

# 基于 YOLOv8-M 的海洋底栖生物检测算法

薄明迪, 刘勇

(青岛科技大学 信息科学技术学院, 山东 青岛 266000)

**摘要:** 当前的目标检测算法在复杂的海底环境中表现较差, 对海洋底栖生物的识别精度较低; 为此, 针对海洋底栖生物的检测问题进行研究, 提出了一种基于 YOLOv8-M 的改进模型 IFL-YOLO; 针对特征提取不足问题展开研究, 设计 C2f-dcs 模块, 通过结合空洞卷积与注意力机制, 扩大模型感受野, 增强网络特征提取能力, 优化小目标检测性能; 针对传统特征融合方式缺乏上下文信息的问题进行研究, 设计 CGF 模块, 应用自适应特征融合, 有效融合上下文信息, 提升定位精度, 并引入小目标检测头进一步提高检测精度; 采用自适应标签分配方法, 根据不同类别样本的统计特征进行自适应的 IoU 阈值设定, 改善正负样本分配能力; 经实验验证, 改进后模型在 DUO 海洋底栖生物数据集上实现了 73.4% 的检测精度, 较改进前提高了 3.1%, 显著提高模型检测精度。

**关键词:** 海洋底栖生物检测; YOLOv8; 空洞卷积; 自适应特征融合; 自适应标签分配

## Marine Benthos Detection Algorithm Based on YOLOv8

BO Mingdi, LIU Yong

(School of Information Science & Technology, Qingdao University of Science and Technology,  
Qingdao 266000, China)

**Abstract:** Current object detection algorithms have poor characteristics in complex seabed environments, leading to low recognition accuracy for marine benthos; To address this issue, research on marine benthos detection is conducted, and an improved model based on YOLOv8-M, IFL-YOLO, is proposed; For insufficient feature extraction, the C2f-dcs module is designed; The receptive field of the model is expanded by combining the dilated convolution and the attention mechanism, enhancing its feature extraction capabilities and optimizing the detection performance of small targets; To address the issue of traditional fusion methods lacking contextual information, the CGF module is designed; An adaptive feature fusion is used to effectively integrate contextual information and improve localization accuracy, while a small target detection head is introduced to further enhance detection precision; An adaptive label assignment method is used to set the adaptive IoU thresholds based on the statistical characteristics of different sample categories, and to improve the distribution of positive and negative samples; Experimental results show that the improved model achieves a detection accuracy of 73.4% on the DUO marine benthos dataset, with an improvement of 3.1% over the previous version, significantly enhancing detection accuracy.

**Keywords:** marine benthos detection; YOLOv8; dilated convolution; adaptive feature fusion; adaptive label assignment

## 0 引言

海洋具有很多的生物资源, 具有很高的经济价值。对海洋资源合理开发有利于保护海洋生态环境同时促进国民经济发展。检测和定位水下目标是水下目标检测的目的。随着机器人和海洋技术的不断发展, 其应用范围愈发广泛。它在很多方面都有应用, 如海洋养殖、环境

检测和资源管理。其中, 水下机器人可以完成大部分任务中的工作。但由于海洋水下环境复杂, 水下生物识别仍然是一个艰巨的挑战。提升算法性能可以使水下机器人更高效地完成上述任务。

海洋底栖生物检测算法主要是基于深度学习的算法。具体可以分为基于卷积神经网络 (CNN, convolutional neural network) 的算法和基于 Transformer 的算

收稿日期: 2025-02-24; 修回日期: 2025-03-17。

基金项目: 中国科学院海洋大科学中心 (KEXUE2019GZ04); 山东省重点研发计划 (2023RKY02009)。

作者简介: 薄明迪 (2000-), 男, 硕士研究生。

通讯作者: 刘勇 (1971-), 女, 博士, 副教授。

引用格式: 薄明迪, 刘勇. 基于 YOLOv8-M 的海洋底栖生物检测算法[J]. 计算机测量与控制, 2026, 34(3): 25-33.

法。基于卷积神经网络的算法主要分为两阶段算法 (Faster-RCNN<sup>[1]</sup>) 和一阶段算法 (YOLOv1<sup>[2]</sup>)。两阶段算法利用区域建议网络产生候选区域, 然后提取区域中的目标特征并使用分类器完成目标预测, 与一阶段算法相比具有更高的检测精度。一阶段算法简化流程, 通过直接训练网络获得预测物体的类别概率和坐标信息。由于网络结构更加简洁, 所以具有更快的检测速度。其次, 目标检测算法根据是否使用锚点可以分为有锚点和无锚点。基于锚的算法 (SSD<sup>[3]</sup>) 通过在不同分辨率的特征图上设置多尺度的锚框来检测不同大小的目标。无锚算法在训练过程中不会引入预定义的锚框。例如, CenterNet<sup>[4]</sup> 将目标检测任务视为关键点 (目标中心点) 检测任务, 避免了使用锚框。它使用中心点的位置预测来定位目标。FCOS<sup>[5]</sup> 在每个像素位置直接预测目标的类别、大小和位置。

除 CNN 算法外, Transformer 算法也广泛应用于计算机视觉中。Vision Transformer<sup>[6]</sup> 是 Transformer 应用到视觉领域, 是处理视觉任务的开山之作, 它可用于作为目标检测任务的特征提取网络。DETR<sup>[7]</sup> 是端对端的目标检测框架首次将 Transformer 应用于目标检测领域, 是卷积神经网络和 Transformer 相结合的结构。Swin Transformer<sup>[8]</sup> 是 Transformer 模型在视觉领域的又一次碰撞, 该模型一经发表就已在多项视觉任务中霸榜。

此外针对水下检测中的难点, 基于单阶段目标检测算法改进的水下检测目标算法也层出不穷。文献 [9] 提出了一种在 YOLOx 的基础上改进的海洋底栖生物检测算法, 通过改进特征融合和检测头的方式提高了检测精度。文献 [10] 将 YOLOv4 与深度可分离卷积结合, 改进特征融合模块从而更好融合语义信息, 提高了对海洋生物的检测能力。文献 [11] 优化了上采样模块以便更好检测小型鱼类, 增加了空间金字塔池化模块来减少错误检测。文献 [12] 在 YOLOv3 的基础上改善了上采样步长并增加了空间金字塔, 提高了对多尺度鱼类的检测性能。文献 [13] 使用大型卷积内核和自注意力机制增强模型特征提取能力, 提高模型对海洋生物的检测精度。但目前海洋底栖生物的检测算法较少, 文献 [14] 专注于研究特定物种, 应用场景较为有限。针对海洋水下环境面临的多重尚未提出合适的解决方案, 需要进一步研究。

目前, 现有的模型用于海洋底栖生物检测存在以下问题:

- 1) 海洋底栖生物尺寸过小、背景复杂, 模型提取特征能力不足, 导致模型检测精度下降;
- 2) 海洋底栖生物大多成群存在, 存在堆叠、遮挡情况, 模型对于边缘信息提取困难, 降低了模型的检测

精度和敏感性;

3) 海洋底栖生物尺寸差异较大, 传统的标签分配方式通常为同一场景中不同尺度的目标分配相同数量的正样本, 这种方式容易导致小目标出现正负样本不平衡的问题。

针对上述问题, 本文提出一种基于 YOLOv8-M 的海洋底栖生物检测算法 IFL-YOLO, 主要贡献如下:

1) 针对特征信息提取不足的问题, 设计了 C2f-dcs (C2f-dilated convolution small) 模块, 并将其替换模型中的 C2f 模块。通过引入空洞卷积和注意力机制, C2f-dcs 模块有效融合了特征中的位置信息和长距离依赖, 同时抑制了无关的背景信息, 从而优化了小目标的检测性能。

2) 针对海洋底栖生物存在堆叠、遮挡情况的问题, 设计了上下文信息自适应融合模块 (CGF, context guide fusion), 替换特征融合网络中的 Concat 操作, 增强不同层特征图之间的信息交互, 并根据特征中不同元素的重要性进行加权, 从而提升了网络对遮挡、堆叠目标的检测能力。

3) 针对训练过程中正负样本不均衡的问题, 引入自适应标签分配 (ATSS, adaptive training sample selection), 通过将目标的统计特征 (如方差和均值) 应用于阈值设定, 可以根据每种类别的统计特性自适应地调整不同种类样本的阈值, 有助于更精准地划分正负样本, 提高模型对海洋底栖生物的检测精度。

为提高模型的检测性能, C2f-dcs 模块、CGF 模块和 ATSS 方法在本模型中协同工作。在特征提取部分, C2f-dcs 模块通过空洞卷积和注意力机制扩大模型的感受野并优化了特征提取, 为后续的特征融合提供了更丰富的信息。CGF 模块则通过自适应特征融合有效结合了上下文信息, 进一步提升了特征的表达能力, 改善了模型的定位能力, 为海洋底栖生物中的小目标检测提供更精确的边界信息。而 ATSS 方法通过自适应地调整正负样本的分配, 使得模型在处理不同类别时更加精准。这些模块通过信息传递和相互作用, 确保了模型对海洋底栖生物的高效识别, 显著提高了模型的检测精度。

## 1 YOLOv8 算法

YOLOv8<sup>[15]</sup> 是一种最前沿的目标检测模型, 它是一种单阶段目标检测算法。使用 CSPDarkNet 作为特征提取部分 (Backbone), PANet<sup>[16]</sup> 作为特征融合部分 (Neck), 使用解耦头 (Decoupled-Head) 作为预测部分。基于精度和参数量考虑, 本文选择使用模型精度和参数量均衡的 YOLOv8-M 作为基线模型。YOLOv8 的网络结构图如图 1 所示。

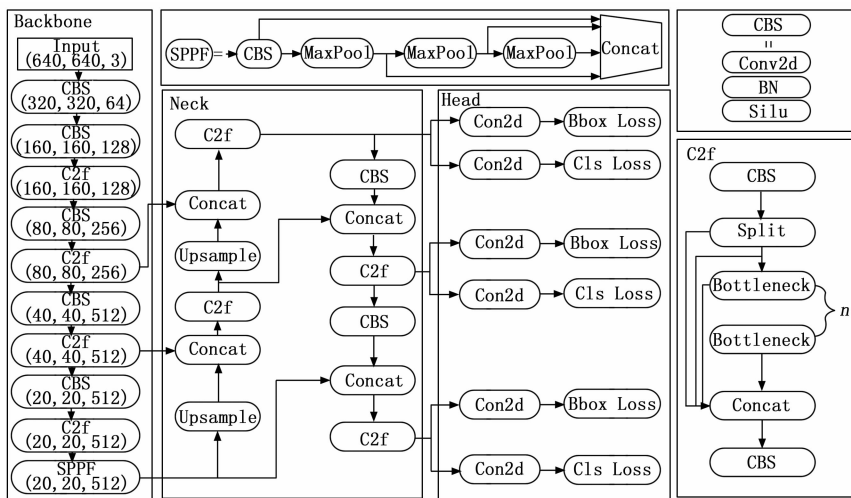


图 1 YOLOv8 网络结构图

YOLOv8 模型网络主要由以下部分组成。

1) 特征提取网络 (Backbone):

特征提取网络使用修改后的 CSPDarkNet53, 主要由 CBS 模块、C2f 模块和 SPPF (Spatial Pyramid Pooling-Fast) 模块组成。将原骨干网络中的 CSP (Cross Stage Partial) 模块替换为 C2f 模块, C2f 模块的结构如图 1 所示 ( $n$  表示残差块数量)。C2f 模块参考了 YOLOv7 ELAN<sup>[17]</sup> 结构的思想采用梯度分流连接, 在保持轻量化的同时丰富了特征提取网络的信息流。CBS 模块由卷积层、BN 层和 SiLU<sup>[18]</sup> 激活函数组成。SPPF 模块由两个 CBS 块和 3 个内核大小为 5 串联的池化层组成。该块的特征图融合第一个 CBS 模块和 3 个池化层的输出, 然后放入下一个 CBS 块中。

2) 特征融合网络 (Neck):

特征融合网络采用路径聚合网络 (PANet), 其中包含特征金字塔网络<sup>[19]</sup> (FPN, feature pyramid network)。FPN 通过多次上采样、下采样和横向连接, 实现在深层与浅层特征图之间的融合。而 PANet 则通过自下而上的路径进行特征融合, 利用元素级加法操作将底层与顶层特征结合, 从而增强网络对位置信息的学习能力, 提升模型在多尺度目标检测中的表现。

3) 检测部分 (Head):

检测部分在特征融合网络之后, 使用 3 个检测头分别检测不同尺度的目标, YOLOv8 的解耦头结构使用两个独立的分支进行对象分类和预测边界框回归, 并针对这两类任务使用不同的损失

函数。对于分类任务, 使用二元交叉熵损失<sup>[20]</sup>, 对于预测边界框回归任务, 采用分布焦点损失 (DFL Loss<sup>[21]</sup>) 和 CIoU。

2 改进的 YOLOv8-M 海洋底栖生物检测算法

针对海洋底栖生物检测任务中存在的不足, 为解决这些问题, 本文提出一种基于 YOLOv8-M 的海洋底栖生物检测算法 IFL-YOLO, 本文所提出的模型网络结构如图 2 所示。

2.1 空洞卷积特征融合模块 C2f-dcs

在海洋底栖生物检测任务中, 大多数样本都是小目标。且受复杂水下环境影响, 图像中的关键特征无法被有效提

取, 导致模型检测精度的下降。

在目标检测模型的计算过程中, 目标的信息通常存储在特征图的通道中, 模型通过这些通道来提取信息。然而, 小目标由于其尺寸较小, 信息较难获取, 在频繁的卷积操作中容易出现信息丢失的问题。这也表明, 针对小目标的检测, 在深度学习中, 关键不在于增加通道数, 而在于如何更有效地捕捉和保留小目标的特征信息以提高检测性能。

为解决上述问题, 本文设计了空洞卷积特征融合模

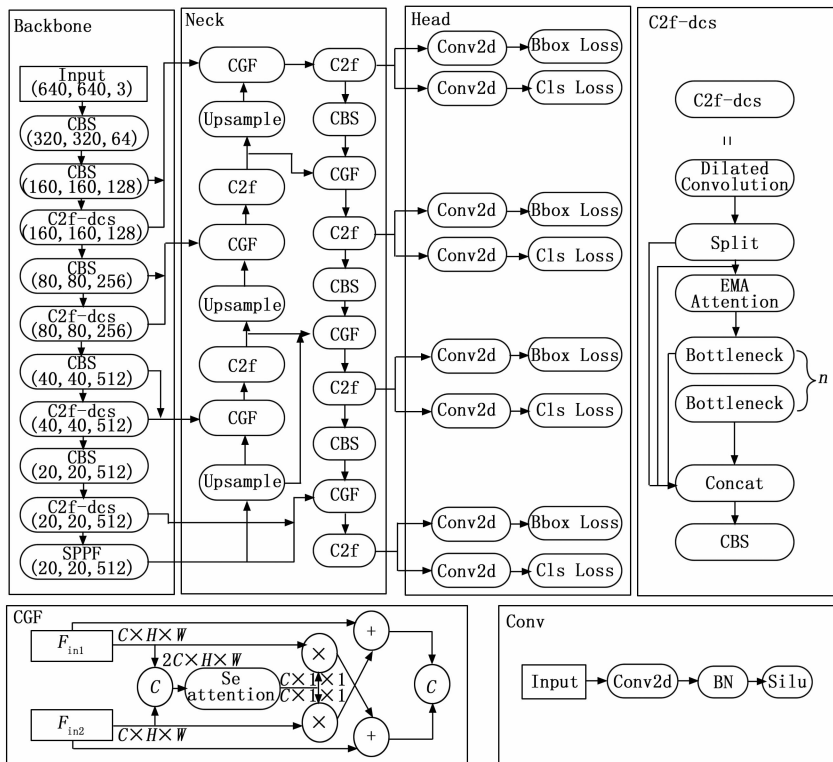


图 2 IFL-YOLO 网络结构

块 C2f-dcs 以替代 Backbone 中的 C2f 模块进行特征提取，以扩大模型感受野来改善模型的特征提取能力。

C2f-dcs 模块输入特征图  $X \in R^{C \times H \times W}$ ，其中  $C$ 、 $H$ 、 $W$  分别表示输入通道数、特征图的高度和宽度。C2f-dcs 主要由空洞卷积和 EMA<sup>[22]</sup> (Efficient Multi-Scale Attention) 注意力机制组成。C2f-dcs 结构如图 3 所示。

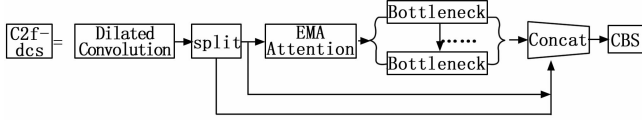


图 3 C2f-dcs 模块结构图

该模块通过将 C2f 模块中的第一个  $3 \times 3$  普通卷积替换为 dilation 为 2 的空洞卷积，增加了模型的感受野，保存了各区域中的特征信息。同时应用 EMA 注意力机制通过突出关键特征的表示，使得网络能够更加关注图像中与目标检测任务密切相关的部分，如目标的形状和大小等特征，同时有效抑制背景噪声和无关信息的干扰，从而减少漏检和误检的发生。

EMA 在水平方向和垂直方向上分别执行一维全局平均池化，与传统的通道注意力方法不同，后者通过二维全局池化将输入转换为单一的特征向量。EMA 采用跨空间学习策略，将并行子网络的学习注意力进行融合，并通过矩阵点积运算捕捉像素级的配对关系，同时强调全局上下文信息。其结构如图 4 所示。

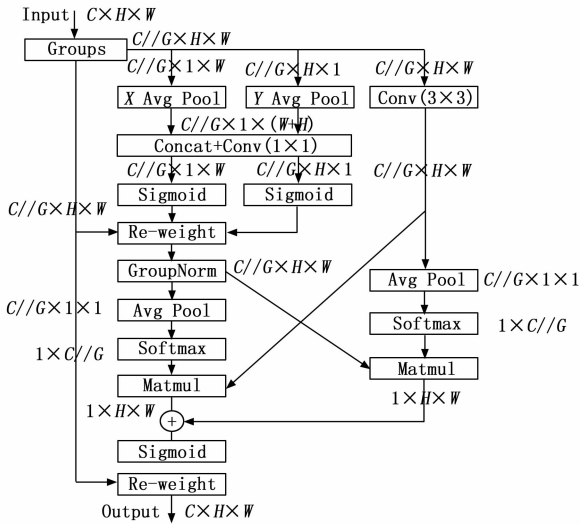


图 4 EMA 注意力结构图

EMA 注意力机制的实现主要分为特征分组、并行子网和跨空间学习阶段。首先输入特征图，将特征图按照不同通道分为  $G$  个子特征，单独学习不同通道中的每个区域。EMA 通过  $X$  方向平均池化、 $Y$  方向平均池化和一个  $3 \times 3$  的标准卷积 3 个分支来提取特征信息。计算公式如公式 (1) 和 (2) 所示：

$$z_c^w = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w) \quad (1)$$

$$z_c^h = \frac{1}{W} \sum_{0 \leq j < W} x_c(h, i) \quad (2)$$

然后，模型将  $X$  方向和  $Y$  方向的输出特征图进行特征融合，输入  $1 \times 1$  的卷积核中进行不同通道间的信息交互，以获得具有全局信息的特征，计算公式如公式 (3) 所示：

$$f = P[(z^h, z^w)] \quad (3)$$

然后模型将特征信息分解为两个方向的特征张量，在不同方向上进行权重计算，计算公式如公式 (4) 和 (5) 所示：

$$g^h = \sigma(f^h) \quad (4)$$

$$g^w = \sigma(f^w) \quad (5)$$

式中， $\sigma$  表示 sigmoid 激活函数， $g^w$  和  $g^h$  是中间权重。

最后 EMA 应用全局池化的操作来处理两个方向的特征图信息，进行特征融合。计算公式如公式 (6) 所示：

$$z_c = \frac{1}{H \times W} \sum_{j=1}^w \sum_{i=1}^H x_c(i, j) \quad (6)$$

式中， $z_c$  表示第  $c$  个通道的输出， $x_c(i, j)$  表示第  $c$  个通道中，位于高度为  $i$ 、宽度为  $j$  位置的输入特征。

下一步特征图输入到 Softmax 激活函数，同时并行分支的输出与 Softmax 输出的注意力权重相乘，融合不同尺度的特征信息。最终，将不同分支输出的注意力权重融合，输入到 Sigmoid 激活函数，得到最终的注意力权重。

通过空洞卷积与 EMA 注意力机制的结合，C2f-dcs 模块具有更强的特征信息提取能力，同时能将网络聚焦于所检测目标。因此，C2f-dcs 可以有效的提高小目标的检测精度。

## 2.2 上下文信息自适应融合模块 CGF

海洋底栖生物通常成群分布，这种生活习惯导致图像中常出现堆叠和遮挡现象。在这种情况下，检测这些堆叠或遮挡的生物需要更丰富的空间信息和上下文信息。YOLOv8 的主干网络采用的是 CNN 结构，通过多次下采样来提取特征。然而，每经过一次下采样，网络的分辨率会逐渐压缩，从而丢失不同分辨率特征图中的信息。CNN 网络的顶层特征图具有较小的分辨率和更强的语义信息，而底层特征图则保留了更大的分辨率和更丰富的细节信息。然而，随着网络层数的增加，底层特征图中的位置信息逐渐丢失，导致难以有效识别图像中的目标。

YOLOv8 采用了 PANet 双层特征金字塔结构，通过双流信息交互，获得了不同尺度的特征图，显著提升了多尺度目标的检测精度。然而，在现有的 FPN 结构中，主要通过人为固定几个尺度特征图来表示不同尺度

的物体 (例如 YOLO 系列中通常固定为大、中、小 3 个尺度), 这就无法确保所有不同尺度的物体信息都能充分覆盖和有效表示。

为了解决特征图缺乏上下文信息的问题, 本节从增强不同特征层信息交互的角度出发, 提出了上下文信息自适应融合模块 (CGF) 替换特征融合部分的 Concat 模块。其结构如图 5 所示。CGF 模块在多尺度特征融合过程中, 既能够引导上下文信息的整合, 又能自适应地进行调整。首先, 模块通过 SE 注意力机制, 在特征融合过程中捕捉并利用关键的上下文信息, 从而提升特征表示的有效性, 并有效引导模型学习与目标检测相关的重要信息, 进而提高检测精度。接着, 模块对特征进行加权重组操作, 增强重要特征的表现, 并抑制不相关特征, 从而提升特征图的判别能力。

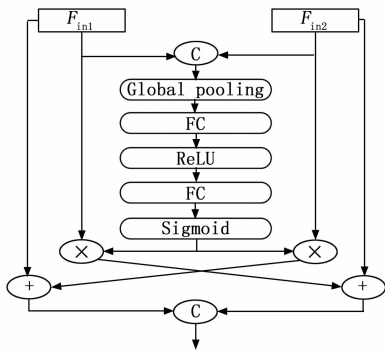


图 5 CGF 模块结构图

CGF 模块通过增强不同层特征图之间的信息交互, 并根据特征中不同元素的重要性进行加权, 从而提升了网络对目标的检测能力。该模块有效提高了海洋底栖生物的检测精度, 进一步增强了信息交互能力。

此外, 网络浅层的特征信息包含更多有助于小目标检测时所需的特征信息, 而 YOLOv8 的特征融合网络使用的更多是深层特征信息进行特征融合。针对这

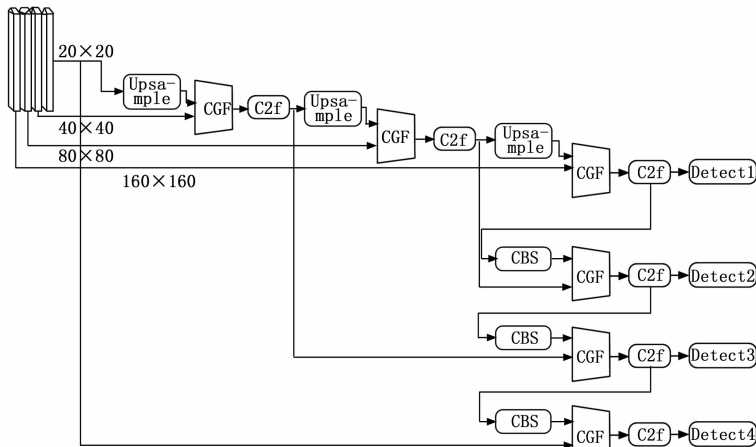


图 6 改进后的特征融合网络结构

一特点, 本节提出了新的小目标检测头, 如图 6 所示 detect1。小目标检测头利用了浅层特征, 富含更多浅层语义信息和位置信息, 提高了对小尺寸目标的检测精度。

### 2.3 自适应标签分配 ATSS

在 YOLOv8 中, 作者延续了 TOOD 中的 TAL 软标签分配策略 (TAL, task-aligned one-stage object detection)。TAL 策略旨在解决分类任务和回归任务在目标检测中关注重点不同的问题。其核心思想是将对齐的锚点视为正样本, 而对未对齐的锚点进行抑制。该策略的对齐度量策略如公式所示:

$$t = s^\alpha \times \mu^\beta \quad (7)$$

式中,  $s$  是类别预测分数,  $\mu$  是回归预测分数,  $\alpha$  和  $\beta$  是超参数。TAL 策略将归一化后的对齐度量作为正负样本的分类权重, 有效缓解了分类头和回归头关注点不一致的问题。但本文所采用的海洋底栖生物数据集中, 海参、海胆之类的生物形状、大小各不相同, 并且相比海星之类的生物, 海参的尺寸也更小。使用传统的正负样本分配方案通常为同一场景下不同尺度尺度的目标分配相同数量的正样本, 这样会造成小目标存在很多低质量的正样本, 或者正负样本不平衡的问题。导致对海洋底栖生物进行正负样本判定时, 训练过程中存在正负样本不平衡的问题, 影响训练效果。

针对上述问题, 本文提出自适应标签分配 ATSS。通过将目标的统计特征 (如方差和均值) 应用于阈值设定, 可以根据每种类别的统计特性自适应地调整不同种类样本的阈值。这种方法有助于更精准地划分正负样本, 提高分类的准确性。

ATSS 的操作主要包括以下几个步骤: 首先, 对于每个图像的真实框  $g$ , 在每个金字塔层级上, 根据  $L2$  距离选择距离  $g$  中心最近的  $K$  个锚框。假设特征金字塔有  $L$  层, 则每个真实框将有  $K \times L$  个候选正样本。具体操作如式 (8) ~ (11) 所示, 其中  $g$  代表一个真实框,  $G$  代表图像中的所有真实框,  $S_i$  代表从每层特征金字塔中根据  $L2$  距离选取的距离中心最近的  $K$  个锚框:

$$\text{for each ground truth } g \in G \quad (8)$$

$$\text{for each level } i \in [1, L] \quad (9)$$

$$S_i = \text{select } k \text{ anchors} \quad (10)$$

$$C_g = C_g \cup S_i \quad (11)$$

下一步, 计算候选框与真实框  $g$  之间的  $IoU$ , 并将其记为  $D_g$ , 同时计算均值  $m_g$  和方差  $v_g$ , 据这些统计数据, 通过简单的加法操作来计算  $IoU$  阈值。最终, 选择  $IoU$  大于或等于阈值  $t_g$  的候选框作为正样本。具体操作如式 (12) ~ (15) 所示:

$$D_g = IoU(C_g, g) \quad (12)$$

$$m_g = Mean(D_g) \quad (13)$$

$$v_g = Std(D_g) \quad (14)$$

$$t_g = m_g + v_g \quad (15)$$

此外, 如果一个 anchor box 被分配给多个 ground-truth box, 则 IoU 最高的一个正样本将被选择, 其余判定为负样本。ATSS 增强了模型对不同物体的正负样本判别能力, 为海洋底栖生物的检测提供了更高效的正样本选择方式, 从而显著提升了检测性能。在后续实验部分, 也证明了此模块的有效性。

### 3 实验结果

#### 3.1 实验设置

##### 1) 实验环境:

本文的实验平台和参数设置如下: 本节中的实验用到的实验环境设置如下: 服务器运行设备为 Linux Ubuntu 18.04。CPU 型号为 12th Gen Intel (R) Core (TM) i7-12700 KF。运行内存为 64 GB RAM。GPU 配置为 NVIDIA Geforce RTX3090-24 G, 使用 Pytorch 1.10 和 CUDA 11.1。

##### 2) 实验参数设置:

本文的训练和测试输入图像大小均为  $512 \times 512$ 。为了保证实验的有效性, 本文进行了 300 epochs 的训练, batch size 大小为 8, 这一大小能够平衡计算效率与内存使用, 且能够确保模型在每次训练中获取足够的样本信息。初始学习率设为 0.01, 学习率衰减率设为 0.0005。当学习率设置为 0.01 时, 模型能够较快收敛并且取得较好的精度, 但在进一步增大学习率后, 模型精度开始下降, 表明过高的学习率导致了训练的不稳定。对于消融实验, 以及对比试验中的部分算法, 都是在 Pytorch 框架中实现的。

##### 3) 实验数据集:

本文使用海洋底栖生物数据集 DUO<sup>[23]</sup> (detecting underwater objects) 数据集进行了训练和评估。DUO 数据集包含了海参、海胆、扇贝和海星 4 个类别的底栖生物, 共计 6 671 张图片用于训练, 1 111 张图片用于测试。该数据集涵盖了来自不同海洋环境的底栖生物图像, 确保了模型在各种海洋环境下的适应性和泛化能力。为了进一步验证所提方法的有效性, 本文还使用了 URPC2020 数据集进行实验。该数据集包含 4 929 张训练集图片和 985 张测试集图片, 同样具有代表性, 包含海参、海胆、扇贝和海星 4 个类别。

在样本选择标准上, DUO 和 URPC2020 数据集都被精心挑选, 确保包含的图片具备不同海洋环境条件的多样性。例如, 数据集中包括了不同的光照条件、海底复杂度、物体大小和位置等因素, 以保证训练模

型能够应对实际应用中可能遇到的多变环境。此外, 考虑到模型对小目标的敏感性, 数据集中包括了较小的底栖生物 (如海参和海胆), 以评估模型在小物体检测中的性能。

对于样本量的确定, 本研究使用的 6 671 张训练图像和 1 111 张测试图像已经满足了现有 YOLOv8-M 模型的训练需求, 并能有效评估模型性能。实验中使用的样本量考虑了数据集类别的平衡性, 确保了不同类别的样本数量足够支持训练过程中的有效学习。在数据集选择上, 既考虑了数量的充足性, 也考虑了类别和环境的多样性, 确保模型能够在不同海洋环境和物种上保持较高的准确性和鲁棒性。

##### 4) 评估指标:

本文使用了 DUO 和 URPC 两个不同的海洋底栖生物数据集进行验证, 具体结果分别如表 1 所示。统计了各个模型在不同数据集上的  $mAP_{0.5:0.95}$  精度表现, 以及模型的参数量和计算量。

#### 3.2 对比实验

为了验证算法的普适性。本节选用了不同类型的算法模型, 在海洋底栖生物数据集上进行对比。这些算法包括二阶段算法 Faster-RCNN, 一阶段基于锚点的算法 YOLOv5-S、YOLOv5-M<sup>[24]</sup>、YOLOv7<sup>[25]</sup> 和 APAN<sup>[26]</sup>。一阶段无锚点检测算法 YOLOv6-S<sup>[27]</sup>、YOLOv8-S、YOLOv8-M。如表 1 所示。

表 1 对比实验结果

方法	图片尺寸	$mAP_{0.5:0.95}/\%$		FLOPs /G	Param /M
		DUO	URPC2020		
Faster R-CNN	640	62.3	44.9	268.5	41.4
YOLOv5-S	640	64.5	45.2	48	20.9
YOLOv5-M	640	66.9	48.1	48	20.9
YOLOv7	640	67.2	48.5	103.5	36.5
APAN	640	57.7	38.4	33	70
YOLOv6-S	640	64.1	48.1	44.1	17.2
YOLOv8-S	640	66.8	49.1	28.8	11.2
YOLOv8-M	640	70.3	50.5	78.7	25.8
Ours	640	<b>73.4</b>	<b>53.1</b>	107.5	29.9

表 1 表明, 通过对不同数据集和指标的对比分析, 可以明显看出, IFL-YOLO 在海洋底栖生物数据集上显著优于其他目标检测算法的性能。即使在不同的海洋底栖生物数据集上, IFL-YOLO 也表现出明显的精度提升。尽管模型的改进导致了计算量和参数数量的增加, 但这种增长在可接受范围内。相较于检测精度的提升, 参数量和计算量的增加是可以接受的。

为了更直观地展示 IFL-YOLO 的检测提升效果, 我们展示了 IFL-YOLO 与部分算法在 DUO 数据集上的推理结果, 如图 7 所示。可以明显看出, IFL-YOLO 在

密集目标和小目标的检测上, 相比于基线模型具有更高的准确性。因此, 从实际效果来看, IFL-YOLO 相较于同类算法, 更适合用于海洋底栖生物图像的目标检测任务。

本节通过选择包含更多小尺寸、密集和重叠目标的数据集来比较各模型的检测效果。图 7 展示了 YOLOv5、YOLOv7、YOLOv8-M 和 IFL-YOLO 的检测结果对比。与其他算法相比, IFL-YOLO 在检测小尺寸的密集重叠目标时表现更为出色。在背景复杂的水下环境中 (如图 7 (a) 所示), 水下的岩石等特征与海胆的外形相似, 其他算法存在不同程度的漏检, 而 IFL-YOLO 能够更好地区分背景和目标之间的特征。在远距离的小尺寸目标检测中, YOLOv5 (d-1) 和 YOLOv7 (d-2) 对海胆的检测效果较差, 而 IFL-YOLO 则能够更好地提取小型目标的信息。对于石缝中的海胆, YOLOv7 (b-2) 和 YOLOv8-M (b-3) 未能检测到, 而 IFL-YOLO 能够更有效地提取与背景相似的目标信息。此外, IFL-YOLO 能够识别堆叠的海星和海胆 (b-4), 而 YOLOv5 和 YOLOv8-M 未能检测到。在图 7 (b) 左下角, YOLOv5 和 YOLOv8-M 将重叠的海胆误识别为一个目标, 而 IFL-YOLO 能够准确识别海胆的边缘, 实现对两个海胆的精确检测。与其他算法相比, IFL-YOLO 具有更低的漏检率和误检率, 但仍然存在一些错误的正样本和负样本。例如, 在图 7 (a-4) 左上角, IFL-YOLO 产生了误检, 而其他算法没有出现此问题。然而, 与其他算法相比 (如图 (a-2) 和 (a-3) 所示), IFL-YOLO 在检测中上部分海胆时, 成功检测到更多的遗漏对象。因此, 少量的误检和漏检是可以接受的。

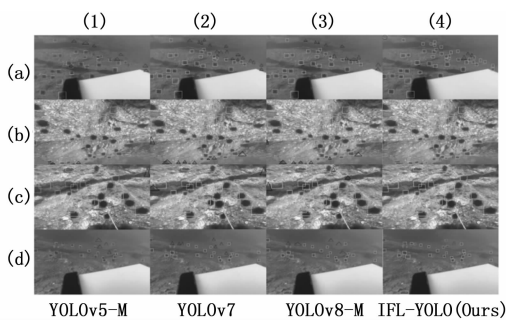


图 7 IFL-YOLO 与其他算法在 DUO 数据集上的检测结果对比

### 3.3 消融实验

为了进一步验证本节提出的改进的有效性, 本文在 DUO 数据集上进行了消融实验。对本节提出的 3 个模块分别进行消融实验, 以验证改进和参数的有效性。

表 2 中的实验结果表明, 通过空洞卷积获取不同尺度的特征图并进行融合, 可以有效提升检测器对小目标

的检测性能, 表现为模型的  $mAP$  得到了提升。针对空洞卷积的应用数量, 进行了实验比较。根据表 2 中的数据, 虽然两种配置在检测精度上的表现相似, 但使用单个空洞卷积在模型的参数量和计算复杂度上具有明显优势。考虑到在保持类似精度的同时显著减少计算资源消耗, 我们认为在本研究中采用单个空洞卷积是一个更为优选的方案。这一发现不仅减轻了模型的计算负担, 还提高了其实用性, 特别是在计算资源受限的场景下。实验结果确认了 C2f-dcs 模块有效增强了小目标的检测效果, 并提升了对所有尺度目标的检测能力。图 8 也验证了该模块在减少小目标漏检方面的改善。

表 2 不同空洞卷积块的效果对比

数量	$mAP/\%$	Param/M	FLOPs/G
1	71.0	26.8	85.8
2	71.2	27.1	87.2

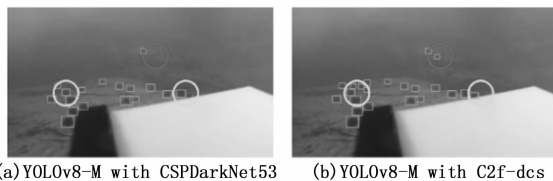


图 8 两种主干网络检测效果的对比

表 3 展示了改进后的 Neck 在包含大量小目标的图像数据集上的实验结果。通过使用 PANet 和改进后的 Neck 进行检测效果对比, 如图 9 所示, 改进后的 Neck 能够更好地融合上下文信息, 同时引入小目标检测头后, 对小尺寸目标的检测效果显著提升, 从而提高了对底栖生物中小尺寸目标的检测精度。

表 3 两种特征融合网络检测效果的对比

Neck	$mAP/\%$	Param/M	FLOPs/G
PANet	70.3	25.8	78.7
Ours	71.1	26.6	79.4

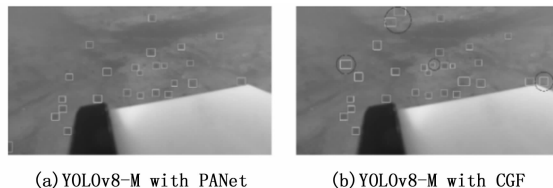


图 9 两种特征融合网络检测效果的对比

表 4 展示了 Task-Aligned Assigner 和 ATSS 的实验对比, 选取了图像尺寸差异较大的数据集进行效果评估。如图 10 所示, 改进后的标签分配方式更充分地考虑了物体形状差异的影响, 从而提升了模型对海洋底栖生物的检测效果。

表 4 两种标签分配方式检测效果的对比

标签分配	$mAP_{0.5:0.95}/\%$
Task-Aligned Assigner	70.3
ATSS	71

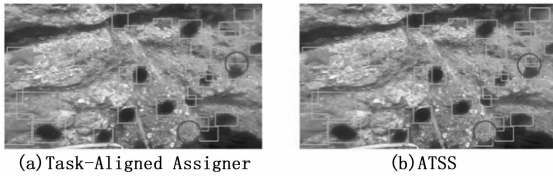


图 10 两种标签分配方式检测效果的对比

此外，为更全面地评估各个模块的作用，对 3 个模块进行交叉消融实验，以验证不同模块的有效性。具体结果如表 5 所示。可见本文提出的 3 个模块提升模型检测精度的效果显著，且可自由组合，具有鲁棒性。

表 5 不同模块的消融实验

C2f-dcs	CGF	ATSS	$mAP/\%$
✓	✗	✓	72.4
✓	✓	✗	72.5
✗	✓	✓	72.2

最后，为了验证提出方法的有效性，进行了消融实验，具体结果如表 6 所示。与基础模型 YOLOv8-M 的结构相比，提出的 IFL-YOLO 算法将  $mAP_{0.5:0.95}$  结果提高了 3.1%。

表 6 IFL-YOLO 消融实验的结果

方法	$mAP_{0.5:0.95}/\%$	Param/M	FLOPS/G
YOLOv8-M	70.3	25.8	78.7
+C2f-dcs	71.6(+1.3)	27.6	92.5
+CGF	72.5(+0.9)	28.5	101.2
+ATSS	73.4(+0.9)	30.2	107.5

实验证明，相较于基线 YOLOv8-M，IFL-YOLO 在海洋底栖生物图像的目标检测效果上有显著提升。同时，如图 11 所示，IFL-YOLO 在训练过程中的准确度曲线相比 YOLOv8-M 表现出稳定的提升。这些结果表明，IFL-YOLO 能够有效解决海洋底栖生物图像中小物体难以检测和物体密集导致漏检的问题。

综合来看，模型检测精度为 73.4%，虽然相对其他算法来说已经取得了不错的成果，但由于海洋底栖生物数据集中存在图像质量低、小目标检测困难等问题，相对于一些类似研究中 80% 以上的检测精度，依然存在提升空间。通过进一步优化数据集、增加数据量、调整超参数、改进模型架构等方法，有望进一步提升检测精度。同时，增加对复杂背景、遮挡物体的适应性以及加强小目标检测能力，是未来研究的重要方向。

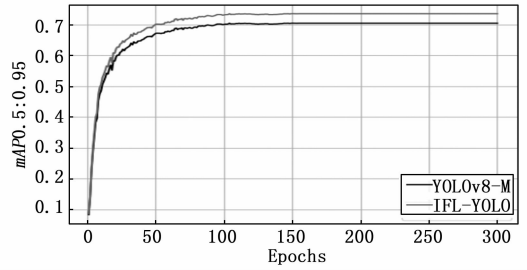


图 11 YOLOv8-M 模型与 IFL-YOLO 在 DUO 数据集上的精度曲线

### 4 结束语

本文提出 IFL-YOLO 用于海洋底栖生物检测，提高了 YOLOv8 的特征提取和边界信息定位能力。针对真实水下场景中出现的物体尺寸小、密集、遮挡重叠等问题，主要从 3 个方面进行了创新实践。首先，提出 C2f-dcs 模块来替换原本的 C2f 模块。它增大了模型的感受野，增强对小目标特征的提取能力，以实现检测小尺寸物体。其次，特征融合部分提出 CGF 和小目标检测头。使得模型能够增强重要特征，同时抑制不重要特征，提升特征图的判别能力，有利于堆叠、遮挡情况下的目标检测。最后提出使用 ATSS 来代替 Task-Aligned Assigner，以提高对原始形状不佳目标的检测能力。ATSS 具有根据样本统计特征自适应划分阈值的能力，增强了对物体的分类和定位能力。分别在 DUO、UR-PC2020 数据集上将 IFL-YOLO 与最先进的目标检测算法进行比较。结果表明 IFL-YOLO 实现了最好的检测效果，同时具有较低的模型复杂度。DUO 数据集上的  $mAP_{0.5:0.95}$  提升到 73.4%，与 YOLOv8-M 相比， $mAP_{0.5:0.95}$  提高了 3.1%。因此 IFL-YOLO 融合了深度学习中最先进的方法和思想，在海洋底栖生物检测任务中表现出色，检测效果准确。IFL-YOLO 提出的模块可以用于 YOLO 系列的其他模型。

### 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 [C] //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] // Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779 - 788.
- [3] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: single shot multibox detector [C] //Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11 - 14, 2016, Proceedings, Part I

14. Springer International Publishing, 2016: 21–37.
- [4] ZHOU X, WANG D, KRÄHENBÜHL P. Objects as points [J]. Arxiv Preprint Arxiv: 1904.07850, 2019.
- [5] TIAN Z, CHU X, WANG X, et al. Fully convolutional one-stage 3d object detection on lidar range images [J]. Advances In Neural Information Processing Systems, 2022, 35: 34899–34911.
- [6] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale [J]. Arxiv Preprint Arxiv: 2010.11929, 2020.
- [7] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers [C] //European Conference on Computer Vision. Cham: Springer International Publishing, 2020: 213–229.
- [8] LIU Z, LIN Y, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows [C] // Proceedings of The IEEE/CVF International Conference on Computer Vision, 2021: 10012–10022.
- [9] XU X, LIU Y, LYU L, et al. Mad-yolo: a quantitative detection algorithm for dense small-scale marine benthos [J]. Ecological Informatics, 2023, 75: 102022.
- [10] ZHANG M, XU S, SONG W, et al. Lightweight underwater object detection based on yolo v4 and multi-scale attentional feature fusion [J]. Remote Sensing, 2021, 13 (22): 4706.
- [11] AL MUKSIT A, HASAN F, EMON M F H B, et al. Yolo-fish: a robust fish detection model to detect fish in realistic underwater environment [J]. Ecological Informatics, 2022, 72: 101847.
- [12] SALMAN A, SIDDIQUI S A, SHAFAIT F, et al. Automatic fish detection in underwater videos by a deep neural network-based hybrid motion learning system [J]. ICES Journal Of Marine Science, 2020, 77 (4): 1295–1307.
- [13] ZHANG J, ZHOU K, et al. An improved yolov5-based underwater object-detection framework [J]. Sensors, 2023, 23 (7): 3693.
- [14] XUAN K, DENG L, XIAO Y, et al. So-yolov5: small object recognition algorithm for sea cucumber in complex seabed environment [J]. Fisheries Research, 2023, 264: 106710.
- [15] TERVEN J, CORDOVA-ESPARZA D M, ROMERO-GONZÁLEZ J A. A comprehensive review of yolo architectures in computer vision: from yolov1 to yolov8 and yolo-nas [J]. Machine Learning And Knowledge Extraction, 2023, 5 (4): 1680–1716.
- [16] WANG K, LIEW J H, ZOU Y, et al. Panet: few-shot image semantic segmentation with prototype alignment [C] //Proceedings Of The IEEE/CVF International Conference on Computer Vision, 2019: 9197–9206.
- [17] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. Yolov7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C] //Proceedings of The IEEE/CVF Conference on Computer Vision And Pattern Recognition, 2023: 7464–7475.
- [18] ELFWING S, UCHIBE E, DOYA K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning [J]. Neural Networks, 2018, 107: 3–11.
- [19] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C] //Proceedings of The IEEE Conference on Computer Vision And Pattern Recognition, 2017: 2117–2125.
- [20] ZHANG Z, SABUNCU M. Generalized cross entropy loss for training deep neural networks with noisy labels [J]. Advances In Neural Information Processing Systems, 2018, 31.
- [21] LI X, WANG W, WU L, et al. Generalized focal loss: learning qualified and distributed bounding boxes for dense object detection [J]. Advances In Neural Information Processing Systems, 2020, 33: 21002–21012.
- [22] OUYANG D, HE S, ZHANG G, et al. Efficient multi-scale attention module with cross-spatial learning [C] // ICASSP 2023–2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2023: 1–5.
- [23] LIU C, LI H, WANG S, et al. A dataset and benchmark of underwater object detection for robot picking [C] //2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), IEEE, 2021: 1–6.
- [24] KIM J H, KIM N, PARK Y W, et al. Object detection and classification based on yolo-v5 with improved maritime dataset [J]. Journal of Marine Science And Engineering, 2022, 10 (3): 377.
- [25] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. Yolov7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C] //Proceedings of The IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464–7475.
- [26] YU H, LI X, FENG Y, et al. Multiple attentional path aggregation network for marine object detection [J]. Applied Intelligence, 2023, 53 (2): 2434–2451.
- [27] LI C, LI L, JIANG H, et al. Yolov6: a single-stage object detection framework for industrial applications [J]. Arxiv Preprint Arxiv: 2209.02976, 2022.