

基于样本迭代融合的海洋生物检测

吴立栋¹, 彭宗举^{1,2*}, 李欣², 苏涛², 陈芬², 王晓东¹

¹宁波大学信息科学与工程学院, 浙江 宁波 315000;

²重庆理工大学电气与电子工程学院, 重庆 310027

摘要 海洋生物相互聚集形成遮挡现象是误检和漏检的重要原因。为了解决这个问题,提出一种采用样本迭代融合辅助网络训练的海洋生物检测方法。首先,选用改进后的深度空洞残差结构作为特征提取网络,提升了网络的特征提取能力;然后,结合海洋生物图像目标遮挡、密集的特点,改进损失函数避免发生误检、漏检现象;最后,为了进一步解决目标遮挡、数据不平衡的问题,提出样本迭代融合方法,生成模拟图像扩充训练集,提高了网络训练的有效性和对小样本量海洋生物的检测能力。实验结果表明,所提海洋生物检测方法在 URPC2018 和台湾鱼类数据集上的准确率分别达 91.36% 和 90.27%,检测准确率和速度高于现有目标检测算法。

关键词 海洋生物检测; 样本迭代融合; 深度学习; 水下目标检测识别; 数字图像处理

中图分类号 TP391

文献标志码 A

DOI: 10.3788/LOP212567

Marine Creature Detection Based on Sample Iterative Fusion

Wu Lidong¹, Peng Zongju^{1,2*}, Li Xin², Su Tao², Chen Fen², Wang Xiaodong¹

¹Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo 315000, Zhejiang, China;

²School of Electrical and Electronic Engineering, Chongqing University of Technology, Chongqing 310027, China

Abstract Occlusion caused by gathering of marine creatures together is an important reason for false and missed detections. Therefore, this study proposes a marine creature detection method based on iterative fusion of sample-assisted network training. First, an improved deep hole residual structure is selected as the feature extraction network, which improves the feature extraction ability of the network. Second, because of the occlusion and dense characteristics of marine creature images, the loss function is improved to avoid false and missed detections. Finally, to solve the problems of target occlusion and data imbalance, a sample iterative fusion method is proposed to generate an extended training set of simulated images. This improves the effectiveness of network training and the ability to detect marine creatures with a small sample size. The experimental results show that the proposed method can achieve a detection accuracy of 91.36% on the URPC2018 dataset and 90.27% on the Taiwan fish dataset. The detection accuracy and speed of the proposed method are higher than those of existing target detection algorithms.

Key words marine creature detection; sample iterative fusion; deep learning; underwater target detection and recognition; digital image processing

1 引言

海洋生物检测在水产养殖等方面有着至关重要的作用。海洋成像环境复杂多变,水下图像往往存在对比度低、纹理失真、光照不均匀等问题,检测效率低下。目前,设计适应全球各地海洋环境及不同成像条件的海洋生物检测方法面临巨大挑战,且少有一种增强方法可以同时处理不同退化类型的水下图像^[1-2]。Fish4Knowledge 项目^[3]公开了珊瑚礁鱼类数据集,获

得了较为准确的渔情并制定了相应的捕捞计划,推动了海洋生物检测技术的发展^[4]。

海洋生物检测属于图像目标检测的应用范畴。近年来,深度学习在目标检测领域得到了广泛的应用^[5]。基于卷积神经网络(CNN)的目标检测方法主要分为两大类:第一类为两步目标检测方法,先生成一系列样本候选框,再通过 CNN 进行样本分类,常见方法有 R-CNN、Fast R-CNN 和 Faster R-CNN^[6]等;第二类为单步检测方法,直接通过回归得到类别概率和位置坐

收稿日期: 2021-09-22; 修回日期: 2021-10-18; 录用日期: 2021-11-08; 网络首发日期: 2021-11-18

通信作者: *pengzongju@126.com

标值,典型方法有 SSD^[7]、RFB^[8]、YOLO^[9-11] 和 EfficientDet^[12]等。两步目标检测方法检测精度较高,但运算速度慢;单步目标检测方法速度快但检测精度较低。

现有的海洋生物检测方法大多在目标检测网络框架上加入图像预处理模块以提升检测精度。Zheng 等^[13]提出一种基于时频特征的深度神经网络的浅海鱼类检测方法,并分析了不同学习速率对检测性能的影响。张有波等^[14]利用多粒度剪枝策略对 YOLOv4 进行通道和卷积层的双向压缩,平均检测速度可以达到 18.2 frame/s。Guan 等^[15]提出一种对水体散射噪声与鱼类目标信号的偏振特性进行正确估计的方法。Zhang 等^[16]提出一种基于 Faster R-CNN 的水下鱼类检测方法,实现了鱼类数据采集自动化,并使用改进的非最大抑制来减小部分鱼类的分类错误,提高了平均精度均值(mAP)。Mandal 等^[17]将 Faster R-CNN 作为主干网络,结合 ZFNet、CNN 和 VGG16 三个网络,对共计 50 种鱼类和甲壳类动物进行检测,mAP 达 82.4%。Rathi 等^[18]在 Fish4Knowledge 数据集中对 21 种热带鱼进行分类,前景增强去噪预处理后使用 YOLOv3 完成识别,mAP 达 96.29%。Chen 等^[19]以 SSD 网络为基本框架,提出多类逆变 Adaboost 方法,用于提高对小物体的检测精度。Lin 等^[20]提出 RoIMix 增强方法来模拟鱼类间的遮挡,在目标聚集的 URPC2018 数据集上的 mAP 达 74.92%,在 VOC2007 上的 mAP 达 80.8%。Han 等^[21]采用灰度阴影法实现了水下视觉增强,对

CNN 结构进行调整,与 Faster R-CNN 和 YOLOv3 相比,mAP 得到较大提升。

上述海洋生物检测方法为了提升检测性能,大部分增加了图像增强预处理模块,往往局限于某种水体环境,不能适用于多种水下场景的海洋生物检测。其中大多数方法^[13-19]忽略了水下场景中引起漏检和误检的海洋生物间的遮挡现象,从而导致目标检测失败^[22]。此外,这类检测方法不能很好地平衡检测精度和实时性。针对现有方法的缺点,本文提出了一种基于样本迭代融合的海洋生物检测方法。一方面,建立轻量化的海洋生物检测网络模型,针对遮挡现象设计了新的损失函数;另一方面,采用水下样本迭代融合方法扩充训练数据集,以解决目标遮挡、样本不平衡的问题。通过在两个数据集上的实验,结果证明所提方法不仅有效地解决了鱼群遮挡导致的漏检、误检等问题,而且平衡了样本数量,在保证检测实时性的同时获得了较高的检测精度。

2 算法描述

所提方法的结构如图 1 所示,主要包括检测网络和水下样本迭代融合两部分。在检测网络中,构建了轻量化的海洋生物识别循环网络模型,融合了目标重合惩罚因子,以降低鱼群遮挡场景中目标的重检、误检概率。在水下样本迭代融合模块中,采用水下样本迭代融合法,分配融合权重因子,以降低样本量不平衡、目标遮挡等对检测效果的影响。

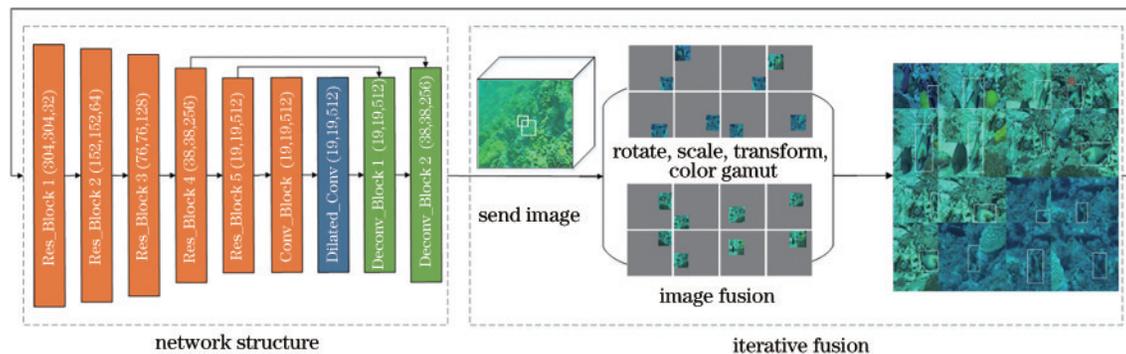


图 1 基于样本迭代融合的海洋生物检测网络模型

Fig. 1 Marine biological detection network model based on sample iterative fusion

2.1 样本权重更新检测网络

目标检测不仅需要精确地识别目标种类,还需精确定位目标位置。R-CNN 及其衍生的 Fast R-CNN、Faster R-CNN^[8]成功地利用 CNN 下采样操作时具有的强语义信息,将目标检测问题转换为分类问题。然而,CNN 的多次下采样操作虽然提高了对目标种类的识别,但深层特征图空间分辨率的降低却导致其无法精确定位水下环境中小目标的位置。为了解决上述问题,本文构建了一种包含丰富语义信息的循环网络体系。

首先,采用跨阶段局部网络(CSPNet)^[9-11]的残差结构提升检测网络的深度,获取待检测图像的高频信息,其中 Res_Block 的具体结构如图 2 所示。Res_Block 由一次下采样卷积模块 Conv_BN_Mish 和多次残差结构堆叠构成,其中 Conv_BN_Mish 卷积模块由单层卷积、BatchNormalization 标准化、Mish 激活函数三部分组成。残差结构处进行了左右拆分:特征提取主干采取通常残差块堆叠的模式;另一部分作为额外残差边,绕过了主干部分的残差结构直接与主干部分的输出进行整合。然后,通过空洞卷积层在不牺牲特

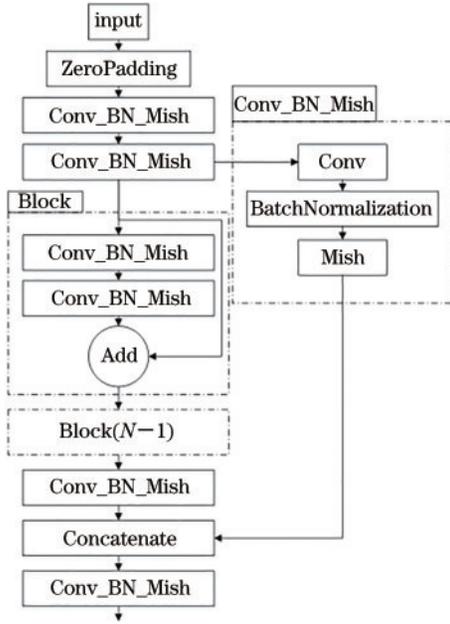


图 2 Res_Block 结构

Fig. 2 Res_Block structure

征图分辨率的情况下获得较大的感受野,从而获得更高层次的语义信息,其中 Dilated_Block 的具体结构如图 3 所示。通过空洞卷积,将特征图的感受野从 3×3 扩大至 15×15 。最后采用两种尺寸的上采样层,即解码层 Deconv Block1 和 Deconv Block2, Deconv_Block 的具体结构如图 4 所示。上采样后的结果与同一尺度的 Res_Block 输出进行特征图融合,使网络能够预测多尺度的目标。同时,为了获取含有丰富语义信息的特征图,通过低频信息与高频信息融合的方式,即上采样得到的尺寸为 $(38, 38, 256)$ 的特征图与编码层特征图通过跳远连接进行融合。考虑到水下图像中多尺度目标检测精度差距过大的问题,还采用了特征金字塔的方法读取目标语义信息及位置信息,并在不同尺度的特征图中生成目标的预测边界框。

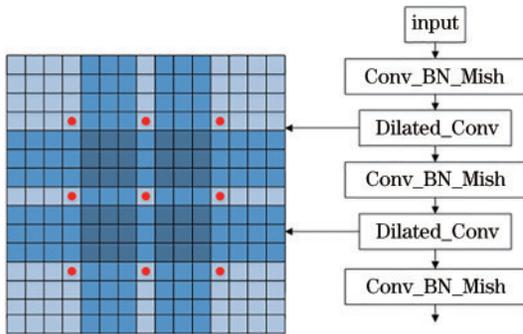


图 3 Dilated_Block 结构

Fig. 3 Dilated_Block structure

在检测网络构建后,基于水下图像中海洋生物的漏检、误检问题提出一种新的损失函数设计方法。以待检测目标与对应真实标签(GT)之间的交并比(IoU)为检测依据,过滤出待选目标框与干扰检测框

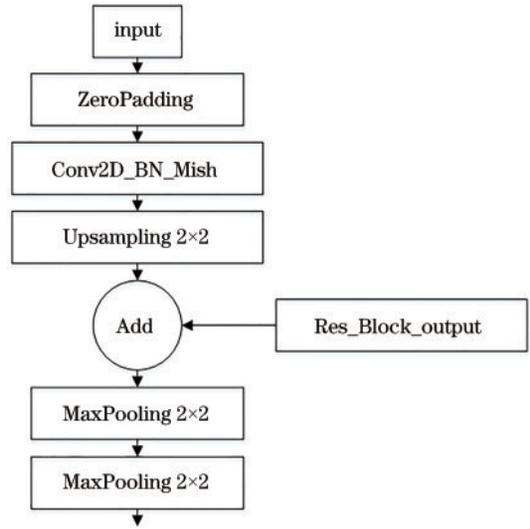


图 4 Deconv_Block 结构

Fig. 4 Deconv_Block structure

用于计算损失函数。最大 IoU 值的预选框为待选目标框,令 C_i 是第 i 个海洋生物待选目标框,获取方法为

$$C_i = \max \text{IoU}(P_i, L_{GT_i}), \quad (1)$$

式中: P_i 和 L_{GT_i} 分别是第 i 个海洋生物的预测框和真实标签。 P_i 和 L_{GT_i} 的 IoU 计算方法为

$$\text{IoU}(P_i, L_{GT_i}) = \frac{\text{area}(P_i \cap L_{GT_i})}{\text{area}(L_{GT_i})}. \quad (2)$$

L_{true} 是评判检测目标是否发生重合的损失函数,公式为

$$L_{\text{true}} = \frac{L_1(P, L_{GT})}{I} = \frac{\sum_{i \in I} |P_i - L_{GT_i}|}{I}, \quad (3)$$

式中: L_1 表示预测框与标签之间的距离; I 是水下图像中所有海洋生物的集合; P 是所有目标预测框的集合。

为了降低漏检率,取获得次大 IoU 的框体作为干扰检测框,使用损失函数 L_{punish} 用于衡量干扰检测框与 GT 间的逼近程度,根据逼近的程度进行惩罚。在训练过程中,干扰检测框尺寸大幅度变换会导致交集突变,从而对损失函数造成很大的影响。针对此问题,先采用 Smooth 函数对 IoU 指标计算的结果进行平滑处理,再计算 L_{punish} ,表达式为

$$L_{\text{punish}} = \frac{\sum_{i \in I} \text{Smooth}[\text{IoU}(D, L_{GT})]}{I}, \quad (4)$$

$$\text{Smooth}[\text{IoU}(D, L_{GT})] =$$

$$\frac{1}{N_{\text{cls}}} \sum_i L_{\text{cls}}[\text{IoU}(D_i, L_{GT_i}), p_i] +$$

$$\lambda \frac{1}{N_{\text{reg}}} \sum_i p_i L_{\text{reg}}(D_i, L_{GT_i}), \quad (5)$$

式中: p_i 为目标存在标记,有物体存在时 p_i 为 1,否则为 0; D_i 为第 i 个海洋生物干扰检测框。置信度损失 L_{reg} 与分类损失 L_{cls} 分别为

$$L_{cls} = -\log \left\{ \text{IoU}(D_i, L_{GT_i}) + [1 - \text{IoU}(D_i, L_{GT_i})] (1 - p_i) \right\}, \quad (6)$$

$$L_{reg} = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & |x| \geq 1 \end{cases} \quad (7)$$

当 Smooth 函数值越大, 干扰检测框与 GT 的逼近程度越高, 干扰检测框越易影响检测精度。因此, L_{punish} 越大, 漏检现象越严重。

进一步地, 针对两个目标候选框紧贴导致回归单一框体的情况, 设计了损失函数 L_{false} 用于衡量干扰检测框 D 与待选目标框 C 间尺寸差异程度, 公式为

$$L_{false} = \frac{\sum_{i \in I} \text{Central}(D_i, C_i) + \text{Constru}(D_i, C_i)}{I}, \quad (8)$$

$$\text{Constru}(D_i, C_i) = \frac{\sum_i (x_i - x'_i) \sum_i (y_i - y'_i)}{(\omega_i + \omega'_i)(h_i + h'_i)}, \quad (9)$$

式中: x_i 和 y_i 分别是第 i 个目标的干扰检测框 D_i 的左上顶点的横坐标和纵坐标; ω_i 和 h_i 分别是第 i 个目标的干扰检测框 D_i 的宽度和高度; x'_i 和 y'_i 、 ω'_i 和 h'_i 分别是第 i 个目标的预测框体 C_i 左上顶点的横坐标和纵坐标、框体的宽度和高度; $\text{Central}(\cdot)$ 是框体中心点位置偏移函数; $\text{Constru}(\cdot)$ 是框体形状偏移函数。Central 与 Constru 函数值越小, 说明干扰检测框 D_i 与预测框体 C_i 的坐标偏移量与尺寸偏移量越小, 导致干扰检测框 D_i 与待选目标框 C_i 重合, 从而影响检测结果。

综上所述, 针对鱼群遮挡导致的漏检、误检等现象, 提出损失函数 L_{water} :

$$L_{water} = L_{true} + \alpha L_{punish} + \epsilon L_{false}, \quad (10)$$

式中: α 是惩罚因子; ϵ 是偏置因子。通过训练, α 和 ϵ 均设置为 0.3。 L_{false} 为负标量, L_{punish} 为正标量。

2.2 水下样本迭代融合方法

针对水下训练集中存在目标遮挡、数据集不平衡的问题, 提出一种水下样本迭代融合的方法。首先, 通过初次水下样本迭代融合获得重叠、遮挡和模糊情况

下的模拟图像扩充训练集的数据量, 以平衡数据集。然后, 更新训练中融合权重因子, 自适应调节各类样本在融合过程的权重, 减小目标遮挡的影响。

具体来说, 水下样本迭代融合需从初始训练集中抽取 16 张图像作为素材生成模拟图像用于丰富训练集。其中对 8 张图像采用旋转、缩放、色域变换方式, 对另外 8 张图像根据每次网络检测效果的优劣给予不同的权重进行融合, 其中, 检测效果的优劣取决于损失值的大小。这种基于损失值进行融合的方式生成的图像能反映出检测效果较差目标的特性。对处理后的 16 张图像进行乱序组合, 生成模拟图像。

另外 8 张图像的融合过程分为水下背景融合与目标框体融合两个阶段。水下背景融合根据目标检测框所包含的坐标信息提取目标物体, 并分离出背景图像。将背景图像平均融合得到新的水下背景图, 然后通过色域变换得到新的图像, 用于模拟不同光学环境下的水下场景。

目标框体融合的过程如图 5 所示, T_1 和 T_2 为两张含有不同海洋生物的图像, 根据各图像内目标框体具有的位置信息提取目标物图像, 对提取出的目标物图像在原图上进行缩放、翻转、平移操作, 随后生成新的训练图像 T' , 模拟遮挡、模糊现象。具体生成方法为

$$T' = \lambda T_1 + (1 - \lambda) T_2, \quad (11)$$

式中: λ 为融合权重因子, 用于平衡各个目标物在新图像中的显示占比。在初次训练时 $\lambda = 0.5$, 后续训练中表达式为

$$\lambda = 1 + \cos(\pi \times L_{IoU \min}). \quad (12)$$

$L_{IoU \min}$ 的表达式为

$$L_{IoU \min} = \begin{cases} e^{\min(L_{IoU_{x_1}}, L_{IoU_{x_2}})}, & L_{IoU_{x_1}}, L_{IoU_{x_2}} > 0.5 \\ \min(L_{IoU_{x_1}}, L_{IoU_{x_2}}), & L_{IoU_{x_1}}, L_{IoU_{x_2}} \leq 0.5 \end{cases}, \quad (13)$$

式中: $L_{IoU_{x_1}}$ 和 $L_{IoU_{x_2}}$ 分别代表 T_1 和 T_2 对应的损失函数值。通过分段函数, 取使识别率较低的图像在样本迭代融合中占比更大的权重, 以便在后续训练中得到识别率的提升。实验总结出 $L_{IoU \min}$ 小于 0.9 较佳。



图 5 图像融合过程

Fig. 5 Image fusion process

结合检测网络和水下样本迭代融合方法两部分内容, 将检测网络训练结果与水下样本迭代融合方法结合进行对抗训练, 提高网络对复杂目标的检测能力。过多使用识别率欠佳的图像会导致网络模型对训练集

拟合程度过高, 降低泛化能力。因此, 将单幅图像参与样本权重更新的次数控制在 5 次之内, 防止过拟合, 使网络充分学习检测效果欠佳的目标特征并具有一定的鲁棒性。另外, 在基于样本迭代融合的海洋生物检测

网络模型中,提供了先验损失值,在一定程度上减少了网络计算量,为网络模型的实时性提供了保障。

3 实验结果及分析

3.1 实验设置

本实验在 Windows10(64 位)环境,配置 3700X CPU, RTX2070Super GPU, 16 G 内存的机器下运行,基于 keras 复现各个算法的深度学习框架。为了评估所提方法的性能,采用 mAP 评估检测精度。由于各个算法提供的配置环境、超参数存在差异,在本实验平台上较难得到较好的结果,对比检测精度的任务存在挑战性。为确保列举的各个对比方法的检测精度范围统一,对各个方法的超参数、运行环境进行调整。为了解决上述问题,并保证对比实验的公平性,对超参数的调整具体如下。

在训练阶段,各个算法均采用退火余弦的学习率衰减方法进行横向对比,学习率最高值为 10^{-3} ,初始学习率为 10^{-5} ,步长为默认值。在验证阶段,对各个算法的置信度阈值和 IoU 阈值进行统计调整,具体值为 0.05 和 0.45。为了符合水下样本迭代融合所需的图片数,缩减遍历数据集的迭代次数,增大并行处理量,使损失函数有效收敛,在训练过程中每次从数据集中选取 16 张样本输入(BatchSize 设置为 16),各个算法一次训练选取的样本数均为 16 张。由于 BatchSize 的调整,不同算法的单个 epoch 的迭代次数发生改变,为获得较高的识别精度,实验中将各个算法的 epoch 统一设置为 160。所提方法采用批标准化扩充数据量后相比于其余几种方法能获得更好的优化效果。此外,采用水下样本迭代融合扩充数据集,网络中还应用了随机改变图像亮度和对比度、随机翻转图像等方法扩充训练数据。在网络后几层中采用随机失活的方法,去掉部分神经元之间的连接,使网络在训练过程中更具有鲁棒性,避免过拟合的现象。

3.2 数据集建立

由于目前水下较常用的检测数据集 Fish4 Knowledge 缺少了目标重合时的海洋生物图像,本文在 URPC2018 和台湾鱼类库的基础上制作了数据集并进行了测试。URPC2018 水下机器人目标抓取大赛官方数据集(分辨率为 720×405)包括海胆、海星和贝壳三种海洋生物,如表 1 所示。

台湾鱼类库中共有 463 种鱼类的视频。视频分辨率低,各种鱼类间纹理信息、颜色信息区别不明显。如果直接选用这些数据会导致检测网络对单一目标进行重复检测时产生多个检测结果,进而影响检测网络的分类能力。此外,识别精度随着训练样本种类增多而呈下降趋势。因此如表 2 所示,从台湾鱼类库中选取了 18 种颜色形态差异较大的鱼类作为检测对象。

首先,对台湾鱼类库中的视频作分帧处理,得到分辨率范围从 352×240 至 1440×810 的图像;然后,通过

表 1 URPC2018 数据集中样本数量统计

Table 1 Number of samples on URPC2018 dataset

Sample	Total	Train	Test
Echinus	453	2672	622
Scallop	343	783	232
Starfish	463	9271	2337

表 2 台湾鱼类库数据集中样本数量统计

Table 2 Number of samples on Taiwan fish dataset

Sample	Total	Train	Test
Abudefduf bengalensis	453	317	136
Abudefduf lorenzi	463	324	139
Abudefduf septemfasciatus	343	240	103
Abudefduf sexfasciatus	208	146	62
Abudefduf vaigiensis	469	328	141
Acanthuridae	284	199	85
Acanthurus blochii	250	175	75
Acanthurus dussumieri	358	251	107
Acanthurus japonicus	349	244	105
Acanthurus lineatus	207	145	62
Acanthurus nigricauda	470	329	141
Acanthurus nigrofuscus	271	190	81
Acanthurus olivaceus	249	174	75
Acanthurus olivaceus	366	256	110
Acanthurus pyroferus	469	328	141
Acanthurus thompsoni	285	200	86
Acanthurus xanthopterus	493	345	148
Aeoliscus strigatus	320	224	96

图像超分辨率重建技术^[23]将其分辨率统一为 608×608 ;最终,得到 37003 张带自制标签的水下图像。两个数据集按照 7:3 划分为训练集和测试集,在训练过程中取 10% 的训练样本作为验证集。从表 1 和表 2 可看出,各个类别的训练样本与测试样本数据量不平衡。所提方法通过水下样本迭代融合扩充数据量得到更多的分布信息,使各类物体数据量达到均衡。

3.3 消融实验

采用消融实验验证检测算法各模块的有效性,实验结果如表 3 所示。迭代融合和损失函数调整能提高对海胆的检测精度,AP 分别提高 5.22 个百分点和 9.27 个百分点。由于海胆容易与其他物体相互遮挡,所提方法中的水下样本迭代融合以及检测网络中损失函数的优化能解决海洋生物检测中的目标遮挡问题。水下样本迭代融合方法和检测网络中损失函数的优化均能提升网络模型的总体检测速度。检测网络的空洞卷积层使得总体检测精度 mAP 提高了 1.16 个百分点,但总体检测速度有所下降。这是因为单独运用空洞卷积层导致参数量提升,影响了网络对贝壳的检测效果,增大了计算量。检测网络的解码层加深了网络

表 3 消融实验
Table 3 Ablation test

Iterative fusion	L_{water}	Dilated_Block	Deconv_Block	AP / %			mAP / %	Speed / (frame·s ⁻¹)
				Echinus	Scallop	Starfish		
				81.08	89.65	89.35	86.69	36.33
✓				86.30	90.24	89.82	88.79	49.31
	✓			90.35	90.84	89.65	90.28	39.97
		✓		84.71	88.63	90.21	87.85	35.89
			✓	83.66	89.01	89.38	87.35	32.38
		✓	✓	85.42	88.34	90.17	88.31	30.15
✓	✓	✓	✓	92.24	90.37	91.46	91.36	49.30

的层数,使得总体检测精度 mAP 提高了 0.66 个百分点,这是因为解码层的特征金字塔结构强化了网络对各个尺寸目标的特征提取能力。当结合空洞卷积层和解码层时,网络的总体检测精度提高了 1.62 个百分点。卷积操作、参数量的增加使得总体检测速度大幅下降。不同类别检测精度不同,贝壳、海星识别率略微下降,海胆的识别率呈直线上升。这是因为贝壳与海星特征简单,通过浅层特征层提取就拥有较高的检测精度,而海胆自身纹理复杂度高,边缘信息模糊导致在网络深度不足时识别率欠佳,需要深层网络提取特征来提高检测精度。当结合所有模块时,对三类目标的检测精度分别提升 11.16 个百分点、0.72 个百分点和 2.11 个百分点,总体检测精度提升 4.67 个百分点,由此显示了迭代融合、损失函数 L_{water} 、空洞卷积层、解码层对所提网络模型的检测精度的提高具有重要意义。

3.4 对比实验

在 URPC2018 数据集及处理后的台湾鱼类数据集上测试所提检测网络模型的识别能力,并与现有的 6 个通用目标检测方法 Faster R-CNN、SSD、YOLOv3、YOLOv4、RFB 和 Efficientdet 进行比较,并基于 Faster R-CNN 对 RoIMix 进行复现。首先验证检测网络的损失函数 L_{water} 是否有助于提升网络的收敛能力,在 URPC2018 数据集上对各个目标检测方法分别进行相同 epoch 的训练,各个算法的损失函数收敛曲线如图 6 所示。

所提检测网络的损失函数 L_{water} 的收敛速度相比于 YOLOv4 大幅提升,这是因为水下样本迭代融合模块提供了先验损失。而相比于 YOLOv4 方法的初始损失值,所提检测网络的初始损失值大约下降了 98.36,

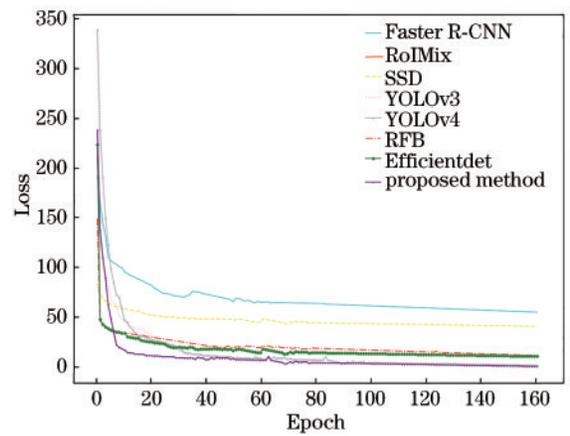


图 6 各目标检测方法的损失函数曲线

Fig. 6 Loss function curves of each target detection method

初始损失值的下降同样利于收敛速度的提升。相比于 Faster R-CNN、SSD、YOLOv3、RoIMix、RFB 和 Efficientdet 目标检测方法,所提检测网络的损失函数收敛后振荡幅度较小,损失波动在 0.02~0.19 之间,损失值更为平滑。

表 4 为 URPC2018 数据集上各个目标检测方法得到的检测精度,黑体和斜体分别表示最优和次优的结果。所提方法的检测精度相比于 Faster R-CNN、SSD、YOLOv3、RFB 和 Efficientdet 分别提高了 25.12 个百分点、10.57 个百分点、9.70 个百分点、18.52 个百分点和 5.17 个百分点,而相比于 YOLOv4 方法提升了 0.30 个百分点,相比 RoIMix 略低。因为 URPC2018 数据集中存在着大量遮挡、重叠目标,其他检测方法产生了不同程度的漏检、误检的现象。RoIMix 通过混合候选区域来模拟目标的重叠、遮挡和模糊,提高检测精度。所提方法针对漏检、误检现象优

表 4 不同方法在 URPC 数据集上的检测精度
Table 4 Detection accuracy of different methods on URPC dataset

Sample	Faster R-CNN	SSD	YOLOv3	YOLOv4	RoIMix	RFB	Efficientdet	Proposed method
Echinus	47.32	76.23	85.39	91.77	93.73	53.71	87.21	<i>92.24</i>
Scallop	75.23	79.77	85.82	<i>92.82</i>	93.67	80.39	81.01	91.46
Starfish	76.16	86.38	73.76	88.58	90.04	84.42	<i>90.36</i>	90.37
mAP	66.24	80.79	81.66	91.06	92.48	72.84	86.19	<i>91.36</i>

化损失函数,并运用水下样本迭代融合方法生成遮挡、重叠图像,进行训练,提高了对此类现象的适应能力。

表 5 为在台湾鱼类库上所提方法与其目标识别方法的检测精度比较结果,黑体和斜体分别表示最优和次优的结果。可以看出,所提方法的检测精度大部分是最优和次优的,其中对部分鱼类的检测精度可以达 100%。这是因为这些鱼类拥有独特颜色及轮廓纹理特征,或此类目标与背景颜色区分明显,有助于网络进行特征的提取、训练;其次,这些鱼类目标在测试集中大多只存在一个鱼类对象,便于执行检测任务;此

外,这些鱼类目标拥有较多的训练样本量,便于网络进行反复的训练。所提方法在台湾鱼类库上的检测结果如图 7 所示。

从表 5 可以看出:所提方法对部分鱼类的检测精度相对较低,例如球吻鼻鱼、褐斑刺尾鲷等,这是因为在待检测对象数量增加的情况下,发生的重叠、遮挡现象导致一定程度的漏检,使检测精度下降,这种情况在没有考虑重叠、遮挡问题的 Faster R-CNN 中尤其明显;RFB 采用三种不同尺寸的空洞卷积(1×1、3×3、5×5)提高感受野,对此类目标的检测精度较强,与之

表 5 不同方法在台湾鱼类库上的检测精度

Table 5 Detection accuracy of different methods on Taiwan fish dataset

unit: %

Sample	Faster R-CNN	SSD	YOLOv3	YOLOv4	RoIMix	RFB	Efficientdet	Proposed method
Abudefduf bengalensis	57.99	73.71	52.90	87.38	99.13	98.20	72.53	99.36
Abudefduf lorenzi	57.99	99.29	94.74	99.95	99.98	99.60	62.02	100.00
Abudefduf septemfasciatus	2.71	88.48	86.62	90.75	96.43	99.19	79.52	100.00
Abudefduf sexfasciatus	55.42	77.05	88.23	93.18	92.72	99.25	83.36	84.94
Abudefduf vaigiensis	19.61	56.37	72.01	67.02	80.98	90.42	77.43	85.54
Acanthuridae	1.42	43.24	60.00	62.55	88.13	76.84	44.91	46.97
Acanthurus blochii	6.29	88.12	58.63	82.62	99.95	98.42	89.46	100.00
Acanthurus dussumieri	0.17	41.92	38.81	89.03	36.49	82.72	64.31	85.97
Acanthurus japonicus	4.30	62.00	73.54	90.65	95.46	94.97	75.96	96.15
Acanthurus lineatus	0.68	82.03	67.44	45.57	86.65	95.55	62.82	90.97
Acanthurus nigricauda	4.80	93.78	71.37	99.09	98.67	95.77	94.41	100.00
Acanthurus nigrofuscus	8.83	60.78	63.18	82.07	85.69	86.38	41.43	69.47
Acanthurus olivaceus	13.49	74.77	73.31	86.43	84.98	95.23	58.84	81.56
Acanthurus olivaceus	44.30	85.18	81.39	99.39	99.97	94.65	80.20	100.00
Acanthurus pyroferus	58.40	82.82	85.16	94.89	94.61	97.34	89.49	96.56
Acanthurus thompsoni	35.01	81.22	83.61	89.19	100.00	97.00	67.69	100.00
Acanthurus xanthopterus	55.29	45.50	74.88	88.12	69.78	66.18	72.94	87.42
Aeoliscus strigatus	41.19	87.42	88.54	91.98	91.13	99.46	81.37	100.00
mAP	25.99	73.54	73.02	85.55	88.93	92.62	72.15	90.27

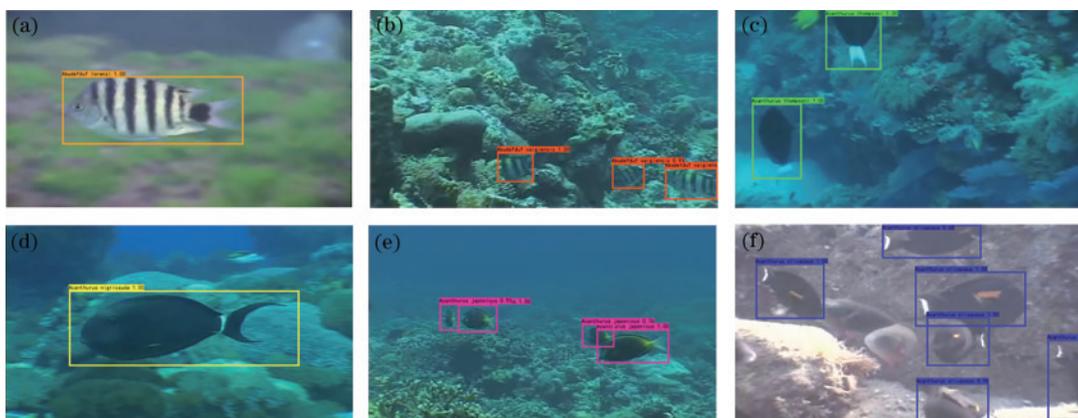


图 7 所提方法在台湾鱼类库上的检测结果。(a)劳伦氏豆娘鱼;(b)布氏刺尾鲷;(c)孟加拉豆娘鱼;(d)黑尾刺尾鲷;(e)日本刺尾鲷;(f)球吻鼻鱼

Fig. 7 Detection results of the proposed method on Taiwan fish dataset. (a) Abudefduf lorenzi; (b) Acanthurus blochii; (c) Abudefduf bengalensis; (d) Acanthurus nigricauda; (e) Acanthurus japonicus; (f) Acanthuridae

采用类似结构的 YOLOv4 也获得较好的检测精度。所提方法为了保证检测速度只采用一组 3×3 的空卷积扩大感受野,因此在面对此类目标时,检测精度不够理想,如图 7(f)所示。

所提方法在台湾鱼类库上得到的平均精度优于 Faster R-CNN、SSD、YOLOv3、YOLOv4、RFB 和 Efficientdet 方法,稍微低于 RFB。从检测精度上看,所提方法比 Faster R-CNN、SSD、YOLOv3、YOLOv4、RoIMix 和 Efficientdet 分别提高了 64.28 个百分点、16.73 个百分

点、17.25 个百分点、4.72 个百分点、1.34 个百分点和 18.12 个百分点。由此说明所提方法对海洋生物的检测精度比其他目标检测方法高。表 6 为不同方法的检测速度比较,所提方法的检测速度相比 YOLOv4 方法提高了 12.97 frame/s,和其他目标检测方法相比获得了显著性的提升。这是因为训练集中样本数量较少,使得其他检测方法对特征学习不完全。而所提水下样本迭代融合方法通过平衡训练样本中各个目标类别的样本数量,使网络模型对小样本的识别率有显著性的提升。

表 6 不同方法的检测速度比较

Table 6 Comparison of detection speed of different methods

Parameter	Faster R-CNN	SSD	YOLOv3	YOLOv4	RoIMix	RFB	Efficientdet	Proposed method
Speed / (frame · s ⁻¹)	23.64	40.80	30.23	36.33	23.70	10.38	18.54	49.30

因此,所提方法在视频库上不仅获得很高的速度,也维持了高水准的 mAP。说明检测网络结构的调整以及水下样本迭代融合在保证检测网络检测效率的同时,并未对网络运行速度造成较大影响。

3.5 结果分析

图 8 为所提方法和 YOLOv4、RFB 两种目标检测方法的检测结果对比。测试图片包含密集鱼群和小尺度鱼类。除所提方法外,其余目标检测方法都存在漏检的情况。由于海洋生物相互遮挡以及水下图像蓝色偏色程度大,检测效果易受水下光照折射影响,导致边

缘模糊,对比度低。YOLOv4 与 RFB 检测方法均为自然场景泛用性检测方法,所以对光照不均、含有成群鱼类、小尺度目标的水下图像的应对能力较差。本文针对海洋生物的特征,设计了用于检测聚集海洋生物的损失函数 L_{water} ,使得所提检测网络模型更适用于对海洋生物的检测,在一定程度上提升了网络的检测精度。对比 YOLOv4 与 RFB 的检测结果可知,在鱼群目标相对密集的情况下,RFB 方法的防漏检能力较强,但仍然不能正确检测出重叠目标。显然,从正确检测的数量上看,所提检测方法均优于其余方法。

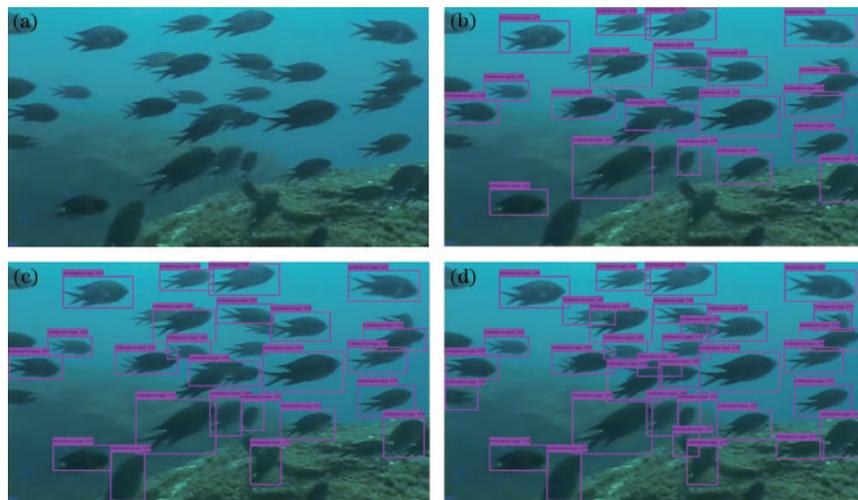


图 8 不同方法的检测结果对比。(a)原始图像;(b)YOLOv4;(c)RFB;(d)所提方法

Fig. 8 Detection result comparison of different methods. (a) Original image; (b) YOLOv4; (c) RFB; (d) proposed method

4 结 论

主要提出了一种神经网络框架用于对小型水下物体的检测,旨在保证一定的实时检测性的条件下提高方法对水下场景目标的检测精度。提出水下样本迭代融合方法对训练图片进行数据扩充,并且模拟了海洋生物遮挡、重叠的训练样本以供网络学习。此外,提出了一种权值更新方法,使训练结果与预处理方法形成

一个反馈链,以解决训练不充分的问题。所提方法在具有挑战性的数据集上实现了良好的性能,而且在复杂度高的数据集上也取得了较好的效果,相比于现有的方法,速度更快,能满足对水下视频实时检测的要求。

参 考 文 献

- [1] Chen X Y, Lu Y, Wu Z X, et al. Reveal of domain

- effect: how visual restoration contributes to object detection in aquatic scenes[EB/OL]. (2020-03-04)[2021-05-06]. <https://arxiv.org/abs/2003.01913>.
- [2] 林森, 赵颖. 水下光学图像中目标探测关键技术研究综述[J]. 激光与光电子学进展, 2020, 57(6): 060002.
Lin S, Zhao Y. Review on key technologies of target exploration in underwater optical images[J]. Laser & Optoelectronics Progress, 2020, 57(6): 060002.
- [3] Boom B J, Huang P X, He J Y, et al. Supporting ground-truth annotation of image datasets using clustering [C]//Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), November 11-15, 2012, Tsukuba, Japan. New York: IEEE Press, 2012: 1542-1545.
- [4] 陈华杰, 白浩然. 基于子空间投影的复杂水下环境运动小目标检测前跟踪方法[J]. 电子与信息学报, 2021, 43(3): 826-833.
Chen H J, Bai H R. Subspace projection based track-before-detect scheme for small moving target in complex underwater environment[J]. Journal of Electronics & Information Technology, 2021, 43(3): 826-833.
- [5] 赵永强, 饶元, 董世鹏, 等. 深度学习目标检测方法综述[J]. 中国图象图形学报, 2020, 25(4): 629-654.
Zhao Y Q, Rao Y, Dong S P, et al. Survey on deep learning object detection[J]. Journal of Image and Graphics, 2020, 25(4): 629-654.
- [6] 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, 2016, 39(6): 1137-1149.
- [7] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.
- [8] Liu S T, Huang D, Wang Y H. Receptive field block net for accurate and fast object detection[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11215: 404-419.
- [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), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.
- [10] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2021-05-02]. <https://arxiv.org/abs/1804.02767>.
- [11] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2021-05-06]. <https://arxiv.org/abs/2004.10934>.
- [12] Tan M X, Pang R M, Le Q V. EfficientDet: scalable and efficient object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 10781-10790.
- [13] Zheng Y L, Gong Q Y, Zhang S F. Time-frequency feature-based underwater target detection with deep neural network in shallow sea[J]. Journal of Physics: Conference Series, 2021, 1756(1): 012006.
- [14] 张有波, 郭威, 周悦, 等. 基于多粒度剪枝的水下遗迹实时目标检测[J]. 激光与光电子学进展, 2021, 58(14): 1410019.
Zhang Y B, Guo W, Zhou Y, et al. Real-time target detection of underwater relics based on multigranularity pruning[J]. Laser & Optoelectronics Progress, 2021, 58(14): 1410019.
- [15] Guan J G, Zhao Y, Zheng Y Q, et al. Optical polarization imaging for underwater target detection with non-scatter background[J]. Journal of Measurement Science and Instrumentation, 2020(4): 335-342.
- [16] Zhang D, Kopanas G, Desai C, et al. Unsupervised underwater fish detection fusing flow and objectiveness [C]//2016 IEEE Winter Applications of Computer Vision Workshops, March 10-10, 2016, Lake Placid, NY, USA. New York: IEEE Press, 2016.
- [17] Mandal R, Connolly R M, Schlacher T A, et al. Assessing fish abundance from underwater video using deep neural networks[C]//2018 International Joint Conference on Neural Networks (IJCNN), July 8-13, 2018, Rio de Janeiro, Brazil. New York: IEEE Press, 2018.
- [18] Rathi D, Jain S, Indu S. Underwater fish species classification using convolutional neural network and deep learning[C]//2017 Ninth International Conference on Advances in Pattern Recognition (ICAPR), December 27-30, 2017, Bangalore, India. New York: IEEE Press, 2017.
- [19] Chen L, Zhou F X, Wang S K, et al. SWIPENET: object detection in noisy underwater images[EB/OL]. (2020-10-19) [2021-04-05]. <https://arxiv.org/abs/2010.10006>.
- [20] Lin W H, Zhong J X, Liu S, et al. ROIMIX: proposal-fusion among multiple images for underwater object detection[C]//2020 IEEE International Conference on Acoustics, Speech and Signal Processing, May 4-8, 2020, Barcelona, Spain. New York: IEEE Press, 2020: 2588-2592.
- [21] Han F L, Yao J Z, Zhu H T, et al. Underwater image processing and object detection based on deep CNN method[J]. Journal of Sensors, 2020, 2020: 6707328.
- [22] 周维, 唐华龙, 李观德, 等. 基于遮挡检测机制的DDAT目标跟踪算法[J]. 激光与光电子学进展, 2020, 57(24): 241501.
Zhou W, Tang H L, Li G D, et al. DDAT target tracking algorithm based on occlusion detection mechanism[J]. Laser & Optoelectronics Progress, 2020, 57(24): 241501.
- [23] 刘可文, 马圆, 熊红霞, 等. 基于残差通道注意力网络的医学图像超分辨率重建方法[J]. 激光与光电子学进展, 2020, 57(2): 021014.
Liu K W, Ma Y, Xiong H X, et al. Medical-image super-resolution reconstruction method based on residual channel attention network[J]. Laser & Optoelectronics Progress, 2020, 57(2): 021014.