5 模型不同组件的消融实验结果

Table 5 Ablation experiment results of different components in model

Model	AP	AP50	AP75	Echinus	Starfish	Holothurian	Scallop
Faster R-CNN	63.1	83.3	71.0	71.2	73.3	64.0	44.1
+SAC	65.7	84.9	74.5	73.7	74.0	66.0	49.3
+RFP	65.8	83.4	73.1	74.1	74.7	66.5	48.0
+GA-RPN	65.5	84.4	73.8	72.2	74.9	66.8	48.0
sg_Faster	68.7	86.2	76.4	75.3	77.5	68.6	53.6

3.4 通用目标检测的可行性

为了验证提出的水下目标检测算法不只是在水下图像数据集上具有良好的表现,于是将本文

水下目标检测算法应用于通用的目标检测数据集 VOC 上进行测试,实验结果如表 6 所示。

实验结果表明,该算法在通用数据集上的检测

性能也具有较好的表现,证明了本文算法在处理不同光照条件、背景复杂度以及目标特征变化时具备较强的适应能力,从而证实了其不仅适用于水下环境,也能在更广泛的目标检测领域中发挥稳定有效的作用。

表 6 不同模型在 VOC 数据集上检测效果

Table 6 Detection Performance of Different Models on the VOC Dataset

Model	Backbone	mAP
Faster R-CNN	ResNet50	79.5
RetinaNet	ResNet50	77.3
Cascade R-CNN	ResNet50	80.00
Boosting R-CNN	ResNet50	81.6
sg_Fster R-CNN	ResNet50	83.7

4 结论

水下环境存在诸多挑战,包括光线衰弱、色彩失真、复杂的背景干扰、目标尺度多样性以及目标特征模糊等问题。针对这些问题,本文提出了一种基于 Faster R-CNN 改进的算法 sg_Faster R-CNN,旨在解决水下目标检测中尺度多样性和特征模糊等问题。本文算法在水下图像数据集进行验证,结果表明在处理不同尺度的目标和特征模糊目标定位问题上表现良好,证明了本文算法在水下目标检测领域具有优越的性能。

参考文献:

- [1] REDMON J, FARHADI A. Yolov3: an incremental improvement[EB/OL]. [2018-04-08]. <https://arxiv.org/abs/1804.02767>.
- [2] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Berlin: Springer International Publishing, 2016: 21–37.
- [3] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[EB/OL]. [2016-01-06]. <https://arxiv.org/pdf/1506.01497>.
- [4] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[EB/OL]. [2016-01-06]. <https://arxiv.org/pdf/1506.01497>.
- [5] CAI Z, VASCONCELOS N. Cascade R-CNN: delving into high quality object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 6154–6162.
- [6] LIN W H, ZHONG J X, LIU S, et al. Roimix: proposal-fusion among multiple images for underwater object detection[C]//ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New York: IEEE, 2020: 2588–2592.
- [7] WANG Z, LIU C, WANG S, et al. UDD: an underwater open-sea farm object detection dataset for underwater robot picking[EB/OL]. (2020-03-08)[2021-08-28]. <https://arxiv.org/abs/2003.01446>.
- [8] LIU H, SONG P, DING R. WQT and DG-YOLO: towards domain generalization in underwater object detection[C]//IEEE International Conference on Image Processing. New York: IEEE, 2020: 1971–1975.
- [9] CHEN L, ZHOU F, WANG S, et al. SWIPENET: object detection in noisy underwater scenes[J]. *Pattern Recognition*, 2022, 132: 108926.
- [10] FAN B, CHEN W, CONG Y, et al. Dual refinement underwater object detection network[C]//Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XX 16. Berlin: Springer International Publishing, 2020: 275–291.
- [11] YE H C H, LIN C H, KANG L W, et al. Lightweight deep neural network for joint learning of underwater object detection and color conversion[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 33(11): 6129–6143.
- [12] ZHAO Z, LIU Y, SUN X, et al. Composited FishNet: Fish detection and species recognition from low-quality underwater videos[J]. *IEEE Transactions on Image Processing*, 2021, 30: 4719–4734.
- [13] LIANG X, SONG P. Excavating roi attention for underwater object detection[C]//2022 IEEE International Conference on Image Processing (ICIP). New York: IEEE, 2022: 2651–2655.
- [14] SONG P, LI P, DAI L, et al. Boosting R-CNN: reweighting R-CNN samples by RPN's error for underwater object detection[J]. *Neurocomputing*, 2023, 530: 150–164.
- [15] QIAO S, CHEN L C, Yuille A. Detectors: detecting objects with recursive feature pyramid and switchable atrous convolution[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

- New York: IEEE, 2021: 10213-10224.
- [16] WANG J, CHEN K, YANG S, et al. Region proposal by guided anchoring[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2019: 2965-2974.
- [17] ZHANG H, CHANG H, MA B, et al. Dynamic R-CNN: Towards high quality object detection via dynamic training[C]//Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part XV 16. Berlin: Springer International Publishing, 2020: 260-275.
- [18] PANG J, CHEN K, SHI J, et al. Libra R-CNN: towards balanced learning for object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 821-830.
- [19] WANG J, ZHANG W, CAO Y, et al. Side-aware boundary localization for more precise object detection[C]//Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part IV 16. Berlin: Springer International Publishing, 2020: 403-419.
- [20] TIAN Z, SHEN C, CHEN H, et al. Fully convolutional one-stage object detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). New York: IEEE, 2019: 9626-9635.
- [21] ZHANG S, CHI C, YAO Y, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2020: 9759-9768.
- [22] ZHU B, WANG J, JIANG Z, et al. Autoassign: differentiable label assignment for dense object detection[EB/OL]. [2020-11-25]. <https://arxiv.org/pdf/2007.03496>.
- [23] ZHU X, SU W, LU L, et al. Deformable detr: Deformable transformers for end-to-end object detection[EB/OL]. (2020-10-08)[2021-03-18]. <https://arxiv.org/abs/2010.04159>.
- [24] DAI J, QI H, XIONG Y, et al. Deformable convolutional networks[C]//Proceedings of the IEEE International Conference on Computer Vision. New York: IEEE, 2017: 764-773.
- [25] ZHU X, HU H, LIN S, et al. Deformable convnets v2: more deformable, better results[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2019: 9308-9316.