DOI: 10.13382/j.jemi.B2104168

结合数据增强和改进 YOLOv4 的水下目标检测算法*

史朋飞 韩 松 倪建军 杨 鑫
(河海大学物联网工程学院 常州 213022)

摘 要:针对水下低质量成像、水下目标形态大小各异、以及水下目标重叠或遮挡导致水下目标检测精度低的问题,提出一种结合数据增强和改进 YOLOv4(you look only once)的水下目标检测算法,在 YOLOv4 的主干特征提取网络 CSPDarknet53 中添加卷积块注意力机制(convolutional block attention module,CBAM),以提高网络模型特征提取能力;在路径聚合网络(path aggregation network,PANet)中添加同层跳接和跨层跳接结构,以增强网络模型多尺度特征融合能力;通过数据增强方法 PredMix(prediction-mix)模拟水下生物重叠、遮挡等显示不完全的情形,以增强网络模型鲁棒性。实验结果表明,结合数据增强和改进 YOLOv4 的水下目标检测算法在 URPC2018(underwater robot picking control 2018)数据集上的检测精度提升到了 78.39%,比 YOLOv4 高出 7.03%,充分证明所提算法的有效性。

关键词:深度学习;目标检测;YOLOv4;水下检测
中图分类号:TP391.4 **文献标识码**:A **国家标准学科分类代码**:520.2060

Underwater object detection algorithm combining data enhancement and improved YOLOv4

Shi Pengfei Han Song Ni Jianjun Yang Xin
(College of Internet of Things Engineering, Hohai University, Changzhou 213022, China)

Abstract: Aiming at the problem of low underwater object detection accuracy caused by low-quality underwater imaging, different shapes or sizes of underwater objects, and overlapping or occlusion of underwater objects, an underwater object detection algorithm combining data enhancement and improved YOLOv4 is proposed. By adding CBAM (convolutional block attention module) to the backbone of YOLOv4—CSPDarknet53, the feature extraction ability of network model is improved. In order to enhance the multi-scale feature fusion ability, the same-layer skip connections and cross-layer skip connections are added to PANet (path aggregation network). To enhance the robustness of the network model, the data enhancement method PredMix (prediction mix) is used to simulate the incomplete display of underwater organisms such as overlap or occlusion. The experimental results show that the detection accuracy of the underwater object detection algorithm combining data enhancement and improved YOLOv4 on URPC2018 dataset is improved to 78.39%, 7.03% higher than YOLOv4, which fully proves the effectiveness of the proposed algorithm.

Keywords: deep learning; object detection; YOLOv4; underwater detection

0 引 言

海洋生物是海洋生态环境的重要组成部分,一直以来都备受关注。海洋生态保护组织通过人工潜水、水下机器人拍摄等方式对海洋生物的分布情况、生活习性进行研究。但是,水下低质量成像严重影响研究人员对海洋生物的研究。因此,目前亟待一种代替肉眼来检测海洋生物的方法。

传统的目标检测方法存在识别效果差、识别速度慢等缺点,难以进行有效的水下目标检测。近年来,深度学习的快速发展给目标检测领域带来巨大突破。目标检测是一种与计算机视觉和图像处理相关的计算机技术,处理的是数字图像或者视频中特定类别的语义对象。由于近年的技术突破,目标检测在学术界得到充分关注,在现实中也得到广泛应用,例如安全监控[1]、自动驾驶[2]、无人机场景分析[3]等。大多数的目标检测都是以深度学习网络为主干,从输入图像中提取特征、分类和定位。目标

收稿日期:2021-04-13 Received Date:2021-04-13
*基金项目:国家自然科学基金(61801169,61873086)、中央高校基本科研业务费(B220202020)项目资助

检测的研究领域包括边缘检测[4-5]、多目标检测[6]、显著性目标检测[7]等。

许多场景下,基于深度学习的目标检测算法都取得了不错的成效。但是,文献[8-10]中指出:水下成像质量低下,水底环境复杂,海洋生物大小不一、形态各异等原因都会干扰水下场景的目标检测效果。并且,海洋生物互相重叠、遮挡(在本文中,"重叠"表示同一类对象之间的覆盖,"遮挡"则表示不同类对象的同种情况)也会干扰检测效果。

目前,基于深度学习的目标检测框架主要有两阶段(two-stage)和单阶段(one-stage)两种。两阶段的目标检测框架最典型的是R-CNN系列:2014年Ross B. Girshick提出了R-CNN[11],先通过选择性搜索生成候选区域,然后对候选区域使用CNN提取特征,从而取代了传统的滑窗法,提高了目标检测的精确度。但是R-CNN有大量重复计算,严重影响检测性能;Fast-RCNN[12]针对R-CNN的计算冗余,选择先将输入图像通过CNN进行特征提取,再通过选择性搜索提取出候选区域,这样只需经过一次CNN就可以得到全部的候选区域,减少了重复计算,但是检测速度依然不理想;Faster-RCNN[13]在Fast-RCNN的基础上,使用区域建议网络(RPN)替代了选择性搜索算法来筛选出候选区域,检测速度进一步提升;单阶段的目标检测框架最典型的是YOLO系列:YOLOv1[14]摒弃了两阶段目标检测算法的生成候选区域步骤,使用先验的锚定框(anchor box)直接将目标区域预测和目标类别预测整合为一步,极大地提升了检测速度,但是每个单元格只能预测一个对象,检测精度也不高;YOLOv2[15]在YOLOv1的基础上进行改进,通过给卷积层添加批标准化来降低模型过拟合,有效提高模型收敛能力,YOLO9000在YOLOv2的基础上利用一种使用分层视图进行对象分类的方法,可以同时检测9 000多种类别;YOLOv3[16]在YOLOv2的基础上,取消了所有的池化层,使用增强卷积核的步长来达到池化层的效果,从而大幅提升检测速度,并且可以同时输出多尺度特征图,强化了小目标的检测能力;YOLOv4[17]在YOLOv3的基础上,筛选了一些从YOLOv3发布至今,被用在各种检测器上能够提升检测精度的技巧(trick),并且将YOLOv3的主干网络Darknet-53替换成CSPDarkNet53,更进一步提升检测速度和精度。

数据增强是扩充数据样本规模的一种方法,可以增强模型的泛化能力。常规的数据增强方法有空间几何变换、像素颜色变换、高斯模糊、随机擦除等。AutoAugment[18]通过搜索算法找到适合特定数据集的图像增强方案;RandAugment[19]直接在数据集上搜索针对该数据集的最优策略;CutOut[20]对输入图像进行遮挡,模拟检测对象被部分遮挡的场景;HideAndSeek[21]将输入图像分成若干区域,对每个区域以一定概率生成掩码;MixUp[22]基于邻域风险最小化原则,使用线性插值将两张输入图像进行融合,得到新的样本数据;RoIMix[23]针对水下图像对比度低、互相遮挡等问题,对RPN产生的候选框使用线性插值。

本文的贡献主要是在YOLOv4的基础上做出以下3点改进,使之更适合水下目标检测。

1)针对水下低质量成像、水下复杂环境导致检测精度低的问题,在YOLOv4的CSPDarknet53主干特征提取网络中添加了CBAM特征注意力模块,使得网络模型在通道维度上学习关注内容和在空间维度上学习关注位置,增强网络模型特征提取能力。

2)针对水下目标大小不一、形态各异导致网络模型检测精度低下的问题,在YOLOv4中的PANet模块中添加同层跳接和跨层跳接结构,从而更充分地结合语义信息丰富的深层特征和位置信息、细节信息丰富的浅层特征,以提升多尺度特征融合能力。

3)针对海洋生物互相重叠、遮挡导致检测精度低下的问题,提出一种应用于YOLOv4的新型数据增强方法PredMix,通过线性加权的方式混合目标图像以模拟水下目标的重叠和遮挡。

1 结合数据增强和改进YOLOv4的水下目标检测算法

1.1 CBAM-CSPDarknet53

光学散射导致水下成像质量低下,并且水下场景往往非常复杂,水底的岩石、水草等都会干扰对目标特征的提取,因此让模型更针对性地提取特征十分有意义。本文选择在YOLOV4的CSPDarknet53主干特征提取网络中添加CBAM注意力机制。YOLOv4的具体结构如图1所示。

CBAM由Woo等[24]于2018年提出的一种用于前馈卷积神经网络的简单而有效的注意力机制结构。对于该注意力机制,给定一张特征映射,CBAM沿着通道和空间两个单独的维度依次推断注意力映射,然后将注意力映射和输入特征映射相乘,进行自适应特征细化。CBAM注意力机制不仅"告诉"模型应该关注哪里,还会提升关键区域的特征表达,只关注重要的特征而抑制或忽视无关的特征。CBAM主要由通道注意力和空间注意力模块构成,其结构如图2所示。

通道注意力模块主要是利用特征的通道之间的关系,生成通道注意力映射。在网络模型中,对于输入的C维的任意特征映射,分别通过平均池化层和最大池化层来聚合空间信息,得到两个C维池化特征映射,然后将这两个池化特征映射送入一个包含隐藏层的多层感知器

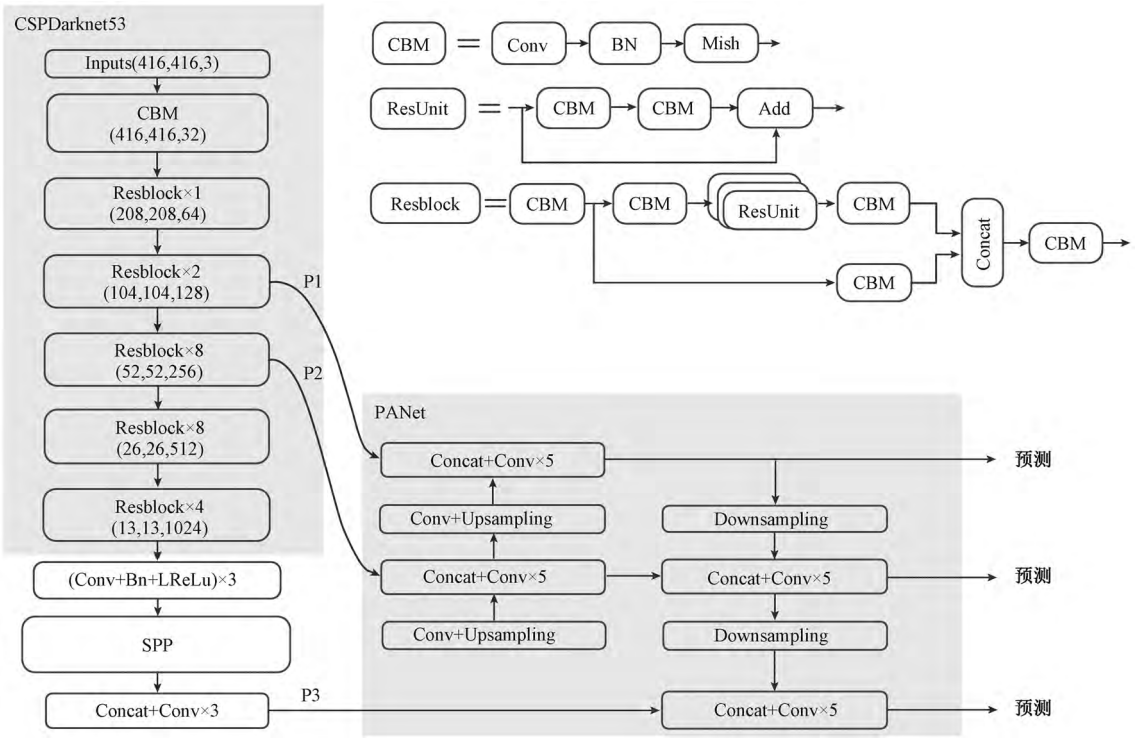

图 1　YOLOv4 整体结构

Fig. 1　Overall structure of YOLOv4

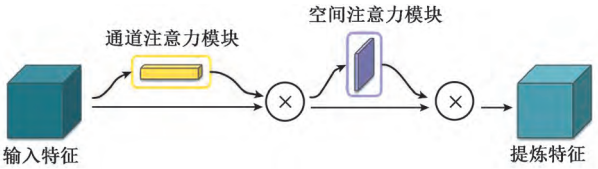

图 2　CBAM 整体结构

Fig. 2　Overall structure of CBAM

(multi-layer perceptron, MLP)中,得到两个 1×1×C 的通道注意力映射,最后将经过多层感知器得到的两个通道注意力映射进行逐元素相加得到最终的通道注意力映射 $M_c \in R^{1\times1\times C}$,通道注意力结构如图 3 所示。

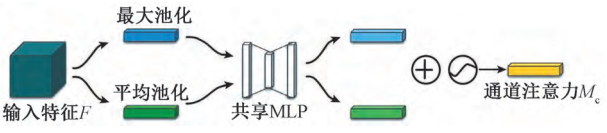

图 3　通道注意力模块结构

Fig. 3　Structure of channel attention module

空间注意力模块主要是利用特征之间的空间关系来生成空间映射。在网络模型中,对经过通道注意力映射细化后的特征映射 F',首先将其沿通道方向进行最大池化和平均池化,得到两个二维的特征映射,然后将这两个 1×H×W 的特征映射进行维度上的拼接操作,对于拼接后的特征映射,利用尺度为 7×7 的卷积层生成空间注意力映射 $M_s \in R^{H\times W\times1}$,空间注意力结构如图 4 所示。

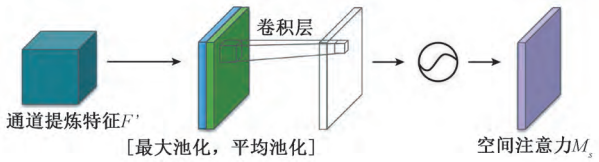

图 4　空间注意力模块结构

Fig. 4　Structure of spatial attention module

CSPDarknet53 中有连续 5 个残差块(Resblock),5 个残差块中 ResUnit 堆叠的数量依次是[1,2,8,8,4]个。本文选择将 CBAM 插入到每个残差块中的首个 ResUnit 中,在经过两次 CBM 之后先添加通道注意力模块,再添加空间注意力模块达成串联结构。具体实现结构如图 5 所示。

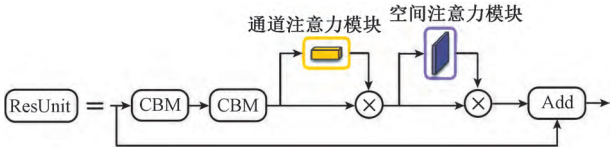

图 5　添加 CBAM 的 ResUnit 结构

Fig. 5　ResUnit structure with CBAM added

1.2　DetPANet

由于海洋生物的生长状况不一、拍摄角度不同等原因,所呈现的海洋生物大小形态各异。YOLOv4 原有的路径聚合网络 PANet 难以应对复杂多样的海洋生物,所以本文在原有的 PANet 模块中添加同层跳接结构与跨层跳接结构,将语义信息丰富的深层特征和位置信息、细节信息丰富的浅层特征充分结合起来。语义信息丰富的深层特征有助于目标的分类,位置、细节信息丰富的浅层特征有助于目标的定位及小目标的检测。具体结构对比如图 6 所示,本文将修改后的 PANet 命名为 DetPANet(Detailed-PANet)。

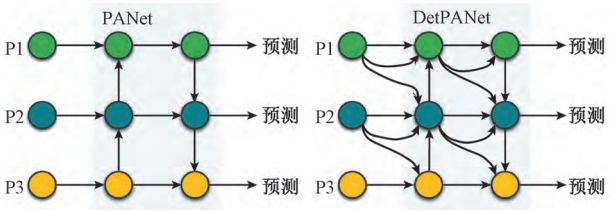

图 6　PANet 与 DetPANet 结构对比

Fig. 6　Structure comparison between PANet and DetPANet

在 YOLOv4 的 CSPDarknet53 特征提取网络中已经使用大量的卷积层去提取图像特征,使得分辨率大大降低,这样会导致图像细节的丢失,并不利于将来精确的目标定位与正确分类。因此,为避免细节的大量丢失,本文在同一水平层中添加跳层连接(skip connection),这样可以把较浅层的卷积层特征引用到更深层,其丰富的浅层信息可以有助于更精准的目标定位与正确分类。在训练过程中,跳层连接使得梯度更容易流动到深层网络。

并且,为了将低分辨率、语义强的特性与高分辨率、语义强的特性相结合,本文在不同大小的预测层之间添加 4 个最大池化层,从而实现深浅层的特征融合与多尺度目标的精确预测。通过底层特征与高层特征的融合,网络能够保留更多高层特征图蕴含的高分辨率细节信息,从而提高了图像定位精度。具体实现的网络结构如图 7 所示。

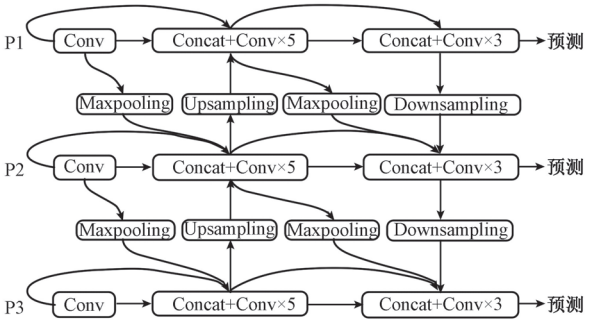

图 7　DetPANet 具体结构

Fig. 7　Specific structure of DetPANet

1.3　PredMix

图 8 展示了水下目标的互相重叠、遮挡现象,这种现象会对目标检测会造成一定的干扰,如果模型只是在原有的数据集上进行训练,那么在目标对象较密集的场景就容易出现漏检、误检问题,并且检测精度也将会有一定程度降低。本文提出将数据增强方法 PredMix 应用到 YOLOv4 的输入端,以解决该问题。

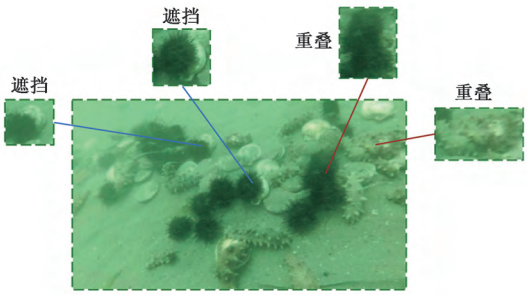

图 8　水下目标的遮挡、覆盖

Fig. 8　Overlaps and occlusions of underwater objects

整个数据增强方法分为两步。

第 1 步,生成训练集的预测数据。

将整个数据集分为 4 个部分:训练集 A、训练集 B、验证集和测试集,在训练集 A、B 上分别单独训练改进 PANet 的 YOLOv4 网络,利用训练完毕的两个网络模型 A、B 分别对另一部分训练集交叉预测,将每张训练集图片的检测结果(具体坐标位置)保存用于后续的数据增强操作,具体操作方式如图 9 所示。

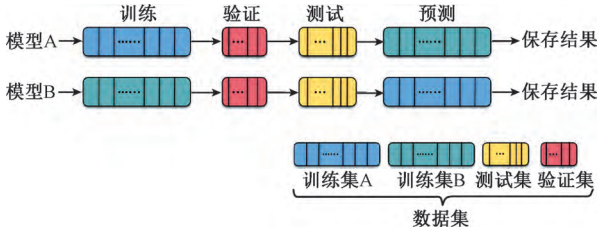

图 9　数据集预操作流程

Fig. 9　Pre operation flow of the dataset

第 2 步,将标签图像与预测图像以随机权重比混合生成新的训练样本参与训练。

令 $x^{truth} \in R^{H\times W\times C}$ 代表训练中某个图像的一个标签部分,y^{truth} 代表标签;$x^{pred} \in R^{H\times W\times C}$ 代表第 1 步预测结果中的某张图像的一个预测矩形框部分,y^{pred} 代表预测分类。PredMix 旨在通过将一张数据集图像中随机一个标签部分图像(x_i^{truth},y_i^{truth})和另一张数据集图像上随机一个被预测的矩形框部分图像(x_j^{pred},y_j^{pred})相结合,从而产生新的训练样本。预测矩形框和标签矩形框的大小通常是不一致的,所以首先将 x_j^{pred} 的大小调整至和 x_i^{truth} 一致。生

成的训练样本（$\tilde{x}$，$\tilde{y}$）用于模型训练。

具体操作定义如下：

$$\tilde{x} = \lambda' x_i^{truth} + (1-\lambda') x_j^{pred},\ \tilde{y} = y_i^{truth} \tag{1}$$

因为认为 y_i^{pred} 是 y_i^{truth} 的遮挡物，所以 y_i^{truth} 是 $\tilde{y}$ 标签。其中，λ' 是标签矩形框部分图像与预测矩形框部分图像的混合比例。λ' 的生成公式如下：

$$\lambda' = \max(\lambda, 1-\lambda) \tag{2}$$

其中，max( ) 是选取两者中较大值的函数。使用 PredMix 来增强数据，可以模拟水下目标重叠、遮挡的情况。本文使用这些新的混合区域代替原来的 x_i^{truth} 区域，并生成新的训练样本。

λ 定义如下：

$$\lambda = \beta(\alpha, \alpha) \tag{3}$$

选择将预测结果与真实标签混合而不是直接将两个真实标签混合，是为了模拟水下多种多样的重叠、覆盖、遮挡、模糊情况，如图 10 所示。预测的矩形框有时会有一定偏差，有时会错把其他物体（石块、水草等）当成标签对象进行预测。不同的预测结果与标签图像混合可以模拟水下对象各种不完全显示的情况。

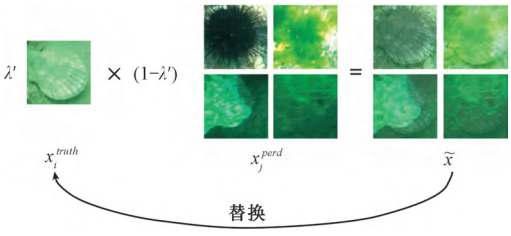

图 10　PredMix 操作示例

Fig. 10　Examples of PredMix operation

通过 PredMix 模拟目标的重叠、遮挡，可以提高模型对密集目标和显示不完全目标的检测能力。从统计学角度来看，PredMix 是预测框选区域和标签框选区域的一种线性插值，可以使得决策边界更加平滑，而不会出现突然过渡。PredMix 遵循邻域风险最小化原则（VRM 原则）而不是经验风险最小化原则（ERM 原则），这可以使得模型具有更好的鲁棒性和泛化能力。根据 ERM 原理训练的模型将经验风险最小化，可以帮助模型更好地拟合训练数据。将经验风险定义为：

$$R_\delta(f) = \frac{1}{n}\sum_{i=1}^{n} l(f(x_i), y_i) x_j^{pred} f \tag{4}$$

其中，f 代表 x 和 y 之间的非线性表达式，n 是样本数量，l 是测量 $f(x_i)$ 和 y_i 之间距离的损失函数。然而这种训练策略使得决策边界对训练数据拟合过多，导致过拟合。PredMix 遵循 VRM 生成训练数据的邻域分布。然后，可以用邻域数据（$\tilde{x}$，$\tilde{y}$）去替换训练数据（x_i^{truth}，y_i^{truth}）。近似预期风险 R_v 定义如下：

$$R_v(f) = \frac{1}{n}\sum_{i=1}^{n} l(f(\tilde{x}), \tilde{y})(\tilde{x}, \tilde{y}) \tag{5}$$

因此，训练过程转化为最小化预期风险。PredMix 使用邻域数据进行训练可以有效增强模型鲁棒性。

2　实验

2.1　实验配置及数据集

网络训练在配备 Intel Xeon（R）CPU E5-2620 v4 @ 2.10 GHz 处理器，32 GB 内存，NVIDIA GeForce RTX2080 显卡的工作站上进行，使用的深度学习平台是 Pytorch。

本文使用的是 URPC2018 水下目标检测大赛官方提供的数据集。URPC2018 数据集共有 5 543 张图像，包括海参、海胆、扇贝和海星 4 类。在 PredMix 操作第 1 步时按训练集 A、训练集 B、验证集、测试集顺序将数据集划分为 4∶4∶1∶1。在之后将训练集 A、B 合并，即按训练集、验证集、测试集顺序数据集被划分为 8∶1∶1。PredMix 数据增强操作只在训练集中进行，不对验证集和测试集操作。

本文选择 YOLOv4 在 VOC 数据集上预训练的权重文件作为预训练权重。PredMix 中的超参数 α 设置为 0.2，输入图像大小设置为 416×416。学习率设置采用余弦退火衰减调整策略，一次学习率周期迭代次数设置为 5，初始学习率为 1×10^{-4}，最小学习率设置为 1×10^{-5}。整个模型训练过程分为两步：第 1 步：冻结 CSPDarknet53 参数训练 80 个 epoch，以避免训练初期权值被破坏，batch size 设置为 16；第 2 步：解冻 CSPDarknet53 参数训练 80 个 epoch，batch size 设置为 4。所做实验网络模型均在 160 个 epoch 之前达到收敛。

2.2　实验结果

本文在实验过程中在控制训练参数一致的情况下将 3 处改进与 YOLOv4 算法做了详细对比，同时还测试了 SSD 模型和 Faster-RCNN 模型在 URPC2018 数据集上的检测效果。在默认 IoU 为 0.5、置信度为 0.5 的前提下，SSD 模型的 *mAP* 为 63.43%，Faster R-CNN 的 *mAP* 为 75.25%，YOLOv4 模型的 *mAP* 为 71.36%。可以发现，同样作为单阶段目标检测算法，SSD 模型的检测效果要远逊于 YOLOv4 模型，而双阶段目标检测算法 Faster R-CNN 的检测效果略强于 YOLOv4，仅在扇贝一类中不及 YOLOv4。这是由于扇贝较小，而 Faster R-CNN 的小目标检测能力较弱。

在经过 PredMix 数据增强之后，YOLOv4 的检测 *mAP* 比之前提升了 2.06%，对 4 个种类的检测效果均有一定提升，其中 *AP* 提升最多的是海胆，为 3.30%。因为海胆

是纯黑色,在实际检测中容易与背景的阴影部分相混淆,而 PredMix 充分模拟了海胆被其他物体遮挡的情况,网络模型对该种情况得到了充分学习。在添加 CBAM 特征注意力模块后 YOLOv4 的检测 *mAP* 提升了 2.43%,其中 *AP* 提升最多的是海参,这是由于海参在有些场景下与岩石特征近似,在检测时容易将两者混淆,添加 CBAM 注意力机制之后,网络模型对海参的特征得到了充分的关注。DetPANet 在 3 处改进中 *mAP* 提升最多,提升了 3.26%,对四种类别的检测效果中,对贝壳提升达到了 4.10%,这是由于贝壳多为小尺度目标,在检测时极易忽略,DetPANet 充分结合了深浅层特征,语义信息、位置信息、细节信息充分融合,从而有效提高了对小目标的检测效果。

本实验还充分验证了任意两种改进的叠加可以继续在原有的单种改进的基础上继续提升检测效果。CBAM 与 PredMix 的混合可以提升 5.06%,在 4 种类别的检测中提升最多的是海参;DetPANet 和 CBAM 的混合可以提升 *mAP* 3.66%,在 4 种类别的检测中提升最多的是贝壳,为 4.06%;DetPANet 和 PredMix 的混合对 *mAP* 的提升最多,为 5.31%。

最终本文提出的改进的 YOLOv4 网络与原 YOLOv4 网络相比, *mAP* 提升了 7.03%,比 SSD 网络的检测 *mAP* 高出 14.29%,比 Faster R-CNN 网络的检测 *mAP* 高出 3.14%,对 4 种类别的检测 *AP* 提升幅度为 4.71% ~ 9.58%。由实验结果可知,本文提出的结合数据增强和改进 YOLOv4 的水下目标检测算法在水下目标检测中能够实现更高精度的目标检测。

表 1 是详细的实验结果,其中 SSD 和 Faster R-CNN 表示 SSD 模型和 Faster R-CNN 模型的测试结果,①表示在 YOLOv4 上单独使用 PredMix 数据增强方法,②表示在 YOLOv4 上单独使用 CBAM 特征注意力机制,③表示在 YOLOv4 上单独使用 DetPANet 网络,①+②、②+③、①+③则表示数字对应的两种改进的结合,①+②+③则是本文提出的结合数据增强和改进 YOLOv4 的水下目标检测算法。

YOLOv4 改进前后的检测结果对比如图 11 所示。可以发现,在多种情况下,例如:对密集目标的检测、对多尺度目标的检测,结合数据增强和改进 YOLOv4 的水下目标检测算法均拥有更好的检测精度。

CBAM 特征注意力机制使得网络模型提取目标特征能力更强,在图像模糊或者目标特征与背景相似的情况下具有更好的检测效果:在图 11(c)组对比中,由于两只海参特征与岩石十分相似,YOLOv4 并未发现它们的存在。但是,本文提出的算法却能以很高的置信度检测出这两只海参。在图 11(g)组中改进后的 YOLOv4 发现了与背景极为相似的海参,而改进前的 YOLOv4 并未发现这只海参,这也能体现 CBAM 对网络模型特征提取能力的充分提升。

表 1 URPC2018 数据集测试结果

Table 1 Test results of URPC2018 dataset

算法	mAP/%	海胆	海星	扇贝	海参
baseline	71.36	83.40	79.66	61.46	60.93
SSD	63.43	76.68	69.44	49.15	58.47
Faster R-CNN	75.25	85.39	**86.62**	55.81	**73.17**
①	73.38	86.70	80.71	63.99	62.12
②	73.79	85.31	82.27	63.36	64.23
③	74.62	86.09	81.79	65.56	65.07
①+②	76.42	87.18	83.61	65.99	68.91
②+③	75.02	86.94	83.27	65.52	64.35
①+③	76.67	86.37	83.39	67.37	69.05
①+②+③	**78.39**	**88.11**	85.97	**68.95**	70.51

DetPANet 通过跨层跳接结构和同层跳接结构增强了网络模型的多尺度特征融合能力,并且由于深浅层特征的充分融合,使用 DetPANet 的 YOLOv4 网络模型对大小不一、形态各异的目标都能达到更好的检测效果:图 11(b)组对比中,由于一只海星只显示了侧面,YOLOv4 轻易就忽略了它,还有一只海星由于尺度太小也被忽略。而本文提出的算法能以极高的置信度检测出只显示侧面的海星,并且也检测出尺度很小的海星。在图 11(h)组中远处大量附着在岩石上的海胆,由于尺度太小并且过于密集,YOLOv4 产生了大量的漏检,而改进后的 YOLOv4 却能精确地区分彼此。这种对不同尺度、不同形态目标的有效检测能力在图 11(f)、(j)组中也同样有所体现。

PredMix 可以充分模拟水下生物的重叠、遮挡等显示不完全现象,从而增强模型鲁棒性,使得模型在检测显示不完全对象时具有更强的检测能力:在图 11(a)组对比中,YOLOv4 并未检测到最上方的两只贝壳,并且在图中央的检测也出现分类错误和定位不精确的问题,而本文提出的算法可以检测到最上方的两只贝壳,在图中央密集生物对象检测中也能较为精准地区分各个生物。在图 11(d)组中,改进后的 YOLOv4 能够以很高的置信度发现被岩石遮挡一半的海胆。这种对显示不完全对象的有效检测能力在图 11(e)、(g)以及(i)组对比中也有明显体现。

以上检测结果充分证明,本文提出的结合数据增强和改进 YOLOv4 的水下目标检测算法能够在水下目标检测中达到更高的检测精度。

本文还针对 PredMix 数据增强做了对比试验,以证明将预测矩形框与标签矩形框进行混合的必要性。具体实验结果如表 2 所示,其中 GTMix(ground truth mix)表示

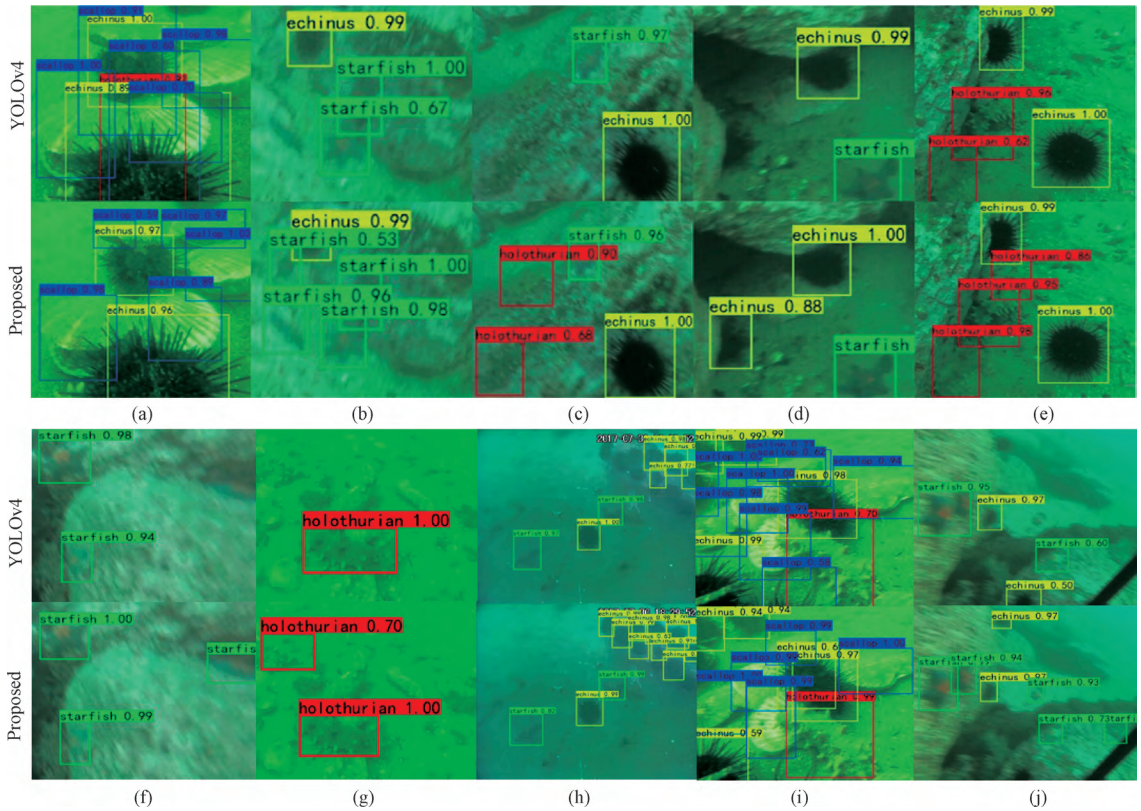

图 11　检测结果对比

Fig. 11　Comparison of test results

将来自不同训练集图片的两个标签矩形框进行混合，RandMix(random mix)则表示将来自不同训练集图片的随机区域和标签矩形框进行混合。

表 2　PredMix 对比实验

Table 2　Comparison test of PredMix

算法	mAP/%	海胆/%	海星/%	扇贝/%	海参/%
baseline	71.36	83.40	79.66	61.46	60.93
GTMix	72.30	84.05	79.72	64.02	61.42
RandMix	72.22	84.93	79.71	63.29	60.98
PredMix	**73.38**	**86.70**	**80.71**	**63.99**	**62.12**

表 3 展示了各改进对网络运行性能的影响，由于 PredMix 是先于网络模型，在数据集上进行的操作，不会对网络运行性能造成影响。CBAM 和 DetPANet 改进降低了网络模型的检测速度，最终，本文提出的算法检测速度为 10.42 帧。同时本文也对 SSD 模型和 Faster R-CNN 模型的检测速度进行了测试，这两种模型的速度均慢于原 YOLOv4 模型以及本文改进的 YOLOv4 模型。

表 3　各改进对网络运行性能的影响

Table 3　The impact of improvements on network performance

算法	FPS
baseline	**15.62**
SSD	9.72
Faster R-CNN	3.19
①	/
②	11.60
③	13.37
①+②	/
②+③	10.84
①+③	/
①+②+③	10.42

3　结　论

本文提出了一种结合数据增强和改进 YOLOv4 的水下目标检测算法，在 YOLOv4 的主干特征提取网络 CSPDarknet53 中添加 CBAM 特征注意力机制，以提高网络模型特征提取能力；在路径聚合网络 PANet 中添加同层跳接和跨层跳接结构，以增强网络模型多尺度特征融合能力；提出一种新型的数据增强方法来增强模型鲁棒性。实验结果表明，结合数据增强和改进 YOLOv4 的水

下目标检测算法有效提高了水下目标检测精度。

参考文献

[1] YU W D, LIAO H C, HSIAO W T, et al. Automatic safety monitoring of construction hazard working zone: A semantic segmentation based deep learning approach [C]. Proceedings of the 2020 the 7th International Conference on Automation and Logistics (ICAL), 2020: 54-59.

[2] HUVAL B, WANG T, TANDON S, et al. An empirical evaluation of deep learning on highway driving[J]. arXiv Preprint,2015, arXiv:1504.01716.

[3] ZHAO Y, MA J, LI X, et al. Saliency detection and deep learning-based wildfire identification in UAV imagery[J]. Sensors, 2018, 18(3): 712.

[4] WANG T, CHEN Y, QIAO M, et al. A fast and robust convolutional neural network-based defect detection model in product quality control[J]. The International Journal of Advanced Manufacturing Technology, 2018, 94(9): 3465-3471.

[5] 伊欣同,单亚峰. 基于改进 Faster R-CNN 的光伏电池内部缺陷检测[J]. 电子测量与仪器学报, 2021, 35(1):40-47.
YIN X T, DAN Y F. Photovoltaic cell internal defect detection based on improved faster R-CNN[J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(1):40-47.

[6] LI X, LIU Y, ZHAO Z, et al. A deep learning approach of vehicle multitarget detection from traffic video[J]. Journal of Advanced Transportation, 2018(11):1-11.

[7] GAO S H, TAN Y Q, CHENG M M, et al. Highly efficient salient object detection with 100k parameters [C]. European Conference on Computer Vision. Springer, Cham, 2020: 702-721.

[8] 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 Progess, 2020, 57(6): 060002.

[9] WANG S H, ZHAO J W, CHEN Y Q. Robust tracking of fish schools using CNN for head identification[J]. Multimedia Tools and Applications, 2017, 76(22): 23679-23697.

[10] SHORTEN C, KHOSHGOFTAAR T M. A survey on image data augmentation for deep learning[J]. Journal of Big Data, 2019, 6(1): 1-48.

[11] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.

[12] GIRSHICK R. Fast R-CNN[C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.

[13] 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.

[14] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.

[15] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.

[16] REDMON J, FARHADI A. YOLOV3: An incremental improvement [J]. arXiv Preprint, 2018. arXiv:1804.02767.

[17] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv Preprint,2020, arXiv:2004.10934.

[18] CUBUK E D, ZOPH B, MANE D, et al. Autoaugment: Learning augmentation strategies from data[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 113-123.

[19] CUBUK E D, ZOPH B, SHLENS J, et al. Randaugment: Practical automated data augmentation with a reduced search space[C]. Proceedings of the IEEE/CVF Confer- ence on Computer Vision and Pattern Recognition Workshops, 2020: 702-703.

[20] DEVRIES T, TAYLOR G W. Improved regularization of convolutional neural networks with cutout[J]. arXiv Preprint,2017, arXiv:1708.04552.

[21] SINGH K K, LEE Y J. Hide-and-seek: Forcing a network to be meticulous for weakly-supervised object and action localization [C]. 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, 2017: 3544-3553.

[22] ZHANG H, CISSE M, DAUPHIN Y N, et al. Mixup: Beyond empirical risk minimization[J]. arXiv Preprint, 2017,arXiv: 1710.09412.

[23] 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). IEEE, 2020: 2588-2592.

[24] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional

block attention module[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3-19.

作者简介

史朋飞,2008 年于南京信息工程大学获得学士学位,2011 年于河海大学获得硕士学位,2016 年于河海大学获得博士学位,现为河海大学副教授,主要研究方向为水下探测与成像、信息获取与处理、机器视觉等。E-mail: shipf@ hhu. edu. cn

Shi Pengfei received his B. Sc. degree in 2008 from Nanjing University of Information Science & Technology, received his M. Sc. degree and Ph. D. degree from Hohai University in 2011 and 2016 respectively, and now he is an assistant professor of College of IOT Engineering in Hohai University. His main research interests include underwater detection and imaging, information acquisition and processing and machine vision.

倪建军(通信作者),1999 年于中国矿业大学获得学士学位,2002 年于中国矿业大学获得硕士学位,2005 年于中国矿业大学获得博士学位,现为河海大学教授,主要研究方向为机器学习、模式识别、机器人等。E-mail:njjhhuc@ gmail. com

Ni Jianjun (Corresponding author) received his B. Sc. degree in 1999 from China University of Mining and Technology, received his M. Sc. degree in 2002 from China University of Mining and Technology, received his Ph. D. degree from China University of Mining and Technology, and now he is a professor of College of IOT Engineering in Hohai University. His main research interests include machine learning, pattern recognition, and robotics.