基于改进 YOLOv5 的水下废弃物红外检测算法

高永奇，袁志祥
（安徽工业大学 计算机科学与技术学院，安徽 马鞍山 243032）

摘要：针对水下废弃物红外目标检测中出现的检测目标边界细节模糊、图像质量低和存在各种不规则形状或损坏的覆盖物等问题，本文提出了一种基于 YOLOv5 的改进目标检测方法（EFDCD-YOLO）。在主干网络中选择 InceptionNeXt 网络，以增强模型的表达能力和特征提取能力。其次，在特征融合层中通过加入 EffectiveSE 注意力机制，自适应地学习特征通道的重要性，并进行选择性加权。采用可变形卷积替代原模型中的 C3 模块，使模型能够更好地感知目标的形状和细节信息。此外，将 CARAFE 算子替代上采样模块，增强对细粒度特征的表现能力，避免信息丢失。在损失函数方面，采用 Focal-EIOU 损失函数，以提高模型对目标定位和边界框回归的准确性。最后，引入 DyHead 替换 YOLOv5 中的头部，通过动态感受野机制和多尺度的特征融合方式，提升模型的准确性。将改进后的 EFDCD-YOLO 模型应用于水下废弃物红外目标检测，相比于 YOLOv5 模型，改进后的模型在准确率（*P*）、召回率（*R*）和平均精度（mAP）方面分别提升了 21.4%、9.7%和 13.6%。实验结果表明，EFDCD-YOLO 能够有效地提升水下废弃物红外目标检测场景的性能，更好地满足水下废弃物红外目标检测的需求。

关键词：水下废弃物红外目标检测；注意力机制；可变形卷积；动态感受野
中图分类号：X52；TN219；TP391.41 **文献标志码：**A **文章编号：**1001-8891(2024)09-0994-12

Improved YOLOv5-based Underwater Infrared Garbage Detection Algorithm

GAO Yongqi，YUAN Zhixiang
(*School of Computer Science and Technology, Anhui University of Technology, Maanshan* 243032, *China*)

Abstract：An improved object detection method (YOLO with EffectiveSE, Focal-EIOU, DCNv2, CARAFE, and DyHead) is proposed based on YOLOv5 to address issues in underwater waste infrared target detection, such as blurred boundary details, low image quality, and the presence of various irregular or damaged coverings. The InceptionNeXt network is selected as the backbone network to enhance the model's expressive power and feature extraction capability. Additionally, the EffectiveSE attention mechanism is introduced in the feature fusion layer to adaptively learn the importance of feature channels and selectively weight them. Deformable convolutions are used to replace the C3 module in the original model, enabling it to better perceive the shapes and details of the targets. Moreover, the CARAFE operator is employed to replace the upsampling module, thereby enhancing the representation ability of the fine-grained features and avoiding information loss. In terms of the loss function, the Focal-EIOU loss function is adopted to improve the accuracy of the model in target localization and bounding box regression. Finally, DyHead is introduced to replace the head of YOLOv5, thereby enhancing the model accuracy via dynamic receptive field mechanisms and multiscale feature fusion. The improved EFDCD-YOLO model is applied to underwater waste infrared target detection and compared to the YOLOv5 model. The model achieves a 21.4% improvement in precision (*P*), 9.7% improvement in recall (R), and 13.6% improvement in mean average precision (mAP). The experimental results demonstrate that EFDCD-YOLO effectively enhances the detection performance in underwater waste infrared target detection scenarios and effectively meets the requirements of underwater infrared target detection.

收稿日期：2023-09-27；**修订日期：**2023-12-13.
作者简介：高永奇（1999-），男，硕士研究生，研究方向为计算机视觉、目标检测。E-mail：gyqgaoyongqi@foxmail.com。
通信作者：袁志祥（1973-），男，副教授，研究方向为机器学习、Petri 网理论。E-mail：zxyuan@ahut.edu.cn。
基金项目：国家自然科学基金（61806005）；安徽高校协同创新项目（GXXT-2020-012）；安徽省高校科学研究重点项目（KJ2021A0373）。

Key words：underwater waste infrared target detection, attention mechanism, deformable convolution, dynamic receptive field

0 引言

水下废弃物红外目标检测是一项具有挑战性的任务；由于受到水下环境中光照不均匀、水质模糊、散射等因素的影响，导致收集到的水下图像质量差；因此，水下废弃物的边界和细节往往难以清晰地展现，给目标检测增加了难度。此外，水下废弃物的多样性和复杂性也给模型在特殊环境下的适应能力和泛化能力带来了挑战；模型需要具备识别不同尺度、形态、材质废弃物的能力。同时也需要具备识别光照和水流等因素引起的物体形态变化的能力。

水下废弃物与周围环境之间存在着极高的相似性，目标与背景之间具有相似的颜色、纹理和形状，使得模型很难区分目标和背景。因此会出现误检和漏检的现象，从而降低目标检测的准确性。

为了克服这些挑战，研究人员采用了各种方法去解决上述问题。例如，Schechner 等人[1]介绍了一种基于偏振信息的图像去雾方法，以改善水下图像的清晰度。然而，该方法需要使用偏振摄影设备来获取偏振信息，因此极大地增加了硬件成本以及实施的复杂性。除此之外，还有一些方法尝试利用水下图像中的颜色信息进行目标检测，如 Bazeilles 等人[2]通过对水下图像中的颜色进行判别来检测水下物体。然而，由于水下环境和海洋生物的颜色相似性较高，这种方法在精确度方面存在一定的限制。也有研究人员通过增强水下图像的对比度和细节来改善目标检测的效果。例如，Li 等人[3]提出的一种基于最小信息损失和直方图分布先验的水下图像去雾方法。该方法对先验信息的准确性和适用性有一定要求，由于水下环境的复杂性和图像特征的多样性导致先验信息不准确，从而影响去雾结果的准确性。

随着深度学习在目标检测领域的发展，一阶段目标检测模型[4-10]和二阶段目标检测模型[11-15]逐渐被应用到水下目标检测当中。一阶段目标检测模型能够直接从输入图像中预测目标的位置和类别，无需进行额外的候选区域生成步骤。在这方面，陈鑫林[9]提出了一种自适应亮度与水下图像处理方法，对 PP-YOLO 模型进行改进，引入了自适应多尺度融合和优化损失函数，提高 YOLO 模型在水下目标检测中的鲁棒性。另外，袁红春等人[10]在 YOLOv5 检测部分嵌入 CBAM 注意力机制，并将卷积模块替换为 Ghost 卷积模块，以减少模型的计算量。除了一阶段目标检测模型以外，在水下目标检测中，二阶段模型通常具有较高的准确性，能够更好地处理目标与背景相似性较高的问题。在这方面，吕晓倩[14]首先通过生成对抗网络增强图像，修正图像的色彩和细节，并利用 STN 数据增广来弥补数据集数量的不足，最后采用 Faster-RCNN[11]网络进行目标的检测。此外王蓉蓉等人[15]利用 HRNet 替代 CenterNet[13]中的骨干网络，并引入瓶颈注意力模块，最后构建特征融合模块以丰富模型的语义信息和空间位置信息。虽然这些方法都对模型进行了改进，但在水下目标检测环境中，仍存在精度方面的问题尚未得到解决。

本文针对水下红外目标检测中存在的问题和目前算法存在的局限性，提出了一种基于 YOLOv5 的改进模型，该模型主要包含以下几个部分：

1）为提升模型在水下目标检测中的精度，利用 InceptionNeXt[16]网络替换 YOLOv5 中的主干网络，以增强网络的表达能力。

2）为解决红外水下目标的多样性和种类繁多的问题，在特征融合层引入 EffectiveSE[17]（Effective Squeeze-Excitation）注意力机制，通过学习通道之间的相互依赖关系，将通道注意力权重进行自适应的融合，以更好地捕捉特征之间的相关性。

3）为解决红外水下物体与背景具有极高相似度的问题，改用 Focal-EIOU[18]（Focal and Efficient IOU）损失函数，提升定位目标和边界框之间的准确性。

4）为增强红外水下目标检测模型的鲁棒性，采用 DCNv2[19]（Deformable Convolutional Network），通过可变形卷积操作，使得模型能够更好地感知目标的形状和细节信息。

5）为解决红外水下目标检测中图片像素低的问题，利用 CARAFE[20]（Content-Aware Reassembly of Features）替代原模型中上采样的工作，通过可学习的组装操作，对特征图进行更细粒度的重组，从而增强模型对细节信息的感知能力。

6）为应对红外水下目标检测中存在的误检和漏检的情况，引入 DyHead[21]（Dynamic Head）替换 Head 部分，根据目标的大小和形状自适应地调整感受野的大小。使得模型能够更好地适应不同尺度的目标，增强模型对目标的感知能力。

1 本文方法

1.1 提高特征表达能力的 InceptionNeXt 网络引入

为了增强网络的表达能力和特征提取能力，本文将 InceptionNeXt 作为主干网络。InceptionNeXt 是一种结合了 Inception 模块和 ConvNeXt 模块的网络架构，旨在更好地捕捉输入特征的多尺度信息和多层次特征表示。

通过式(1)～(8)组合不同尺度和层次的特征表示，InceptionNeXt 网络可以提高模型的表达能力和特征提取能力。同时由于模块内部的并行操作和模块之间的串联连接，InceptionNeXt 网络具有较少的参数量和计算量，从而在实际应用中具备较高的效率和实用性，其中 InceptionNeXt 的结构如图 1 所示。

$$a=F(\mathrm{BN}(C(x,k_1),\ \mathrm{gamma},\ \mathrm{beta})) \tag{1}$$

$$b_1=F(\mathrm{BN}(C(x,k_2),\ \mathrm{gamma},\ \mathrm{beta})) \tag{2}$$

$$b_2=F(\mathrm{BN}(C(b_1,\ k_3,\ \mathrm{stride}=s),\ \mathrm{gamma},\ \mathrm{beta})) \tag{3}$$

$$c_1=F(\mathrm{BN}(C(x,k_4),\ \mathrm{gamma},\ \mathrm{beta})) \tag{4}$$

$$c_2=F(\mathrm{BN}(C(c_1,k_5,\ \mathrm{pad}=p),\ \mathrm{gamma},\ \mathrm{beta})) \tag{5}$$

$$c_3=F(\mathrm{BN}(C(c_2,k_6,\ \mathrm{stride}=s),\ \mathrm{gamma},\ \mathrm{beta})) \tag{6}$$

$$d=F(\mathrm{BN}(C(x,\ k_7,\ \mathrm{stride}=s),\ \mathrm{gamma},\ \mathrm{beta})) \tag{7}$$

$$\mathrm{Inception}=\mathrm{Concatenate}(a,\ b_2,\ c_3,\ d) \tag{8}$$

式(1)～(8)中：a、b、c、d 分别是 Inception 层内不同分支的输出；k 代表卷积核大小；s 代表步长；n 代表输入的通道数。F 表示 ReLU 激活函数；C 表示 Convolution 卷积运算；BN 是批归一化操作。通过这样的多个 Inception 模块的堆叠，可以构建出一个深度、准确率较高的图像特征提取网络。

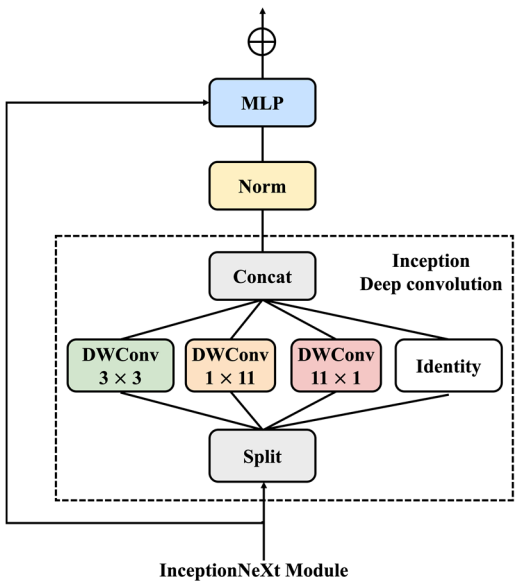


图 1 InceptionNeXt 网络结构

Fig.1 InceptionNeXt network structure

1.2 基于自适应学习通道权重的注意力机制改进

为了提高模型性能并增强表达能力，本文在特征融合层加入了 EffectiveSE 注意力机制。EffectiveSE 是一种轻量级的注意力机制，旨在增强神经网络的特征表达能力。

EffectiveSE 的过程可以用公式(9)～(11)表示：

首先，对于输入的特征图 X_{ijc}，进行全局平均池化得到每个通道的权重：

$$W_c=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}X_{ijc} \tag{9}$$

然后，通过一个全连接层将权重 W_c 映射到一个新的激活值 $f'(w_c;\theta)$（其中 θ 是可学习的参数），并对激活值进行 sigmoid 激活：

$$S_c=\sigma(f'(w_c;\theta)) \tag{10}$$

最后，将每个通道的权重 S_c 乘以原始的特征图 X，生成加权特征图 Y。

$$Y_{ijc}=S_cX_{ijc} \tag{11}$$

式中：H 和 W 分别是输入的特征图的高和宽；c 是特征图的通道数。通过这样的过程，EffectiveSE 的注意力机制可以增强有用的特征并减弱无用的特征，从而提高模型的性能和效率。EffectiveSE 的结构如图 2 所示。

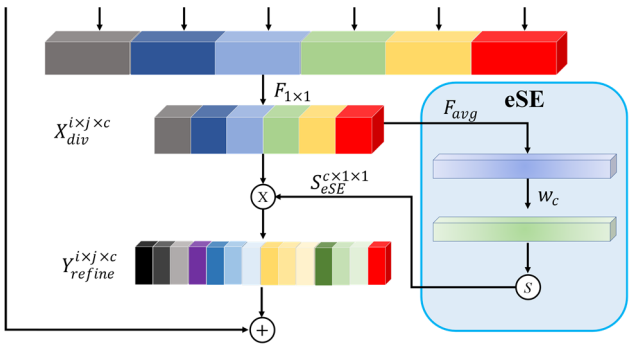


图 2 EffectiveSE 模块结构

Fig.2 EffectiveSE module structure

1.3 针对背景高相似度的损失函数改进

为了解决损失函数对难例样本的关注度不足的问题，将 Focal-EIOU 指标替代原有的 CIOU 模块。Focal-EIOU 是一种用于衡量目标检测模型性能的指标，它将焦点因子（focal factor）引入到 EIOU（enhanced intersection over union）指标中，从而提升难例样本的识别能力。

其中，Focal-EIOU 如公式(12)所示：

$$L_{\text{focal-EIOU}}=\mathrm{IoU}^{\gamma}L_{\mathrm{EIOU}} \tag{12}$$

式(12)中：γ 是一个用于控制曲线弧度的超参；L_{EIOU} 的公式如(13)所示：

$$L_{\mathrm{EIOU}}=L_{\mathrm{IOU}}+L_{\mathrm{dis}}+L_{\mathrm{asp}}=$$
$$1-\mathrm{IoU}+\frac{\rho^2\left(b,b^{\mathrm{gt}}\right)}{c^2}+\frac{\rho^2\left(w,w^{\mathrm{gt}}\right)}{c_w^2}+\frac{\rho^2\left(h,h^{\mathrm{gt}}\right)}{c_h^2} \tag{13}$$

式(13)中：c_w 和 c_h 分别是两个矩形的闭包的宽和高。从中可以看出，EIOU 将损失函数分成了 3 个部分：IOU 损失 L_{IOU}，距离损失 L_{dis}，边长损失 L_{asp}，其中，EIOU 如图 3 所示。总的来说，Focal-EIOU 指标能够更加准确地衡量目标检测的性能，同时加强对目标的位置、大小、姿态和形状的感知能力。

通过焦点因子的设计，使得 Focal-EIOU 指标能够增加对难例样本的关注度，使模型更加专注于难以检测的目标实例。提升模型在目标实例上的检测性能，并增强模型对小目标和遮挡目标的检测能力。

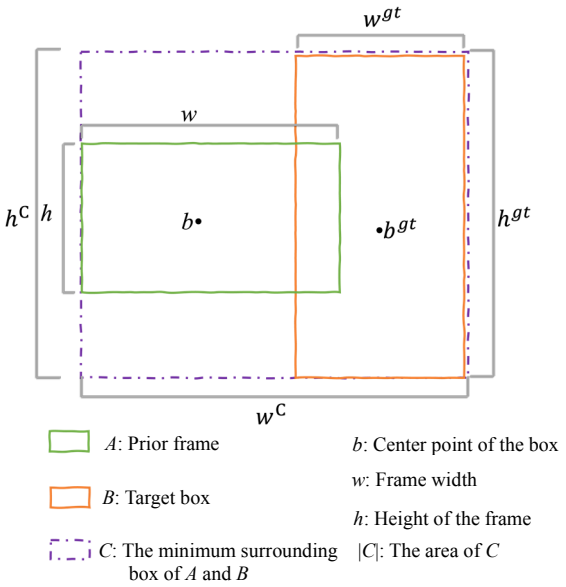


图 3 EIOU 模块结构

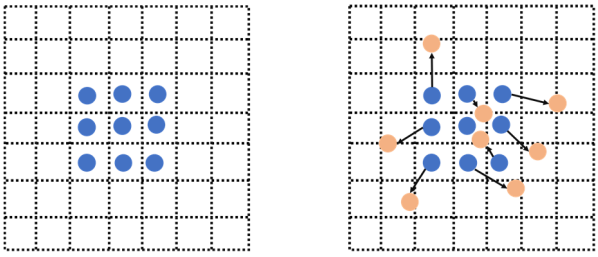
Fig.3 EIOU module structure

1.4 针对目标尺度变化的建模能力改进

为了提升模型的感知能力，文中将 DCNv2 替代原模型的 C3 模块。DCNv2 能够自适应地学习卷积核的形状和位置，进一步提高目标检测的准确性和性能。通过 DCNv2，模型能够在输入特征图的关键区域进行非整形采样，从而更好地适应图像中形变的物体；其中可变形卷积的示意图如图 4 所示。

DCNv2 中主要提出了两种新的卷积核学习机制，分别是 Deformable point-wise 卷积和 Deformable dilated 卷积。Deformable point-wise 卷积用于学习动态卷积核的位置。Deformable dilated 卷积则通过学习非整数采样点的空洞卷积计算，从而识别物体的边缘。

DCNv2 的公式和计算过程如下，其中对于 Deformable point-wise 卷积的过程，如公式(14)所示。



(a) 常规卷积 (b) 可变形卷积

(a) Conventional convolution (b) Deformable convolution

图 4 可变形卷积示意图

Fig.4 Deformable convolutional network sketch map

首先，假设输入特征图为 $X\in R^{C\times H\times W}$，正常的卷积核为 $K\in R^{C\times k_h\times k_w}$，即将卷积核放在输入特征图上进行卷积运算。

然后沿着通道维度把输入特征图 X 分成形状为 $C'\times k_h\times k_w$ 的 patch，设为 $X_{n,c'}\in R^{k_h\times k_w}$。那么第 c'个输出通道上的值 $Y_{n,c'}$为：

$$Y_{n,c'}=\sum_{i=1}^{k_h}\sum_{j=1}^{k_w}W_{c'}(i,j)\cdot X_{n,c'}(j,i) \tag{14}$$

式中：c'是输出通道的索引；$W_{c'}\in R^{k_h\times k_w}$ 是卷积核中各个位置的权重，即可学习的参数，用于捕获不同位置的特征。

为了学习每个位置的权重，DCNv2 增加了偏置 Δp。对于每个位置(i,j)，新的偏置$\Delta p_{i,j}$ 需要用于更新位置(i,j)的权重 $W_c(i,j)$。通过计算沿着每个位置的梯度来获得$\Delta p_{i,j}$，然后使用反向传播来更新 $W_c(i,j)$和$\Delta p_{i,j}$。

总体而言，DCNv2 利用动态卷积核的位置和非整形采样技术，实现了高精度和高性能的物体检测和分割，并且在不增加计算成本的情况下降低了大规模图像数据的标注和训练成本。这一步提升了模型对于不规则和变形物体的感知能力。

1.5 基于细粒度重组的 CARAFE

为了增强模型对细粒度特征的感知能力，本文引入了 CARAFE 替代原有的上采样操作。CARAFE 由两个步骤组成：第一步是根据每个目标位置的内容预测一个重组核，第二步是用预测的核对特征进行重组。给定一个尺寸为 $C\times H\times W$ 的特征图 X 和上样本比σ（假设σ为整数），CARAFE 将生成一个尺寸为 $C\times\sigma H\times\sigma W$ 的新特征图 X_0。

其中，CARAFE 的计算过程如公式(15)～(16)所示：

$$W_{l'}=\psi(N(x_l,k_{\mathrm{encoder}})) \tag{15}$$

$$x'_{l'}=\phi(N,(x_l,k_{up}),W_{l'}) \tag{16}$$

式(15)、(16)中：ψ表示核预测模块；l'表示相应的位置；$W_{l'}$表示在 l'位置方面的核；ϕ表示将 x_l 的与内核 $W_{l'}$重组为 $x'_{l'}$。

CARAFE 模块能够对特征进行细粒度的重组，从而提高模型对细节信息的感知能力。

1.6 基于自适应调整感受野的 DyHead

为了提升模型的精度和泛化能力，本文引入了 DyHead 替代原有模型的 Head 部分。

DyHead 的核心思想是将注意力机制应用于检测模型的头部；通过在检测头中引入注意力模块，DyHead 能够自适应地调节卷积核的大小和形状，以解决传统固定卷积核大小和形状所带来的物体尺寸和形状变化的问题。

DyHead 的注意力机制主要包括 3 个部分：Scale-aware 注意力机制、Spatial-aware 注意力机制、Task-aware 注意力机制。

Scale-aware 注意力：通过引入尺度感知注意力基于其语义重要性对不同尺度特征进行融合，过程如公式：

$$\pi_L(F)\cdot F=\sigma\left(f\left(\frac{1}{SC}\sum_{S,C}F\right)\right)\cdot F \tag{17}$$

式(17)中：$f(\cdot)$为线性函数，采用 1×1 卷积近似；$\sigma(x)$为 hard-sigmoid 激活函数。

Spatial-aware 注意力引入另一个空间位置感知注意力模块以聚焦不同空间位置的判别能力。首先采用形变卷积对注意力学习稀疏化，然后进行特征跨尺度集成，过程如公式：

$$\pi_L(F)\cdot F=\frac{1}{L}\sum_{l=1}^{L}\sum_{k=1}^{K}w_{l,k}\cdot F\left(l;p_k+\Delta_{k;c}\right)\cdot\Delta m_k \tag{18}$$

式(18)中：K 为稀疏采样位置数。

Task-aware 注意力机制为促进联合学习与目标表达能力的泛化性，我们设计了一种任务感知注意力。它可以动态开关特征通道以辅助不同任务，过程如公式：

$$\pi_C(F)\cdot F=\max(\alpha^1(F)\cdot F_c+\beta^1(F),\ \alpha^2(F)\cdot F_c+\beta^2(F)) \tag{19}$$

式(19)中：$[\alpha^1,\alpha^2,\beta^1,\beta^2]^{\mathrm{T}}=\theta(\cdot)$为超参数，用于控制激活阈值，而$\theta(\cdot)$为 ReLU 函数。

DyHead 通过组合不同的卷积核，来处理不同尺度的物体。为了保持计算效率，DyHead 还增加了自适应的卷积核的数量选择机制，以根据物体的尺寸和形状选择不同数量的卷积核来适应不同的物体。其中，DyHead 的结构如图 5 所示。

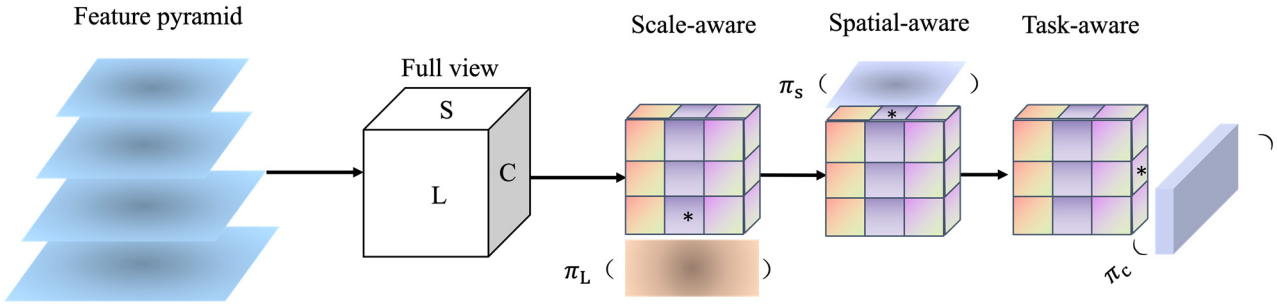


图 5 DyHead 模块结构

Fig.5 DyHead module structure

1.7 改进后的 EFDCD-YOLO 网络模型

为了解决水下废弃物红外目标检测中的复杂问题，本文对 YOLOv5 模型进行了改进[22]，并提出了 EFDCD-YOLO 模型，模型的主要流程如图 6 所示。从图 6 中可以看出，模型在主干网络部分引入 InceptionNeXt 替代原有 YOLOv5 的 CSPDarknet53 主干网络，通过多尺度和多层次的特征提取，以及参数共享和并行操作的设计，提高了模型的表达能力和特征提取能力。其次，在 Neck 的特征融合层中，采用 CARAFE 算子对一层卷积后的特征图进行上采样操作，然后利用 EffectiveSE 注意力机制对特征图的空间信息进行建模，将空间注意力权重与通道注意力权重相乘，以增强特征的判别能力。此外，将 DCNv2 替代 C3 模块，以提升目标检测的准确性和性能，同时保持较快的推理速度。最后，在头部分采用了 DyHead 的思想，将经过可变形卷积处理的特征图输入，通过多尺度特征融合的方式，有效地结合来自不同层级的特征，提取丰富的语义信息，并增强模型对目标的表示能力。通过上述改进，提高了模型在目标检测任务中的泛化能力。

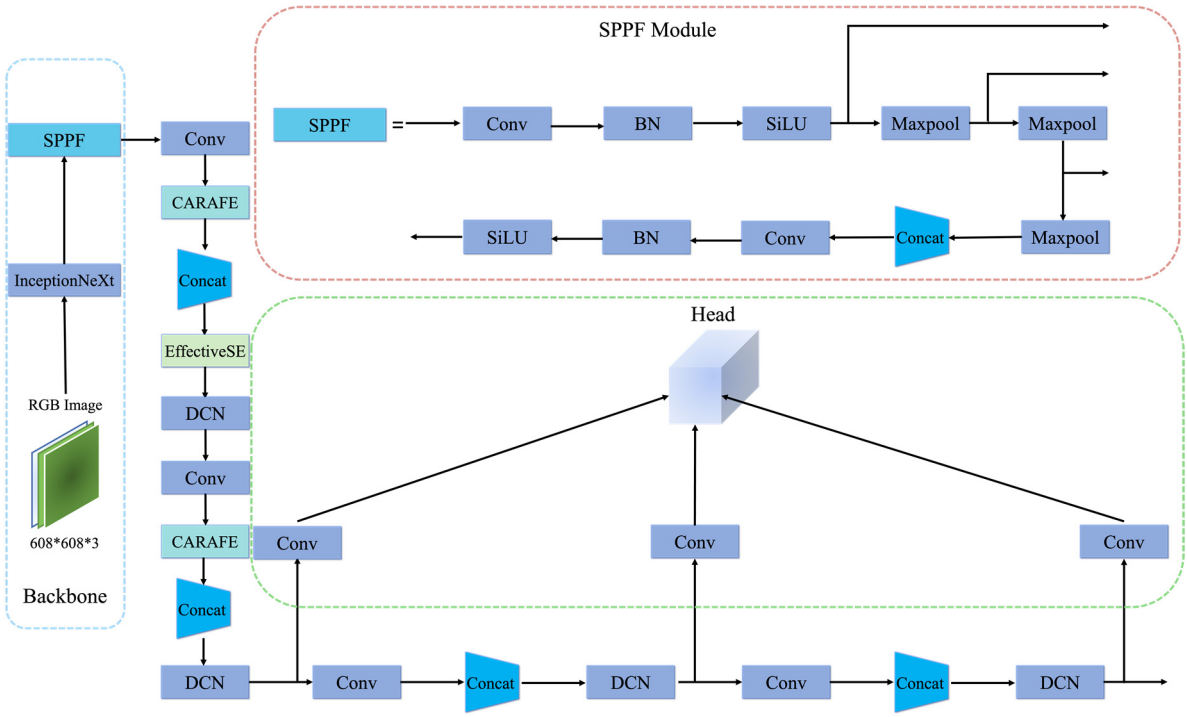


图 6 EFDCD-YOLO 网络结构

Fig.6 EFDCD-YOLO network structure

2 实验结果与分析

2.1 实验环境

在模型训练的过程中将模型 epoch 设置为 100，其他参数都是 YOLOv5 6.0 的默认版本。其中，实验环境如表 1 所示。

表 1 实验环境配置

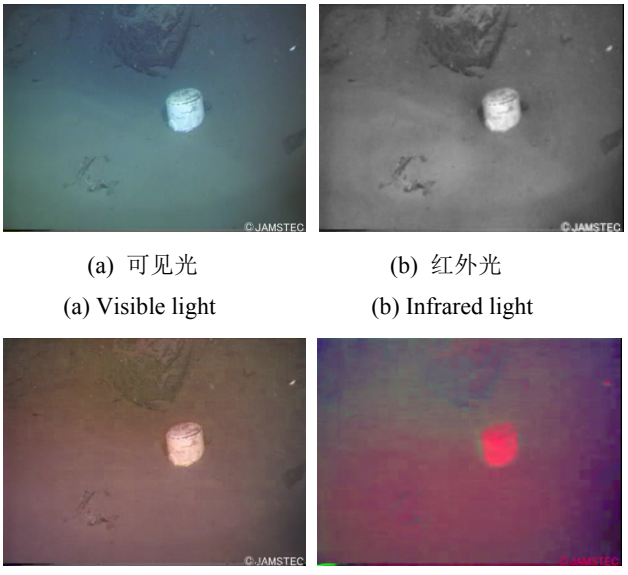
Table 1 Experimental environment configuration

Configuration item	Configuration item parameter
CPU	Intel(R)Core(TM)i9-10900X CPU@3.70GHz
GPU	NVIDIA RTX2080ti
Graphics card	12G
OS version	Ubuntu20.04
CUDA	10.2
Compiling environment	Python3.8+Pytorch1.12.1

2.2 数据集介绍

为了评估文中提出的 EFDCD-YOLO 模型在水下废弃物红外目标检测中的性能，本文选择了 Trash-ICRA19[23] 数据集作为模型测试的基准。Trash-ICRA19 数据集源自 J-EDI 海洋废弃物数据，其中包含了从真实环境中捕捉到的多种类型的海洋碎片图像。研究人员对收集到的图像进行处理，并提取出了 5700 张图像。而这些图像总共涵盖了 3 大类别分别是：plastic（塑料）、bio（生物）、rov（遥控潜水器），并为它们标注了边界框。为了进行训练和测试，对红外图像图 7(b)，通过 HSV 变换得到图 7(c)，经过综合变换得到图 7(d)。

文中按照训练集和测试集 9:1 的比例将这些图像划分为 5130 张训练集和 570 张测试集。通过在上述数据集上进行测试，能够更为全面地评估文中改进模型在水下废弃物红外目标检测任务中的性能。



(a) 可见光 (b) 红外光

(a) Visible light (b) Infrared light

(c) HSV 变换 (d) 综合变换

(c) HSV transformation (d) Comprehensive transformation

图 7 Trash-ICRA19 数据集

Fig.7 Trash-ICRA19 Dataset

2.3 评价指标

为评估模型在水下红外目标检测中的性能，文中采用 P（精确率）、R（召回率）、mAP（各类别的 AP 平均值）、GFLOPs（神经网络的计算量）、Params（参数量）的评价指标，各个指标的计算方法如公式(20)～(23)所示：

$$P=\frac{\mathrm{TP}}{\mathrm{TP+FP}} \tag{20}$$

$$R=\frac{\mathrm{TP}}{\mathrm{TP+FN}} \tag{21}$$

式(20)、(21)中：TP（True Postives）为正样本被正确检测的数量；FP（False Postives）为正样本被错误预测的数量；FN（False Negatives）为负样本被错误检测的数量。

$$\mathrm{AP}=\int_0^1 P(R)\mathrm{d}R \tag{22}$$

$$\mathrm{mAP}=\frac{\sum_1^n \mathrm{AP}_l}{n} \tag{23}$$

式(22)、(23)中：为检测目标的类别数；AP_l 为类别 l 的平均准确度。

2.4 实验结果分析

为了平衡不同尺度和层次的特征表示，以提升模型的表达能力和特征提取能力，实验对 YOLOv5 模型的主干网络模块进行了替换，并比较了 CSPDarknet53 和 InceptionNeXt 在 YOLOv5 模型上精度、计算量和参数量的变化。结果如表 2 所示。

表 2 替换主干网络实验

Table 2 Replacing the backbone network experiment

	mAP/%	GFLOPs/G	Params/M
CSPDarknet53	43.8	15.8	7.0
InceptionNeXt	53.0(+9.2)	75.2	32.2

从表 2 可以看出，通过替换主干网络，网络 InceptionNeXt 相较于原有主干网络 CSPDarknet53 提升了 9.2%的精度。尽管这些替换会增加模型的计算量和参数量，但在精度方面却取得了性能上的提升。实验结果表明，从实验的计算量和参数量上来看，在 Trash-ICRA19 数据集上引入 InceptionNeXt 网络可以扩展网络的宽度和深度，从模型的平均精度上来看，InceptionNeXt 加强了改进算法的表达能力和特征提取能力，提升了模型在水下目标检测任务中的检测精度。

为了评估在特征融合层中引入注意力机制对模型的影响，实验以引入 InceptionNeXt 网络的 YOLOv5 模型为基准，在 Trash-ICRA19 数据集上对比了引入 CoordAttention[24]、GAM[25]、BiFormer[26]、SGE[27]、EffectiveSE 注意力机制后模型在精度、计算量和参数量上的变化。实验结果如表 3 所示。

表 3 添加注意力机制实验

Table 3 Add attention mechanism experiment

	mAP/%	GFLOPs/G	Params/M
BaseLine	53.0	75.2	32.2
CoordAttention$^+$	53.2(+0.2)	75.2	32.2
GAM$^+$	53.6(+0.6)	107.9	42.4
SGE$^+$	53.6(+0.6)	75.1	32.2
BiFormer$^+$	55.0(+2.0)	139.6	33.8
EffectiveSE$^+$	55.0(+2.0)	75.4	32.6

从表 3 可以观察到，在引入 GAM 和 BiFormer 的注意力机制后，模型的精度有一定的提升，但极大程度地增加了模型的计算量和参数量。加入 CoordAttention、SGE 和 EffectiveSE 注意力机制后，虽然在计算量和参数量上有着小幅度增加，但是分别实现了 0.2%、0.6%和 2%的精度提升。

尽管 EffectiveSE 和 BiFormer 注意力机制都提升了 2%的精度，但综合考虑到计算资源和参数量的变化，因此选择在特征融合层中引入 EffectiveSE 注意力机制。

EffectiveSE 注意力的优势在于提升模型精度的同时在计算量和参数量上的增加相对较小，更有助于模型捕捉关键特征。

为了提高目标定位精度和边界框回归的准确性，实验选择引入 InceptionNeXt 主干网络和 EffectiveSE 注意力机制为基线模型，并将模型中的 CIOU 损失函数替换为 DIOU[28]、SIOU[29]、WIOU[30]和 FocalEIOU，以评价改进后的 IOU 对模型精度、计算量和参数量的影响。实验结果如表 4 所示。

表 4 损失函数改进实验

Table 4 Improvement experiment of loss function

	mAP/%	GFLOPs/G	Params/M
BaseLine	55.0	75.4	32.6
SIOU$^+$	55.0(+0.0)	75.4	32.6
WIOU$^+$	55.0(+0.0)	75.4	32.6
DIOU$^+$	55.6(+0.6)	75.4	32.6
FocalEIOU$^+$	55.7(+0.7)	75.4	32.6

从表 4 可以观察到，更改模型的损失函数以后，模型的计算量和参数量并未发生变化。但是通过替换 DIOU 和 FocalEIOU 损失函数，模型的精度分别提升了 0.6%和 0.7%，因此选择将 FocalEIOU 作为改进算法的损失函数。

将损失函数替换为FocalEIOU后模型在未增加额外计算量和参数量的情况下，能够更好地衡量目标定位的准确性，从而提高检测结果的质量。

为了增强模型的感受野、改善目标定位、增加模型对细粒度特征的感知能力，实验采用可变形卷积 DCNv2 替代 YOLOv5 模型中的卷积操作，引入 CARAFE 替代原模型的上采样操作，并将模型中的 Detect模块改为利用多尺度特征融合的DyHead模块。实验以 InceptionNeXt 为主干网络、EffectiveSE 注意力和 FocalEIOU 损失函数为基线模型。并评估了在引入 DCNv2、CARAFE 和 DyHead 后，模型在 Trash-ICRA19 数据集上精度、计算量和参数量的变化，实验结果如表 5 所示。

表 5 添加 DCNv2、CARARE 和 DyHead 的实验结果

Table 5 Add DCNv2, CARARE, and DyHead experiment results

DCNv2	CARAFE	DyHead	mAP/%	GFLOPs/G	Params/M
-	-	-	55.7	75.4	32.6
√			55.8(+0.1)	74.0	32.7
	√		55.9(+0.2)	75.8	32.7
		√	56.4(+0.7)	76.2	32.8
√	√		55.8(+0.1)	74.3	32.9
√		√	56.5(+0.8)	74.8	33.0
	√	√	57.3(+1.6)	76.5	33.0
√	√	√	57.4(+1.7)	75.0	33.2

从表 5 可以观察到将 DCNv2、CARAFE、DyHead 进行不同的配置实验后，模型在平均精度、计算量和模型参数量上的变化。在改变模型的卷积操作后，模型在提高精度的同时，减少了模型的计算量。另一方面，模型选择 CARAFE 算子，在提升模型的精度和鲁棒性的同时，还增加了对细粒度特征的感知能力[20]。引入 DyHead 使模型的精度提升了 0.8%，虽然计算量和参数量小幅度提高，但得到了更准确的目标定位和分类结果，减少了误检和漏检的情况，从而提高了模型的准确性和鲁棒性。

在同时引入 DCNv2、CARAFE 和 DyHead 的情况下，模型的性能得到了进一步的提升。平均精度达到了最高值 57.4%，高于其他的实验配置。

通过上述改进，模型在一定程度上增强了处理复杂环境的能力。增加了模型的感受野使其能够更好地处理远距离目标，改善了目标定位和性能；提高了模型的特征表达能力和感知能力，使其能够更好地适应各种细粒度特征和复杂场景。

2.5 对比实验

为了验证本文所提出模型的有效性，实验将 EFDCD-YOLO 模型与 YOLOv5-LeakyReLU、YOLOv5-transformer、PP-YOLO、YOLO-Ghost、YOLOv5、YOLOv7 和 YOLOv8 等模型在 Trash-ICRA19 数据集上进行对比，观察不同模型在 *P*、*R*、mAP、GFLOPS 和 Params 方面的性能，其中对比实验的结果如表 6 所示。

表 6 对比实验

Table 6 Comparative experiment

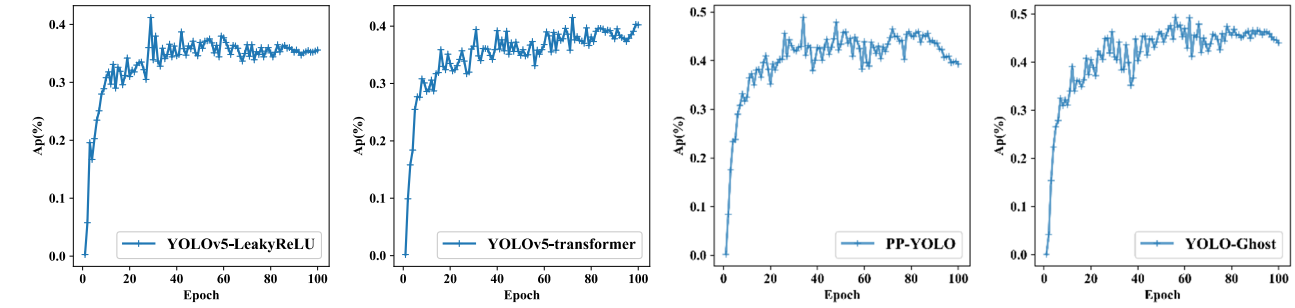
	P/%	*R*/%	mAP/%	GFLOPs/G	Params/M
YOLOv5-LeakyReLU	59.8	34.9	38.0	15.8	7.02
YOLOv5-transformer	46.3	40.7	40.1	15.6	7.02
PP-YOLO	56.8	44.5	46.5	16.1	12.3
YOLO-Ghost	57.8	46.1	46.7	8.0	3.68
YOLOv5	43.8	44.0	43.8	15.8	7.02
YOLOv7	45.2	53.2	51.8	103.2	36.5
YOLOv8	51.0	44.4	47.9	78.7	25.8
EFDCD-YOLO	65.2	53.7	57.4	75.0	33.2

从表 6 可以观察到，EFDCD-YOLO 在 *P*、*R*、mAP 方面分别达到了 65.2%、53.7%和 57.4%的精确度，对比于基线模型 YOLOv5 分别提升了 21.4%、9.7%和 13.6%。同时相较于其他模型，在 *P*、*R*、mAP 方面都具有较大的优势，尽管在计算量和参数量上有一定的增加，但这种权衡是合理的，EFDCD-YOLO 的优势在于设计的特征融合和上下文感知机制，使其能够有效地提高目标检测的准确性和鲁棒性[16-21]。同时在另一方面模型能够克服水下环境的挑战，提取更具有辨别力的特征，从而实现更精确的目标检测结果。

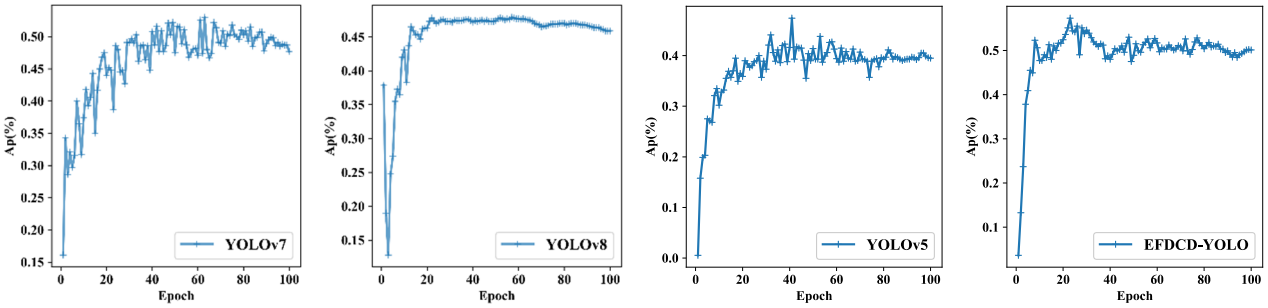
为了更好地展示 EFDCD-YOLO 相对于其他模型的优势，实验对比了各模型在经过 100 轮训练后精度的变化，如图 8 所示。

从图 8 中可以看出，EFDCD-YOLO 在经过 100 轮训练后的模型精度始终高于其他相对比的模型，即其曲线更靠近图像的上方。此外，还可以观察到，在 0～10 轮的训练中，EFDCD-YOLO 模型的精度增长速度较其他模型更快，这进一步验证了模型的有效性。

为了验证模型的性能，实验绘制了不同模型的 *P-R* 曲线，如图 9 所示。*P-R* 曲线反映了准确率与召回率之间的关系，横坐标表示召回率，纵坐标表示精确率。曲线越靠近右上方，表示性能越好，同时当一个曲线完全包含另一个曲线时，后者的性能优于前者。



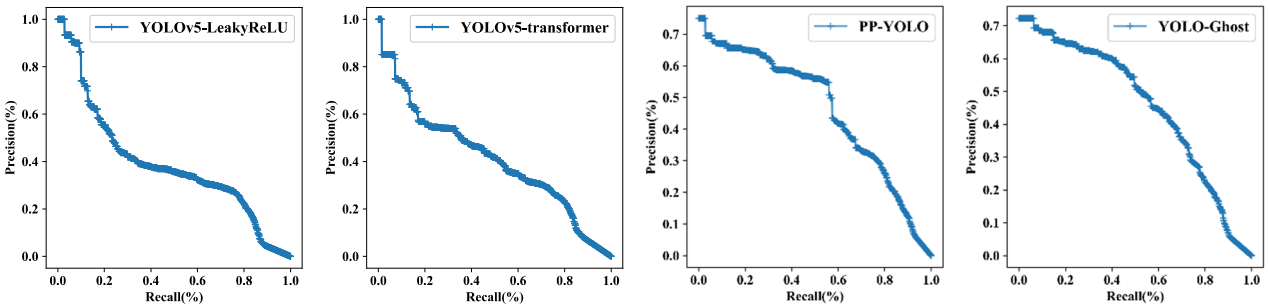
(a) YOLOv5-LeakyReLU (b) YOLOv5-transformer (c) PP-YOLO (d) YOLO-Ghost



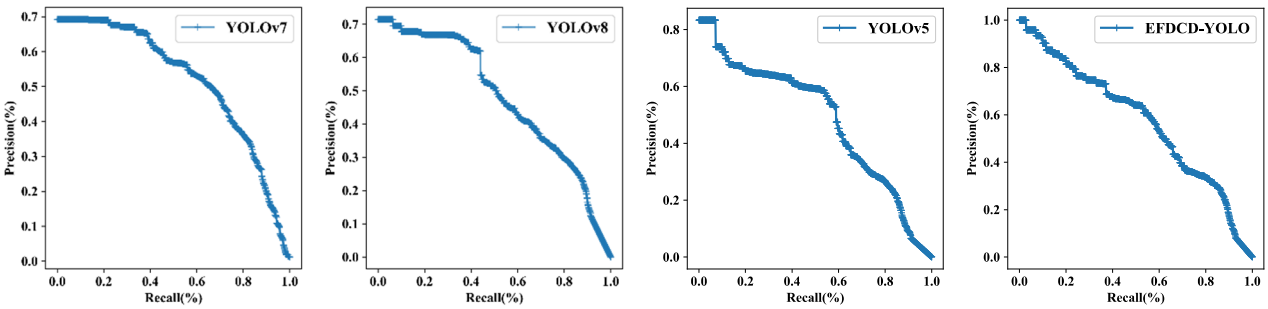
(e) YOLOv7 (f) YOLOv8 (g) YOLOv5 (h) EFDCD-YOLO

图 8 mAP 曲线对比

Fig.8 mAP curves comparison



(a) YOLOv5-LeakyReLU (b) YOLOv5-transformer (c) PP-YOLO (d) YOLO-Ghost



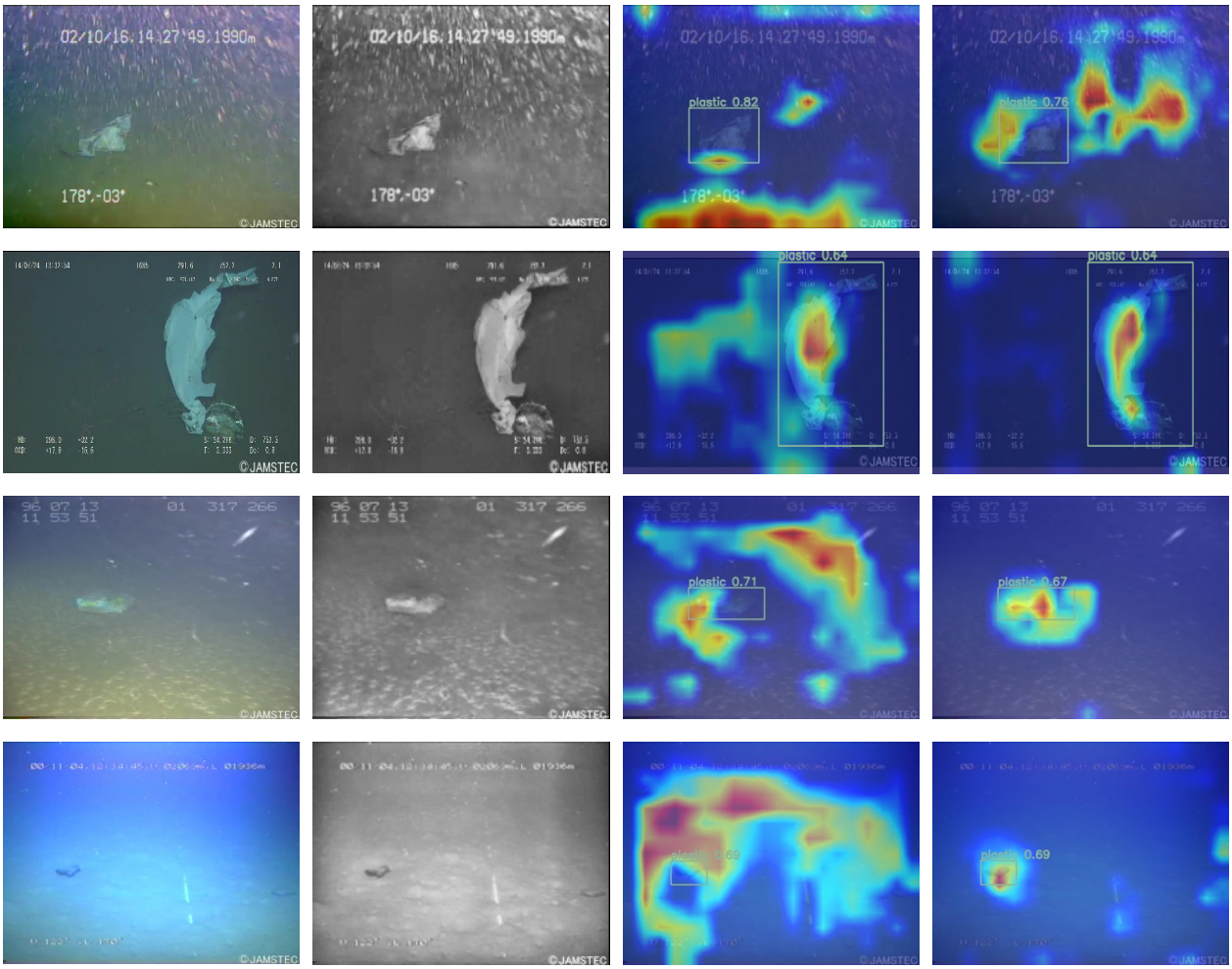
(e) YOLOv7 (f) YOLOv8 (g) YOLOv5 (h) EFDCD-YOLO

图 9 *P-R* 曲线对比

Fig.9 *P-R* curves comparison

从图 9 中可以观察到，EFDCD-YOLO 的曲线更靠近图像的右上角，并且其曲线下面积也包含了其他模型的曲线。这表明 EFDCD-YOLO 的模型性能高于相对比的模型。

同时为了进一步验证模型在数据集上的注意力分布，实验选择了测试集中的 4 张图像，并绘制了 EFDCD-YOLO 和 YOLOv5 模型的注意力对比热力图，如图 10 所示。



(a) 可见光图像　(b) 红外图像　(c) YOLOv5　(d) EFDCD-YOLO

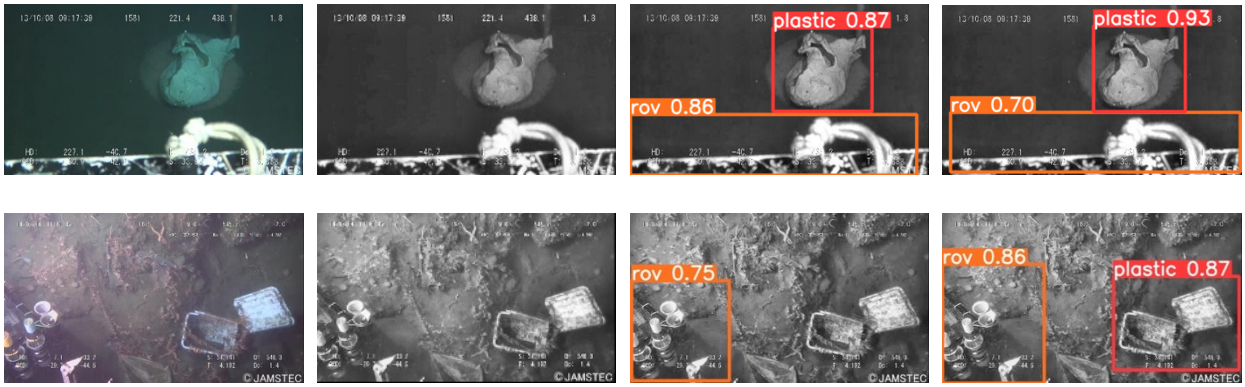
(a) Visible light image　(b) Infrared image　(c) YOLOv5　(d) EFDCD-YOLO

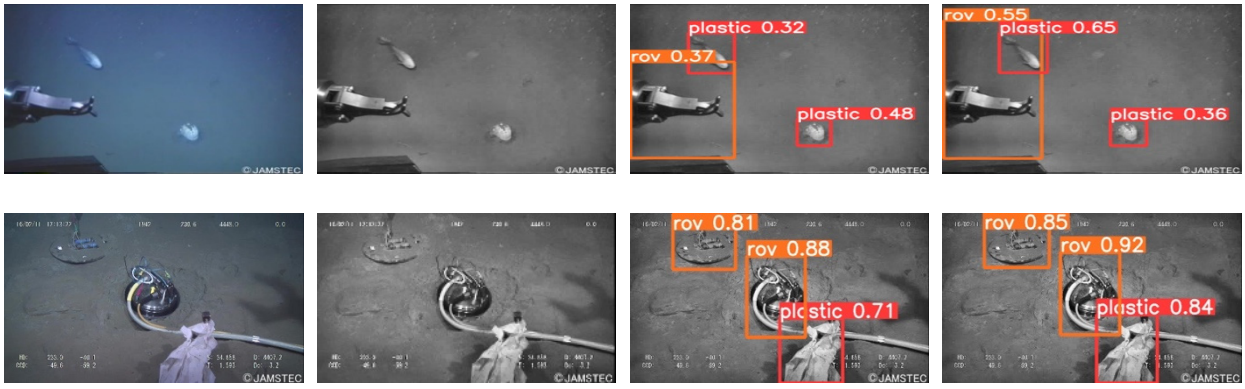
图 10　注意力对比热力图

Fig.10　Attention comparison heat chart

从图 10 中可以观察到，EFDCD-YOLO 能够更好地将注意力聚焦于检测目标上，并且相较于 YOLOv5 模型，EFDCD-YOLO 能够更准确地定位目标的形状和具体位置。EFDCD-YOLO 的注意力更加集中，而 YOLOv5 的注意力在目标形状和位置的定位上更为分散。这表明了 EFDCD-YOLO 在水下目标检测任务中对目标位置定位的精确性以及对目标形状和大小的识别能力。

为了进一步对比和验证 EFDCD-YOLO 模型的有效性，实验对比了 YOLOv5 和 EFDCD-YOLO 在水下目标检测中的检测结果，如图 11 所示。





(a) 可见光图像 (b) 红外图像 (c) YOLOv5 (d) EFDCD-YOLO

(a) Visible light image (b) Infrared image (c) YOLOv5 (d) EFDCD-YOLO

图 11 模型效果对比

Fig.11 Comparison of model effects

从图 11 中可以观察到，在聚焦于相同物体的情况下，EFDCD-YOLO 相较于 YOLOv5 模型表现出更高的检测精度和置信度，并且减少了漏检的情况。

通过上述实验和可视化结果进一步验证了 EFDCD-YOLO 在水下废弃物红外目标检测中的有效性和适用性。

3 结语

模型通过引入 InceptionNeXt 作为主干网络，并且在特征融合层中采用 CARAFE 算子替代上采样操作，加入 EffectiveSE 注意力机制，利用 DCNv2 可变形卷积替换原有的 C3 模块，同时在 Head 部分采用 DyHead 的思想，提出了一种基于 YOLOv5 的 EFDCD-YOLO 水下废弃物红外目标检测模型。通过对 EFDCD-YOLO 在 Trash-ICRA19 数据集上的性能评估，展示了该模型在水下废弃物红外目标检测中的优势。

实验结果表明，EFDCD-YOLO 在精确率、召回率和平均精确度方面均取得了显著的提升，分别达到了 65.2%、53.7%和 57.4%。尽管 EFDCD-YOLO 增加了模型的计算量和参数量，但由于其特征融合和上下文感知机制，使得该模型能够更好地应对水下红外目标检测中的各种挑战，例如低质量水下图像、目标与背景的区分困难以及位置和形态变化等问题。此外，通过对实验结果的可视化分析，进一步验证了 EFDCD-YOLO 相对于其他改进模型的优势。

因此 EFDCD-YOLO 在水下废弃物红外目标检测领域具有一定的应用价值，也为改善水下废弃物红外目标检测的准确性和鲁棒性提供了相应的参考。

参考文献：

[1] Schechner Y Y, Narasimhan S G, Nayar S K. Instant dehazing of images using polarization[C]//*Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE*, 2001, **1**: I-I.

[2] Bazeilles, Quidui, Jaulinl. Identification of underwater man-made object using a colour criterion[J]. *Proceedings of the Insitute of Acoustics*, 2007, **29**(6): 25-52.

[3] LI C Y, GUO J C, CONG R M, et al. Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior[J]. *IEEE Transactions on Image Processing*, 2016, **25**(12): 5664-5677.

[4] 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.

[5] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 7263-7271.

[6] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.

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

[8] LIU W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//*Computer Vision–ECCV*, 2016: 21-37.

[9] 陈鑫林. 基于深度学习的水下垃圾检测[D]. 贵阳: 贵州师范大学, 2022.
CHEN Xinlin. Underwater Garbage Detection Based on Deep Learning [D]. Guiyang: Guizhou Normal University, 2022.

[10] 袁红春, 臧天祺. 基于注意力机制及 Ghost-YOLOv5 的水下垃圾目标检测[J]. 环境工程, 2023, **41**(7): 214-221. DOI:10.13205/j.hjgc.

202307029.

YUAN Hongchun, ZANG Tianqi. Underwater garbage target detection based on attention mechanism and Ghost-YOLOv5[J]. *Environmental Engineering*, 2023, **41**(7): 214-221. DOI: 10.13205/ j.h JGC.202307029.

[11] JIANG H, Learned Miller E. Face detection with the faster R-CNN[C]// 12th *IEEE International Conference on Automatic Face & Gesture Recognition*, 2017: 650-657.

[12] CAI Z, Vasconcelos N. Cascade R-CNN: High quality object detection and instance segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, **43**(5): 1483-1498.

[13] ZHOU X, WANG D, Krähenbühl P. Objects as points[J]. arXiv preprint arXiv:1904.07850, 2019.

[14] 吕晓倩. 基于 Faster R-CNN 的水下目标检测方法研究与实现[D]. 哈尔滨: 哈尔滨工业大学, 2019.

LYU Xiaoqian. Research and Implementation of Underwater Target Detection Method Based on Faster R-CNN [D]. Harbin: Harbin Institute of Technology, 2019.

[15] 王蓉蓉, 蒋中云. 基于改进 CenterNet 的水下目标检测算法[J]. **激光与光电子学进展**, 2023, **60**(2): 239-248.

WANG Rongrong, JIANG Zhongyun. Underwater target detection algorithm based on improved CenterNet[J]. *Progress in Laser and Optoelectronics*, 2023, **60**(2): 239-248.

[16] YU W, ZHOU P, YAN S, et al. Inceptionnext: When inception meets convnext[J]. arXiv preprint arXiv:2303.16900, 2023.

[17] Lee Y, Park J. Centermask: Real-time anchor-free instance segmentation[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 13906-13915.

[18] ZHANG Y F, REN W, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. *Neurocomputing*, 2022, **506**: 146-157.

[19] WANG R, Shivanna R, CHENG D, et al. DCN v2: Improved deep & cross network and practical lessons for web-scale learning to rank systems[C]//*Proceedings of the Web Conference*, 2021: 1785-1797.

[20] WANG J, CHEN K, XU R, et al. Carafe: Content-aware reassembly of features[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019: 3007-3016.

[21] DAI X, CHEN Y, XIAO B, et al. Dynamic head: Unifying object detection heads with attentions[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 7373-7382.

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

[23] Fulton M, HONG J, Islam M J, et al. Robotic detection of marine litter using deep visual detection models[C]//*International Conference on Robotics and Automation* (ICRA). *IEEE*, 2019: 5752-5758.

[24] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 13713-13722.

[25] LIU Y, SHAO Z, Hoffmann N. Global attention mechanism: retain information to enhance channel-spatial interactions[J]. arXiv preprint arXiv: 2112.05561, 2021.

[26] ZHU L, WANG X, KE Z, et al. BiFormer: vision transformer with bi-level routing attention[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 10323-10333.

[27] LI X, HU X, YANG J. Spatial group-wise enhance: Improving semantic feature learning in convolutional networks[J]. arXiv preprint arXiv: 1905.09646, 2019.

[28] 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*, 2020, **34**(7): 12993-13000.

[29] Gevorgyan Z. SIoU loss: More powerful learning for bounding box regression[J]. arXiv preprint arXiv: 2205.12740, 2022.

[30] TONG Z, CHEN Y, XU Z, et al. Wise-IoU: bounding box regression loss with dynamic focusing mechanism[J]. arXiv preprint arXiv: t2301.10051, 2023.