采用注意力机制与改进 YOLOv5 的水下珍品检测

林 森[1]，刘美怡[2]，陶志勇[2]

（1. 沈阳理工大学自动化与电气工程学院，沈阳 110159；2. 辽宁工程技术大学电子与信息工程学院，葫芦岛 125105）

摘 要：海胆、海参、扇贝等水下珍品在渔业中具有重要意义和价值，最近，利用机器人捕捞水下珍品成为发展趋势。为了探测水下珍品的数量及分布情况，使水下机器人获得更加可靠的数据，该研究提出基于注意力机制与改进 YOLOv5 的水下珍品检测方法。首先，使用 K-means 匹配新的锚点坐标，增加多个检测尺度提升检测精度；其次，将注意力机制模块融入特征提取网络 Darknet-53 中获得重要特征；然后，利用 Ghost 模块的轻量化技术优势，引入由 Ghost 模块构成的 Ghost-BottleNeck 代替 YOLOv5 中的 BottleNeck 模块，大幅度降低网络模型的参数与计算量；最后，将 IOU_nms 修改为 DIOU_nms 以优化损失函数。采用基于实际水下环境建立的数据集，样本数量为 781 幅图像，按照 9∶1 的比例随机划分训练与测试集，对改进的网络进行验证。结果表明，该研究算法可获得 95.67%平均准确率，相比 YOLOv5 算法可提升 5.49 个百分点，试验效果良好，研究结果可以为水下珍品的检测捕捉提供更加准确快捷的方法。

关键词：机器视觉；图像识别；水下珍品；轻量化；YOLOv5；注意力机制；多尺度

doi：10.11975/j.issn.1002-6819.2021.18.035

中图分类号：TP391 **文献标志码**：A **文章编号**：1002-6819(2021)-18-0307-08

林森，刘美怡，陶志勇．采用注意力机制与改进 YOLOv5 的水下珍品检测[J]．农业工程学报，2021，37(18)：307-314.
doi：10.11975/j.issn.1002-6819.2021.18.035 http://www.tcsae.org

Lin Sen, Liu Meiyi, Tao Zhiyong. Detection of underwater treasures using attention mechanism and improved YOLOv5[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(18): 307-314. (in Chinese with English abstract) doi：10.11975/j.issn.1002-6819.2021.18.035 http://www.tcsae.org

0 引 言

在农业养殖生产作业中，水下珍品（海参、海胆、扇贝等）一直深受渔民的喜爱。早期渔民们对其捕捞方式主要为撒网捕捞和人工抓取[1-2]两种形式。撒网捕捞虽然可以有效减少渔民成本，但是长期如此会严重损害海底的生态环境。人工抓取虽然解决了海底环境大范围被破坏的问题，但也给渔民带来更高的捕捞成本，同时增加了人身安全隐患。近年来，中国海洋科技水平不断提高，渔业、水产养殖等海洋经济也愈发依赖水下目标探测技术的发展。目前，有部分研究者把基于卷积神经网络的目标检测框架应用到渔业生产中，获得了一定效果[3-5]。

传统目标检测中根据检测对象的颜色、纹理和边缘等特征进行识别。如 Hsiao 等[6]提出一种基于稀疏表示分类（Sparse Representation-based Classification, SRC）的最大概率局部排序方法，称为 SRC-MP，用于实际鱼类识别。特征面和鱼面通过鱼类数据库提取特征数据，采用特征空间维数和部分排序值两个参数对方案进行优化，识别率达到 81.8%。Fabic 等[7]利用斑点计数和形状分析从水下视频序列中进行鱼类检测，采用预处理使珊瑚变黑进一步去除珊瑚背景，使用 Canny 边缘检测来提取鱼类轮廓。吴一全等[8]提出一种基于 Krawtchouk 矩、灰度共生矩阵、蜂群优化多核最小二乘支持向量机的识别方法，可以快速准确地识别淡水鱼的种类，对 5 种淡水鱼识别精度均达 83.33%以上。崔尚等[9]提出基于 Sobel 改进算子的海参图像识别研究，采用直方图均衡对图像进行预处理，利用 Sobel 改进算子将增强后的图像进行分割处理，经过多次膨胀、腐蚀处理和小目标移除算法处理，得到只含有海参目标的二值化图像。马国强等[10]提出改进的 K-均值聚类算法可以精准识别人工养殖的石斑鱼，该算法在输入图像清楚和干扰小等情况下分割准确率可以达到 98%。此类方法识别效果较好，但通常仅能检测单一目标且需要人工设计算子进行特征提取，工作量较大。

近年来，越来越多的深度学习方法被应用于水下目标检测。李艳君等[11]提出一种立体视觉下动态鱼体尺寸测量方法，该研究使用双目立体视觉技术获取三维信息，通过 Mask-RCNN（Mask-Region Convolution Neural Network）网络进行鱼体检测与精细分割，试验平均相对误差分别在 4.7%和 9.2%左右。董鹏等[12]提出一种水下海参自动检测与尺寸测量的方法，其在左目矫正图像上，利用预先训练的 YOLOv3 海参检测模型，进行海参自动检测和感兴趣区域定位，所提方法在 0.5～1.5 m 范围内平均误差为 1.65%。赵德安等[13]采用优化的 Retinex 算法提高了图像对比度，增强图像细节，利用卷积神经网络 YOLOv3 识别出河蟹，准确率为 96.65%。郭祥云等[14]提出一种基于深度残差网络的水下海参实时识别算法，采用颜色变换方法进行数据增强，该方法具有较高识别准确率，可达到 97.25%。Li 等[15]改进 Faster R-CNN 网络结

收稿日期：2021-03-02 修订日期：2021-09-09
基金项目：国家重点研发计划（2018YFB1403303）
作者简介：林森，博士，副教授。研究方向：图像处理与机器视觉，模式识别与人工智能等。Email：lin_sen6@126.com

构，提出适用于水下鱼类目标检测的轻型 R-CNN，准确率可达 89.95%。徐建华等[16]提出一种基于 YOLOv3 算法的目标识别模型。通过降采样重组、多级融合、优化聚类候选框、重新定义损失函数等方式优化网络结构，水下目标识别的准确率为 75.1%。王小宇等[17]提出了适用于水下目标检测识别场景的卷积神经网络结构，该方法水下目标识别准确率要高于传统卷积神经网络和高阶统计量特征的传统方法，可达到 91.7%。Mandal 等[18]通过 Faster R-CNN 与 3 个分类网络（ZFNet、CNN-M 和 VGG-16）相结合，进一步对 50 种鱼类和甲壳类动物进行检测，其平均准确率为 82.4%。Chuang 等[19]基于完全无监督的特征学习以及错误弹性分类器提出水下鱼类识别框架，可以较好地识别不同环境下的鱼类，平均准确率为 92.1%。Luo 等[20]利用人工神经网络去除图像中的噪音并准确识别鱼群，准确率为 89.6%。

以上基于卷积神经网络的方法检测单一品种时，准确率较高，但针对多品种检测时效果不理想，平均准确率较低。为了解决上述问题，实现水下珍品的精确捕捞，在 YOLOv5 的基础上提出一种基于注意力机制与改进 YOLOv5 的水下珍品检测方法，称为 CG-YOLOv5。本文方法的优势主要在于：1）在特征提取网络 Darknet-53 上融合注意力机制（Convolutional Block Attention Module，CBAM）结构，提升特征提取网络性能；2）使用 K-means 匹配新的锚点坐标，将 YOLOv5 算法中的 3 个检测尺度扩展为 4 个，提高模型对水下目标的检测精度；3）利用 Ghost 模块的轻量化技术优势，引入由 Ghost 模块构成的 Ghost-BottleNeck 代替 YOLOv5 中的 BottleNeck 模块，大幅度降低网络模型的参数与计算量。本文算法针对多品种检测提高平均准确率，可通过大量水下珍品图像检测试验进行验证，为后续的现代化珍品捕捞提供参考。

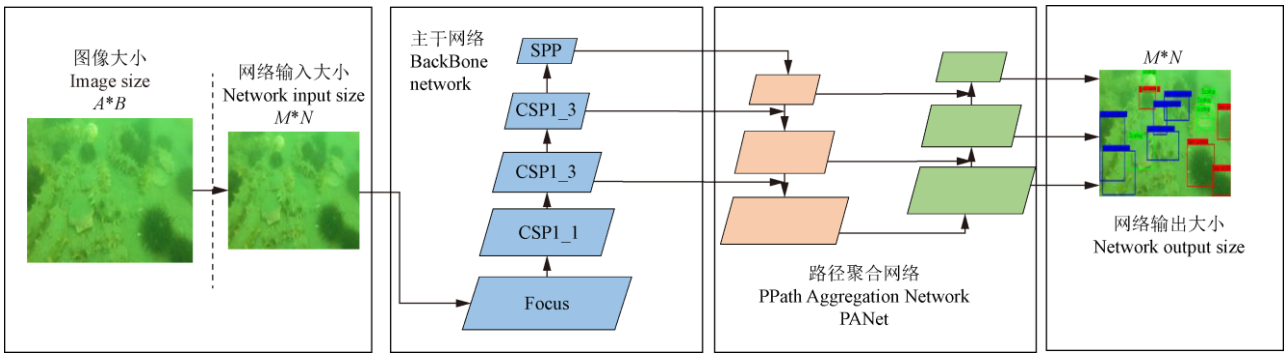
1 理论基础

1.1 YOLOv5 算法原理

YOLOv5 具有速度快、灵活性高的特点，网络结构主要包括 Darknet-53 主干网络、路径聚合网络（Path Aggregation network，PANet）[21]，如图 1 所示。

主干网络采用 CSP1_*X* 结构，主要包括两个分支，分支一由 *X* 个 Bottleneck 模块串联，分支二为卷积层，然后两个分支拼接到一起，使网络深度增加，特征提取能力大幅增强。

PANet 结构是由卷积操作、上采样操作、CSP2_X 构成的循环金字塔结构，可以使图像不同特征层之间相互融合，以进行掩模预测，经非极大值抑制（Non-Maximum Suppression，NMS）[22]获得最终预测框。



注：*A*、*B* 表示数据集中的图像分辨率大小；*M*、*N* 表示网络输入的图像分辨率大小；SPP 表示空间金字塔池化结构；CSP1_1、CSP1_3 表示 YOLOv5 中的瓶颈层；Focus 表示切片操作。

Note: *A* and B represent the image resolution of the data set; *M* and *N* represent the image resolution of the network input; SPP represents the spatial pyramid pooling structure; CSP1 _ 1 and CSP1 _ 3 represent the bottleneck layer in YOLOv5; Focus represents slicing operation.

图 1 YOLOv5 算法网络结构

Fig.1 Network structure of YOLOv5 algorithm

1.2 注意力机制模块

CBAM 是一种结合空间和通道的卷积注意力机制模块，给定中间特征图 $F=\mathbb{R}^{C\times H\times W}$ 作为输入，CBAM 模块会沿着两个独立的维度（通道和空间）依次推断注意力图，然后将注意力图与输入特征图相乘以进行自适应特征优化。

为了有效提取检测目标的轮廓特征，获取检测目标的主要内容，引入通道注意力模块，其计算方法如下

$$Mc(F)=\sigma\left(W_1\left(W_0\left(F_{\text{avg}}^c\right)\right)+W_1\left(W_0\left(F_{\max}^c\right)\right)\right) \quad (1)$$

式中 σ 表示 Sigmoid 函数，$W_0\in\mathbb{R}^{c/r\times c}$，$W_1\in\mathbb{R}^{c\times c/r}$，两个输入共享权重 W_0 和 W_1，ReLU 激活函数后接 W_0，F_{avg}^c、$F_{\max}^c$ 表示利用平均池化和最大池化在空间上生成的特征映射，H 为高度，W 为宽度，C 为通道数，r 为减少率。

为了精准定位检测目标的位置，提高目标检测准确率，引入空间注意力模块关注重点特征，其计算方法如下

$$Ms(F)=\sigma(f^{7\times7}([F_{\text{avg}}^s;F_{\max}^s])) \quad (2)$$

式中 F_{avg}^s、$F_{\max}^s$ 表示通道的平均池化特征和最大池化特征，$f^{7\times7}$ 表示滤波器尺寸为 7×7 的卷积运算。

1.3 Ghost 模块

本文基于 Ghost 模块的轻量化优势，提出 Ghost-BottleNeck 模块，如图 2 所示，该模块类似于 ResNet[23]中的基本剩余块，由两个堆叠的 Ghost 模块组成。左半部分充当 Ghost-BottleNeck 的扩展层，用于增加通道数量，从而增加特征维度，右半部分是为减少特征维度使其与输入一致，通过跳跃连接将左右两部分的输

入与输出相加，可以清晰的看到左右两部分的区别在于左半部分引入了 Relu 激活函数，目的是防止在输入数据后网络向后传播的过程中产生梯度消失现象。右半部分没有引入 Relu 激活函数的原因是：经过 Relu 激活函数后下一层和前一层输入数据的分布不同，从而需要不断适应不同的输入分布，导致网络训练速度降低。

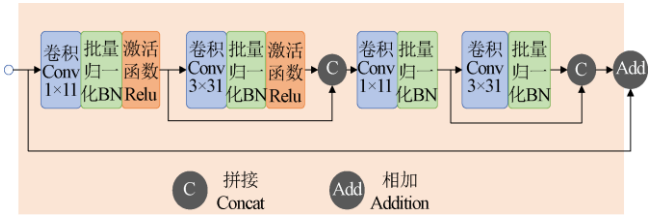


图 2　Ghost-BottleNeck 模块

Fig.2　Ghost-BottleNeck module

2　CG-YOLOv5 水下目标检测算法

虽然 YOLOv5 致力于水下目标检测，但在复杂水下环境中，许多目标区域信息容易丢失，不利于水下珍品的检测。为了提高检测精度，提出 CG-YOLOv5 水下目标检测算法，参数如表 1 所示，模型如图 3 所示。本文对于该模型的主要改进为：将注意力机制 CBAM、Ghost-BottleNeck 与 DarkNet-53 融合组成新的特征提取网络 CGDarkNet-53；使用 K-means 匹配新的锚点坐标，将 YOLOv5 检测尺度扩展为 4 个，提高水下目标检测精度。本文采用 CGDarkNet-53 作为 CG-YOLOv5 的主干网络，与 YOLOv5 相比，CG-YOLOv5 仅有一种 CSP_X 结构，可将梯度变化完整的集成到特征图中，加强网络特征融合能力，从而保证准确率。CG-YOLOv5 增加一个新的检测尺度用于提升目标检测精度，即网络层 15 输出得到的 Yolo head1。

2.1　CGDarkNet-53 主干网络

为了抑制网络中无用特征，CGDarkNet-53 引入 CBAM 增加网络深度并提升特征提取能力。CBAM 利用通道注意力机制和空间注意力机制结合的方法对特征向量进行筛选加权，其中通道注意力机制重点描述检测目标的内容，空间注意力机制重点描述检测目标的位置，通过两者结合来体现重要特征信息，弱化一般特征信息，进一步对水下珍品进行更加精确的定位和识别。Ghost-BottleNeck 具有更简易的线性运算，在轻量化的同时保持准确性，与 CBAM 结合生成 CGCSP_*X* 单元，如图 3 所示，Ghost-BottleNeck 右侧增加 Leaky Relu 函数，避免负值输入的梯度为 0，进而解决部分神经元不学习的问题，更充分地学习图像特征。

2.2　多尺度输入

YOLOv5 算法使用 K-means 对数据集中的 bounding box 聚类以获取合适锚点，锚点框的选取会直接影响目标检测的效果。为进一步提高检测精度，需要对水下数据集的标签重新聚类获取新的锚点。CG-YOLOv5 将 YOLOv5 中的 3 个检测尺度扩展为 4 个检测尺度，在执行多尺度检测时，经过 15 层得到第一个检测尺度 Yolo head1。将所得 Yolo head1 上采样的结果与第 5 层进行特征融合，得到第二个检测尺度 Yolo head2。接着将 Yolo head2 进行卷积运算与 Yolo head1 进行特征融合，得到第三个检测尺度 Yolo head3。最后将 Yolo head3 进行卷积运算与第 11 层进行特征融合，得到第 4 个检测尺度 Yolo head4。因此，CG-YOLOv5 具有更好检测不同尺度目标的性能。

表 1　网络参数

Table 1　Network parameters

层数 Layers	网络层 Network layer	输入尺寸 Input size	步长 Step	通道数 Number of channels
1	Focus	640×640×3	—	32
2	Conv3×3	320×320×32	2	64
3	CGBottleNeckCSP_1	160×160×64	—	64
4	Conv3×3	160×160×64	2	128
5	CGBottleNeckCSP_3	80×80×128	—	128
6	Conv3×3	80×80×128	2	256
7	CGBottleNeckCSP_3	40×40×256	—	256
8	Conv3×3	40×40×256	2	512
9	SPP	20×20×512	—	512
10	CGBottleNeckCSP_1	20×20×512	—	512
11	Conv1×1	20×20×512	1	256
12	Upsample	20×20×256	—	—
13	Concat	—	—	—
14	CGBottleNeckCSP_1	40×40×256		256
15	Conv1×1	40×40×256	1	128
16	Upsample	40×40×128	—	—
17	Concat	—	—	—
18	CGBottleNeckCSP_1	80×80×128		128
19	Conv3×3	80×80×128	2	128
20	Concat	—	—	—
21	CGBottleNeckCSP_1	80×80×256	—	256
22	Conv3×3	80×80×256	2	256
23	Concat	—	—	—
24	CGBottleNeckCSP_1	80×80×256	—	512
25	Detect	—	—	—

2.3　损失函数

CG-YOLOv5 的损失函数 L 主要由回归框预测误差 L_{loc}、置信度误差 L_{conf}、目标类别损失函数 L_{cla} 三部分组成。损失函数的计算公式如下

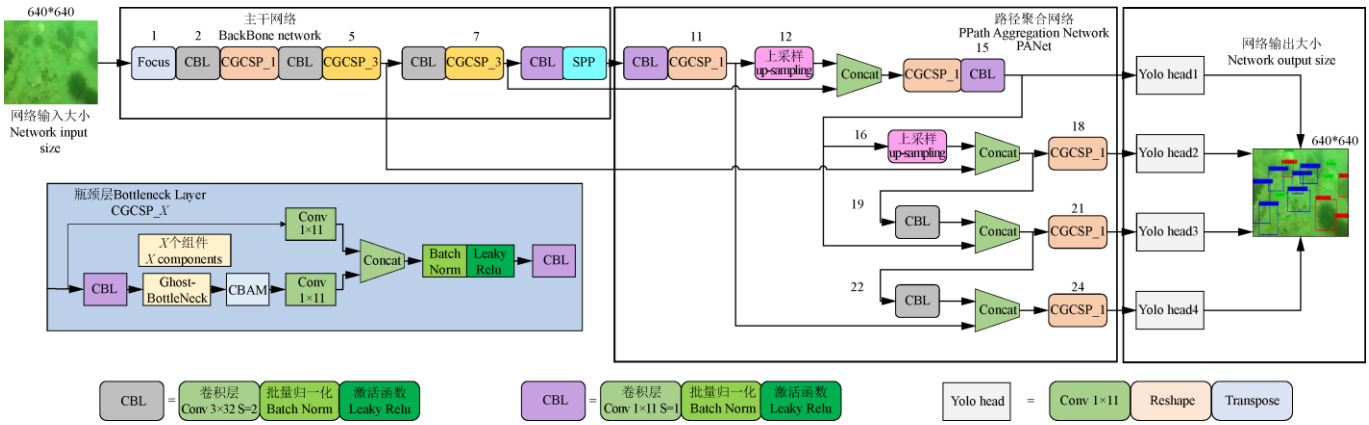
$$L = L_{\text{conf}} + L_{\text{cla}} + L_{\text{loc}} \tag{3}$$

式中 L_{loc} 利用 GIoU Loss[24]（Generalized Intersection over Union）函数，GIoU 计算公式如下

$$\text{GIoU} = \text{IoU} - \frac{\left|C - (A \cup B)\right|}{|C|} \tag{4}$$

式中 IoU 表示为预测框重叠区域，A 表示预测框，B 表示真实框，C 表示 A、B 最小包围框。

在目标检测的后处理过程中，针对很多目标框的筛选，通常需要 NMS 操作。YOLOv4 针对边界框中心点的位置信息，在 CIOU_Loss[25]的基础上采用 DIOU_nms，在重叠目标的检测中，DIOU_nms 的效果优于传统的 NMS。本文采用加权 NMS 的方式，在同样的参数情况下，将 NMS 中 IOU_nms 修改为 DIOU_nms，对于某些遮挡重叠的目标，除了考虑预测框重叠区域的 IoU 外，还考虑两个预测框中心点之间的距离，有效提升检测精度。



注：图中数字表示网络层数；CGCSP_*X* 表示有 *X* 个 Ghost-BottleNeck 模块；Conv 表示卷积操作；3×3、1×1 表示池化区域大小；*S* 表示步长。
Note: Numbers in the graph represent the number of network layers; CGCSP_*X* represents *X* Ghost-BottleNeck modules; Conv represents convolution operation; 3 ×3 and 1 ×1 represent the size of the pooling area; *S* represents step.

图 3 CG-YOLOv5 算法网络结构

Fig.3 Network structure of CG-YOLOv5 algorithm

3 结果与分析

3.1 试验平台

试验基于 Ubuntu18.04、Python3.7.7 和 PyTorch1.6.0 搭建的深度学习框架，试验相关硬件配置和模型参数如表 2 所示。CG-YOLOv5 可以自适应图片缩放，选取 640×640 大小的图像作为输入，可获得等比例大小的特征图作为检测尺度。通过多次试验得出，学习率选取 0.01 可以较快达到局部收敛，批量大小为 16 时训练速度较快。

表 2 试验相关硬件配置和模型参数

Table 2 Test related hardware configuration and model parameters

名称 Name	配置 Configuration	名称 Name	取值 Values
GPU	RTX2080Ti	图片大小/像素 Size of images/pixel	640×640
CPU	Core I7-6850k	学习率 Learning rate	0.01
CUDA	10.2	优化器 Optimizer	Adam
CuDNN	7.6.5	批量大小 Batch size	16

3.2 试验数据集

本文试验采用湛江水下机器人比赛数据集（http://uodac.pcl.ac.cn/），该数据集共有三个类别：海参、海胆、扇贝，数据集中的图像是由水下机器人在真实海底环境中拍摄的视频通过按帧截取所得，其图像分辨率为 1 920×1 080 像素，部分图像由于拍摄角度或没有水下珍品被人工删除，挑选后的数据集共包括 781 张图像，以 9∶1 比例随机划分训练集和测试集，抽样后再次统计标注信息、类别比例和大小分布，使训练集与验证集分布相似，达到划分目的。为了满足试验所需要求，首先，把数据集转变成 VOC2007 格式；然后，借助 Labelimg 软件对转换的数据集进行标注，手动设置类别为 Sea cucumber（海参）、Sea urchin（海胆）、Scallop（扇贝）三类。此外，为体现算法鲁棒性，即可在复杂水下环境中进行珍品的检测抓捕，数据集中图像为原始图像，没有进行任何清晰化等预处理。

3.3 评价指标

为了验证提出模型的有效性，从定性和定量两方面进行评估。对于定性评价，通过对比 CG-YOLOv5 和比较方法的检测图像差异来评估模型性能，即比较目标框的定位精确度，以及是否存在漏检、误检情况。定量评价方面，主要选取的指标为：准确率（Precision，*P*）、召回率（Recall，*R*）、平均准确率（Average Precision，AP）、平均精度均值 mAP（mean Average Precision）。公式如下

$$P=\frac{T_P}{T_P+F_P}\times 100\% \tag{5}$$

$$R=\frac{T_P}{T_P+F_N}\times 100\% \tag{6}$$

$$\mathrm{AP}=\int_0^1 P(r)\mathrm{d}r \tag{7}$$

以检测的海胆类别为例，式中 T_P 表示检测模型正确识别的数量，F_P 表示识别错误或未识别的数量，F_N 表示误把海胆目标检测为海参或扇贝的数量，P 为准确率，r 为召回率，$P(r)$ 是以 r 为参数的函数。平均准确率 AP 表示准确率对召回率的积分，通常使用 *P* 和 *R* 两个指标来衡量模型的好坏。所有类别的 AP 取平均值就是平均精度均值 mAP，mAP 可以衡量整个模型的性能。

3.4 结果与分析

3.4.1 定性结果与分析

为了更直观的体现 CG-YOLOv5 的性能，试验中随机抽取图像，将本文算法与 SSD[26]、Faster R-CNN[27]、YOLOv4[28]、YOLOv5、PP-YOLO[29]、PP-YOLOv2[30]、YOLOX[31]等目标检测算法进行对比，8 种检测算法均在同一试验平台进行训练与测试，算法结果对比如图 4 所示。

由图 4 观察可知 CG-YOLOv5 有效降低了漏检，提高了精度。如 Image3 所示，SSD 算法针对于小目标的检测效果较差，图像上方所示的海胆并未被准确检测；Image2 中，Faster R-CNN 在小目标检测上优于 SSD 算法，但对海参的检测效果并不好，Image3 中，图像右下方海参存在漏

检情况，Image1 中，海参检测正确但仅检测出某一小部分；YOLOv4 和 PP-YOLO 对海胆检测效果较差，如 Image3 和 Image4 上方海胆未被准确检测；YOLOv5 的海参检测效果优于 Faster R-CNN 算法，对小目标检测优于 SSD 算法，但也存在误检和漏检的情况，如 Image4 右上方扇贝存在误检；PP-YOLOv2 未准确检测海胆和扇贝，如 Image3 和 Image4 左上方，Image1 和 Image2 右上方均存在漏检情况；YOLOX 在 Image1 上方漏检海胆，Image3 左侧漏检海参；与其他算法相比，CG-YOLOv5 成功检测出 Image2 中下方海参和 Image3 中左上方较小海参和海胆，不但检测精度较高，而且能够适应复杂的水下环境，提升小目标和被遮挡目标的检测率，鲁棒性强。

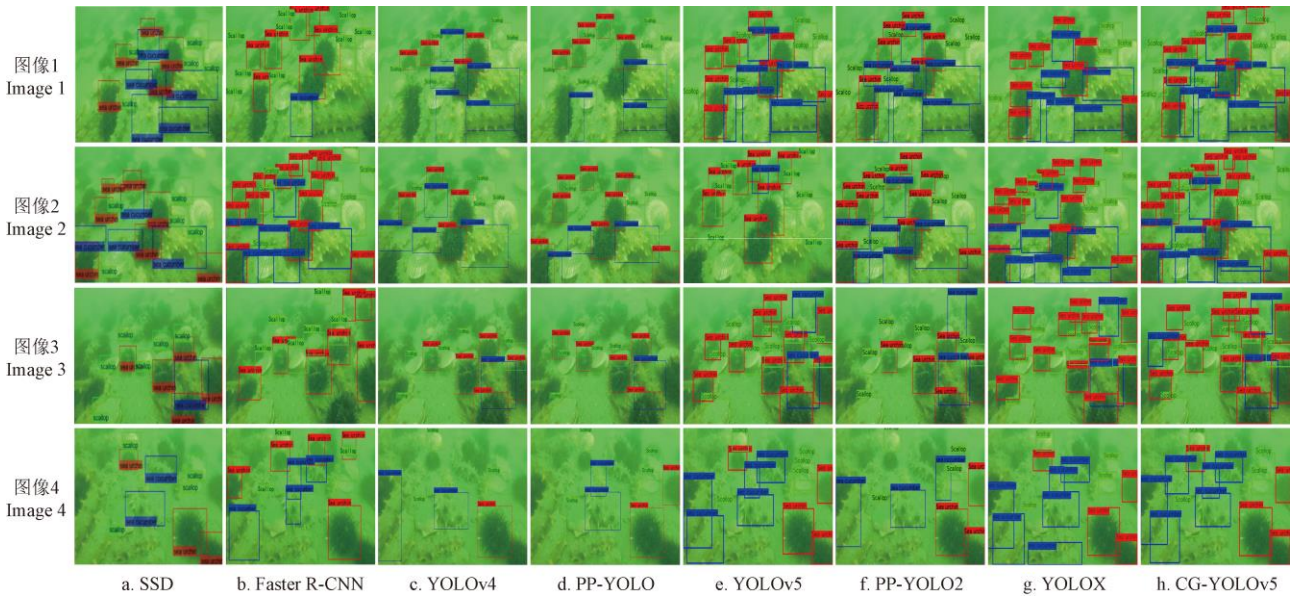


图 4　不同算法结果对比

Fig.4　Comparison of different algorithm results

3.4.2　消融试验

为了更好地验证本文算法的有效性，进行了消融试验，共验证 8 组网络，同样使用湛江水下机器人比赛数据集进行测试，试验模型客观评价对比结果如表 3 所示，加粗字体为算法最优值。

模型的网络复杂度可以用参数量（Parameters）和浮点运算量（Floating Point operations，FLOPs）来衡量，两者共同描述了数据经过复杂网络的计算量，参数量和浮点运算量数值越小，网络复杂度越低。如表 3 所示，采用 Ghost 模块的检测网络，其参数量和浮点运算量均低于其他相同参数条件下未采用 Ghost 模块的检测网络，从而说明 Ghost 模块具有轻量化作用，可有效提升网络性能。

表 3　消融试验

Table 3　Ablation experiment

序号 No.	模型 Models	平均准确率 Average precision/%			平均精度均值 Mean average precision/%	平均检测时间 Average detection time/s	参数量 Parameters/10^6	浮点运算量 Floating Point operations FLOPs /10^9
		海胆 Sea urchin	扇贝 Scallop	海参 Sea cucumber				
0	YOLOv5	88.25	87.45	94.84	90.18	0.015	7.3	16.9
1	YOLOv5+CBAM	95.68	92.91	95.82	94.80	0.021	26.1	58.1
2	YOLOv5+多尺度	92.54	93.94	95.22	93.90	0.020	7.3	16.9
3	YOLOv5+Ghost	92.13	89.28	95.22	92.21	**0.014**	**5.3**	**11.3**
4	YOLOv5+CBAM+多尺度	**96.15**	94.29	95.98	95.48	0.025	26.1	58.1
5	YOLOv5+Ghost+CBAM	94.63	91.51	95.07	93.74	0.019	24.2	52.4
6	YOLOv5+Ghost+多尺度	93.27	91.58	93.47	92.77	0.017	5.3	11.3
7	CG-YOLOv5	95.73	**94.35**	**96.93**	**95.67**	0.023	24.2	52.4

由表 3 数据可知，引入 CBAM 的目标检测网络相比于 YOLOv5 网络 mAP 值提升了 4.62 个百分点，检测时间增加 0.006 s，虽检测时间有所增加，但精度较高，说明注意力机制网络能抑制无用特征，有效提高 CNN 性能。将 YOLOv5 的 3 个检测尺度扩展为 4 个，可提升模型检测精度，mAP 值提升 3.72 个百分点。引入 Ghost-BottleNeck 替代 BottleNeck 可降低网络参数量和计算量，mAP 提升 2.03 个百分点。试验结果表明，本文提出的每个措施在性能方面均有所提升，CG-YOLOv5 与 YOLOv5 相比 mAP 值提升 5.49 个百分点，海胆、扇贝、海参的平均准确率提升 7.48、6.90、2.09 个百分点。虽然检测效率有少量降低，但检测精度得到较大的提升。

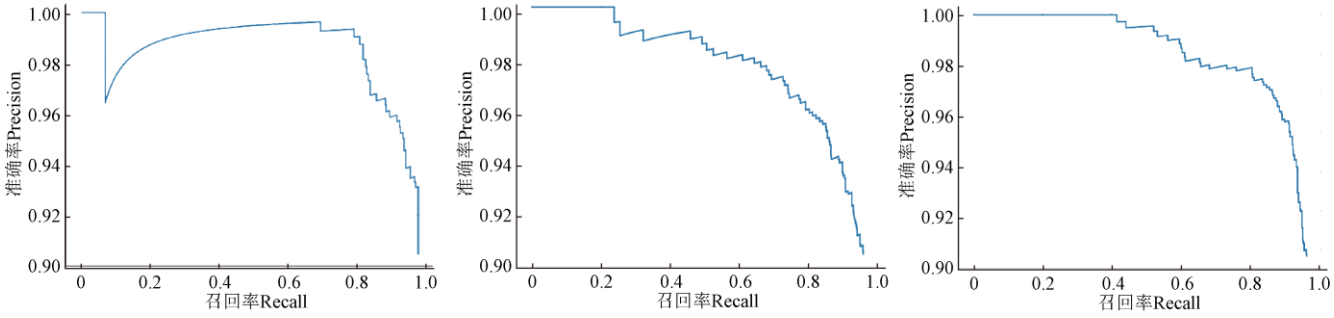
3.4.3　定量结果与分析

CG-YOLOv5 算法检测结果的 P-R 曲线如图 5 所示，

直角坐标系中横坐标是召回率，纵坐标是准确率。通过计算坐标系中 P-R 曲线下方部分面积便可以获得该类别的 AP 值，海参、扇贝和海胆三种类别的 AP 值依次为 96.93%、94.35%、95.73%，未完全识别主要原因为复杂的水下环境对检测造成干扰，因此后续工作将考虑增加图像的预处理，用于提升图像的清晰率和识别率。

本文算法与 SSD、Faster R-CNN、YOLOv4、YOLOv5、PP-YOLO、PP-YOLOv2、YOLOX 检测算法的性能指标对比结果如表 4 所示，加粗字体为算法最优值，从表中数据可知，CG-YOLOv5 算法可以获得更高的检测精度。在时间上，100 张图片平均检测时间虽高于 YOLOv5 和 SSD，但相比其他 5 种算法，具有良好的检测速度，以牺牲较少的时间为代价，获取更高的精度。因此，结合精度和速度综合考虑，本文算法更适合完成水下机器人对珍品的检测任务。

根据上述结果分析，与其他 7 种算法相比，CG-YOLOv5 算法在性能上具有更高的优势。改进模型更加充分利用了低层次的特征信息，进而提升在小目标检测方面的检测率；同时模型的注意力机制缩减无用特征对模型的干扰和影响，改善了遮挡目标检测效果，较大幅度提高模型的性能。



a. 海参
a. Sea cucumber

b. 扇贝
b. Scallop

c. 海胆
c. Sea urchin

图 5 CG-YOLOv5 算法各类目标 P-R 曲线

Fig.5 P-R curve of various targets of CG-YOLOv5 algorithm

表 4 不同检测算法性能指标对比

Table 4 Comparison of performance indicators for different detection algorithms

模型 Models	平均准确率 Average precision/%			平均精度均值 Mean average precision/%	平均检测时间 Average detection time/s
	海胆 Sea urchin	扇贝 Scallop	海参 Sea cucumber		
Faster R-CNN	62.40	68.33	7.28	46.00	0.026
SSD	87.46	86.50	76.68	83.55	**0.015**
YOLOv4	44.56	43.91	40.23	42.9	0.046
YOLOv5	88.25	87.45	94.84	90.18	**0.015**
PP-YOLO	41.34	41.45	42.31	41.70	0.034
PP-YOLOv2	88.90	91.70	85.50	88.66	0.077
YOLOX	88.31	87.06	88.94	88.10	0.169
CG-YOLOv5	**95.73**	**94.35**	**96.93**	**95.67**	0.023

4 结 论

本文提出基于注意力机制与改进 YOLOv5 的水下珍品检测方法，以便渔民利用水下机器人对水下珍品进行识别和捕捞。首先采用注意力机制 CBAM 对 YOLOv5 的特征提取网络进行改进。然后，利用 K-means 匹配新的锚点坐标，将 YOLOv5 算法中的 3 个检测尺度扩展为 4 个，提高了模型对水下目标的检测精度。最后，利用 Ghost 的优势，引入 Ghost-BottleNeck 代替 YOLOv5 中的 BottleNeck 模块，降低 YOLOv5 卷积神经网络的计算成本。

试验结果表明，所提出的 CG-YOLOv5 算法平均精度均值可以达到 95.67%，具有更好的准确性，故应用于水下珍品的检测时具有良好成效，与 SSD、Faster R-CNN、YOLOv4、YOLOv5、PP-YOLO、PP-YOLOv2、YOLOX 算法相比，本文算法平均检测时间为 0.023s ，检测速度较快。结合速度和精度综合考虑，具有较高应用价值，为后续的自动化珍品捕捞提供参考。

［参 考 文 献］

[1] Choe S, Ohshima Y. On the morphological and ecologica differences between two commercifl forms, “Green”and “Red”, of the Japan common sea cucumber, Stichopus japonicus Selenka[J]. Bull Jpn Soc Sci Fish, 1961, 27: 97-106

[2] Mitsunaga N, Matsumura S. Growth and survival of hatchery produced juveniles of sea cucumber Apostichopus japonicus in different size[J]. Bull Nagasaki Prefect Inst Fish, 2004, 30: 7-13.

[3] 郑一力，张露. 基于迁移学习的卷积神经网络植物叶片图像识别方法[J]. 农业机械学报，2018，49(S)：354-359.
Zheng Yili, Zhang Lu. Image recognition method of plant leaves based on transfer learning by convolutional neural network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(S): 354-359. (in Chinese with English abstract)

[4] 薛金林，闫嘉，范博文. 多类农田障碍物卷积神经网络分类识别方法[J]. 农业机械学报，2018，49(S1)：42-48.
Xue Jinlin, Yan Jia, Fan Bowen. Convolutional neural network classification and recognition method for multi-class farmland obstacles[J]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(S1): 42-48. (in Chinese with English abstract)

[5] 姜红花，王鹏飞，张昭，等. 基于卷积网络和哈希码的玉米田间杂草快速识别方法[J]. 农业机械学报，2018，49(11)：30-38.
Jiang Honghua, Wang Pengfei, Zhang Zhao, et al. Quick recognition method of weeds in corn field based on convolutional network and hash code[J]. Transactions of the

Chinese Society for Agricultural Machinery, 2018, 49(11): 30-38. (in Chinese with English abstract)

[6] Hsiao Y H, Chen C C, Lin S I, et al. Real-world underwater fish recognition and identification, using sparse representation[J]. Ecological informatics, 2014, 23: 13-21.

[7] Fabic J N, Turla I E, Capacillo J A, et al. Fish population estimation and species classification from underwater video sequences using blob counting and shape analysis[C]// 2013 IEEE International Underwater Technology Symposium (UT). IEEE, 2013: 1-6.

[8] 吴一全，殷骏，戴一冕，等. 基于蜂群优化多核支持向量机的淡水鱼种类识别[J]. 农业工程学报，2014，30(16)：312-319.
Wu Yiquan, Yin Jun, Dai Yimian, et al. Identification method of freshwater fish species using multi-kernel support vector machine with bee colony optimization[J]. Transactions of the Chinese Society for Agricultural Engineering (Transactions of the CSAE), 2014, 30(16): 312-319. (in Chinese with English abstract)

[9] 崔尚，段志威，李国平，等. 基于 Sobel 改进算子的海参图像识别研究[J]. 电脑知识与技术：学术交流，2018，14(22)：145-146.
Cui Shang, Duan Zhiwei, Li Guoping, et al. Research on sea cucumber image recognition based on Sobel improved operator[J]. Computer Knowledge and Technology: Academic Exchange, 2018, 14(22): 145-146. (in Chinese with English abstract)

[10] 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[J]. 计算机应用与软件，2016，33(5)：192-195.
Ma Guoqiang, Tian Yunchen, Li Xiaolan. Application of K-means clustering algorithm in color image segmentation of grouper on sea water background[J]. Computer Applications and Software, 2016, 33(5): 192-195. (in Chinese with English abstract)

[11] 李艳君，黄康为，项基. 基于立体视觉的动态鱼体尺寸测量[J]. 农业工程学报，2020，36(21)：220-226.
Li Yanjun, Huang Kangwei, Xiang Ji. Measurement of dynamic fish dimension based on stereoscopic vision[J]. Transactions of the Chinese Society for Agricultural Engineering (Transactions of the CSAE), 2020, 36(21): 220-226. (in Chinese with English abstract)

[12] 董鹏，周烽，赵悰悰，等. 基于双目视觉的水下海参尺寸自动测量方法[J]. 计算机工程与应用，2021，57(8)：271-278.
Dong Peng, Zhou Feng, Zhao Congcong, et al. Automatic measurement method of underwater sea cucumber size based on binocular vision[J]. Computer Engineering and Application, 2021, 57(8): 271-278 (in Chinese with English abstract)

[13] 赵德安，刘晓洋，孙月平，等. 基于机器视觉的水下河蟹识别方法[J]. 农业机械学报，2019，50(3)：151-158.
Zhao Dean, Liu Xiaoyang, Sun Yueping, et al. Underwater crab recognition method based on machine vision[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(3): 151-158. (in Chinese with English abstract)

[14] 郭祥云，胡敏，王文胜，等. 基于深度学习的非结构环境下海参实时识别算法[J]. 北京信息科技大学学报：自然科学版，2019(3)：27-31.
Guo Xiangyun, Hu min, Wang Wensheng, et al. Real time identification algorithm of sea cucumber in unstructured environment based on deep learning[J]. Journal of Beijing University of Information Technology: Natural Science Edition, 2019(3): 27-31 (in Chinese with English abstract)

[15] Li X, Tang Y, Gao T. Deep but lightweight neural networks for fish detection[C]//OCEANS 2017-Aberdeen. IEEE, 2017: 1-5.

[16] 徐建华，豆毅庚，郑亚山. 一种基于 YOLO-V3 算法的水下目标识别跟踪方法[J]. 中国惯性技术学报，2020，28(1)：129-133.
Xu Jianhua, Dou Yigeng, Zheng Yashan. An underwater target recognition and tracking method based on YOLO-V3 algorithm[J]. Journal of Chinese Inertial Technology, 2020, 28(1): 129-133. (in Chinese with English abstract)

[17] 王小宇，李凡，曹琳，等. 改进的卷积神经网络实现端到端的水下目标自动识别[J]. 信号处理，2020，36(6)：958-965.
Wang Xiaoyu, Li Fan, Cao Lin, et al. Improved convolutional neural network to realize end-to-end automatic recognition of underwater targets[J]. Signal Processing, 2020, 36(6): 958-965. (in Chinese with English abstract)

[18] 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). Rio de Janeiro: IEEE, 2018: 1-6.

[19] Chuang M C, Hwang J N, Williams K. A feature learning and object recognition framework for underwater fish images[J]. IEEE Transactions on Image Processing, 2016, 25(4): 1862-1872.

[20] Luo S, Li X, Wang D, et al. Automatic fish recognition and counting in video footage of fishery operations[C]// 2015 International Conference on Computational Intelligence and Communication Networks (CICN). Jabalpur: IEEE, 2015: 296-299.

[21] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 8759-8768.

[22] Neubeck A, Van Gool L. Efficient non-maximum suppression[C]// 18th International Conference on Pattern Recognition (ICPR'06). Hong Kong: IEEE, 2006, 3: 850-855.

[23] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770-778.

[24] Rezatofighi H, Tsoi N, Gwak J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 658-666.

[25] Zheng Z, Wang P, Liu W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[C]// Proceedings

of the AAAI Conference on Artificial Intelligence (AAAI). New York Hilton Midtown: 2020: 12993-13000.

[26] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]// European Conference on Computer Vision. Amsterdam: Springer, Cham, 2016: 21-37.

[27] 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.

[28] Ghiasi G, Cui Y, Srinivas A, et al. Simple copy-paste is a strong data augmentation method for instance segmentation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 2918-2928.

[29] Zheng Z, Zhao J, Li Y. Research on Detecting Bearing-Cover Defects Based on Improved YOLOv3[J]. IEEE Access, 2021, 9: 10304-10315.

[30] Li J, Zhang Z, Tian Y, et al. Target-Guided Feature Super-Resolution for Vehicle Detection in Remote Sensing Images[J]. IEEE Geoscience and Remote Sensing Letters, 2021: 1-5.

[31] Zhu X, Lyu S, Wang X, et al. TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 2778-2788.

Detection of underwater treasures using attention mechanism and improved YOLOv5

Lin Sen[1], Liu Meiyi[2], Tao Zhiyong[2]

(1. *School of Automation and Electrical Engineering, Shenyang Ligong University, Shenyang* 110159, *China*; 2. *School of Electronic and Information Engineering, Liaoning Technical University, Huludao* 125105, *China*)

Abstract: Underwater treasures, such as sea urchins, sea cucumbers, and scallops, have always been preferred in fish production, due mainly to the high value-added industry. However, two conventional approaches, including net fishing and manual catching, cannot meet the application requirements of rapid detection in the actual large-scale cultivation in modern agriculture, particularly on time-consuming, labor-intensive, and severe destruction of submarine environments in the early days. Alternatively, deep learning has widely been characterized by high resolution and fast speed in recent years. Therefore, it is a promising application potential to the target detection framework using the convolutional neural network in fishery production. It is also highly necessary to improve the detection performance in complex underwater environments. In this study, a YOLOv5 detection of underwater treasure was proposed using the attention mechanism, referred to as CG-YOLOv5, in order to provide a more accurate dataset for underwater robots. The main advantages were as follows: 1) DarkNet-53 was introduced the CBAM to deepen the network for the better performance of feature extraction, further to suppress the worthless features in the network. Specifically, the CBAM combined the channel and spatial attention to filter and weight the feature vectors. The channel attention focused mainly on what the detection target was, whereas, spatial attention was used to determine where the detection target was. As such, the prominent feature information was represented via two combined mechanisms, while weakening the general features. 2) The lightweight Ghost-Bottleneck module was introduced to replace the Bottleneck in YOLOv5. A simpler linear operation in Ghost-Bottleneck was utilized to maintain a higher accuracy with light weights. 3) New anchor points were obtained by clustering the labels of underwater datasets. A new detection scale was also added to the original three detections for higher detection accuracy. CG-YOLOv5 network mainly included CGDarknet-53 backbone network, Focus structure, Spatial Pyramid Pooling structure (SPP), and Path Aggregation Network (PANet). Focus served as a benchmark network with down sampling to change the input size of 640×640×3 to 320×320×32. Only one CSP structure was involved in the CG-YOLOv5 to integrate gradient changes completely into the feature map for feature fusion enhancement. The SPP structure was used to maximize the pooling of the feature layer. Four scales were utilized in the pooling layers with the pooling core sizes of 1×1, 5×5, 9×9, and 13×13, respectively. As such, the SPP effectively increased the perception field, while isolating significant contextual features. Furthermore, path aggregation networks were used to fuse different feature layers of an image. A specific dataset was also selected to verify the model using the actual underwater environment. There were 781 underwater images, 90% of which were employed as training datasets, and the rest were for testing. The experimental results demonstrated that the model fully met the requirement of detection and recognition for the treasures in complex underwater environments, compared with the current deep learning. The average accuracy was 95.67%. Compared with YOLOv5, the average precision of sea urchin, scallop and sea cucumber increased by 7.48, 6.90 and 2.09 percentage points, and mAP increased by 5.49 percentage points base point. Compared with other classical algorithms, the method has better accuracy and lower complexity. The finding can provide a more accurate and fast way to detect and capture aquatic products.

Keywords: computer vision; image recognition; underwater treasures; lightweight; YOLOv5; attention mechanism; multi-scale