

# 基于空间特征选择的水下目标检测方法

蔡 达, 范保杰

南京邮电大学自动化学院、人工智能学院, 江苏 南京 210023

基金项目: 国家自然科学基金(61876092, U2013210); 中国人工智能协会, 华为 Mindspore 学术奖励基金(BJ221005)

通信作者: 范保杰, jobfbj@gmail.com 收稿/录用/修回: 2021-12-30/2022-02-09/2022-03-17

## 摘要

针对传统目标检测方法在水下识别任务中误检率较高的问题, 基于一阶段全卷积检测器(FCOS)引入多尺度特征选择及中心边界特征选择, 实现高精度水下目标检测。模型中的自适应加权融合特征金字塔通过设置可学习权重加权融合所有的特征层级, 实现多尺度空间特征选择。此外, 为了处理检测中分类和回归任务之间的特征耦合问题, 并分离不同任务之间的共享特征, 设计了基于空间特征解耦的检测头, 实现了中心和边界区域的特征选择。实验中, 针对水下数据集 URPC2018 和 UWD2021 进行性能测试, 并与先进的目标检测方法进行对比。大量的实验结果表明, 基于空间特征选择的 FCOS 模型在水下检测任务中展现出优异的性能, 在 URPC2018 和 UWD2021 上的类平均精度(mean Average Precision, mAP)分别为 82.7% 和 83.3%。

## 关键词

水下目标检测  
一阶段全卷积检测器  
水下数据集  
特征选择  
中图法分类号: TP18  
文献标识码: A

## Spatial Feature Selection for Underwater Object Detection

CAI Da, FAN Baojie

College of Automation & Artificial Intelligence, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

## Abstract

Due to the high error detection rate in the underwater identification task, we introduce multi-scale feature selection and center and boundary feature selection for a fully convolutional one-stage object detector(FCOS) to make the high-quality underwater object detection. In the adaptive weighted fusion feature pyramid, multi-scale spatial feature selection can be realized by setting learnable weights to integrate all feature levels. In addition, in order to solve the feature coupling entanglement between the classification and regression task, and separate the shared features among different tasks, a feature decoupling detection head based on spatial feature was designed to realize the feature selection of center and boundary regions. Experiments are carried out on the underwater datasets URPC2018 and UWD2021, comparing with other object detection methods. Extensive experiments results demonstrate that the proposed method shows extremely excellent performance, achieving 82.7% mAP and 83.3% mAP on URPC2018 and UWD2021.

## Keywords

underwater objection  
detection;  
fully convolutional one-stage  
object detector;  
underwater dataset;  
feature selection

## 0 引言

目标检测是计算机视觉中的主要任务之一。在

深度学习时代之前, 早期的目标检测方法依赖于手工设计的特征, 很难适应多变的应用场景。现如今由于深度学习的发展, 基于神经网络的目标检测方

法越来越受欢迎。两阶段检测器 Faster R-CNN<sup>[1]</sup> 使用 Region Proposal Network (RPN) 代替选择性搜索, 极大地减少了提取候选区域的时间开销。一阶段检测器 YOLO (you only look once)<sup>[2]</sup>、SSD (single shot multiBox detector)<sup>[3]</sup> 系列算法因它们极快的运行速度被工业界所青睐。李欣健等<sup>[4]</sup> 通过深度可分离卷积和数据增强技术在火灾检测取得进展。邱健数等<sup>[5]</sup> 通过级联卷积神经网络融合特征, 完成目标的精确定位和分类。最近, FCOS<sup>[6]</sup>、RepPoints<sup>[7]</sup> 在速度和精度方面取得了不错的成就, 甚至超越了一些两阶段检测器。ATSS (adaptive training sample selection)<sup>[8]</sup>, PAA (probabilistic anchor assignment)<sup>[9]</sup> 在正负样本分配上进行创新, 弥补了一阶段检测器精度不足的问题。GFocal V2 (generalized focal loss)<sup>[10]</sup> 将分类和定位联合表示, 用向量表示包围框的分布。尽管通用目标检测方法飞速发展, 但直接应用在水下场景时往往表现不佳。

水下目标检测作为目标检测重要的分支, 可以应用于海洋开发、海洋勘探等多个场景。配备智能目标检测系统的水下机器人对智慧海洋战略具有重大意义。不幸的是, 水下图像会受到光线吸收和散射的影响, 且水中的杂质会引入噪声并增加散射效果。这些不利影响限制了水下图像和视频的实际应用。

水下图像一般分为声学图像和光学图像。声学图像是由声纳产生的, 它对物体的几何形状很敏感, 但忽略了物体的光学色彩。光学图像则由相机拍摄产生, 它们可以准确地捕捉物体的形状和颜色。因此, 基于光学图像的水下目标检测方法受到了研究人员的青睐。Spampinato 等<sup>[11]</sup> 应用视觉系统检测鱼类的种类和数目。Chen 等<sup>[12]</sup> 提出了一个新型采样加权网络以及一种鲁棒的训练范式来处理水下噪声多的问题。Fan 等<sup>[13]</sup> 通过引入复合连接的主干网络、双重锚框细化、焦损失 (Focal Loss)<sup>[14]</sup>、可变形卷积 (DCN<sup>[15]</sup>) 等优化方式, 在水下数据集上取得了良好的性能。Akkaynak 等<sup>[16]</sup> 利用具有深度信息的水下图像, 提出了一种基于修正模型的颜色恢复方法, 以处理色彩偏移。Lin 等<sup>[17]</sup> 描述了图像之间的相互作用, 将不同的图像候选区域混合, 以模拟遮挡问题。尽管如此, 它们在水下场景的实际应用中仍然不尽如人意。

如何有效地表示和处理多尺度特征是目标检测的主要难点之一。为了在不同尺度上表示不同大小的物体, Lin 等<sup>[18]</sup> 开创了一种自上而下的特征金字塔结构。随后, Liu 等<sup>[19]</sup> 添加了一个额外的路径来

增强特征。近年来, Ghiasi 等<sup>[20]</sup> 采用了神经架构搜索来寻找更好的网络结构, 但它需要花费大量的时间和 GPU 资源, 且难以解释原因。为了实现更加有效、合理的特征融合, BIFPN (bi-directional feature pyramid network)<sup>[21]</sup> 采用了具有更高准确性和更高效率的特征连接块。它通过多次重复堆叠特征块, 共享高分辨率特征图的细节信息和低分辨率特征图的丰富语义。陈琴等<sup>[22]</sup> 利用扩张卷积建立中心邻域金字塔, 捕获不同级别的对比度信息, 最后直接融合不同层级的信息。然而, 受习性影响, 水下生物易群居, 导致对象目标密集, 在网络下采样时, 信息容易丢失。此外, 上述方法在特征融合时只考虑相邻的特征层级, 这进一步限制了网络的表征能力。

为了建模特征区域和任务之间的关系, 提高检测精度, 研究人员开始注意到分类和回归任务之间相互纠缠。Song 等<sup>[23]</sup> 提出如果检测器试图从同一个空间点或锚点推断分类得分和回归坐标, 它总是会得到一个不完美的权衡结果。Kong<sup>[24]</sup> 等提出使用中心采样的方式, 有效去除了大量包含背景的特征。Gidaris 等<sup>[25]</sup> 预测了对象边界在候选框中的可能性。Wang 等<sup>[26]</sup> 采用了一个额外的阶段, 通过将包围框划分为等分的桶来预测边界的移动, 并精确定位边界在桶中的位置。Qiu 等<sup>[27]</sup> 提出使用边界特征来增强单点特征, 以期取得更好的精度。但是这些方法仍无法有效地解除分类和回归任务之间的纠缠, 这使得检测器在水下场景中的表现不佳。

针对以上研究, 解决通用目标检测算法在水下场景的实际应用中检测精度不高的问题, 本文提出了一种基于空间特征选择的水下目标检测方法, 以建模特征区域和分类、回归任务之间的关系。受人类视觉系统的特点启发, 所提出的模型使用物体的中心特征进行分类识别, 边界特征则用于回归定位任务, 进而消除不同任务之间的耦合性, 提高水下目标检测算法的精度与泛化性能。为了减小特征融合时的信息丢失问题, 本文给出了一种特征自适应加权融合策略。该策略可以简单、有效地应用到特征金字塔网络中。FCOS 模型作为一阶段的无锚检测器, 具有良好的精度与效率。因此, 可以作为本实验的基线。本文主要贡献包括:

- 1) 提出了一种水下目标检测网络, 使用中心特征进行分类识别, 边界特征进行回归定位, 可以解除分类和回归任务之间的耦合性, 提高检测精度。

- 2) 提出了自适应加权融合特征金字塔网络, 考

虑所有特征层级的加权融合,可以有效地减少信息丢失问题,增强特征表征能力。

3)在水下数据集 URPC2018<sup>[12]</sup> 和 UWD2021<sup>[13]</sup> 上取得了最先进的结果,分别取得 82.7% mAP、83.3% mAP。

## 1 基于 FCOS 模型的空间特征选择

本文通过对一阶段检测模型 FCOS 的改进来适应水下目标检测任务。以 ResNet-50 作为主干网络,neck 部分采用改进的自适应加权融合特征金字塔,结合特征解耦的检测头,实现高精度水下目标检测。模型的输入是包含 4 个水下生物种类的图片,包含海参、海星、海胆和贝壳这 4 个类别。模型的输出是图片所包含的类别概率和对应的坐标位置。提出的水下目标检测流程,如图 1 所示。

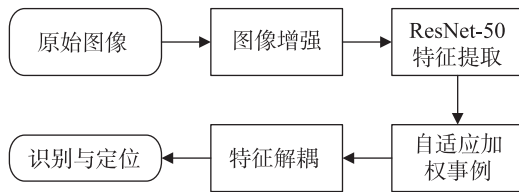


图 1 水下目标检测流程图

Fig.1 The flow chart of the underwater object detection

### 1.1 数据预处理

为了获得更高的精度和泛化性能,与通用目标检测方法一致,对水下数据做数据增强。通过引入颜色抖动、水平翻转、随机旋转和随机裁剪等方法,适应水下目标在不同拍摄条件下的颜色状态变化。输入网络的图片通过插值,将长边固定为 1 333,短边在 640 和 800 之间随机变化。

### 1.2 自适应多尺度特征选择

特征融合的目的是聚合不同分辨率的特征。在特征融合过程中,难免会出现信息丢失和干扰。尤其是在对象密集的水下数据中,简单地堆叠特征会使得问题更加严重。为了防止这一问题,给定一系列多尺度特征,本文的目标是寻找一种新的方法让网络自己学习合适的特征融合方式,来聚合不同级别的特征。BIFPN 使用跨尺度连接和特征加权融合,在数据集<sup>[28]</sup>取得了优异的性能。但在水下数据集中,BIFPN 重复的特征块堆叠,使得冗余信息增多,特征质量下降,导致最终的性能表现不佳,本文将在实验部分详细展示。受 BIFPN 特征加权融合启发,本文让网络学习每个特征层的权值。与 BIFPN 不同的是,在加权融合之前,网络保留了

高层语义特征,且没有堆叠特征块。此外,所提出的方法把所有的特征层级放在一起,而不是只融合相邻层之间的特征。所提出的自适应加权融合特征金字塔网络为不同层级的特征设置可学习的权重,在网络学习过程中去优化该参数,学习每个层级对总体特征的贡献。所设计的自适应加权融合金字塔网络如图 2(b)所示,图 2(a)为 BIFPN 结构。

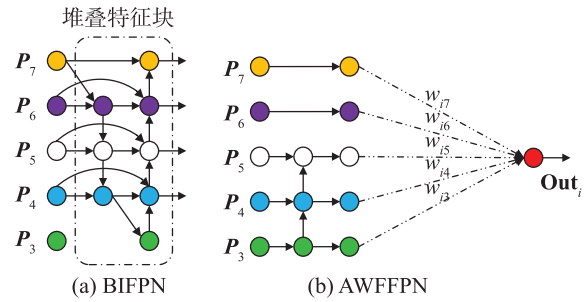


图 2 自适应加权融合特征金字塔网络

Fig.2 The adaptive weighted feature fusion pyramid network

具体来说,该网络将 ResNet-50 输出的特征层 ( $C_3$ 、 $C_4$ 、 $C_5$ ) 作为金字塔网络的输入,分别经过一个步长为 1,卷积核大小为  $3 \times 3$  的卷积层,得到  $P_3$ 、 $P_4$ 、 $P_5$ 。 $P_6$  由  $P_5$  通过一个步长为 2,卷积核大小为  $3 \times 3$  的卷积层得到, $P_7$  则由  $P_6$  经过一个 RE-LU 层后,进行一个步长为 2,卷积核大小为  $3 \times 3$  的卷积计算得到。由此,特征被分为 5 个层级,为了减少信息丢失,并且避免信息干扰,自适应加权融合特征金字塔网络设置了 5 个可学习的权重,分别对应 5 个金字塔层级。这些权重不是固定的数,它们随着网络的训练而优化,并且在得到不同层级的特征输出时,它们的值是不同的。该网络层的输出可用公式表示为

$$\text{Out}_i = w_{i3}P_3 + w_{i4}P_4 + w_{i5}P_5 + w_{i6}P_6 + w_{i7}P_7 \quad (1)$$

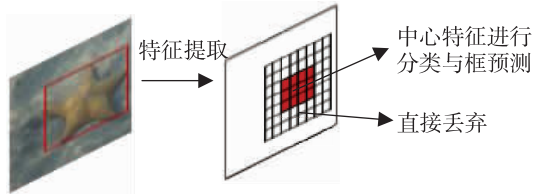
其中,  $\text{Out}_i$  表示该网络层输出,  $w$  表示各层级的可学习权重,  $i$  表示第  $i$  层级。

### 1.3 中心及边界特征选择

精确的目标定位对于目标检测至关重要。目前大多数方法直接应用 RetinaNet<sup>[14]</sup> 的共享检测头结构,FCOS 也不例外。它们使用相同的特征层进行分类和回归任务,忽略了不同任务对特征要求的差异性。共享特征图的检测头虽然被证明是有效的,但这种结构过于简单,无法解除分类任务和回归任务之间的纠缠。

此外,水下生物的形状是不规则的,不能完全覆盖真实标注的边界框。显然,框的边缘区域被大量的背景信息充斥,但同时也包含了物体的边缘信

息。采用中心采样策略, 完全去除样本边缘或者使用边界特征忽略物体中心, 这些显然都是不合理的。在大多数情况下, 分类任务和回归任务共用一个特征层, 这限制了检测器的性能。受人类视觉系统的启发, 视野中心具有较高的视觉敏锐度, 这可



以让人们更快地识别物体。然后, 为了获得更精确的目标位置, 视野中心从目标的中心向周围区域移动, 寻找目标的边缘。因此, 本文提出在分类分支采用对象的中心特征, 回归分支采用边界特征, 如图 3 所示。

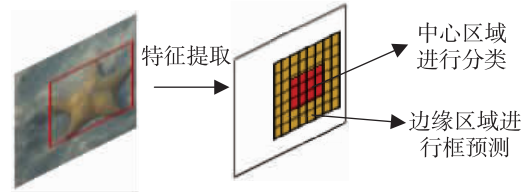


图 3 空间特征选择结构

Fig.3 Spatial feature selection structure

根据以上分析, 为了获得更准确的目标定位, 本文设计了一个空间特征解耦的检测头, 这个检测头是用来实现分类和回归的分离。在分类分支使用中心采样策略, 用公式表示为

$$\begin{aligned} x_1 &= c_x - 0.5w\varepsilon \\ y_1 &= c_y - 0.5h\varepsilon \\ x_2 &= c_x + 0.5w\varepsilon \\ y_2 &= c_y + 0.5h\varepsilon \end{aligned} \quad (2)$$

其中,  $\varepsilon = 0.3$  是缩放系数,  $c_x, c_y$  是中心坐标,  $w$  和  $h$  分别为框的宽和高,  $(x_1, x_2, y_1, y_2)$  为特征提取后, 特征图中对象的真实位置(左上角及右下角坐标值)。它经过了不同倍数的下采样, 经真实框除以  $2l$  后所得数值。 $l$  为网络各层级下采样倍数。将所得的中心特征送入分类结构, 通过连续的卷积层预测分类得分。

回归分支则采用边界特征。受边缘检测启发, 物体的边缘通常是梯度变化大的位置, 而卷积类似于求梯度的过程, 因此设置了一个最大池化层来寻找极值点。具体来说, 对于物体包围框的 4 个边界, 选择边界最大的响应点并且通过插值计算它们的位置。此外, 添加 Group Norm 和 RELU(GN-RELU)层以保证训练阶段的稳定性。在网络最后添加一个卷积核大小为  $3 \times 3$  的卷积层来聚合边界特征。这些边界特征被送入一个简单的卷积层结构, 该结构由 4 个卷积核大小为  $3 \times 3$  的卷积层组成, 且每个卷积层之后都对应一个 GN 层。首先, 边界信息由检测头产生的粗预测框生成。然后, 根据边界信息对包围框进行细化, 提取边界特征点再次预测物体位置。回归分支的输出是由预测框和精细化预测框组合生成。值得注意的是, 与大多数检测器不同

的是, 该网络不限制超出图像边缘的预测框, 并通过插值的方式提供边界特征。边界特征提取结构如图 4 所示。

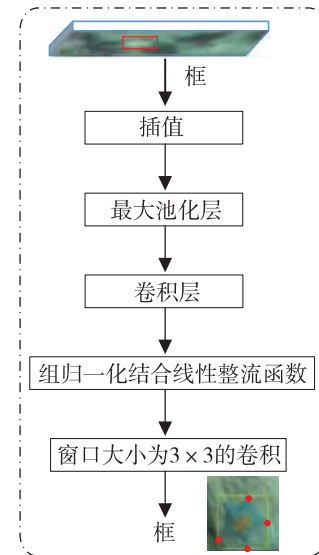


图 4 边界特征提取结构

Fig.4 The structure of the boundary feature extractor

如图 5 所示, 最终结合中心采样分支和边界特征提取分支形成特征解耦检测头, 分离共享特征层, 解除分类任务和回归任务之间的耦合。其中,  $H$  和  $W$  分别指特征图的高和宽。

#### 1.4 损失函数

所提出的水下目标检测网络可以实现端到端的优化, 损失函数被定义为

$$L = \frac{1}{N_{\text{pos}}} \sum_{x, y} (L_{\text{cls}} + L_{\text{reg1}} + L_{\text{reg2}} + L_{\text{cnt}}) \quad (3)$$

其中,  $L_{\text{cls}}$  表示分类损失,  $L_{\text{reg1}}, L_{\text{reg2}}$  表示回归损失,  $L_{\text{cnt}}$  表示中心度损失, 分别使用 Focal loss、GIOU loss、GIOU loss、交叉熵损失 (cross entropy loss)。

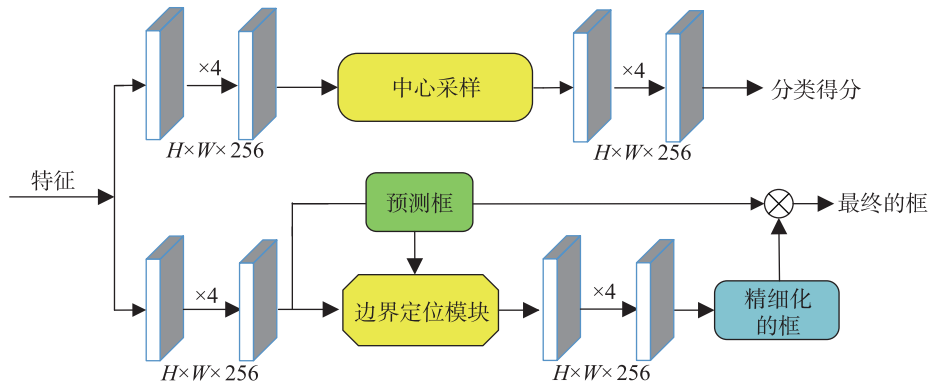


图5 特征解耦检测头

Fig.5 The structure of the feature decoupled head

它们都是关于  $x, y$  的函数,  $x, y$  分别表示预测值和真实值。

## 2 实验结果与分析

本文在 URPC2018<sup>[12]</sup> 和 UWD2021<sup>[13]</sup> 数据集上构建了实验。首先,我们在 URPC2018 上对现有的水下检测算法进行了比较。然后,还将所提出的方法与在水下数据集上的几种最先进的检测器进行了比较。最后,在消融实验中分析了影响检测器性能的关键因素。

### 2.1 实验设置

实验中主要使用 ResNet-50 作为主干网络,使用 ImageNet<sup>[29]</sup> 上预训练的模型作为初始化,并且训练 36 个 epoch。对于特殊的训练周期或使用其他主干网络,实验中将单独标记。网络使用随机梯度下降优化,在总训练轮次的 80% 和 90% 各进行一次学习率衰减,衰减幅度为之前的 10 倍。学习率设置为 0.005,权重衰减系数为 0.000 1,动量为 0.9, Batch Size 设为 2。硬件 CPU 为 i9-9900k,内存大小为 32 G,显卡使用 RTX3090。

### 2.2 数据及评价指标

URPC2018 包含 4 000 张水下图片, UWD2021 包含 12 000 张水下图像。由于 URPC2018 和 UWD2021 水下数据集都使用 VOC<sup>[30]</sup> 数据集中的指标作为评价标准,即 mAP,本文继续沿用了这种标准。指标数值越高,代表检测质量越高,检测器性能越强。公式为

$$\text{mAP} = \frac{\sum_{i=1}^c AP_i}{C} \quad (4)$$

其中,  $AP_i$  表示第  $i$  个类别的平均精度,  $C$  表示检测任务中的类别数量。

### 2.3 结果与分析

**实验 1** 现有的水下目标检测算法较少,表 1 中收录了一些在 URPC2018 数据集上的检测结果。

表 1 与现有的水下检测算法在 URPC2018 上进行比较

Tab.1 Comparison with existing underwater detection algorithms on URPC2018

算法	mAP / %	训练周期 / epoch	主干网络
SWIPENET <sup>[12]</sup>	68.0	120	SWIPENET
FERNet <sup>[13]</sup>	74.2	150	ResNet-50 *
ROIMix <sup>[17]</sup>	74.9	12	ResNet-101
本文方法	79.7	12	ResNet-50
本文方法	82.3	36	ResNet-50

ResNet-50\* 代表复合连接的特征提取主干网络,由 VGG-16<sup>[31]</sup> 联合 ResNet-50 构成。从表 1 可以看出,本文的算法在训练速度和测试精度上有较大优势,超越了上述的水下目标检测算法。

**实验 2** 本文还将一些最先进的检测器纳入对比范围。为了展现本文方法的有效性,实验中为这些检测器选择了结构更复杂的主干网络,例如 ResNeXt<sup>[32]</sup>、Res2Net<sup>[33]</sup>、CBNetV2<sup>[34]</sup>、Swin-Transformer<sup>[35]</sup> 等,结合可变形卷积来增加特征提取能力。而对于本文提出的方法,只使用 ResNet-50,就可以超越上述先进的检测方法。结果如表 2 所示。

从表 2 可以看出,即使对比较大的主干网络,本文的算法依然超过了一些最新的检测器,甚至包括两阶段检测器,例如 Cascade-RCNN、Double-Head-RCNN 等。又如在 COCO 数据集上精度较高的 CenterNetV2,由于未处理特征关系,导致在水下表现出较差的性能。如图 6 所示,为了更加清晰地显示检测的结果,实验中绘制了每个检测类别的 Precision/Recall (PR) 曲线。曲线与坐标轴包围的面积越大,代表检测器性能越好。

表2 与其他检测算法在 URPC2018 上比较  
Tab.2 Comparison with other detection algorithms on URPC2018

算法	mAP / %	主干网络	Hulothurian	Echinus	Scallop	Starfish
FCOS <sup>[6]</sup>	73.4	ResNet-50	68.1	89.7	51.3	84.3
RepPoints <sup>[7]</sup>	76.5	ResNet-50	74.0	89.5	57.0	84.4
ATSS <sup>[8]</sup>	78.0	ResNet-50	78.5	87.1	62.2	84.3
PAA <sup>[9]</sup>	80.3	ResNeXt-101-64x4d-DCN	78.9	91.9	63.4	87.0
GFocalV2 <sup>[10]</sup>	79.8	Res2Net-101-DCN	78.0	91.7	62.6	86.7
RetinaNet <sup>[14]</sup>	75.5	ResNet-50	71.8	90.2	53.9	86.0
Cascade-RCNN <sup>[36]</sup>	81.6	CB-Res2Net-101-DCN	80.6	90.0	68.4	87.5
Sparse-RCNN <sup>[37]</sup>	78.2	Swin-Base	76.2	89.9	62.1	84.6
CenterNetV2 <sup>[38]</sup>	74.5	R2-101-DCN-BIFPN	69.9	86.9	58.3	82.9
DoubleHead-RCNN <sup>[39]</sup>	77.5	ResNet-101	74.9	91.0	58.2	85.8
本文方法	<b>82.3</b>	ResNet-50	80.1	92.2	68.3	88.5
本文方法 <sup>+</sup>	<b>82.7</b>	ResNet-101	80.5	92.8	68.5	88.9

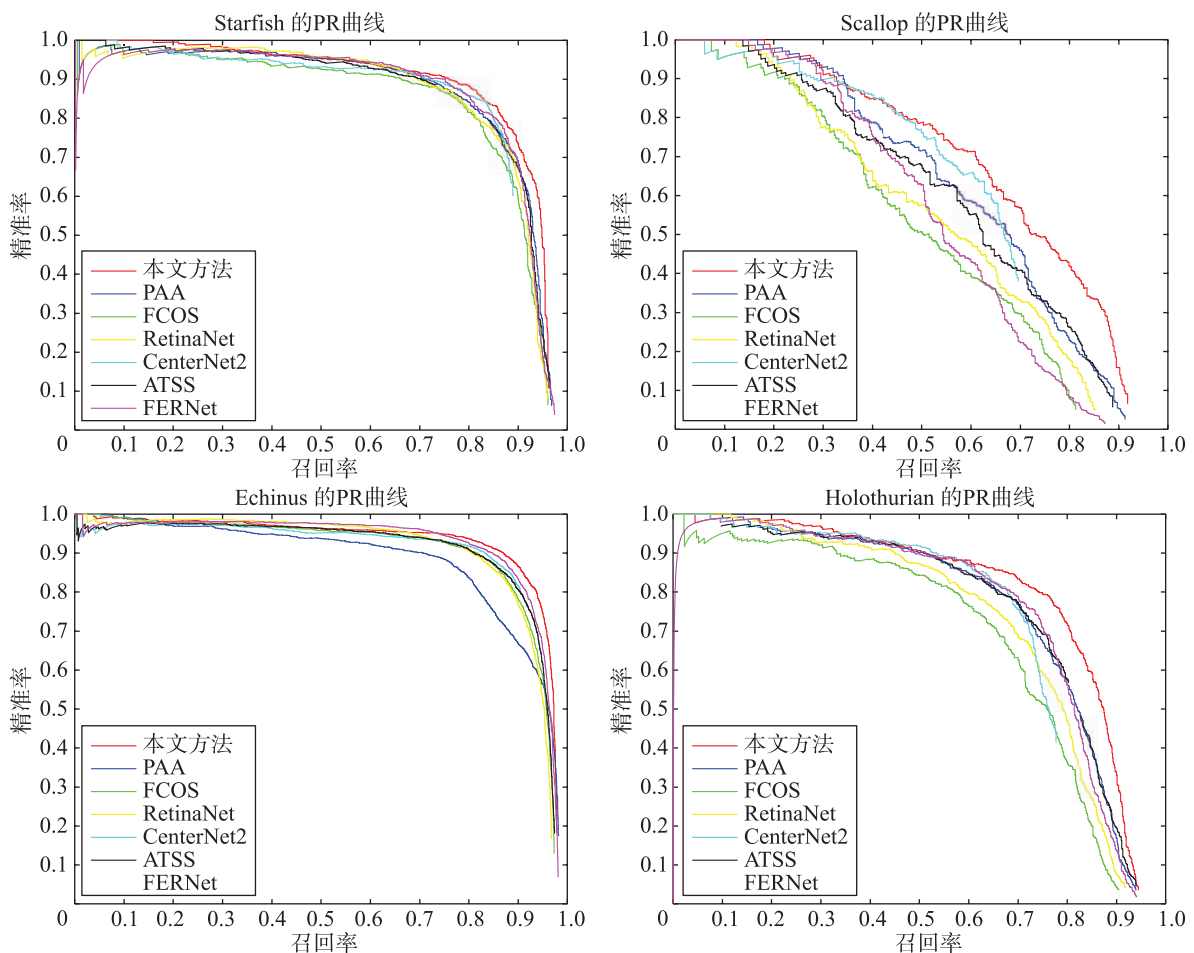


图6 URPC2018 数据集上的 PR 曲线

Fig.6 PR curves on URPC2018

从 PR 曲线中可以看到, 本文提出的方法, 即图中红色线, 每个检测类别与坐标轴包围的面积都高于其他算法, 展现出较好的检测性能。

**实验 3** 此外, 本文还在水下数据集 UWD2021 上进行实验, 与其他先进算法比较。结果如表 3 所

示。

从表 3 中可以看出, 本文的方法优于其他算法。在使用 ResNet-50 作为主干网络的情况下, 具有更小的参数量, 在 UWD2021 上获得了 82.0% mAP。证明了所提出的空间特征选择结构的有效性。

表3 与其他检测算法在 UWD2021 上进行比较

Tab.3 Comparison with other detection algorithms on UWD2021

算法	mAP / %	主干网络	Hulothurian	Echinus	Scallop	Starfish
Faster-RCNN <sup>[1]</sup>	78.2	CB-ResNet-50	78.3	91.0	59.6	85.7
Faster-RCNN * <sup>[1]</sup>	77.4	ResNet-101	76.4	89.7	57.5	86.0
ATSS <sup>[8]</sup>	78.8	ResNet-50	77.2	91.3	60.5	86.3
PAA <sup>[9]</sup>	79.9	ResNeXt-101-64x4d-DCN	79.0	90.4	63.6	87.1
RetinaNet <sup>[14]</sup>	77.3	ResNet-50	76.1	90.8	57.6	84.8
Cascade-RCNN <sup>[36]</sup>	77.2	Swin-Large	76.5	87.7	60.6	84.0
DoubleHead-RCNN <sup>[39]</sup>	77.0	ResNet-101	75.9	90.4	55.9	85.6
VFNet <sup>[40]</sup>	80.7	ResNeXt-101-64x4d-DCN	80.2	91.4	64.3	84.3
FoveaBox <sup>[24]</sup>	78.2	ResNet-50	77.6	90.9	58.4	85.9
RepPointsV2 <sup>[41]</sup>	78.9	ResNet-50	77.5	91.7	59.8	86.7
DetectoRS <sup>[42]</sup>	74.4	ResNeXt-101-32x4d	72.3	86.5	55.1	82.7
NAS-FCOS <sup>[43]</sup>	78.4	ResNet-50	75.4	91.3	60.1	86.6
本文方法	<b>82.0</b>	ResNet-50	80.2	91.8	68.0	88.0
本文方法*	<b>83.3</b>	ResNeXt-101-32x4d	80.9	92.7	69.8	89.7

**实验4** 为了证明所提出模块的有效性,本文设置了消融实验。实验将从多尺度特征选择、中心边界特征选择进行阐述,提供的实验结果都是基于 URPC2018 数据集。实验中逐步增加设计的结构以显示检测器性能变化,实验结果如表4~表6所示。

表4 URPC2018 上的多尺度特征选择

Tab.4 Multiscale feature selection on URPC2018

方法	mAP / %	重复块次数	输出通道
FPN	76.2	-	256
BIFPN	76.5(+0.3)	6	160
BIFPN	76.3(+0.1)	7	288
AWFFPN	<b>79.8(+3.6)</b>	-	256

从表4可知,本文设计的多尺度空间特征选择,有效地处理了不同层级之间的特征关系,能够适应水下目标检测任务,提升了检测器的性能。对于BIFPN中重复的特征块,实验参照FCOSV2和CenterNetV2中的设置。可以看出,重复的连接块,对特征金字塔在水下场景中的特征表述能力的提升有限,而本文所提出的方法相对提高了3.6% mAP的精度。

表5 URPC2018 上的中心边界特征选择

Tab.5 Center and boundary feature selection on URPC2018

方法	mAP / %	仅中心特征	仅边界特征
方法1	73.4	-	-
方法2	76.2(+3.8)	√	-
方法3	78.2(+4.8)	-	√
方法4	<b>80.3(+6.9)</b>	√	√

注:√表示使用了中心特征或者边界特征,-表示没使用。

从表5可知,仅使用中心特征或者边界特征虽然能够提高检测精度,但是检测器性能未被完全释放。使用中心特征可以获得3.8% mAP的精度增益,而使用边界特征可以达到4.8% mAP的增益。通过使用中心特征分类,边界特征回归,解除不同任务之间的耦合性,检测器的性能可以达到80.3% mAP,有了6.9% mAP的精度增幅,有效证明了特征解耦合的必要性。

表6 URPC2018 上的消融实验

Tab.6 Ablation experiments on URPC2018

方法	mAP / %	多尺度特征选择	中心边界特征选择
方法1	73.4	-	-
方法2*	79.8(+6.4)	√	-
方法3*	78.2(+7.1)	-	√
方法4*	<b>82.3(+8.9)</b>	√	√

从表6可以看出,经过空间特征选择,检测器模型逐步提高到82.3% mAP。检测可视化结果如图7所示。

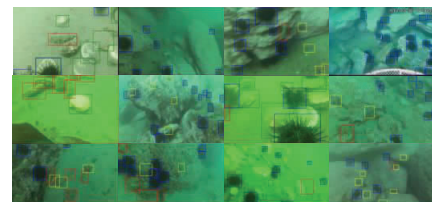


图7 本文方法的检测结果

Fig.7 The detection results of the proposed method

### 3 结论

本文分析了水下场景中限制检测器性能的挑战

性问题, 提出了基于空间特征选择的水下目标检测方法, 考虑多尺度特征选择和中心边界特征选择, 结合自适应加权融合特征金字塔网络和特征解耦检

测头, 有效地提升了检测器的性能。在两个水下数据集 URPC2018 和 UWD2021 上取得了较好的结果, 证明了所提出方法的有效性。

## 参考文献

- [ 1 ] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. *Advances in Neural Information Processing Systems*, 2015, 28: 91 – 99.
- [ 2 ] 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*. Piscataway, USA: IEEE, 2016: 779 – 788.
- [ 3 ] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//*European Conference on Computer Vision*. Berlin, Germany: Springer, 2016: 21 – 37.
- [ 4 ] 李欣健, 张大胜, 孙利雷, 等. 复杂场景下基于 CNN 的轻量火焰检测方法[J]. *模式识别与人工智能*, 2021, 34(5): 415 – 422.  
Li X J, Zhang D S, Sun L L, et al. CNN-based lightweight flame detection method in complex scenes[J]. *Pattern Recognition and Artificial Intelligence*, 2021, 34(5): 415 – 422.
- [ 5 ] 孟志青, 邱健数. 基于级联卷积神经网络的复杂花色布匹瑕疵检测算法[J]. *模式识别与人工智能*, 2020, 33(12): 1135 – 1144.  
Meng Z Q, Qiu J S. Defect detection algorithm of complex pattern fabric based on cascaded convolution neural network[J]. *Pattern Recognition and Artificial Intelligence*, 2020, 33(12): 1135 – 1144.
- [ 6 ] Tian Z, Shen C, Chen H, et al. Fcos: Fully convolutional one-stage object detection[C]//*IEEE Conference on Computer Vision*. Piscataway, USA: IEEE, 2019: 9627 – 9636.
- [ 7 ] Yang Z, Liu S, Hu H, et al. Reppoints: Point set representation for object detection[C]//*IEEE Conference on Computer Vision*. Piscataway, USA: IEEE, 2019: 9657 – 9666.
- [ 8 ] Zhang S, Chi C, Yao Y, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection[C]//*IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway, USA: IEEE, 2020: 9759 – 9768.
- [ 9 ] Kim K, Lee H S. Probabilistic anchor assignment with IOU prediction for object detection[C]//*European Conference on Computer Vision*. Berlin, Germany: Springer, 2020: 355 – 371.
- [ 10 ] Li X, Wang W, Hu X, et al. Generalized focal loss v2: Learning reliable localization quality estimation for dense object detection[C]//*IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway, USA: IEEE, 2021: 11632 – 11641.
- [ 11 ] Spampinato C, Chen-Burger Y H, Nadarajan G, et al. Detecting, tracking and counting fish in low quality unconstrained underwater videos[J]. *VISAPP*, 2008, 2(1): 514 – 519.
- [ 12 ] Chen L, Zhou F, Wang S, et al. SWIPENET: Object detection in noisy underwater images[EB/OL]. arXiv: 2010.10006, 2020. (2020 – 10 – 19) [2021 – 12 – 29]. <https://arxiv.org/abs/2010.10006>.
- [ 13 ] Fan B, Chen W, Cong Y, et al. Dual refinement underwater object detection network[C]//*European Conference on Computer Vision*. Berlin, Germany: Springer, 2020: 275 – 291.
- [ 14 ] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//*IEEE International Conference on Computer Vision*. Piscataway, USA: IEEE, 2017: 2980 – 2988.
- [ 15 ] Dai J, Qi H, Xiong Y, et al. Deformable convolutional networks[C]//*IEEE International Conference on Computer Vision*. Piscataway, USA: IEEE, 2017: 764 – 773.
- [ 16 ] Akkaynak D, Treibitz T. Sea-thru: A method for removing water from underwater images[C]//*IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway, USA: IEEE, 2019: 1682 – 1691.
- [ 17 ] Lin W H, Zhong J X, Liu S, et al. Roimix: Proposal-fusion among multiple images for underwater object detection[C]//*IEEE International Conference on Acoustics, Speech and Signal Processing*. Piscataway, USA: IEEE, 2020: 2588 – 2592.
- [ 18 ] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//*IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway, USA: IEEE, 2017: 2117 – 2125.
- [ 19 ] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]//*IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway, USA: IEEE, 2018: 8759 – 8768.
- [ 20 ] Ghiasi G, Lin T Y, Le Q V. NAS-FPN: Learning scalable feature pyramid architecture for object detection[C]//*IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway, USA: IEEE, 2019: 7036 – 7045.
- [ 21 ] Tan M, Pang R, Le Q V. Efficientdet: Scalable and efficient object detection[C]//*IEEE conference on Computer Vision and*

- Pattern Recognition. Piscataway, USA: IEEE, 2020; 10781 – 10790.
- [22] 陈琴, 朱磊, 后云龙, 等. 基于深度中心邻域金字塔结构的显著目标检测[J]. 模式识别与人工智能, 2020, 33(6): 496 – 506. Chen Q, Zhu L, Hou Y L, et al. Salient object detection based on deep center-surround pyramid[J]. Pattern Recognition and Artificial Intelligence, 2020, 33(6): 496 – 506.
- [23] Song G, Liu Y, Wang X. Revisiting the sibling head in object detector[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2020; 11563 – 11572.
- [24] Kong T, Sun F, Liu H, et al. Foveabox: Beyond anchor-based object detection[J]. IEEE Transactions on Image Processing, 2020, 29: 7389 – 7398.
- [25] Gidaris S, Komodakis N. Locnet: Improving localization accuracy for object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2016; 789 – 798.
- [26] Wang J, Zhang W, Cao Y, et al. Side-aware boundary localization for more precise object detection[C]//European Conference on Computer Vision. Berlin, Germany: Springer, 2020; 403 – 419.
- [27] Qiu H, Ma Y, Li Z, et al. Borderdet: Border feature for dense object detection[C]//European Conference on Computer Vision. Berlin, Germany: Springer, 2020; 403 – 419.
- [28] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: Common objects in context[C]//European Conference on Computer Vision. Berlin, Germany: Springer, 2014; 740 – 755.
- [29] Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2009; 248 – 255.
- [30] Everingham M, Van Gool L, Williams C K I, et al. The Pascal visual object classes (VOC) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303 – 338.
- [31] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. arXiv: 1409.1556, 2014. (2015-4-105)[2021-12-29]. <https://arxiv.org/abs/1409.1556>.
- [32] Xie S, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2017; 1492 – 1500.
- [33] Gao S H, Cheng M M, Zhao K, et al. Res2Net: A new multi-scale backbone architecture[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(2): 652 – 662.
- [34] Liang T, Chu X, Liu Y, et al. CBNetV2: A composite backbone network architecture for object detection[EB/OL]. arXiv: 2107.00420, 2021. (2021-7-12)[2021-12-29]. <https://arxiv.org/abs/2107.00420>.
- [35] Liu Z, Lin Y, Cao Y, et al. Swin Transformer: Hierarchical vision transformer using shifted windows[EB/OL]. arXiv: 2103.14030, 2021. (2021-3-25)[2021-12-29]. <https://arxiv.org/abs/2103.14030>.
- [36] Cai Z, Vasconcelos N. Cascade R-CNN: Delving into high quality object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2018; 6154 – 6162.
- [37] Sun P, Zhang R, Jiang Y, et al. Sparse R-CNN: End-to-end object detection with learnable proposals[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2021; 14454 – 14463.
- [38] Zhou X, Koltun V, Kr? henbühl P. Probabilistic two-stage detection[EB/OL]. arXiv: 2103.07461, 2021. (2021-3-12)[2021-12-29]. <https://arxiv.org/abs/2103.07461>.
- [39] Wu Y, Chen Y, Yuan L, et al. Rethinking classification and localization for object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2020; 10186 – 10195.
- [40] Zhang H, Wang Y, Dayoub F, et al. Varifocalnet: An IOU-aware dense object detector[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2021; 8514 – 8523.
- [41] Chen Y, Zhang Z, Cao Y, et al. RepPoints V2: Verification meets regression for object detection[EB/OL]. arXiv: 2007.08508, 2020. (2020-7-16)[2021-12-29]. <https://arxiv.org/abs/2007.08508>.
- [42] Qiao S, Chen L C, Yuille A. Detectors: Detecting objects with recursive feature pyramid and switchable atrous convolution [C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2021; 10213 – 10224.
- [43] Wang N, Gao Y, Chen H, et al. NAS-FCOS: Fast neural architecture search for object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2020; 11943 – 11951.

## 作者简介

蔡 达(1996 –), 男, 硕士生。研究领域为模式识别, 水下目标检测。

范保杰(1983 –), 男, 博士, 教授。研究领域为机器人视觉和机器学习, 目标检测、跟踪与语义分割。