

# 基于 GA-RetinaNet 的水下目标检测<sup>①</sup>



袁明阳, 宋亚林, 张 潮, 沈兴盛, 李世昌

(河南大学 软件学院, 开封 475004)

通信作者: 宋亚林, E-mail: syl@vip.henu.edu.cn

**摘 要:** 水下目标自动检测方法对海洋智能捕捞工作发挥着重要作用, 针对现有目标检测方法存在的对水下生物检测精度不高问题, 提出了一种 GA-RetinaNet 算法的水下目标检测方法. 首先, 针对水下图像存在密集目标的特点, 通过引入分组卷积替换普通卷积, 在不增加参数复杂度的基础上得到更多特征图, 提高模型的检测精度; 其次, 根据水下生物多为小目标生物的特点, 引入上下文特征金字塔模块 (AC-FPN), 利用上下文提取模块保证高分辨率输入的同时获得多个感受野, 提取到更多上下文信息, 并通过上下文注意力模块和内容注意力模块从中捕获有用特征, 准确定位到目标位置. 实验结果显示, 选用 URPC2021 数据集进行实验, 改进的 GA-RetinaNet 算法比原算法检测精度提高了 2.3%. 相比其他主流模型, 该算法对不同类型的水下目标均获得了较好的检测结果, 检测精度有较大提升.

**关键词:** 目标检测; 水下图像; RetinaNet; 分组卷积; AC-FPN

引用格式: 袁明阳, 宋亚林, 张潮, 沈兴盛, 李世昌. 基于 GA-RetinaNet 的水下目标检测. 计算机系统应用, 2023, 32(6): 80-90. <http://www.c-s-a.org.cn/1003-3254/9134.html>

## Underwater Object Detection Based on GA-RetinaNet

YUAN Ming-Yang, SONG Ya-Lin, ZHANG Chao, SHEN Xing-Sheng, LI Shi-Chang

(School of Software, Henan University, Kaifeng 475004, China)

**Abstract:** Automatic underwater object detection methods play an important role in intelligent marine fishing. To address the problem that the existing object detection methods are not accurate enough for underwater biological detection, this study proposes an underwater object detection method based on the GA-RetinaNet algorithm. Firstly, according to the existence of dense objects in underwater images, the study introduces group convolution to replace ordinary convolution, which can provide more feature information without increasing the complexity of parameters and thereby improve the accuracy of the model. Secondly, according to the characteristic that underwater objects are mostly small objects, the attention-guided context feature pyramid network (AC-FPN) is introduced. The context extraction module is used to obtain more receptive fields while guaranteeing high-resolution inputs and thus extract more contextual information. The context attention module and the content attention module are utilized to capture useful features for the accurate positioning of the object. Experimental results show that the improved GA-RetinaNet algorithm enhances the detection accuracy by 2.3% compared with the original algorithm when the URPC2021 dataset is selected. Compared with other mainstream models, the GA-RetinaNet algorithm achieves better detection results for different types of underwater objects, and the detection accuracy is greatly improved.

**Key words:** object detection; underwater images; RetinaNet; group convolution; attention-guided context feature pyramid network (AC-FPN)

① 基金项目: 河南省科技研发项目 (212102210078); 河南省重大科技专项 (201300210400); 河南省重点研发与推广专项 (科技攻关)(202102210380)

收稿时间: 2022-12-05; 修改时间: 2023-01-06; 采用时间: 2023-01-19; csa 在线出版时间: 2023-04-07

CNKI 网络首发时间: 2023-04-10

## 1 引言

随着人们物质生活水平不断提高,对水产品需求量逐渐加大,引得水产品价格不断升高,使我国水产品行业飞速发展<sup>[1]</sup>.在传统水产养殖业中,依靠潜水员下水捕捞作业的方式人力成本过高,无法进行长时间捕捞,同时对潜水员身体有很大伤害,因此利用水下机器人<sup>[2]</sup>进行智能捕捞已经成为发展趋势.目标检测<sup>[3]</sup>是水下捕捞机器人的关键环节,其检测结果将直接影响后续捕捞作业地规划与控制.

国外研究人员对水下目标检测方面的研究开始较早,Olmos等人<sup>[4]</sup>提出了一种检测无约束水下视频的人造物体检测算法,但其检测速度较慢. Barat等人<sup>[5]</sup>提出了一种基于边界的水下目标识别方法,该方法只在目标区域与背景颜色相差较大时,检测精度较高. Chuang等人<sup>[6]</sup>提出了一种由非监督学习和容错检测组成的水下鱼类识别框架,该框架对类不平衡的水下鱼类目标具有较高检测精度.国内近些年也开始重视水下目标检测方面的研究,王慧斌等人<sup>[7]</sup>提出了一种用于水下目标检测的区域显著度计算模型,通过抑制背景区域的显著度,分割得到目标轮廓,实现水下目标检测.马国强等人<sup>[8]</sup>将K-均值聚类算法应用于水下花斑鱼识别,在背景环境简单的情况下识别率较高. Zhu等人<sup>[9]</sup>提出了一种基于显著性区域合并的水下目标检测和分割技术,对水下图像分割效果显著.

此前水下目标检测大多使用传统目标检测算法,适用于简单水下背景的单目标类目标检测,因此很难泛化地运用在智能捕捞工作中.近年来,随着深度学习的飞速发展,基于深度学习的目标检测相较于传统检测方法,具有较好的鲁棒性和准确度,因此,较多的研究者们将基于深度学习的通用目标检测算法应用于水下检测,并起到较好的检测效果.基于深度学习的目标检测<sup>[10]</sup>主要分为二阶段目标检测(如R-CNN<sup>[11]</sup>、Fast R-CNN<sup>[12]</sup>、Faster R-CNN<sup>[13]</sup>算法)和一阶段目标检测(如YOLO<sup>[14]</sup>、SSD<sup>[15]</sup>、RetinaNet<sup>[16]</sup>算法).一阶段目标检测方法较二阶段目标检测方法有着较快的检测速度,如RetinaNet算法使用不同的特征图对不同尺寸的物体进行,提高了检测效率,在实际作业中实用性较高.所以现有的基于深度学习的水下目标检测算法多属一阶段目标检测. Zhou等人<sup>[17]</sup>提出使用3种图像增强方法对数据集进行扩充,再通过Faster R-CNN网络完成对该数据集中水下目标的检测识别. Chen等人<sup>[18]</sup>提出

了一种新的加权损失函数 invert multi-class AdaBoost,在实验的迭代过程中增加权重去学习分类错误的样本,以此减少噪声影响,提高检测网络的精确度.赵力等人<sup>[19]</sup>提出了一种新的非对称双分支水下生物分类模型,其中卷积神经网络分支有效地学习到全局特征,交互分支增强局部特征学习能力.该算法在常用数据集上达到了良好的分类效果. Fan等人<sup>[20]</sup>在SSD模型中使用复合连接骨干网络,引入预测细化方案,提高网络的特征提取能力,缓解水下图像模糊失真的问题,但该方案会增加网络计算量,降低检测速度. Liu等人<sup>[21]</sup>提出了一种增加水下数据集多样性方法,同时提出了DG-YOLO检测模型,该模型在背景统一的情况下检测性能良好.

RetinaNet网络模型对通用目标的检测效果较为理想,但水下目标的特殊形状特征,水下图像存在的对比度低、蓝绿色差、颜色失真等现象,都提高了水下目标检测的难度.为进一步提高水下图像的检测效果,加强目标检测在水下捕捞方面的应用,本文基于RetinaNet网络模型进行修改,提出了一种改进RetinaNet的水下目标检测算法GA-RetinaNet,能够较好地解决水下图像中主要特征不明显、目标较小且密集性高的问题,提高目标检测精度.

本文相关改进工作主要如下:使用分组卷积替换普通卷积,提高相邻特征层的对角相关性,得到多个关注区域不同的互补特征,可以更完整的表达图像特征,有效检测到图像中的密集目标,提高模型的检测精度;引入AC-FPN模块,解决水下图像对比度低,目标容易收到背景特征干扰的问题.通过CEM模块寻找到目标细粒度特征信息,保留下小目标特征,再通过注意力机制使模型在特征图中获取有效生物的语义信息和空间信息,增强感兴趣区域特征响应,准确找到水下目标的具体位置,解决小目标遗漏问题.

## 2 RetinaNet网络模型

一阶段目标检测方法虽检测速度快,但在特征提取的过程中,容易产生大量负样本,造成正负样本失衡,降低检测算法的精确度.新的一阶段检测方法RetinaNet网络中提出Focal loss损失函数,有效缓解了样本类别不平衡带来的检测精度低问题. RetinaNet<sup>[16]</sup>网络模型是Lin等人提出的一种基于回归的目标检测网络,主要由3个模块构成,其结构如图1所示.图1中的

A 为残差神经网络 ResNet50<sup>[22]</sup>, B 为特征金字塔 (feature pyramid networks, FPN)<sup>[23]</sup>, C 为两个全卷积 (fully convolutional networks, FCN)<sup>[24]</sup>子网络. 预处理后的图像首先经过 ResNet50 进行特征提取, 再进入特征

金字塔得到 5 个不同大小的特征图, 在每个特征图上生成锚框, 使用预先设定的交并比阈值进行锚框与目标的匹配, 随后将带有不同锚框的特征图输入分类子网络和回归框子网络, 预测目标出现概率并确定其具体位置.

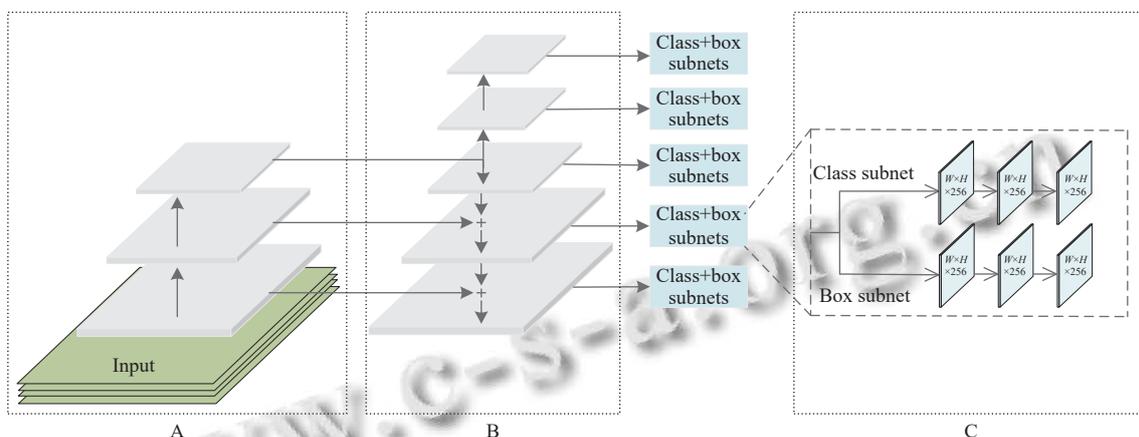


图1 RetinaNet 网络模型

随着神经网络深度的增加, 模型准确率先是不断提高, 到达最大值后, 会毫无预兆地出现大幅下降. 这是由于更深层的神经网络引入过多的非线性转换而导致无法实现简单的线性转换. ResNet 残差网络通过残差块中的快捷连接, 将一部分原始数据直接输出至下一层, 平衡了线性转换和非线性转换之间的矛盾, 不增加额外的网络参数的同时, 解决了网络梯度下降、精确度饱和的问题.

针对图像低层特征的语义信息少, 目标位置准确; 高层特征的语义信息多, 位置信息粗略的特征, 特征金字塔模型通过 top-down 操作和横向连接使高层语义特征和低层位置特征相融合, 图像既保留了高分辨率又扩大了感受野, 增强神经网络的特征提取能力.

样本类别不平衡<sup>[25]</sup>是影响目标检测准确率的常见因素, 数据集中样本数少的类别分类难度高, 就会造成样本不平衡, 导致网络模型训练时出现过拟合现象. RetinaNet 网络模型中的 Focal loss 损失函数, 通过在交叉熵函数前增加一个调节因子, 改进交叉熵函数, 降低训练过程中简单样本的损失权重, 使算法更加关注困难样本的学习, 有利于难检测目标的特征学习.

### 3 改进的 GA-RetinaNet 网络模型的水下目标检测

由于水下环境的特殊性, 图像存在色彩失真、对

比度低的情况, 且水下数据集中检测目标的大小、形状、分布各不相同, 需要从多个方面考虑生成锚框并完成检测任务. 本文针对原 RetinaNet 网络模型中存在的特征不明显、感受野单一、上下文信息联系少的问题提出 GA-RetinaNet 网络模型. 本文网络模型如图 2 所示, 处理后的图像先经过加入分组卷积的 G\_ResNet50 进行特征提取, 获得多个包含不同关注区域的特征图, 通过特征图之间的相互补充, 较好地解决了密集目标不易被检测到的问题; 再将特征图输入到改进的 AC-FPN 模块中进行特征融合, 由 CEM 模块从特征图的多个感受野中找出大量上下文信息, 旨在保留多尺度信息, 再通过 AEM 模块中两个并列的注意力机制从上下文信息获得有效的定位信息, 改进的 AC-FPN 模块缓解了特征图分辨率和感受野之间矛盾, 能捕捉到更多小目标生物, 最后将带有锚框的特征图输入全卷积子网络完成水下目标的检测和定位.

#### 3.1 基于分组卷积的 G\_ResNet50 骨干网络

卷积层的功能是对输入图像进行特征提取, ResNet50 骨干网络在相同时间只通过 1 个卷积层进行特征提取, 但水下图像中会存在密集目标, 使用普通卷积模块无法同时关注到全部特征, 因此需要将原始数据同时进行多次卷积操作. 分组卷积的主要作用是划分输入通道实现多分支效果, 并将所有分支提取

的特征进行级联得到输出结果,在降低参数量的同时不降低模型的训练精度.但本文利用了分组卷积能够加相邻层 filter 之间的对角相关性,通过对原始图像

并列进行多次卷积操作,得到多个关注区域不同的互补特征,再将特征信息进行拼接操作,作为最终的输出特征.

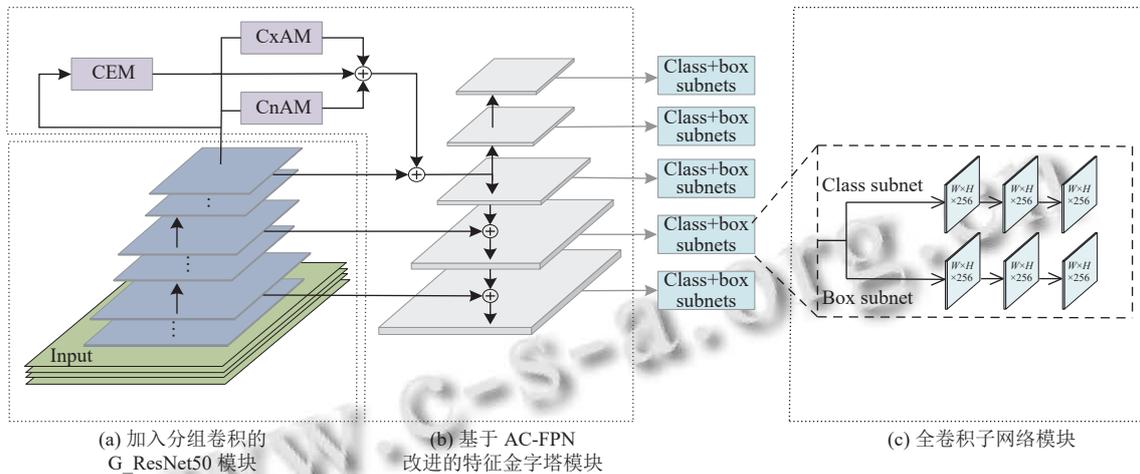


图2 GA-RetinaNet 网络模型

常规的卷积操作通过通道相关性和空间相关性的联合映射提取图像特征.若将卷积层通道相关性和空间性进行解耦合,使其分开映射,可以达到更好的特征提取效果.本文将 ResNet50 残差块中的快捷连接,改为使用分组卷积<sup>[26]</sup>的快捷连接,得到新的基于分组卷积的 G\_ResNet50 骨干网络. G\_ResNet50 网络中不同的卷积路径得到的特征图之间耦合性较低,关注的主要特征也不同,因此特征图可以相互补充,更完整地表达图像特征.

图3为 ResNet50 残差块的快捷连接,  $H(x)$  表示最优解映射,表达式为:

$$H(x) = F(x) + x \quad (1)$$

其中,  $F(x)$  为卷积层的输出,  $x$  为原始输入.

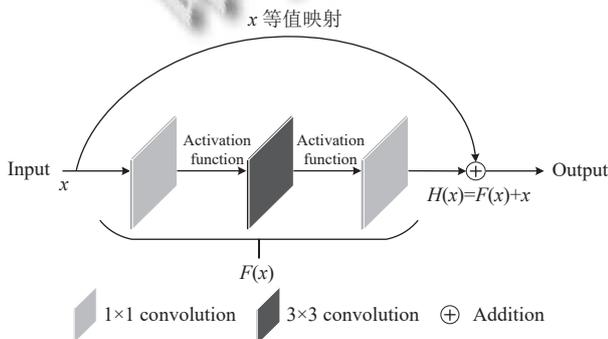


图3 ResNet 残差块的快捷连接

图4为使用分组卷积的快捷连接,此时  $H(x)$  的表达式为:

$$H(x) = \sum_{i=1}^C T_i(x) + x \quad (2)$$

其中,  $T_i(x)$  可以是任意函数,  $C$  为残差块中的分组卷积数量. 本文采用 32 个卷积为一组, 此时使用分组卷积的参数量和原网络参数量大致相同, 能够获得更多特征信息且不增大网络的计算量.

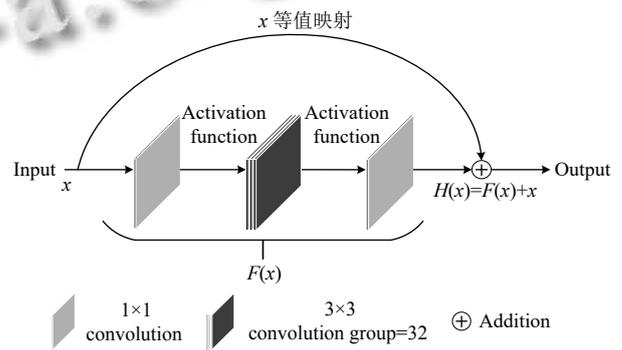


图4 加入分组卷积的快捷连接

### 3.2 基于 AC-FPN 的特征金字塔

目标检测主要通过增加卷积层和下采样操作增大网络感受野,使网络模型能够适用于多尺度检测任务,但这种操作会造成网络参数和显存的增加;网络感受

野增大的同时,特征图尺寸降低,会影响回归框定位的准确性. RetinaNet 网络中特征金字塔通过直接融合不同层级的特征图,缓和大感受野和图像高分辨率的矛盾. 水下图像中的目标多为小目标,特征金字塔中直接元素相加,容易导致不同感受野之间缺乏联系,小目标损失特征信息,无法被检测到<sup>[27]</sup>. 为解决这一问题,本文将注意力引导的上下文特征金字塔模块 (attention-guided context feature pyramid network, AC-FPN)<sup>[28]</sup> 融合进特征金字塔.

AC-FPN 模块能够提取水下目标的细粒度特征,并通过注意力机制使模型偏向于关注目标的位置信息. AC-FPN 模块由两部分组成,分别为上下文提取模块 (context extraction module, CEM) 和注意力引导模块 (attention-guided module, AM). CEM 模块多个感受野

中寻找大量上下文信息,保留更多的有效细节,以此检测到小目标信息; AM 模块使深层特征和浅层特征更好的融合,并校准了空洞卷积引起的位置偏移,从特征图捕获有效的定位信息,进一步提高模型的检测精度.

CEM 模块如图 5 所示,由扩展层和上采样层并列组成. CEM 模块的输入为 G\_ResNet50 网络获得的浅层特征,扩展层采用密集连接<sup>[29]</sup>的空洞卷积模型<sup>[30]</sup>,对输入图像进行不同感受野的特征提取,并在每个连接处引入可变形卷积<sup>[31]</sup>,提高模型的形变能力. 连续的卷积操作能获得丰富的语义信息,但容易造成特征丢失,因此添加上采样层,通过上采样和卷积操作保持输入特征图的粗粒度信息,保留特征图的原始特征. 最后将两个路径的输出拼接,使用卷积操作将粗粒度特征和细粒度特征进行融合.

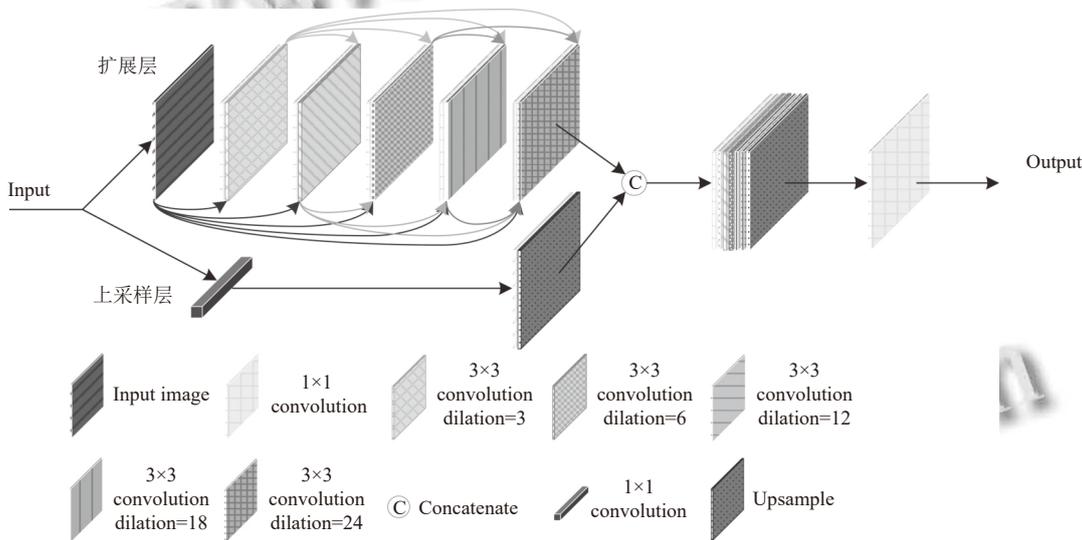


图 5 上下文提取模块

扩展层旨在增强特征图的融合效果同时缓解特征图分辨率和感受野之间矛盾. 根据空洞卷积规律表达式 (式 (3)), 本文选用空洞率为 3, 6, 12, 18, 24 的堆叠空洞卷积.

$$M_i = \max \{M_{i+1} - 2r_i, M_{i+1} - 2(M_{i+1} - r_i), r_i\} \quad (3)$$

其中,  $M_i = r_i$  且  $M_2 \leq k$ ,  $M$  为空洞卷积的卷积率,  $k$  为卷积核大小,  $i$  为卷积层顺序.

通过设置不同空洞率的空洞卷积,由不同感受野进行提取特征,获得多尺度信息,增加找到水下目标的概率;采用密集连接使每一层都接受其前面所有层作为其额外的输入,减轻梯度消失现象,实现浅层特征重

用,增强特征图的传递效果,保留水下小目标特征;通过引入可变形卷积,对卷积核中每个采样点的位置都增加一个偏移量,实现在当前位置附近随意采样但不局限于该固定区域,提高网络对形变的建模能力,更好地提取感兴趣区域特征,解决普通卷积对未知的变化适应性差、泛化能力不强,无法提取出困难水下目标特征的问题.

CEM 模块的算法表达式描述如式 (4), 算法描述如算法 1.

$$Output = Concat(f(x) + b, Upsample(x)) \quad (4)$$

其中,  $f(x)$  为经过扩展层得到的特征,  $b$  为偏置项.

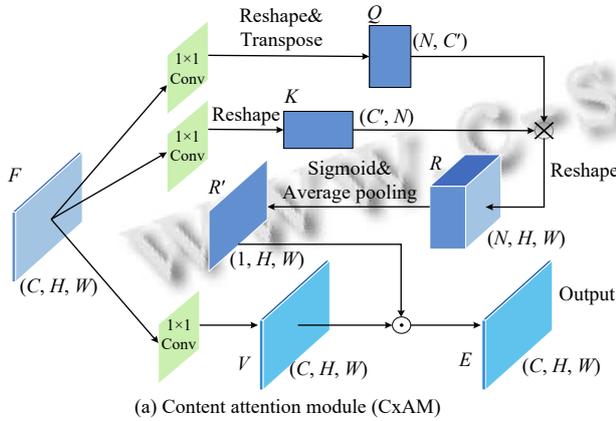
## 算法 1. CEM 算法

输入: 浅层特征 Input

输出: 丰富语义信息特征 Output

1. 将浅层特征 Input 经过扩展层得到  $f(x)$ .
2. 将浅层特征 Input 经过上采样层得到  $g(x)$ .
3. 将  $f(x)$  与  $g(x)$  进行 Concat 操作得到输出特征 Output.

AM 模块如图 6 所示, 由上下文注意力模块 (context attention module, CxAM) 和内容注意力模块 (content attention module, CnAM) 并列组成. CxAM 模块自适应的关注输入进来多尺度、多感受野的语义信



息, 却忽略了特征图的空间信息; CnAM 模块可以缓解 CEM 模块使用可变形卷积造成的目标位置偏移, 但由于其更关注空间信息, 会牺牲相应的语义信息. 通过将 CxAM 模块和 CnAM 模块的输出特征融合, 可获得更全面的特征表示.

CxAM 模块对于 CEM 模块输出的特征图  $F \in \mathbb{R}^{C \times H \times W}$ , 使用卷积层  $W_q$ 、 $W_k$  和  $W_v$  获得新的特征图, 并将其转换为潜在空间, 转换公式为式 (5):

$$Q = W_q^T F, K = W_k^T F, V = W_v^T F \quad (5)$$

其中,  $\{Q, K, V\} \in \mathbb{R}^{C' \times H \times W}$ .

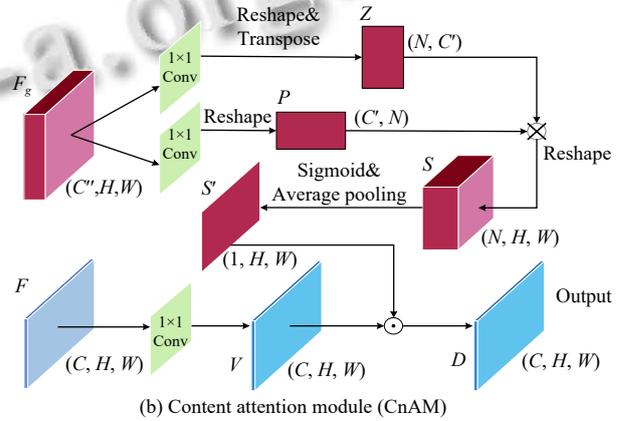


图 6 注意力引导模块

将  $Q$  和  $K$  转换为向量  $\{Q, K\} \in \mathbb{R}^{C' \times N}$ , 其中  $N = H \times W$ . 构造出关系矩阵:  $R = Q^T K$ , 其中  $R \in \mathbb{R}^{N \times N}$ .

将关系矩阵转换为  $R \in \mathbb{R}^{N \times H \times W}$ , 通过 Sigmoid 激活函数 (式 (6)) 和平均池化操作对  $R$  进行归一化, 得到注意力矩阵  $R' \in \mathbb{R}^{1 \times H \times W}$ :

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (6)$$

最后对  $R'$  和  $V$  进行元素相乘, 得到输出特征. 公式如式 (7):

$$E_i = R' \odot V_i \quad (7)$$

CxAM 算法具体描述如算法 2 所示.

## 算法 2. CxAM 算法

输入: 深层特征  $F$ 输出: 语义信息特征  $E$ 

1. 浅层特征  $F$  经过卷积得到特征  $Q$ 、 $K$ 、 $V$ .
2. 特征  $Q$ 、 $K$  经过重构得到新的特征  $Q$ 、 $K$ .
3. 特征  $Q$ 、 $K$  进行求和得到特征  $R$ .
4. 特征  $R$  经过 Sigmoid 函数和平均池化得到特征  $R'$ .
5. 特征  $R'$ 、 $V$  相乘得到输出特征  $E$ .

CnAM 模块和 CxAM 模块的不同在于 CnAM 模块有两个不同的输入特征, 分别为 CEM 模块的输出和 G\_ResNet50 骨干网络的顶层输出, 以此修复被破坏的定位信息. CnAM 使用 G\_ResNet50 骨干网络输出的顶层特征图  $F_g \in \mathbb{R}^{C' \times H \times W}$  生成注意力矩阵, 首先通用卷积层  $W_p$  和  $W_z$  将  $F_g$  转换为潜在空间, 其转换公式为式 (8):

$$P = W_p^T F_g, Z = W_z^T F_g \quad (8)$$

其中,  $\{P, Z\} \in \mathbb{R}^{C' \times H \times W}$ .

将  $P$  和  $Z$  转换为向量  $\{P, Z\} \in \mathbb{R}^{C' \times N}$ , 其中  $N = H \times W$ . 构造出关系矩阵:  $S = P^T Z$ , 其中  $S \in \mathbb{R}^{N \times N}$ .

将关系矩阵转换为  $S \in \mathbb{R}^{N \times H \times W}$ , 通过 Sigmoid 激活函数和平均池化操作对  $S$  进行归一化, 得到注意力矩阵  $S' \in \mathbb{R}^{1 \times H \times W}$ .

最后对  $S'$  和潜在空间  $V$  (见式 (5)) 进行元素相乘, 得到输出特征. 公式如式 (9):

$$D_i = S' \odot V_i \quad (9)$$

CnAM 算法具体描述如算法 3 所示.

## 算法 3. CnAM 算法

输入: 浅层特征  $F_g$ , 深层特征  $F$   
输出: 空间信息特征  $D$

1. 浅层特征  $F_g$  经过卷积得到特征  $Z$ 、 $P$ 、 $V$ 。
2. 深层特征  $F$  过卷积得到特征  $V$ 。
3. 特征  $Z$ 、 $P$  经过重构得到新的特征  $Z$ 、 $P$ 。
4. 特征  $Z$ 、 $P$  进行求和得到特征  $S$ 。
5. 特征  $S$  经过 Sigmoid 函数和平均池化得到特征  $S'$ 。
6. 特征  $V$ 、 $S'$  相乘得到输出特征  $D$ 。

将 CxAM 模块、CnAM 模块和 CEM 模块输出的特征进行相加, 得到更全面的特征表示, 将其输入至全卷积神经网络预测目标出现概率以确定目标的具体位置。

## 3.3 Focal loss 损失函数

GA-RetinaNet 网络模型使用的损失函数为 Focal loss 损失函数, 是基于二分类交叉熵损失 (cross entropy loss, CE) 的动态缩放损失函数, 通过动态缩放因子, 动态的降低训练过程中简单样本的权重, 使算法重心快速聚焦到困难样本上。交叉熵损失的表达式为:

$$CE(p, y) = \begin{cases} -\log(p), & \text{if } y = 1 \\ -\log(1-p), & \text{otherwise} \end{cases} \quad (10)$$

其中,  $y \in \{\pm 1\}$ , 分别代表图像中的前景和背景。  $p \in [0, 1]$  代表 GA-RetinaNet 网络模型预测该物体属于前景的概率。为了表达方便, 对于交叉熵损失表达式中的  $(p, y)$  可以定义为  $p_t, p_t$  表达式为:

$$p_t = \begin{cases} p, & \text{if } y = 1 \\ 1-p, & \text{otherwise} \end{cases} \quad (11)$$

结合式 (10) 和式 (11), 可将交叉熵损失表达式化简为:

$$CE(p, y) = CE(p_t) = -\log(p_t) \quad (12)$$

针对水下数据集中存在类不平衡问题, 在交叉熵损失函数中引入权重因子  $\alpha \in [0, 1]$ , 当样本为正样本时, 权重因子取  $\alpha$ , 当样本为负样本时, 权重因子取  $1-\alpha$ 。针对简单样本和困难样本难以区分的问题, 引入调制因子  $(1-p_t)^\gamma$ , 其中可调聚焦参数  $\gamma \geq 0$ , 可使算法聚焦在困难样本上, 避免整个训练过程过分关注简单样本, 造成目标检测精确度降低。

引入权重因子和调制因子后, Focal loss 损失函数的表达式为:

$$FL(p_t) = -\alpha(1-p_t)^\gamma \log(p_t) \quad (13)$$

当  $p_t \rightarrow 1$ , 说明该样本是简单样本, 像素预测准确, 调制因子  $(1-p_t)^\gamma \rightarrow 0$ , 降低了简单样本的损失比例; 当  $p_t \rightarrow 0$ , 说明该样本被错分为前景, 即像素预测不准确,

$(1-p_t)^\gamma \rightarrow 1$ , 不改变样本的损失比例, 相当于增加了困难样本的损失比例, 使网络更加关注困难样本的学习。表达式中  $\alpha_t$  和  $\gamma$  采用经验值 0.25 和 2.0。

## 4 实验结果与分析

## 4.1 数据集与评价标准

本文针对提出的水下目标检测方法进行实验验证, 数据集选用 2021 年全国水下机器人大赛中的公开数据集 URPC2021, 该数据集包括 7600 张不同尺寸的水下图像以及图像对应的标注文件, 包括海参 (holothurian)、海胆 (echinus)、扇贝 (scallop)、海星 (starfish) 这 4 种生物。将数据按 8:2 的比例随机划分为训练集和验证集。由于数据集中图像分辨率大小不一, 将图像分辨率统一化为  $512 \times 512$  像素, 随机选取训练集中一半图像进行翻转, 用于数据增广。除此之外, 不对图像进行其他处理。

本文主要评估指标采用平均精度 (average precision, AP) 和平均精度均值 (mean average precision, mAP)。平均精度是综合召回率 (recall, R) 和准确率 (precision, P) 的性能指标, 平均精度均值是各目标类别平均精度的平均值。其定义如式 (14) 和式 (15) 所示:

$$AP = \sum_n (R_n - R_{n-1}) P_n \quad (14)$$

$$mAP = \frac{\sum AP}{N(\text{classes})} \quad (15)$$

其中,  $R_n$  和  $P_n$  (式 (16) 和式 (17)) 分别表示第  $n$  个检测阈值下的召回率和准确率,  $N$  为待检测目标的类别数量。AP 值和 mAP 值越大, 说明其检测精度越高。

$$R = \frac{TP}{TP + FN} \quad (16)$$

$$P = \frac{TP}{TP + FP} \quad (17)$$

TP 表示正确检测到水下目标数量; FN 表示漏检的水下目标数量; FP 表示误检为水下目标的数量。

## 4.2 实验环境及过程

本实验具体实验环境如表 1 所示。在训练 URPC2021 数据集时, 本文采用 SGD 作为优化器, 初始学习率设置为 0.005、momentum 动量因子为 0.9、weight-decay 权重衰减因子为 0.0001, batchsize 设置为 2, 总的训练轮次为 50 轮。

表1 实验环境

环境	属性
系统	Ubuntu 16.04
CPU	Intel(R) Xeon(R) CPU E5-2640 v4 @ 2.40 GHz
GPU	NVIDIA GeForce P100
CUDA	10.2
OS	CentOS 7
内存	128 GB
编译环境	PyTorch 1.7.1+ Python 3.8

实验过程如下.

(1) 将数据集按 8:2 随机划分为训练集和数据集并统一分辨率大小, 随机选取一半训练集和全部训练集图像进行翻转.

(2) 将训练集图像输入本文的网络模型进行 1 个轮次训练.

(3) 在每个轮次训练结束后将验证集图像输入评价网络模型并记录评价数据.

(4) 重复步骤 (2) 和步骤 (3) 直至结束整个训练.

### 4.3 实验结果与分析

#### 4.3.1 收敛曲线分析

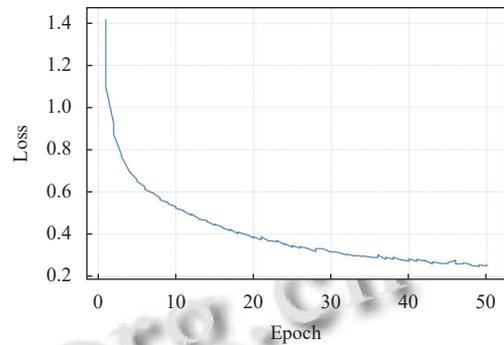
使用 GA-RetinaNet 网络模型在 URPC2021 数据集上经训练后, 实验收敛曲线随迭代次数的变换曲线如图 7. 图 7(a) 为损失函数曲线图, 图 7(b) 为检测精度曲线图. 由图 7 可以看出, 当从开始训练至迭代到第 10 次时, 模型加速收敛, 损失值从 1.4 急剧下降到 0.5 左右, 接着当迭代次数不断增加时, 损失值呈现缓慢下降趋势. 当迭代次数达到 40 次左右时损失值基本达到稳定, 最终本文所使用模型的损失值为 0.244 89. GA-RetinaNet 模型的检测精度随迭代次数的增加逐渐上升, 最终趋于稳定, 本文所使用的模型检测精度达到 61.23%.

#### 4.3.2 不同模型对比分析

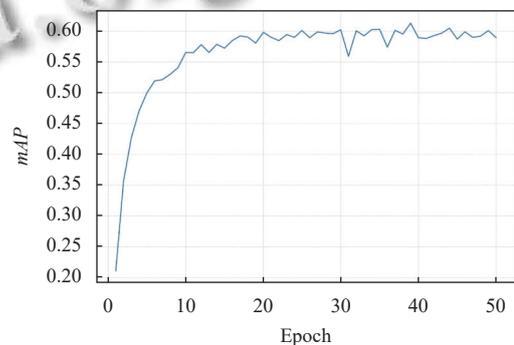
本文选用 Faster R-CNN、SSD、RetinaNet、DG-YOLO、AX-RetinaNet 和 GA-RetinaNet 这 6 种网络模型对 URPC2021 数据集进行测试, 对比不同网络模型的性能与检测结果, 其实验结果如表 2 所示.

Faster R-CNN, SSD, RetinaNet 等主流模型, 对通用目标的检测结果比较理想. 但水下环境造成的蓝绿色差、水下目标不规则的形状特性、废弃垃圾或者海藻引起的多目标遮挡等问题会给水下目标检测带来困难, 从表 2 中可以看出, Faster R-CNN, SSD, RetinaNet 对水下目标检测的检测精度分别为 51.8%、53.06% 和 59.02%, 并未达到理想效果. 且 Faster R-CNN, SSD 的检测速度较慢. 说明主流模型无法直接应用于水下

目标检测, 需要针对检测目标的特点进行模型改进, 提高其检测精度.



(a) Loss



(b) mAP

图7 GA-RetinaNet 网络模型收敛曲线

表2 不同网络模型在 URPC2021 数据集上检测效果

模型名称	AP (%)				mAP (%)	检测速度 (ms)
	holothurian	echinus	scallop	starfish		
Faster R-CNN	22.6	73.8	44.3	66.6	51.81	143.13
SSD	35.2	75.0	34.3	67.8	53.06	71.05
RetinaNet	37.2	76.8	52.7	69.4	59.02	61.18
DG-YOLO	36.4	65.5	42.4	55.5	49.93	131.58
AX-RetinaNet	21.2	72.3	44.8	62.7	50.27	67.11
GA-RetinaNet	38.2	78.4	57.1	71.5	61.29	68.42

DG-YOLO 模型为水下目标检测模型, 但该模型在检测前通过使用 WQT 数据增强, 将海洋、湖泊河流等不同水下环境泛化为统一领域, 统一的背景环境缓和了不同图像之间的色彩差异, 提高了模型检测效果. 当本实验中不对数据集进行 WQT 数据增强, 即不对水下环境进行特殊处理, 使用不同背景环境的图像进行网络的训练和测试, 其目标检测效果不佳, 说明该模型很难满足不同水下环境的实际应用.

AX-RetinaNet 网络模型是 Bao 等人<sup>[32]</sup> 在 2022 年最新提出的 RetinaNet 变体, 通过添加 SE-Net 和 X-module 模块的方法改进网络模型, 用来自动检测和识别茶叶的病虫害情况. 由于 AX-RetinaNet 网络模型的

应用领域并非水下目标检测,且茶叶病虫害和水下目标的形状特点、背景环境差距较大,因此该网络模型的改进并不适用于水下目标检测,其检测精度同样不理想。

较之其他几种网络模型,GA-RetinaNet模型针对水下目标检测这一特殊场景,对RetinaNet网络进行改进,通过加入分组卷积和AC-FPN模块,降低水下偏色对检测造成的影响,不遗漏重叠或密集存在的目标,并捕捉到更多水下小目标,使各种目标的检测精度相较于一些主流的检测模型,都有不同程度的提升,平均检测精度达到了61.3%,有着较高的检测精度,在提高检测精度的同时检测速度略有降低,为68.42 ms。

图8展示了不同网络模型在URPC2021数据集上

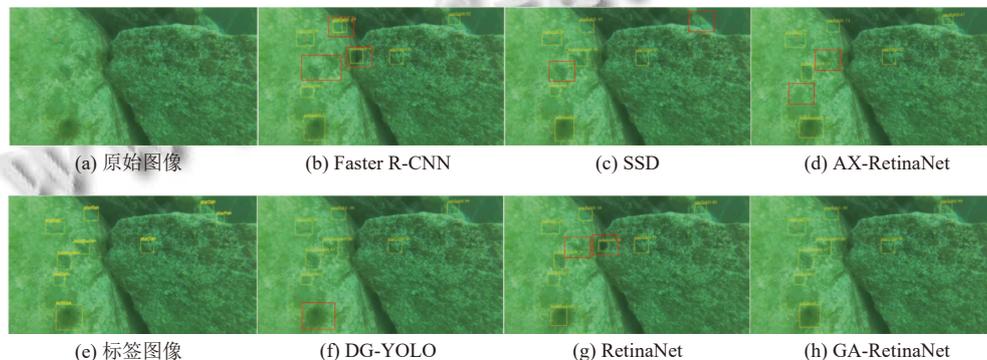


图8 不同网络模型在URPC2021数据集上检测效果图

#### 4.3.3 消融实验分析

为充分说明本文方法提出的改进算法对水下目标检测的有效性,本文在RetinaNet网络模型上依次添加改进模块,对使用G\_ResNet50、AC-FPN模块进行消融实验,对比改进前后模型的检测精确,其实验结果对比如表3所示。

表3 不同改进模块在URPC2021数据集上检测效果

模型名称	AP (%)				mAP (%)	检测速度 (ms)
	holothurian	echinus	scallop	starfish		
RetinaNet	37.2	76.8	52.7	69.4	59.02	61.18
RetinaNet+G_ResNet50	39.4	75.9	54.2	68.7	59.80	61.20
RetinaNet+AC-FPN	36.9	77.4	56.4	69.1	59.94	67.11
GA-RetinaNet	38.2	78.4	57.1	71.5	61.29	68.42

对比RetinaNet网络模型和使用G\_ResNet50替换普通卷积这两组实验,可以看出引入分组卷积使海胆和海星的检测精度降低了0.9%和0.7%,但海参和扇贝的检测精度提升较高,约为2.2%和1.5%,平均检测精度也较原算法提升了0.88%。检测速度与RetinaNet网络

检测效果图。可以看出,GA-RetinaNet模型标注的检测结果更加准确,更接近真实标签。如图8红框标注所示,与GA-RetinaNet网络模型相比,Faster R-CNN、SSD、RetinaNet、DG-YOLO和AX-RetinaNet模型均存在数量不一的漏检、误检现象。如图8(b)–图8(g)所示,由于海水存在蓝绿色差,使岩石和水下目标存在不同程度的偏色,岩石上的海参和海胆都有未被检测出来的情况;如图8(b)和图8(g)所示,岩石中间的阴影部分比较模糊,因此被误认为海胆;GA-RetinaNet网络模型对水下目标的检测更全面,标注的位置也更为精确,因此在这6种网络模型中,GA-RetinaNet的检测效果更好。

模型大致持平,并不会对检测速度造成较大影响。通过向网络模型中引入分组卷积,在同一时间将图像进行多次卷积操作,得到多个不同特征,较好地解决了原网络模型没有考虑到数据集中存在水下目标分布不均且密集存在的问题,减少检测时的遗漏现象,提高了检测精度。

对比RetinaNet网络模型和在特征金字塔中引入AC-FPN这两组实验,可以看出引入AC-FPN使海胆和海星的检测精度稍有降低,各有0.3%,对海参和扇贝的检测精度提高有0.6%和3.7%,平均检测精度提升了0.92%。但检测速度较原网络降低5 ms,说明AC-FPN模块不会引入过多计算量。水下图像中检测目标体积较小,原网络在特征融合时直接将特征元素相加,存在不同感受野之间缺少联系,特征信息损失,造成目标检测精度降低的问题。通过在网络模型中加入AC-FPN,保持特征图分辨率不变的情况下,从不同感受野中提取多个上下文信息进行融合,并由注意力机制从中捕获准确的位置信息,提高目标检测的精度。

对比4项实验结果可以看出,同时引入G\_ResNet50和AC-FPN的网络模型对每个类别的水下目标检测精

度都有较高提升,与原网络模型相比,平均检测精确度提高 2.3%,检测速度降低 7.24 ms,通过略微降低的检测速度提高了检测精度.说明改进方法较原网络模型提取到更多特征信息,减少目标的特征遗漏;缓解了高分辨率和大感受野之间矛盾,增强图像融合效果.综上,GA-RetinaNet 网络模型对水下目标检测有着更准确的检测效果.

表 4 为模型改进前后在 URPC2021 数据集上的性能指标参数表,Params 为网络模型所需参数数量,GFLOPs 为模型的计算量,可以用来衡量算法的复杂度.可以看出本文的算法在引入一定参数数量(+17.34M)的情况下,使精度得到有效提升(+2.3%),且没有过于增大算法复杂度(+0.56 GFLOPs),检测速度与 RetinaNet 模型大致持平.但水下设备的大小受限,运算器的运算能力有限,目前无法实施算法部署,后续对模型进行持续优化后有望进行实际部署.

图 9 展示了不同改进模块在 URPC2021 数据集上检测结果图,从图中可以看出 GA-RetinaNet 模型检测出的图像更接近真实标签,与 RetinaNet 网络模型相比,

分别加入 G\_ResNet50 和 AC-FPN 后检测出的目标数目增多,但误检、漏检和重复选择选中相同目标的情况仍会存在,加入两个模块的结果最理想.如图 9(c) 第 1 行的下方红框所示,加入分组卷积后出现误检现象,选中了没有目标的岩石边缘;如图 9(d) 第 1 行左上方红框所示,加入 AC-FPN 后产生多个明显的重复选框;如图 9(c) 和图 9(d) 第 2 行左侧红框处所示,此处的海胆未被检测出来.如图 9(e) 中第 1 行,通过同时使用两个改进方法,减少了误检、漏检和重复选框的问题,证明本文的改进方法可以使水下目标匹配到合适的锚框;如图 9(e) 第 2 行,GA-RetinaNet 模型能准确地检测出漏检的海胆,证明此时使难检测的水下目标得到相对充分的特征学习.因此本文提出的方法对水下图像的检测效果更好.

表 4 不同改进模块在 URPC2021 数据集上性能指标参数

模型名称	<i>mAP</i> (%)	Params (M)	GFLOPs	检测速度 (ms)
RetinaNet	59.02	36.17	13.16	61.18
RetinaNet+G_ResNet50	59.80	36.17	13.16	61.20
RetinaNet+AC-FPN	59.94	53.51	13.72	67.11
GA-RetinaNet	61.29	53.51	13.72	68.42

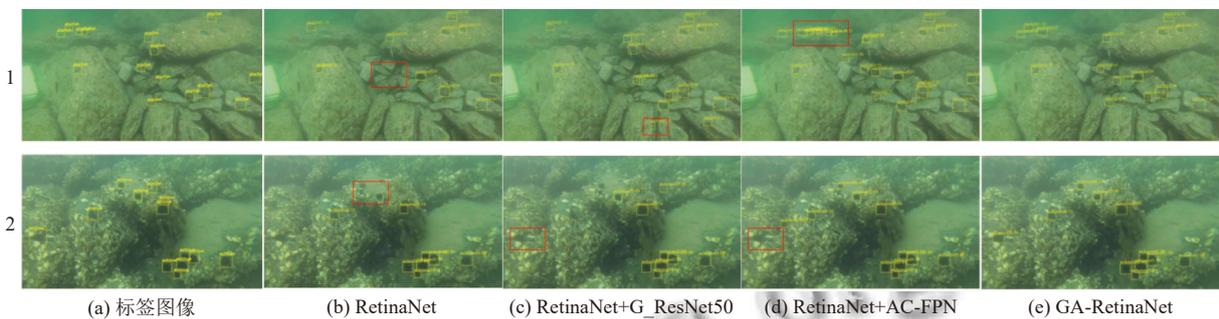


图 9 不同改进模块在 URPC2021 数据集上检测结果图

## 5 结束语

水下目标检测因图像中常有密集存在的目标,且目标形状特征不规则、体积较小,较普通目标检测难度更大.本文设计了一种基于改进 RetinaNet 的水下目标检测算法 GA-RetinaNet,采用分组卷积替换普通卷积,同时在特征金字塔中加入 AC-FPN 模块,通过上下文提取模块丰富语义信息,注意力机制模块增加感兴趣区域的特征响应,提高网络模型的检测精度.实验结果表明,GA-RetinaNet 网络模型对水下目标的平均检测精度可达到 61.3%,能检测出多数的水下目标.

### 参考文献

1 张志锋,贺蓉,吴大千,等.我国海洋生态文明建设和生态

- 环境保护进展、形势与思考.环境与可持续发展,2022,47(3):3-6.[doi:10.19758/j.cnki.issn1673-288x.202203003]
- 2 刘晓阳,杨润贤,高宁.水下机器人发展现状与发展趋势探究.科技创新与生产力,2018,(6):19-20.[doi:10.3969/j.issn.1674-9146.2018.06.019]
- 3 Papageorgiou C, Poggio T. A trainable system for object detection. International Journal of Computer Vision, 2000, 38(1):15-33.[doi:10.1023/A:1008162616689]
- 4 Olmos A, Trucco E. Detecting man-made objects in unconstrained subsea videos. Proceedings of the 2002 British Machine Conference. Cardiff: BMVA Press, 2002. 517-526.
- 5 Barat C, Phlypo R. A fully automated method to detect and segment a manufactured object in an underwater color image. EURASIP Journal on Advances in Signal Processing, 2010, 2010:10.
- 6 Chuang MC, Hwang JN, Williams K. A feature learning and

- object recognition framework for underwater fish images. *IEEE Transactions on Image Processing*, 2016, 25(4): 1862–1872.
- 7 王慧斌, 张倩, 王鑫, 等. 基于区域显著度与水下光学先验的目标检测. *仪器仪表学报*, 2014, 35(2): 387–397.
  - 8 马国强, 田云臣, 李晓岚. K-均值聚类算法在海水背景石斑鱼彩色图像分割中的应用. *计算机应用与软件*, 2016, 33(5): 192–195. [doi: [10.3969/j.issn.1000-386x.2016.05.048](https://doi.org/10.3969/j.issn.1000-386x.2016.05.048)]
  - 9 Zhu YF, Chang L, Dai JL, *et al.* Automatic object detection and segmentation from underwater images via saliency-based region merging. *Proceedings of the OCEANS 2016*. Shanghai: IEEE, 2016. 1–4.
  - 10 Zou ZX, Chen KY, Shi ZW, *et al.* Object detection in 20 years: A survey. *Proceedings of the IEEE*, 2023: 1–20. [doi: [10.1109/JPROC.2023.3238524](https://doi.org/10.1109/JPROC.2023.3238524)]
  - 11 Girshick R, Donahue J, Darrell T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*. Columbus: IEEE, 2014. 580–587.
  - 12 Girshick R. Fast R-CNN. *Proceedings of the 2015 IEEE International Conference on Computer Vision*. Santiago: IEEE, 2015. 1440–1448.
  - 13 Ren SQ, He KM, Girshick R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks. *Proceedings of the 28th International Conference on Neural Information Processing Systems*. Montreal: MIT Press, 2015. 91–99.
  - 14 Redmon J, Divvala S, Girshick R, *et al.* You only look once: Unified, real-time object detection. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016. 779–788.
  - 15 Liu W, Anguelov D, Erhan D, *et al.* SSD: Single shot multibox detector. *Proceedings of the 14th European Conference on Computer Vision*. Amsterdam: Springer, 2016. 21–37.
  - 16 Lin TY, Goyal P, Girshick R, *et al.* Focal loss for dense object detection. *Proceedings of the 2017 IEEE International Conference on Computer Vision*. Venice: IEEE, 2017. 2999–3007.
  - 17 Zhou H, Huang H, Yang X, *et al.* Faster R-CNN for marine organism detection and recognition using data augmentation. *Proceedings of the 2017 International Conference on Video and Image Processing*. Singapore: ACM, 2017. 56–62.
  - 18 Chen L, Liu ZH, Tong L, *et al.* Underwater object detection using invert multi-class AdaBoost with deep learning. *Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN)*. Glasgow: IEEE, 2020. 1–8.
  - 19 赵力, 宋威. 基于非对称双分支交互神经网络的水下生物识别. *计算机应用研究*, 2021, 38(4): 1240–1244, 1255.
  - 20 Fan BJ, Chen W, Cong Y, *et al.* Dual refinement underwater object detection network. *Proceedings of the 16th European Conference on Computer Vision*. Glasgow: Springer, 2020. 275–291.
  - 21 Liu H, Song PH, Ding RW. Towards domain generalization in underwater object detection. *Proceedings of the 2020 IEEE International Conference on Image Processing (ICIP)*. Abu Dhabi: IEEE, 2020. 1971–1975.
  - 22 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016. 770–778.
  - 23 Lin TY, Dollár P, Girshick R, *et al.* Feature pyramid networks for object detection. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017. 936–944.
  - 24 Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Boston: IEEE, 2015. 3431–3440.
  - 25 Oksuz K, Cam BC, Kalkan S, *et al.* Imbalance problems in object detection: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(10): 3388–3415. [doi: [10.1109/TPAMI.2020.2981890](https://doi.org/10.1109/TPAMI.2020.2981890)]
  - 26 Ioannou Y, Robertson D, Cipolla R, *et al.* Deep roots: Improving CNN efficiency with hierarchical filter groups. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017. 5977–5986.
  - 27 Duan KW, Du DW, Qi HG, *et al.* Detecting small objects using a channel-aware deconvolutional network. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020, 30(6): 1639–1652. [doi: [10.1109/TCSVT.2019.2906246](https://doi.org/10.1109/TCSVT.2019.2906246)]
  - 28 Cao JX, Chen Q, Guo J, *et al.* Attention-guided context feature pyramid network for object detection. *arXiv*: 2005.11475, 2020.
  - 29 Huang G, Liu Z, Van Der Maaten L, *et al.* Densely connected convolutional networks. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017. 2261–2269.
  - 30 Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. *arXiv*:1511.07122, 2015.
  - 31 Dai JF, Qi HZ, Xiong YW, *et al.* Deformable convolutional networks. *Proceedings of the 2017 IEEE International Conference on Computer Vision*. Venice: IEEE, 2017. 764–773.
  - 32 Bao WX, Fan T, Hu GS, *et al.* Detection and identification of tea leaf diseases based on AX-RetinaNet. *Scientific Reports*, 2022, 12(1): 2183. [doi: [10.1038/s41598-022-06181-z](https://doi.org/10.1038/s41598-022-06181-z)]

(校对责编: 牛欣悦)