

# 基于DMS-YOLO的水下模糊目标检测算法

霍瑞峰, 郑广海\*

大连交通大学轨道智能工程学院, 辽宁 大连

收稿日期: 2026年3月23日; 录用日期: 2026年4月21日; 发布日期: 2026年4月29日

## 摘要

针对水下目标模糊、边缘对比度低导致检测精度下降的问题, 本文提出了一种改进的水下目标检测模型 DMS-YOLO。该方法首先引入动态上采样模块(Dynamic Sampling, DySample), 通过自适应调整采样位置实现特征重建, 以改善细节信息的恢复效果; 其次构建动态检测头(Dynamic Head, DyHead), 从尺度、空间及任务三个维度提升特征表达的自适应性; 进一步在C2PSA结构中嵌入多尺度膨胀注意力机制(Multi-Scale Dilated Attention, MSDA), 以增强多尺度信息融合能力; 最后采用Slide Loss损失函数, 通过动态加权策略提升模型对分类困难样本的学习能力。在RUOD数据集上的实验结果表明, DMS-YOLO的mAP@0.5达到87.5%, mAP@0.5:0.95为64.0%, 该方法对模糊目标细节表达及多尺度特征融合方面表现出一定改进, 并在检测精度与计算复杂度之间呈现出较为合理的平衡。

## 关键词

水下目标检测, YOLOv11, DySample, DyHead, MSDA, Slide Loss

# DMS-YOLO: An Underwater Blurred Object Detection Algorithm

Ruifeng Huo, Guanghai Zheng\*

School of Rail Transit Intelligent Engineering, Dalian Jiaotong University, Dalian Liaoning

Received: March 23, 2026; accepted: April 21, 2026; published: April 29, 2026

## Abstract

To address the decline in detection accuracy caused by blurred underwater targets and low edge contrast, this paper proposes an improved underwater object detection model, DMS-YOLO. The method first introduces a dynamic upsampling module (Dynamic Sampling, DySample), which

\*通讯作者。

reconstructs features by adaptively adjusting sampling positions to improve the recovery of fine-grained details. Next, a dynamic detection head (Dynamic Head, DyHead) is constructed to enhance the adaptability of feature representation from three dimensions: scale, spatial, and task. Furthermore, a multi-scale dilated attention mechanism (Multi-Scale Dilated Attention, MSDA) is embedded into the C2PSA structure to strengthen multi-scale feature fusion. Finally, the Slide Loss function is adopted, which improves the model's learning capability for difficult classification samples through a dynamic weighting strategy. Experimental results on the RUOD dataset show that DMS-YOLO achieves an mAP@0.5 of 87.5% and an mAP@0.5:0.95 of 64.0%. The results indicate that the proposed method shows improvements in the representation of blurred target details and multi-scale feature fusion, while maintaining a relatively balanced trade-off between detection accuracy and computational complexity.

## Keywords

Underwater Object Detection, YOLOv11, DySample, DyHead, MSDA, Slide Loss

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

随着水下视觉技术在海洋资源调查、水下机器人巡检及生态环境监测等领域的发展,目标检测任务正由单一类别向多类别复杂场景演进。在真实海洋环境中,检测对象不仅包括底栖生物,还涵盖鱼类及潜水人员等多种目标,类别多样且尺度变化显著,对模型特征表达能力提出更高要求。同时,受水体吸收与散射、光照衰减及悬浮颗粒等因素影响,水下图像普遍存在对比度低、细节模糊及背景干扰强等问题[1],进一步增加了检测难度。此外,水下机器人及自主潜航器受限于嵌入式平台的计算与存储资源,对模型轻量化与推理效率提出严格要求[2]。因此,在有限计算开销下提升检测精度成为关键问题。

近年来,基于深度学习的目标检测方法取得了显著进展[3][4]。根据检测流程不同,主流方法可分为双阶段与单阶段检测器。双阶段方法如 Fast R-CNN [5]和 Faster R-CNN [6],通过候选区域生成与精细分类实现较高精度,但流程复杂、计算开销较大,难以满足实时性需求。相比之下,单阶段方法如 SSD [7]和 YOLO [8]系列,采用端到端预测,在特征图上直接完成分类与定位,具有结构简洁、推理速度快等优势,更适用于实时检测任务。近年来,YOLO 系列持续演进,在精度与效率之间取得良好平衡,已成为轻量化目标检测的重要基础方法。

而针对水下环境中的难题,已有研究从特征建模、注意力机制以及面向小目标的轻量化结构设计等方面对水下目标检测方法进行了改进。文献[9]提出 DP-FishNet,构建了融合 Transformer 全局感知与 CNN 局部特征的双路径结构,提升特征融合质量及类别不平衡处理能力。文献[10]在 YOLOv7 中引入 ACmix 模块与全局注意力机制,通过融合卷积与注意力操作增强特征提取能力,从而提升检测精度与推理效率。此外,文献[11]针对水下小目标密集场景,提出 YOLOv8-LA 模型,通过引入轻量化部分卷积模块并结合 CARAFE 上采样,有效提升了小目标检测性能与实时性。

尽管上述方法在一定程度上提升了水下目标检测性能,但在复杂水下场景目标检测中仍存在目标细节特征恢复不足,自适应调节能力弱,对小目标多尺度信息融合不充分,对边界模糊及难分样本关注不足等问题。为此,本文在 YOLOv11n 基础上提出了一种新的检测框架 DMS-YOLO,本文的主要改进如下:

(1) 为缓解上采样过程中细节信息丢失问题,引入 DySample 动态上采样模块,通过学习采样位置实

现自适应特征重采样, 从而提升水下模糊目标的细节恢复能力与小目标特征表达效果。

(2) 针对水下目标特征自适应能力不足的问题, 引入 **Dynamic Head** 动态检测头模块。从尺度、空间及任务三个维度对特征进行动态调节, 增强模型在水下目标尺度变化剧烈时的适应能力。

(3) 针对小目标多尺度信息融合不充分的问题, 构建改进型 **C2PSA\_MSDA** 多尺度注意力模块。通过多尺度膨胀注意力机制融合局部细节与全局上下文信息, 提高对水下模糊及微小目标的判别能力。

(4) 针对边界模糊及难分样本关注不足的问题, 采用 **Slide Loss** 损失函数。依据动态调整样本权重, 促使模型重点关注水下场景中难以分类的边界样本, 有效缓解了难易样本不均衡带来的训练偏差, 减少误检与漏检。

## 2. DMS-YOLO 算法设计

YOLOv11 是 Ultralytics 在 YOLOv8 基础上进一步优化的目标检测模型, 在保持轻量化的同时提升了检测性能。其核心改进包括: 在网络结构中采用 **C3K2** 模块替代原有的 **C2f** 结构, 以增强特征表达能力; 在 **SPPF** 模块后引入 **C2PSA** 模块, 通过金字塔分割注意力机制强化关键空间信息建模; 同时在检测头中引入深度可分离卷积(**DWConv**), 以降低计算开销并提升推理效率。

在此基础上, 针对水下模糊场景中中小目标检测面临的细节信息丢失、特征表达能力不足以及困难样本关注不足等问题, 本文提出一种基于 YOLOv11n 改进的水下目标检测模型 **DMS-YOLO**。其中, **D** 表示 **DySample** 动态上采样模块和 **DyHead** 动态检测头; **M** 表示 **MSDA** 多尺度膨胀注意力机制; **S** 表示 **Slide Loss** 损失函数。这些改进分别用于增强细节信息恢复能力、提升模型对水下目标的自适应表征能力、加强多尺度特征融合能力, 并提高模型对困难样本的学习能力整体结构如图 1 所示。

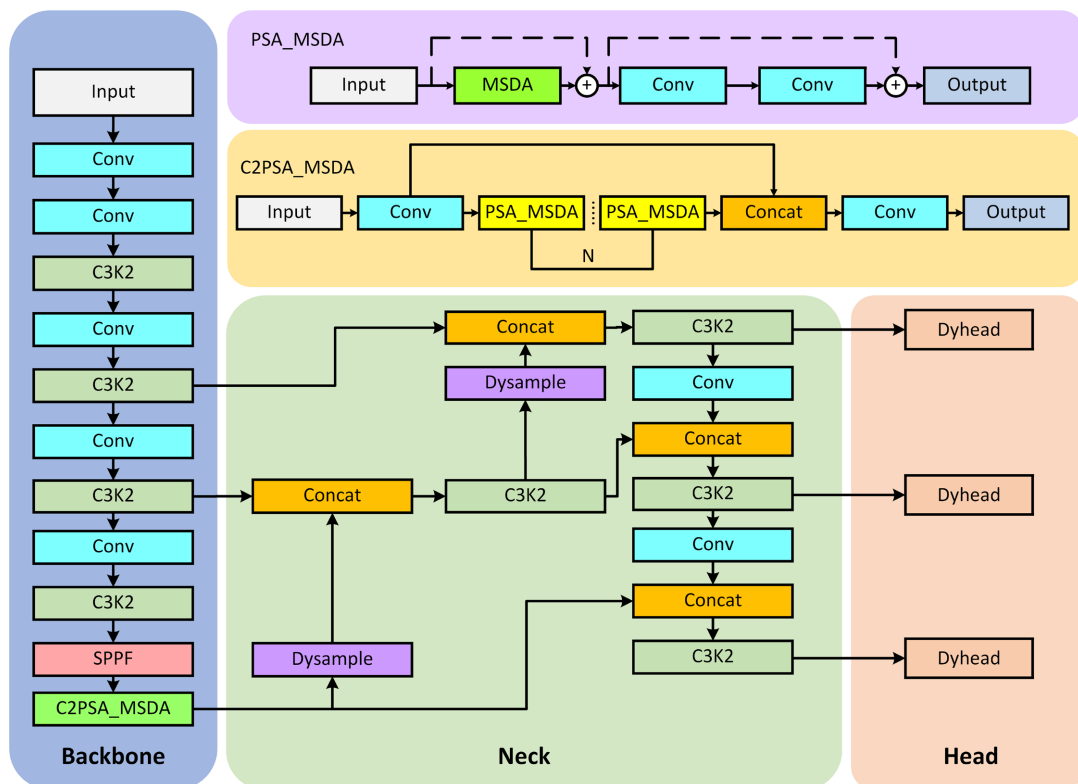


Figure 1. DMS-YOLO network architecture  
图 1. DMS-YOLO 网络结构

在主干网络(Backbone)部分, 构建了 C2PSA\_MSDA 模块, 其在原有 C2PSA 结构中嵌入多尺度膨胀注意力机制, 通过多尺度膨胀注意力机制融合局部细节与全局上下文信息, 以增强不同尺度特征的表达能力。在颈部网络(Neck)设计中, 引入 DySample 动态上采样模块, 通过自适应采样位置对特征进行重采样, 从而在提升分辨率的同时保留关键细节信息。在检测头(Head)部分, 引入 Dynamic Head 动态检测头, 从尺度、空间和任务三个维度对特征进行动态加权, 增强模型在水下目标尺度变化剧烈时的适应能力。此外, 在训练阶段引入 Slide Loss 损失函数, 通过动态加权机制增强模型对边界模糊及难分样本的关注。通过上述改进, 模型在模糊水下环境中实现了对小目标更稳定且更精确的检测, 在保持有限计算开销的同时有效提升了整体检测性能与鲁棒性。

## 2.1. DySample 动态上采样模块

水下成像受散射与吸收效应影响, 易出现对比度低、图像模糊及尺度变化显著等问题。YOLOv11 通过特征金字塔结构实现多尺度特征融合, 但其上采样过程多依赖固定插值方法, 缺乏对特征内容的自适应建模能力, 在复杂水下场景中难以有效恢复空间细节信息。这种局限性容易造成小目标特征表达不足甚至退化, 进而影响模型对小目标的检测精度与整体性能表现。

为了解决上述问题, 本文引入动态上采样模块(Dynamic Sampling, DySample) [12]。DySample 以点采样(point sampling)机制为核心, 通过动态回归采样位置来实现特征重采样, 而不是像传统方法那样生成卷积核, 从而在几乎不增加计算开销的情况下显著提升特征重建质量。该模块能够根据输入特征结构自适应预测采样偏移量(sampling offsets), 避免了类似 CARAFE 等方法中动态卷积带来的额外计算成本。DySample 模块的整体结构如图 2 所示。

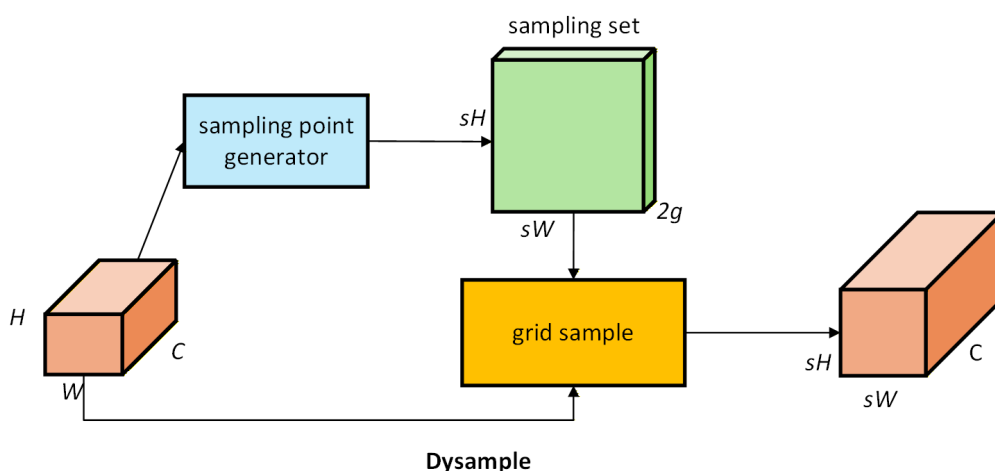


Figure 2. DySample module  
图 2. DySample 模块

给定输入特征图  $\mathcal{X} \in \mathbb{R}^{C \times H_1 \times W_1}$  和采样网格  $\mathcal{S} \in \mathbb{R}^{2 \times H_2 \times W_2}$ , 其中通道维度 2 表示空间坐标, 则输出特征  $\mathcal{X}'$  通过双线性插值计算得到:

$$\mathcal{X}' = \text{grid\_sample}(\mathcal{X}, \mathcal{S}) \quad (1)$$

其中,  $\text{grid\_sample}(\cdot)$  表示利用采样网格  $\mathcal{S}$  中坐标进行插值的采样操作, 该过程的核心在于通过学习得到的采样位置, 对原始特征进行自适应重构, 从而更好地保留关键空间信息。

给定上采样倍率  $s$ , 首先通过一个线性层从输入特征中预测偏移张量  $\mathcal{O} \in \mathbb{R}^{2 \times s^2 \times H \times W}$ , 该张量用于编码

每个采样位置的空间位移:

$$\mathcal{O} = \text{linear}(\mathcal{X}) \tag{2}$$

随后, 通过像素混洗(Pixel Shuffling)操作将偏移量重新排列, 使其与目标分辨率对齐, 并与规则采样网格  $\mathcal{G}$  相加, 从而得到最终的采样位置:

$$S = \mathcal{G} + \mathcal{O} \tag{3}$$

其中,  $\mathcal{g}$  表示均匀分布的基础采样网格, 而  $\mathcal{O}$  提供自适应位移, 使采样过程能够更加精准地聚焦于信息丰富的区域, 最终输出特征的分辨率为  $C \times sH \times sW$ , 通过上述过程, DySample 能够在放大特征图的同时有效恢复空间细节信息。

在整体网络结构中, DySample 模块被集成于 neck 部分用于上采样操作, 从而在特征融合阶段增强空间细节恢复能力, 同时保持轻量级计算开销, 使其特别适用于计算资源受限的水下部署场景。

### 2.2. Dynamic Head 动态检测头

在复杂水下环境中, YOLOv11 原始检测头对多尺度特征处理采用固定结构, 缺乏对输入特征的自适应能力, 并且在水下场景中, 浑浊度、光散射等因素常导致目标边界模糊、前景与背景混叠, 使得目标特征表达更加困难。同时, 标准卷积的均匀采样易引入冗余背景信息, 从而削弱特征表达能力。此外, 原始检测头同时承担分类与回归任务, 在共享特征的情况下缺乏解耦机制, 易引发任务冲突, 限制整体性能的进一步提升。

针对上述问题, 本文引入动态检测头(Dynamic Head, Dyhead) [13], 其整体结构如图 3 所示。该模块将注意力机制解耦为尺度感知、空间感知与任务感知三个独立分支, 从不同维度对特征进行自适应优化, 从而增强模型在复杂场景下的特征表达能力。

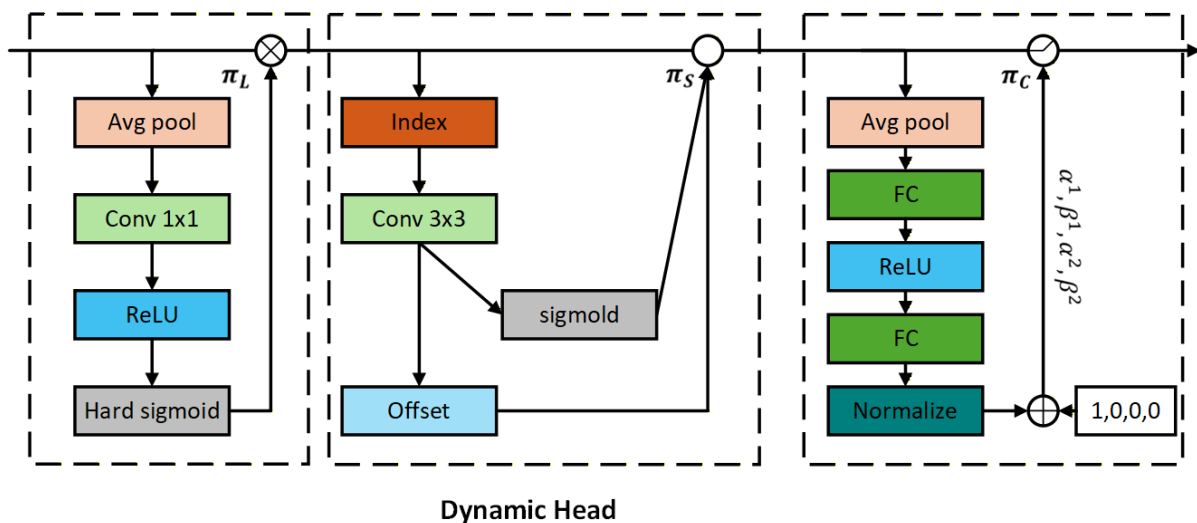


Figure 3. Dynamic Head module  
图 3. Dynamic Head 模块

尺度感知注意力  $\pi_L$  在不同特征层之间建立依赖关系, 根据目标尺度自适应分配权重:

$$\pi_L(F) \cdot F = \sigma \left( f \left( \frac{1}{SC} \sum_{s,c} F \right) \right) \cdot F \tag{4}$$

其中,  $f(\cdot)$  表示  $1 \times 1$  卷积,  $\sigma(x)$  代表 Hard Sigmoid 激活函数,  $S = H \times W$  代表所有空间位置的总数,  $C$  是通道维度。该分支通过对全局特征进行压缩建模, 实现对不同尺度特征的重要性评估。

空间感知注意力  $\pi_s$  结合可变形采样与多层特征聚合机制, 聚焦于信息密集区域:

$$\pi_s(F) \cdot F = \frac{1}{L} \sum_{l=1}^L \sum_{k=1}^K w_{l,k} \cdot F(l; p_k + \Delta p_k; c) \cdot \Delta m_k \quad (5)$$

其中,  $K$  表示稀疏采样位置的数量,  $p_k + \Delta p_k$  表示该区域的偏移量,  $\Delta m_k$  是位置  $p_k$  处的自学习重要性标量, 两者均从  $F$  中间层级的输入特征中学习得到。该机制通过对关键区域进行稀疏采样, 有效抑制背景噪声干扰。

任务感知注意力  $\pi_c$  通过动态通道调制实现分类与回归任务的有效解耦:

$$\pi_c(F) \cdot F = \max(\alpha^1(F) \cdot F_c + \beta^1(F), \alpha^2(F) \cdot F_c + \beta^2(F)) \quad (6)$$

其中,  $F_c$  是第  $c$  个通道的特征切片,  $[\alpha^1, \alpha^2, \beta^1, \beta^2]^T = \theta(\cdot)$  构成控制激活阈值的超函数。该分支结合了特征的全局平均池化和归一化, 通过自适应调整不同通道的响应, 实现分类与定位信息的有效分离。

通过引入 Dynamic Head 模块, 联合学习尺度感知、空间感知与任务感知注意力, 模型在应对水下场景中的大尺度变化、复杂背景干扰及目标模糊问题时, 能够更有效地提取和表达关键特征, 从而显著提升检测精度与整体鲁棒性。

### 2.3. MSDA 多尺度膨胀注意力机制

在 YOLOv11 中, C2PSA 模块通过 PSA 注意力机制筛选重要特征, 但其注意力主要集中在通道信息上, 对水下目标的位置、边缘等空间细节关注不足。除此之外, 水下图像中的小目标经多层下采样后细节损失严重, 仅靠通道重标定难以提取有效特征, 进而限制检测性能。

为此, 本文在 C2PSA 结构中引入多尺度膨胀注意力机制(Multi-Scale Dilated Attention, MSDA) [14], MSDA 的整体结构如图 4 所示。MSDA 机制的核心思想是利用具有不同膨胀率的膨胀卷积并行提取特征, 从而增强网络对多尺度上下文信息的感知能力。该自适应设计使模型能够同时捕获精细的局部特征与跨感受野的全局上下文信息, 实现局部细节与全局特征的有效融合。

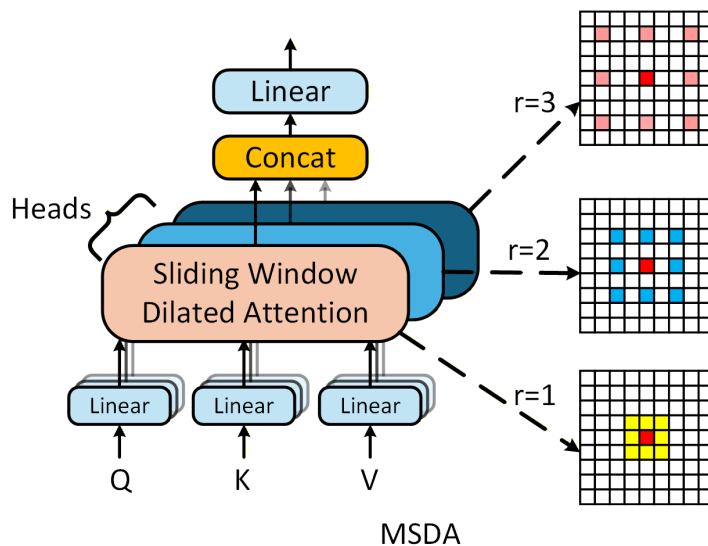


Figure 4. Multi-scale dilated attention  
图 4. MSDA 注意力机制

具体而言, MSDA 首先通过线性映射生成查询( $Q$ )、键( $K$ )和值( $V$ )三个特征表示, 并沿通道划分为多个注意力头, 每个头负责处理不同的特征子空间。通过每个头引入不同膨胀率  $r$ , 执行多尺度滑动窗口膨胀注意力(Sliding Window Dilated Attention, SWDA), 以构建多尺度感受野。默认设置为  $r = \{1, 2, 3\}$ , 分别对应不同空间范围, 从而实现局部细节与全局语义的互补, 具体流程如下。

对于原始特征图中坐标为  $(i, j)$  的查询, SWDA 并不会与整个特征图中的所有位置建立注意力关系, 而是仅在以  $(i, j)$  为中心的局部滑动窗口中选择部分键和值进行计算。若窗口大小为  $w \times w$ , 膨胀率为  $r$ , 则参与注意力计算的键和值所在位置  $(i', j')$  的集合可表示为:

$$\{(i', j') \mid i' = i + p \times r, j' = j + q \times r\}, \quad -\frac{w}{2} \leq p, q \leq \frac{w}{2} \quad (7)$$

该机制通过间隔采样扩大感受野, 在不增加窗口规模的情况下获取更广泛上下文信息。在此基础上, 对于位置  $(i, j)$  的输出特征  $x_{ij}$  计算过程可表示为:

$$x_{ij} = \text{Attention}(q_{ij}, K_r, V_r) = \text{Softmax}\left(\frac{q_{ij} K_r^T}{\sqrt{d_k}}\right) V_r, \quad 1 \leq i \leq W, 1 \leq j \leq H \quad (8)$$

其中  $K_r, V_r$  为上述按膨胀率选取的局部键值集合。

在此基础上, MSDA 进一步结合多头注意力机制以实现多尺度特征提取, 通过线性映射分别生成查询、键和值三个特征矩阵  $Q$ 、 $K$  和  $V$ 。再将特征通道划分为  $n$  个注意力头, 每个注意力头在独立的子空间中执行 SWDA 操作, 并设置不同的膨胀率  $r_i$ 。第  $i$  个头的输出  $h_i$  可表示为:

$$h_i = \text{SWDA}(Q_i, K_i, V_i, r_i), \quad 1 \leq i \leq n \quad (9)$$

其中  $r_i$  为第  $i$  个头的膨胀率,  $Q_i, K_i, V_i$  为输入到第  $i$  个头的特征图切片。最后, 通过线性投影层融合从不同头提取的多尺度特征, 将信息提炼并集成为一个全面的特征表示:

$$X = \text{Linear}(\text{Concat}[h_1, \dots, h_n]) \quad (10)$$

引入 C2PSA\_MSDA 后, 模型能够有效融合多尺度语义信息, 提升特征稳定性与判别性, 使其在复杂水下环境中获得更准确的检测与定位结果。该模块在应对模糊、遮挡及光照不均等问题时表现出更强鲁棒性, 同时兼顾局部细节与全局上下文, 有效提升检测精度与定位能力。

#### 2.4. Slide Loss 损失函数

在水下目标检测中, 样本不平衡严重影响模型性能。训练数据通常由高 IoU 的易分类正样本和大量低 IoU 的简单负样本构成, 而真正关键的是接近分类阈值的边界样本, 如遮挡目标、小尺度物体及边界模糊样本。这类样本数量较少, 却对模型判别能力至关重要。然而, 由于简单样本在数量上占优, 其在损失中的累积贡献主导梯度更新, 使模型偏向学习易样本, 忽视困难样本, 进而削弱泛化能力, 在复杂水下环境中易产生误检与漏检。

为缓解该问题, 本文引入 Slide Loss 损失函数[15], 通过基于 IoU 的动态加权机制增强模型对困难样本的关注。首先计算所有样本 IoU 的均值作为自适应阈值  $\mu$ :

$$\mu = \frac{1}{N} \sum_{i=1}^N \text{IoU}_i \quad (11)$$

其中,  $N$  表示样本总数,  $\text{IoU}_i$  表示第  $i$  个样本的交并比。  $\mu$  用于自适应评估当前训练数据的整体匹配质量。

在此基础上构建权重函数:

$$f(x) = \begin{cases} 1, & x \leq \mu - 0.1 \\ e^{1-\mu}, & \mu - 0.1 < x < \mu \\ e^{1-x}, & x \geq \mu \end{cases} \quad (12)$$

其中,  $x$  表示样本的 IoU。当  $x$  位于边界区间  $(\mu - 0.1, \mu)$  时, 赋予更高权重以强化学习; 对于易区分的负样本, 保持权重为 1; 对于匹配较好的正样本 ( $x \geq \mu$ ), 其权重随着 IoU 增大而逐渐衰减, 以防止模型过拟合。

最终 Slide Loss 定义为:

$$L_{\text{Slide}} = \frac{1}{N} \sum_{i=1}^N f(x_i) \cdot L_{\text{BCE}}(p_i, y_i) \quad (13)$$

$$L_{\text{BCE}}(p, y) = -[y \cdot \log(\sigma(p)) + (1 - y) \cdot \log(1 - \sigma(p))] \quad (14)$$

其中,  $p_i$  为模型对第  $i$  个样本的预测得分,  $y_i$  为对应的真实标签,  $\sigma(\cdot)$  为 sigmoid 函数,  $L_{\text{BCE}}(p, y)$  为二分类交叉熵损失。通过对每个样本先计算该基础损失, 再乘以动态权重  $f(x_i)$ , 最后取平均, 该设计使模型在训练中重点关注 IoU 处于边界区域的困难样本, 有效缓解了样本不平衡对梯度更新的主导问题, 有效提升模型在复杂环境中的鲁棒性与检测精度。

### 3. 实验结果与分析

#### 3.1. 实验平台

所有实验均在配备 NVIDIA GeForce RTX 4090D GPU (24 GB 显存) 和 24 核 AMD EPYC 9754 CPU 的实验平台上进行。实验环境为 CUDA 12.1 和 python 3.10, 深度学习框架为 PyTorch 2.1.0, 所有实验均未使用预训练权重。在训练阶段, 所有输入图像统一缩放至  $640 \times 640$  像素。网络采用 SGD 优化器训练 300 轮, 批大小设为 32, 初始学习率为 0.01。

#### 3.2. 数据集

RUOD 数据集[16]是大规模水下目标检测数据集, 专注于多类别、复杂场景下的目标识别任务。该数据集包含从真实水下环境中采集的 14,000 张高清图像, 涵盖鱼类、潜水员、海星、珊瑚、海龟、海胆、海参、扇贝、鱿鱼和水母等 10 个常见水下目标类别。图像采集自不同水域、光照条件和拍摄距离, 充分涵盖了水下成像中的低对比度、颜色衰减、悬浮粒子干扰以及目标尺度变化剧烈等典型挑战。本文按照 8:1:1 的比例将数据集随机划分为训练集、验证集和测试集。

#### 3.3. 评价指标

为全面评估模型的检测性能, 本文采用精确率(Precision,  $P$ )、召回率(Recall,  $R$ )、平均精度(Average Precision, AP)以及平均精度均值(mean Average Precision, mAP), 参数量及计算量 GFLOPs 作为主要评价指标。其中, 精确率用于衡量预测结果的准确性, 召回率用于反映实际目标被正确检测出的比例:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (15)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (16)$$

对于单一类别, 其检测性能通过平均精度 AP 进行评估, 定义为精确率 - 召回率( $P$ - $R$ )曲线下的积分面积:

$$AP = \int_0^1 P(R) dR \quad (17)$$

为了全面衡量模型在所有类别上的整体性能, 平均精度均值 mAP 定义为所有  $N$  个类别的 AP 的平均值:

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (18)$$

### 3.4. 消融实验

为验证 DMS-YOLO 算法中各改进模块对水下目标检测性能的实际贡献, 本文基于 RUOD 数据集设计并开展消融实验。将所引入的四个模块分别记为: M1 (DySample 动态上采样模块)、M2 (Dynamic Head 动态检测头模块)、M3 (C2PSA\_MSDA 多尺度注意力模块)和 M4 (Slide Loss 损失函数)。实验结果如表 1 所示。

**Table 1.** Ablation experiment

**表 1.** 消融实验

YOLOv11	M1	M2	M3	M4	mAP@0.5/%	mAP@0.5:0.95/%	Params (M)	GFLOPs (G)
√					85.5	62.6	2.6	6.3
√	√				86.7	63.2	2.61	6.3
√		√			86.8	63.4	3.10	7.5
√			√		86.4	63.1	2.62	6.3
√				√	86.0	63.3	2.6	6.3
√	√	√			87.2	63.8	3.11	7.5
√	√	√	√		87.4	64.0	3.13	7.5
√	√	√	√	√	87.5	64.0	3.13	7.5

实验结果表明, 各模块均能带来不同程度的性能提升。引入 DySample 动态上采样模块后, mAP@0.5 由 85.5% 提升至 86.7%, 说明自适应特征重采样能够有效恢复下采样过程中丢失的细节信息, 提升模型在模糊水域中的检测能力。替换 Dynamic Head 后, mAP@0.5 达到 86.8%, 为单模块最优, 表明其在尺度、空间及任务维度的动态建模能够增强模型对复杂场景的适应性。加入 C2PSA\_MSDA 多尺度注意力模块后, mAP@0.5 提升至 86.4%, 说明多感受野特征提取有助于兼顾局部细节与全局信息, 从而提升特征判别能力。在采用 Slide Loss 损失函数后, mAP@0.5 提升至 86.0%, 虽然单模块增益相对较小, 但在多模块组合中表现出良好的协同效应, 表明该损失函数通过动态调整困难样本权重, 能够优化模型对难例的学习能力。

在模块组合实验中, DySample 与 Dynamic Head 联合使用时, mAP@0.5 提升至 87.2%; 进一步引入 C2PSA\_MSDA 后达到 87.4%。当四个模块全部集成时, 模型性能最优, mAP@0.5 和 mAP@0.5:0.95 分别达到 87.5% 和 64.0%。从模型复杂度来看, 完整模型的参数量由基线的 2.6M 增至 3.13M, 计算量由 6.3 GFLOPs 增至 7.5 GFLOPs, 仍保持在轻量化模型范围内。

综合而言, DMS-YOLO 在有限计算开销增加的前提下, 实现了检测精度的稳定提升, 各模块协同作用使模型在水下小目标检测任务中取得精度与复杂度之间的良好平衡, 其可视化示例如图 5 所示。

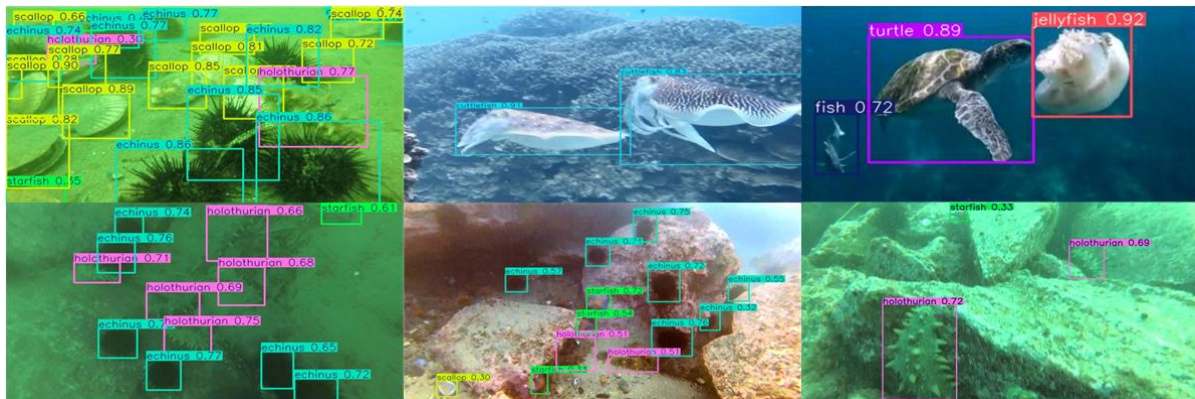


Figure 5. Visualization detection results  
图 5. 可视化检测结果

### 3.5. 对比实验

为全面评估所提出的 DMS-YOLO 算法在模糊水下环境中对小目标检测的有效性, 本文在统一实验条件下基于 RUOD 数据集开展对比实验。选取的主流基线模型涵盖当前先进的轻量化检测器, 包括 YOLOv5n、YOLOv8n、YOLOv10n [17]、YOLOv11n [18]、YOLOv12n [19], 以及基于 Transformer 的 RT-DETR-r18 [20]。各模型在 RUOD 数据集上的对比结果如表 2 所示。

Table 2. Comparison of different algorithms  
表 2. 不同算法的对比实验

算法模型	mAP@0.5/%	mAP@0.5:0.95/%	Params (M)	GFLOPs (G)
RT-DETR-r18	86.9	63.4	20.0	57.0
YOLOv5n	85.0	61.0	2.5	7.1
YOLOv8n	85.3	62.2	3.0	8.0
YOLOv10n	83.6	59.6	2.69	8.2
YOLOv11n	85.5	62.6	2.6	6.3
YOLOv12n	85.4	62.6	2.51	5.8
DMS-YOLO	87.5	64.0	3.13	7.5

在 RUOD 数据集上的定量对比结果表明, 所提出的 DMS-YOLO 在检测性能上整体优于同类算法。在核心指标方面, DMS-YOLO 的 mAP@0.5 达到 87.5%, mAP@0.5:0.95 达到 64.0%, 均为所有模型中的最优结果。相比于当前主流模型, DMS-YOLO 相较 YOLOv11n 和 YOLOv12n 在 mAP@0.5 上分别提升 2.0% 和 2.1%, 在 mAP@0.5:0.95 上也有稳定提升, 表明所提出方法在复杂水下环境下具有更优的特征表达能力和检测鲁棒性。

从模型规模角度分析, DMS-YOLO 的参数量为 3.13M, 虽略高于 YOLOv5n (2.5M)、YOLOv11n (2.6M) 及 YOLOv12n (2.51M), 但仍处于轻量化模型范围内。同时, 其计算量为 7.5 GFLOPs, 略高于 YOLOv11n (6.3 GFLOPs) 和 YOLOv12n (5.8 GFLOPs), 显著低于 YOLOv8n (8.0 GFLOPs)、YOLOv10n (8.2 GFLOPs) 和 RT-DETR-r18 (57.0 GFLOPs) 等复杂模型。其中, 在参数量远小于 RT-DETR-r18 (20.0M) 的情况下, DMS-YOLO 仍在 mAP@0.5 上实现了 0.6% 的性能提升, 体现出更优的参数利用效率和部署潜力。

为进一步分析所提方法在不同类别水下目标上的检测性能, 本文将 DMS-YOLO 与基线模型 YOLOv11n 在 RUOD 数据集中十类水下目标的平均精度(AP)进行了对比分析, 对比结果如表 3 所示。

**Table 3.** Comparison of accuracy of ten types of underwater targets before and after improvement  
**表 3.** 改进前后十类水下目标的精度对比

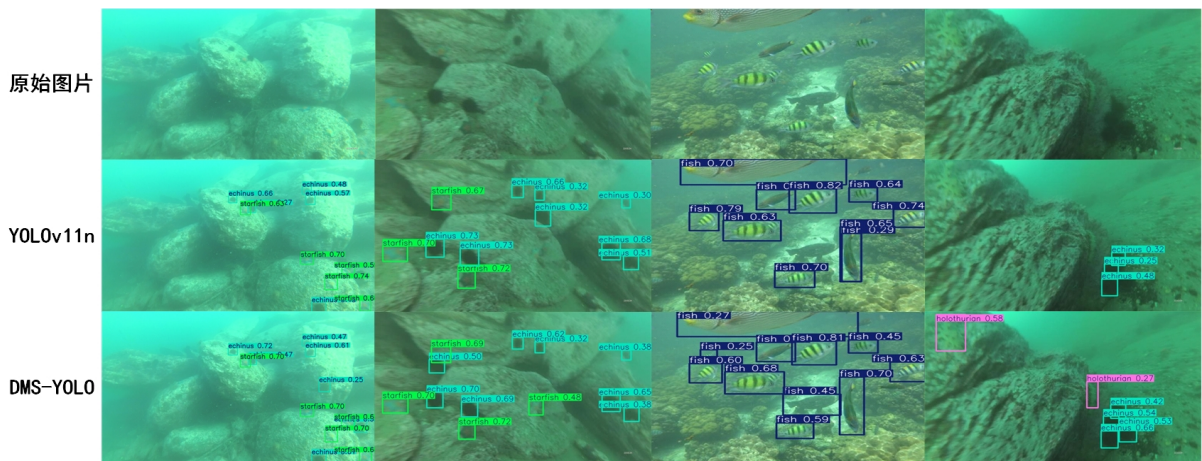
算法模型	AP/%									
	corals	cuttlefish	diver	echinus	fish	holothurian	jellyfish	scallop	starfish	turtle
YOLOv11n	78.7	96.9	93	90.6	76.3	80.6	76.5	78.8	86.6	96.9
DMS-YOLO	80.7	97.9	95.6	91.5	79.8	83.9	79.5	79.7	88.2	97.8

从表中我们可以看出, 对于体型较小或边缘细节较复杂的目标, 改进效果较为明显。例如鱼类(fish)的 AP 由 76.3% 提升至 79.8% (+3.5%), 海参(holothurian)由 80.6% 提升至 83.9% (+3.3%), 水母(jellyfish)由 76.5% 提升至 79.5% (+3.0%)。这类目标通常尺寸较小或边界模糊, 容易受到水下环境噪声和光照变化的影响, 而所引入的动态上采样与多尺度注意力机制在一定程度上增强了细节信息恢复能力和多尺度特征表达能力, 因此检测性能提升较为显著。

对于形态结构较复杂或具有明显轮廓特征的目标, 模型同样表现出稳定提升。例如潜水员(diver)的 AP 由 93.0% 提升至 95.6% (+2.6%), 海星(starfish)由 86.6% 提升至 88.2% (+1.6%), 海胆(echinus)由 90.6% 提升至 91.5% (+0.9%), 表明改进后的模型在复杂形态目标的特征表达方面具有更好的适应能力。

此外, 对于轮廓较清晰且原本检测精度较高的目标, 如乌贼(cuttlefish)和海龟(turtle), 模型仍保持一定提升, 其中乌贼(cuttlefish)的 AP 从 96.9% 提升至 97.9% (+1.0%), 海龟(turtle)从 96.9% 提升至 97.8% (+0.9%)。同时, 珊瑚(corals)与扇贝(scallop)等目标的检测精度也分别提升 2.0% 和 0.9%。总体来看, DMS-YOLO 在不同尺寸和形态特征的水下目标上均表现出稳定的性能改进, 说明所提出的改进策略在复杂水下场景中具有良好的适应性。

综合来看, DMS-YOLO 在有限计算开销的前提下, 实现了检测精度的显著提升, 在精度、参数规模与计算复杂度之间取得了良好的平衡。这表明该方法在水下模糊小目标检测任务中具有较强的适应能力与实际应用潜力。



**Figure 6.** Visualization detection results of different algorithms  
**图 6.** 不同算法的可视化检测结果

### 3.6. 可视化分析

为更加直观地评估各模型的检测效果, 选取水下真实场景下的图像进行可视化对比, 如图 6 所示, 分别展示了原始图像、YOLOv11 的检测结果以及本文所提出的 DMS-YOLO 的检测结果。可以看出 DMS-YOLO 在目标检测效果上表现更优, 能够更准确地定位水下模糊小目标并减少漏检与误检现象, 直观展现了本文算法在水下模糊退化场景下的检测优势与鲁棒性。

## 4. 结论

针对水下目标检测中细节恢复能力不足、自适应调节能力有限、多尺度信息融合不充分以及对模糊边界和困难样本关注不足等问题, 本文提出一种改进的检测模型 DMS-YOLO。首先, 引入 DySample 动态上采样模块, 通过自适应重建采样位置, 在一定程度上改善了下采样过程中细节信息的恢复效果; 其次, 采用 Dynamic Head 动态检测头, 从尺度、空间和任务三个维度对特征进行自适应调节, 以提升模型在复杂场景下的特征表达灵活性; 进一步构建 C2PSA\_MSDA 模块, 在 C2PSA 结构中引入多尺度膨胀注意力机制, 用于融合不同感受野的信息, 从而加强对多尺度目标的表征能力; 最后, 引入 Slide Loss, 通过动态调整样本权重, 使模型在训练过程中更加关注困难样本。在 RUOD 数据集上的实验结果表明, 在仅增加少量参数与计算开销的前提下, DMS-YOLO 的 mAP@0.5 达到 87.5%, mAP@0.5:0.95 达到 64.0%, 相较基线模型取得一定提升。与多种现有方法的对比结果显示, 该模型在检测精度与计算复杂度之间表现出较为平衡的性能。未来工作将进一步结合模型蒸馏与通道剪枝等方法, 以探索在保持检测性能的同时进一步降低模型复杂度的可能性。

## 参考文献

- [1] Xu, S., Zhang, M., Song, W., Mei, H., He, Q. and Liotta, A. (2023) A Systematic Review and Analysis of Deep Learning-Based Underwater Object Detection. *Neurocomputing*, **527**, 204-232. <https://doi.org/10.1016/j.neucom.2023.01.056>
- [2] Fan, Y., Zhang, L. and Li, P. (2024) A Lightweight Model of Underwater Object Detection Based on YOLOv8n for an Edge Computing Platform. *Journal of Marine Science and Engineering*, **12**, Article 697. <https://doi.org/10.3390/jmse12050697>
- [3] Wang, X., Gao, H., Jia, Z. and Li, Z. (2023) BI-YOLOv8: An Improved Road Defect Detection Model Based on YOLOv8. *Sensors*, **23**, Article 8361. <https://doi.org/10.3390/s23208361>
- [4] Liu, Y., Huang, Z., Song, Q. and Bai, K. (2025) PV-YOLO: A Lightweight Pedestrian and Vehicle Detection Model Based on Improved YOLOv8. *Digital Signal Processing*, **156**, Article 104857. <https://doi.org/10.1016/j.dsp.2024.104857>
- [5] Girshick, R. (2015) Fast R-CNN. 2015 *IEEE International Conference on Computer Vision (ICCV)*, Santiago, 7-13 December 2015, 1440-1448. <https://doi.org/10.1109/iccv.2015.169>
- [6] Ren, S., He, K., Girshick, R. and Sun, J. (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1137-1149. <https://doi.org/10.1109/tpami.2016.2577031>
- [7] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., et al. (2016) SSD: Single Shot MultiBox Detector. In: Leibe, B., Matas, J., Sebe, N. and Welling, M., Eds., *Lecture Notes in Computer Science*, Springer International Publishing, 21-37. [https://doi.org/10.1007/978-3-319-46448-0\\_2](https://doi.org/10.1007/978-3-319-46448-0_2)
- [8] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 779-788. <https://doi.org/10.1109/cvpr.2016.91>
- [9] Liu, Y., An, D., Ren, Y., Zhao, J., Zhang, C., Cheng, J., et al. (2024) DP-FishNet: Dual-Path Pyramid Vision Transformer-Based Underwater Fish Detection Network. *Expert Systems with Applications*, **238**, Article 122018. <https://doi.org/10.1016/j.eswa.2023.122018>
- [10] Liu, K., Sun, Q., Sun, D., Peng, L., Yang, M. and Wang, N. (2023) Underwater Target Detection Based on Improved YOLOv7. *Journal of Marine Science and Engineering*, **11**, Article 677. <https://doi.org/10.3390/jmse11030677>
- [11] Qu, S., Cui, C., Duan, J., Lu, Y. and Pang, Z. (2024) Underwater Small Target Detection under YOLOv8-LA Model.

- Scientific Reports*, **14**, Article No. 16108. <https://doi.org/10.1038/s41598-024-66950-w>
- [12] Liu, W., Lu, H., Fu, H. and Cao, Z. (2023) Learning to Upsample by Learning to Sample. 2023 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, 1-6 October 2023, 6027-6037. <https://doi.org/10.1109/iccv51070.2023.00554>
- [13] Dai, X., Chen, Y., Xiao, B., Chen, D., Liu, M., Yuan, L., *et al.* (2021) Dynamic Head: Unifying Object Detection Heads with Attentions. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 7373-7382. <https://doi.org/10.1109/cvpr46437.2021.00729>
- [14] Jiao, J., Tang, Y., Lin, K., Gao, Y., Ma, A.J., Wang, Y., *et al.* (2023) DilateFormer: Multi-Scale Dilated Transformer for Visual Recognition. *IEEE Transactions on Multimedia*, **25**, 8906-8919. <https://doi.org/10.1109/tmm.2023.3243616>
- [15] Yu, Z., Huang, H., Chen, W., Su, Y., Liu, Y. and Wang, X. (2024) YOLO-FaceV2: A Scale and Occlusion Aware Face Detector. *Pattern Recognition*, **155**, Article 110714. <https://doi.org/10.1016/j.patcog.2024.110714>
- [16] Fu, C., Liu, R., Fan, X., Chen, P., Fu, H., Yuan, W., *et al.* (2023) Rethinking General Underwater Object Detection: Datasets, Challenges, and Solutions. *Neurocomputing*, **517**, 243-256. <https://doi.org/10.1016/j.neucom.2022.10.039>
- [17] Chen, H., Chen, K., Ding, G., Han, J., Lin, Z., Liu, L., *et al.* (2024) YOLOv10: Real-Time End-to-End Object Detection. *Advances in Neural Information Processing Systems*, **37**, 107984-108011. <https://doi.org/10.52202/079017-3429>
- [18] Khanam, R. and Hussain, M. (2024) YOLOv11: An Overview of the Key Architectural Enhancements. arXiv:2410.17725.
- [19] Tian, Y., Ye, Q. and Doermann, D. (2025) YOLOv12: Attention-Centric Real-Time Object Detectors. arXiv:2502.12524.
- [20] Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., *et al.* (2024) DETRs Beat YOLOs on Real-Time Object Detection. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 16-22 June 2024, 16965-16974. <https://doi.org/10.1109/cvpr52733.2024.01605>