

`push0 g 0 G`pop FishDet-M: 用于在多样化水生视觉领域中实现鲁棒鱼类检测和 CLIP 引导的模型选择的统一大型基准

Muayad Abujabal, Lyes Saad Saoud, and Irfan Hussain

`push0 g 0 G`pop

Abstract—在水下图像中准确检测鱼类对于生态监测、水产养殖自动化和机器人感知至关重要。然而，由于数据集的分散性、图像条件的多样性以及评估协议的不一致，实际部署仍然有限。为了解决这些问题，我们提出了 FishDet-M，这是针对鱼类检测的最大统一基准，包含了 13 个公开可用的数据集，涵盖从海洋到半咸水、遮挡场景和水族馆场景在内的多种水生环境。所有数据都使用 COCO 格式的注释进行了统一，包括边界框和分割掩码，从而实现一致且可扩展的跨领域评估。我们系统地对 28 种当代目标检测模型进行基准测试，覆盖了从 YOLOv8 到 YOLOv12 系列、基于 R-CNN 的检测器以及基于 DETR 的模型。评估采用标准指标，包括 mAP、mAP@50 和 mAP@75，以及特定尺度分析 (AP_S 、 AP_M 、 AP_L) 和推理性能分析 (包括延迟和参数数量)。结果显示了在 FishDet-M 上训练的不同模型的检测性能差异，以及不同架构模型在准确性和效率之间的权衡。为支持自适应部署，我们引入了基于 CLIP 的模型选择框架，该框架利用视觉-语言对齐动态识别与每个输入图像最语义相关的检测器。这种零次选择策略无需集成计算即可实现高性能，为实时应用提供了可扩展的解决方案。FishDet-M 建立了一个标准化和可重复的平台，用于评估复杂水域场景中的目标检测。所有数据集、预训练模型和评估工具均公开可用，以促进未来在水下计算机视觉和智能海洋系统方面的研究。

Index Terms—fish detection, underwater vision, object detection, benchmark dataset, CLIP-based model selection, YOLO, deep learning, marine robotics, ecological monitoring

I. 简介

在水下环境中准确检测鱼类依然是一个主要挑战，因为浑浊、遮挡、光散射和动态背景导致的视觉退化。这些条件降低了对比度，扭曲了颜色，并引入了杂乱，削弱了基于视觉的检测模型的可靠性。来自植被、岩石或密集鱼群的遮挡进一步遮蔽了目标。尽管深度学习推动了水下感知的发展，其效果受到碎片化数据集的限制，这些数据集往往缺乏多样性、一致的注释格式和复杂真实世界条件的表现。

为了解决这些限制，一些特定领域的数据集应运而生。SmallFish 数据集专注于在能见度差的情况下检测小目标 [1]，DUFish 捕捉了密集的群体行为 [2]，而 DePondFi 则引入了自然条件下具有挑战性的池塘鱼类影像 [3]。同时，检测模型也在发展以应对水下的视觉失真。FishDet-YOLO 整合了增强模块以应对低对比度目标 [4]，YOLOv8 TF 引入了变压器增强的细化和类别敏感学

习 [5]，而 IDLAFD UWSN 框架利用混合架构在模拟遮挡下提高检测能力 [6]。

尽管这些进展取得了显著成效，但该领域仍然缺乏能够在真实水域环境中支持大规模和跨领域评估的统一基准。为了弥补这一差距，我们引入了 FishDet-M，这是一项通过将 13 个公开可用的数据集统一成符合 COCO 协议的单一注释格式而构建的整合基准。FishDet-M 涵盖各种水下环境，包括珊瑚礁、水产养殖池和半咸水区，提供了 296,885 个标注的鱼类实例，分布在 105,000 张图像中。该基准支持对 28 种领先检测模型的评估，包括 YOLO 变体、基于 DETR 的架构和区域提议网络。

本研究的贡献总结如下：

`push0 g 0 G`pop FishDet-M 基准：我们引入了一统一的鱼类检测基准数据集，该数据集将 13 个公开的水下数据集整合为一种与 COCO 标准兼容的注释格式。该数据集涵盖了多样的环境、物种和真实条件。

`push0 g 0 G`pop 检测模型的综合评估：我们评估了包括 YOLOv8 至 YOLOv12、基于 DETR 的模型和 R-CNN 模型在内的 28 个先进目标检测架构。评估使用了标准化指标，如平均精度均值 (mAP)、尺度感分数 (AP_S 、 AP_M 、 AP_L)、推理时间和参数数量。

`push0 g 0 G`pop 基于 CLIP 的自适应模型选择：我们提出了一种上下文感知机制，利用 CLIP 的视觉-语言对齐功能，根据输入图像内容自动选择最适合的检测模型，实现动态和稳健的推理。

`push0 g 0 G`pop 性能洞察和部署指南：我们分析了模型在遮挡和能见度差等条件下的鲁棒性，并提供可操作的建议，以选择适用于实时生态监测、水产养殖管理和机器人部署的模型。

`push0 g 0 G`pop 公共发布以支持可重复性：完整的数据集、源代码、检查点和评估工具通过我们的 GitHub 仓库公开发布，以促进可重复研究并推动进一步的发展。

II. 相关工作

A. 水下鱼类检测的挑战与解决方案

由于海洋环境复杂的光学特性，水下鱼类检测仍然是一项具有挑战性的任务。浑浊、光散射和遮挡会降低能见度并引入背景杂乱 [7], [8]，这使得准确检测变得困难，尤其是对于可变形和低对比度的目标。

为减轻这些影响，最近的方法结合了水下特定图像增强技术，如光照校正和颜色还原，并与改进的深度学习框

`push0 g 0 G`pop

This work was conducted at the Khalifa University Center for Autonomous Robotic Systems (KUCARS), Abu Dhabi, United Arab Emirates.

The authors are with Khalifa University, Abu Dhabi, UAE (e-mails: 100057733@ku.ac.ae; lyes.saoud@ku.ac.ae; irfan.hussain@ku.ac.ae).

`push0 g 0 G`pop

`push0 g 0 G`pop

figures/fish_app.png

figures/fish_app.png
`push0 g 0 G

Fig. 1: 用于比较 FishDet-M 上检测输出的交互式应用界面。支持多种模型、边界框叠加和阈值控制。

`push0 g 0 G
`pop`pop

器结合使用 [9], [4]。然而，在遮挡严重或视觉复杂度高的场景中，性能仍然会下降 [10]。

数据集的限制进一步加剧了这些问题。许多现有的数据集集中于具有有限物种多样性或环境变化的受控设置 [1], [11]。较新的数据集如 DUFish [2] 和 FishTrack23 [12] 提高了多样性和现实性，但更广泛的数据集环境仍然是碎片化的。表 ?? 总结了代表性的数据集。

深度学习通过 YOLO、Cascade R-CNN 和 DETR [14], [15], [16] 等模型提高了检测性能，得益于注意力模块、可变形层和领域适应性 [4], [17], [5]。基于混合卷积和

transformer 的架构进一步提高了鲁棒性 [5]，而在增大数据集上的微调有助于推广到未见过的情况。

由于遮挡、伪装和种间变异性，挑战依然存在 [13]，解决这些问题需要具备上下文感知的训练和强大的推理策略。评估协议正在扩展，超越平均精度和交并比，还包括规模感知指标、延迟和模型大小 [1], [2]，支持实时部署。

尽管取得了进展，跨领域的变化和不一致的基准限制了泛化能力 [19]。克服这些问题对于在不同的海洋环境中实现可靠的鱼类检测仍然是至关重要的。

B. 鱼类检测数据集与基准

检测模型的有效性在很大程度上依赖于多样化且标注良好的数据集。许多早期的数据集仅限于清晰或浅层环境，限制了其普适性。现实世界中的水生场景在浑浊度、光照和物种组成上差异显著，因此需要多样化的基准进行鲁棒评估 [20], [1]。

最近的努力通过扩大生态和视觉的多样性解决了这些限制。DePondFi [20] 提供了带注释的池塘图像，其中 YOLO 模型实现了高的 mAP@50 得分。Fish4Knowledge 支持海洋环境中的检测和分类，FishNet 检测器实现了超过 92% 的检测精度 [21]。SmallFish 专为在浑浊环境中进行小物体检测而设计 [1]。

FishTrack23 [12] 强调跨栖息地的密集跟踪，而 LifeCLEF 2014 [22] 则作为鱼类识别的基准。其他数据集如 UOMT [23] 和 FS48 [24] 解决显著目标检测和重识别的问题。OzFish [25], [26] 包含针对澳大利亚海岸鱼类的物种级别注释。

表格 ?? 提供了这些数据集的比较概述，重点展示了环境、任务和公开可用性。尽管它们做出了贡献，该领域仍然缺乏统一的标注标准、基准协议和跨研究的可比性。即将推出的数据集，如 FishDet-M，旨在通过结构化元数据和现实的测试条件来应对这些挑战。

C. 用于水下物体检测的深度学习方法

深度学习显著提升了水下探测性能，尤其在图像质量因浑浊、遮挡和光照变化而下降的环境中 [4], [17], [25]。像 YOLO [5]、Faster R-CNN [27] 和 DETR 这样的基础模型已被广泛应用于海洋探测工作流程。

为了解决视觉局限性，研究人员引入了注意力机制、可变形层和融合模块，以增强部分可见目标的特征表示。Transformer 已被结合到卷积骨干网络中，从而在低对比度条件下实现远程特征建模 [5], [6]。

专门的变体，如 FishDet-YOLO 和 YOLOv8-TF，专注于水下数据分布并解决类别不平衡问题 [4], [5]。域适应技术，包括对抗学习和风格迁移，进一步帮助弥合合成和真实世界数据集之间的性能差距 [28]。

融合卷积和变压器组件的混合模型利用了 CNN 的空间定位优势和变压器的语义上下文建模能力。这种结合提高了在复杂、遮挡和动态的海洋场景中检测的准确性 [5]。

这些架构的设计越来越注重部署。它们的应用范围包括生态监测和渔业管理到机器人探索以及实时鱼类追踪 [29], [30], [31]。

D. 遮挡处理和物种多样性

遮挡是水下场景中的一个持久挑战，特别是在拥挤的栖息地或礁区。解决方案包括 3D 几何建模、基于平面镜的遮挡缓解，以及旋转边界框回归 [32], [33]。增强型 YOLO 模型的 Transformer 和使用排斥损失的方法在重叠目标上也表现出了改进的性能 [13]。

使用 DCGAN 和 UIEGAN 等模型进行的生成式增强能够提供富含遮挡的样本，增强训练的多样性 [34], [35]。时间跟踪方法通过在不同帧之间关联被遮挡的物体进一步提高了稳健性 [36]。

物种的多样性增加了复杂性，这是由于形态和色彩的多样性。细粒度模型使用注意力机制、局部化补丁和多分支架构来检测物种间细微的差异 [37], [38]。伪装物种利用背景匹配或偏振光，仍然具有特别的挑战性 [39]。

采用主动检测和少样本学习等高级策略，能够通过少量的数据高效地适配模型 [40]。像 MobileNetV2 这样的轻量级模型，当与语义模块进行微调时，在性能和计算负担之间提供了令人满意的权衡 [41]。

表 ?? 概述了检测系统中处理遮挡和物种水平变化的关键方法。

E. 评估指标和基准测试协议

水下检测的可靠基准测试依赖于一致和全面的评价标准。标准指标如 mAP、IoU、精度和召回率常被报告 [44]，但在实现和数据集使用上的变动常常阻碍再现实性 [45]。

为了针对不同目标尺度进行检测，研究人员采用了 APM 和 APL 指标，这些指标反映了在小型、中型和大型目标上的性能。这些指标对于理解浑浊条件下的行为至关重要，因为在这种情况下，更小的鱼更容易被遗漏 [47]。

实际部署还需要运行时指标，如推理时间、内存使用、模型大小。这些指标决定了模型是否可以在计算资源有限的嵌入式设备或水下机器人上实现。[48]。

图像质量评估指标，包括 UWEQM [49] 和 QDE [4] 也很重要。这些指标有助于量化增强和恢复步骤对整体检测可靠性的影响。最近的多曝光和基于融合的方案提供了更丰富的评估，结合了对比度、色度和结构一致性 [50]。

标准化基准测试协议和数据集划分将提高各研究之间的可比性。新 efforts 应该旨在将性能指标与部署考虑和现实质量评估相结合，以指导现实世界海洋应用中的模型选择。

F. 在水产养殖、机器人技术和海洋监测中的应用

鱼类检测模型在水产养殖、机器人和环境监测中发挥着至关重要的作用。在水产养殖中，深度学习使得在浑浊、低光条件下实现自动化的鱼类计数、尺寸估计和健康评估成为可能 [51], [52], [53]。轻量级模型如 AquaYOLO 和 CUIB YOLO [54], [55] 被优化以用于嵌入式部署，并集成到智能喂养和水质监测系统中 [56]。

在水下机器人领域中，鱼类检测支持导航、障碍物检测和目标采样，这对于自主任务如珊瑚礁调查、污染跟踪和海洋测绘至关重要 [57], [58], [59], [60]。最近的研究也强调了使用语义分割和检测管道对水产养殖基础设施进行检测和缺陷分割 [61]。

在保护方面，检测模型可以帮助进行生物多样性跟踪、海洋养殖区映射和生态系统健康监测。与无人机和卫星图像的结合使得能够检测出由环境变化和人类活动驱动的趋势 [62], [63], [64]。

尽管有所进展，物种、形态和水体条件的多样性仍然挑战着稳健性。模型压缩、领域适应和多传感器融合方面的努力正在进行中。像 AASNet 和量化的 YOLO 变体这样的高效模型在海洋环境边缘部署中具有很好的前景 [66]。这些应用展示了鱼类检测在可持续水产养殖、自主探索和生态评估中的日益重要性 [19], [51], [64], [67]。

III. 数据记录

FishDet-M 基准包含 105 , 556 张图像和 296 , 个已标注的鱼实例，使用分层抽样将其划分为训练集、验证集和测试集，以确保在栖息地类型、可见性条件和尺度方面的平衡表征。每个部分的关键统计数据总结在表 ?? 中。

FishDet-M 显示出广泛的视觉异质性, 图像分辨率从 78×53 到 4608×3456 像素不等 (平均值为 715×468), 实例密度从单条鱼到每帧多达 256 只鱼的密集聚集 (平均值为每张图片 2.81 个实例)。

图 ?? 中展示了对对象级统计数据的可视化。图 ?? 中的边框面积分布显示了一条右偏曲线, 以小型和中型对象为主。图 ?? 中的纵横比在 1 到 2 之间达到峰值。图 ?? 中每幅图像的实例直方图, 采用对数刻度绘制, 揭示了一种长尾模式, 其中大多数图像包含很少的鱼, 而一个较小的子集则包括密集的鱼群。

图像层级的视觉属性总结在图 ?? 中。图 ?? 中的图像分辨率直方图展示了空间覆盖范围的多样性。图 ?? 中的 RGB 强度直方图揭示了大多数场景的光线较暗, 并且以绿色-蓝色色调为主, 这是水下拍摄 [68] 的典型特征。图 ?? 中的物体对图像分辨率热图, 以对数-对数比例显示, 强调了在低分辨率和高分辨率图像中小物体的普遍存在性。

图 ?? 中突出了额外的水下特有成像特征。图 ?? 中的饱和度直方图显示了沉闷色调的主导地位, 与光吸收和散射一致。图 ?? 中的对比度分布是通过灰度标准差计算的, 确认了整体对比度低。图 ?? 中的通道比率图说明了相对于红色的强蓝色和绿色偏倚, 证实了水生环境中典型的光谱失真 [68]。

图 ?? 提供了额外的空间洞察。图 ?? 中的平均图像强度直方图可以作为能见度因霾和背散射而下降的代理。图 ?? 中的边界框中心热力图揭示了对象放置的中心偏差, 反映了人工操作数据采集中的典型框架倾向。

为了促进探索性分析和模型透明性, 我们开发了一个如图 1 所示的交互式桌面 GUI。该工具支持对基于 YOLO 和基于 transformer 的模型检测输出进行比较评估, 支持调整置信度阈值并叠加预测的边界框。额外的元数据如 mAP 得分和推理时间实时显示, 帮助研究人员识别模型在不同水域条件下的适用性。

IV. 方法

数据集聚合与标注。FishDet-M 整合了 13 个水下鱼类检测数据集, 结合了公开可用的资源库和数据集。数据来源于珊瑚礁、水族馆、沿海区域和河口地带, 包括静态图片和视频帧。所有标注均被标准化为 COCO 格式 [69], 且边界框被统一为一个与物种无关的鱼类类别。这确保了在不同格式和命名规范下的标注完整性。在质量控制过程中, 去除了无效或模糊的标注。

环境和视觉多样性。FishDet-M 涵盖了广泛的现实世界条件, 包括清澈和浑浊的水域、浅水和深水场景、人工和自然光照、运动模糊以及物体遮挡。这些变化模拟了实际部署场景, 并提高了模型的鲁棒性。该数据集还包括陆地和实验室视图, 以扩展多样性并评估领域转换。

协调与质量保证。数据集整合涉及坐标格式统一、框体尺寸验证, 以及对缺失或错误注释的人工校正。所有边界框均遵循 COCO 格式 $x_{\min}, y_{\min}, \text{width}, \text{height}$ [69]。严格的验证步骤确保了数据集的一致性, 并排除了损坏的样本。

分区策略。一个源感知分层拆分保持了每个数据集在训练集 (80%)、验证集 (10%) 和测试集 (10%) 中的比例代表性。这保留了较小数据集的生态和视觉多样性, 并防止了来自像 FishNet [70] 这样的大贡献者的偏见。

数据集选择标准。选择是基于八个标准: 任务复杂性、模态丰富性、体量、注释质量、生态多样性、视觉挑战性和部署准备度。代表性的数据集包括 DeepFish [25]、`push0 g 0 G` pop30-38 FPS。

FishNet [70]、Brackish-MOT [29] 和 TrashCan 1.0。源数据集的完整列表在表 ?? 中列出。移除了非物理性或重复的帧, 以集中于注释的鱼实例并防止数据泄露。为了可视化代表性, 图 ?? 提供了一个四面板气泡图, 对比了关键维度 (如任务复杂性与模态丰富性), FishDet-M 始终位于右上象限, 突出了其多样性、注释体量和面向现实世界部署的准备度。

基于 CLIP 的模型选择。我们集成了一个基于 CLIP 的模块, 利用图像嵌入和提示文本之间的语义相似性, 自动从 28 个候选检测模型中选择最合适的模型, 如图 ?? 所示。该机制可以在不进行手动选择或使用启发式方法的情况下实现上下文感知的推理。

技术验证。我们对 28 个模型进行了基准测试, 包括 YOLOv8-v12 [14], [71], [72], [73]、YOLO-NAS [74]、Cascade R-CNN [15]、Sparse R-CNN [75]、DETR [16], [76], [77]、RetinaNet [78] 和 MobileNetv2-S [79]。模型在 FishDet-M 上使用默认超参数进行训练。评估采用了 mAP、mAP₅₀、mAP₇₅、AP_S、AP_M、AP_L、精度-召回曲线和 F1 分数。通过 FPS、模型大小和延迟评估效率。

硬件和软件。所有实验均在 Windows 11 系统下, 使用 RTX 4090 GPU 和 Intel i9-14900K CPU 以及 64GB RAM 运行。代码库包括 PyTorch [80]、Ultralytics [81] 和 MMDetection [82]。所有检测指标均使用 pycocotools [83] 进行混合精度评估计算。

V. 结果与讨论

本节对水下鱼类检测的目标检测模型进行了全面评估, 提出了定量的性能指标和定性的模型行为分析。我们在统一的 FishDet-M 基准上评估模型性能, 检验其效率特性, 并研究其在各种子测试划分上的泛化能力, 以反映不同水下条件。

A. 定量评估

我们在 FishDet-M 基准测试中评估了 28 种目标检测模型, 以分析各种架构的性能和计算效率。表 ?? 展示了精度、特定尺度的检测和推理指标。

YOLO 系列始终实现了最高的检测准确度。YOLOv8 达到了 0.491 的 mAP, 紧随其后的是 YOLO12l (0.484) 和 YOLO11l (0.484)。这些模型在小型、中型和大型物体尺度 (AP_S, AP_M, AP_L) 上表现出强劲的性能, 证实了它们对于不同鱼类大小的适应性。YOLO-NAS-1 也提供了强劲的结果 (0.470 mAP), 并且推理效率高。

轻量化模型如 YOLOv8n 和 YOLOv10n 提供了绝佳的速度与准确性平衡。YOLOv8n 仅使用 320 万参数便能达到 251 FPS 的速度, 使其成为实时和资源受限场景的理想选择。

相比之下, 包括 DETR、RT-DETR-l 和 Deformable DETR 在内的基于变压器的模型显示出较低的 mAP (0.317-0.390), 在小物体检测中表现不佳, 并且延迟较高 (15 到 45 毫秒), 这使它们在水下环境中的实时应用不太合适。

区域建议模型如 Cascade R-CNN (0.449 mAP) 提供了可靠的准确性但速度较慢。单阶段的替代方案如 RetinaNet (0.448 mAP) 和 FCOS (0.475 mAP) 在效率上更优,

1) 精确率-召回率分析: 在图 ?? 中的精确率-召回率曲线显示, 大多数模型在开始时具有高精确率, 随着召回率的增加而下降。MobileNetV2-SSD 在召回率为 0.4 时急剧下降, 其他模型如 Sparse R-CNN 和 Deformable-DETR 也表现出更早的退化。在 $\text{IoU}=0.75$ 的图 ?? 中, 精确率开始较低并且更早下降, 特别是对于变换器模型。

如图 ?? 和 ?? 所示, IoU 阈值为 0.50 和 0.75 时的 F1 分数趋势。YOLO 变体早期达到峰值并平稳下降, 而 Sparse R-CNN 和 DETR 则在较晚时候达到峰值。Deformable-DETR 显示出较平缓的曲线, 并早期下降。在更严格的 IoU 阈值下, 各模型表现出更一致的模式, 其中 FCOS 和 MobileNetV2-SSD 在超过 0.6 置信度后急剧下降。

3) 对于测试集的评估: 为了更好地理解每个组成数据集的贡献, 我们对十三个源数据集的单独测试分区进行了模型性能的详细评估。结果表明, 根据数据集的复杂性, 模型表现出显著的变化。例如, YOLO12x 在 DeepFish [25] 上达到了 0.846 的 mAP, 该数据集具有水下场景, 并包含显著的大型鱼实例。相比之下, 同一模型在 FishDataset [11] 上得分为 0.371, 这里的鱼在水箱中重叠和遮挡, 并且外观相似, 造成了显著的视觉挑战。在整个基准测试中, FishDet-M 对最高性能的模型表现一致, 在 0.481 到 0.491 之间 (见表 ??), 展示了它在多样化视觉领域中的泛化能力。

根据环境特征, 各个数据集可大致分为三个性能层级。高性能层级的数据集, 如 DeepFish [25] 和 FishVideo [86], 通常包括大鱼、背景干扰最小且光照条件稳定。中等层级包括如 FishNet [70] 和 Fish4Knowledge [84] 的数据集, 这些数据集涉及可变的鱼类尺寸、混合光照条件和中等遮挡。最具挑战的层级包括如 TrashCan 1.0 [30] 和 Brackish-MOT [29] 的数据集, 在这些数据集中, 视觉噪声、低图像质量、高浑浊度和频繁遮挡显著阻碍了检测准确性。每个数据集上所有评估模型的详细指标已在补充材料中提供 (表 SI 到 SXIII)。这些结果证实, 虽然 FishDet-M 在任何单个数据集上没有达到最高准确性, 但它在不同领域中一致促进了可推广特征学习。这进一步通过各子集间的模型性能方差较小得以证明。平均精确值为 -1 表示在相应子测试分割中, 根据尺寸阈值没有符合条件的对象存在。

B. 定性和泛化分析

除了定量指标外, 视觉检查和跨域测试还可以提供有关模型在不同水下条件下的稳健性进一步的见解。图 ?? 显示了 28 个模型在四幅具有代表性的图像上的输出, 这些图像是为了突出常见的失败场景而选择的。

在伪装条件下, 当鱼类与背景相似时, 大多数模型由于对比度低而产生不准确或过大的边界框。珊瑚或重叠结构的遮挡导致检测遗漏或检测结果破碎, 特别是在背景纹理与鱼体相似的情况下。小物体检测尤其具有挑战性, 有限的像素占用经常导致假阴性。在多云且可见度降低的密集场景中, 模型常常将实例合并或无法精确定位目标, 显示出在拥挤环境中的局限性。

为了评估泛化能力, 我们在一个由 1500 幅来自不同来源的未见图像组成的单独数据集上测试了所有模型。如表 ?? 所示, 来自 YOLO 家族的顶级模型, 包括 YOLO12l 和 YOLO11x, 在各种目标尺度上保持了高精度 (mAP 约为 0.63)。轻量级变种如 YOLOv8n 和 YOLOv10n 也表现出可靠的性能。相比之下, 基于变压器的检测器如 Deformable DETR 的准确率明显下降, 尤其是在小物体

方面, 而像 Faster R-CNN 这样的区域提议模型则显示出中等的可迁移性。单阶段检测器如 RetinaNet 和 FCOS 在领域转换下表现得更为稳定。

这些视觉和跨数据集的评估共同证实了在 FishDet-M 上训练的模型的适应能力, 并突出显示了持续存在的限制, 包括伪装、遮挡、小物体检测和密集场景分离。这些挑战是在自然水下环境中可靠鱼类检测的开放性挑战。

C. 基于 CLIP 的模型选择分析

我们分析了 CLIP 引导的模型选择器的行为, 以了解其在 FishDet-M 数据集上的偏好。如图 ?? 所示, YOLO 是被选择最多的模型, 其次是 YOLOv8l、YOLOv12m 和 YOLOv11n。这些模型很可能由于更强的特征表示和更好地与 CLIP 的语义线索对齐而占据优势。相比之下, 较轻或较轻的模型 (例如 YOLOv10m、YOLOv8m) 被选择频率较低。

为了进一步分析这一趋势, 我们在图 ?? 中展示了每个模型的 CLIP 相似性分数分布。YOLOv8x、YOLOv8l 和 YOLOv12m 一贯获得更高的中位数分数, 并且其四分位距较小, 表明语义对齐可靠。分数分布较广或较低的模型不太受欢迎, 反映了 CLIP 优先考虑稳定语义接近性的倾向。

这些发现证实, CLIP 选择器并非随机作用, 而是倾向于选择具有一致视觉-语言对应关系的模型。这支持了通过语言驱动模型路由来适应多样的水下条件的可行性, 如近期在使用 CLIP 相似性 [93] 的偏差感知水下人工智能研究中所展示的。用于 CLIP 指导选择的具体提示在补充材料中提供, 以实现完全的可重复性。

D. 讨论与见解

对 FishDet-M 的评估强调了 YOLO 系列的一贯性能优势, 特别是 YOLO12x, 在不同目标尺度上提供了很强的精度, 并且在适合实时海洋应用的推理上表现高效 [73], [48]。轻量级的 YOLO 变体也表现良好, 有效地平衡了速度和准确性。

基于 Transformer 的检测器, 如 DETR、Deformable DETR 和 RT DETR-l, 性能有限, 尤其是在小物体检测上并且引发较高的延迟 [16], [77], [76]。基于区域的模型, 如 Cascade R-CNN 以及单阶段检测器如 RetinaNet 和 FCOS [15], [78], [91], 实现了较为适中的准确性且效率高, 但未能达到 YOLO 的最佳结果。

F1 分数和精度召回分析表明, 在较低的 IoU 阈值下, YOLO 模型表现相似, 但在更严格的匹配条件下 (IoU 0.75), 轻量级模型和 Transformer 模型对定位要求更为敏感。这强调了需要在不同阈值下评估模型, 以评估其可靠性。

对个别数据集的评估显示出从鱼类数据集 [11] 的 0.371 mAP 到 DeepFish [25] 的 0.846 的广泛性能变化。这反映了水下条件的多样性。合并的 FishDet-M 基准提供了平衡的训练分布, 使模型具有普遍适用性, 避免了对特定领域的过拟合。

在未见数据集上进行的泛化测试证实了这种鲁棒性。许多 YOLO 变体在新数据上的表现相当或更好, 这进一步强调了多样化的、综合训练集在实际水下应用中部署的价值。[92]

我们还评估了一个自适应模型选择系统, FishDet-M 使用 CLIP, 它使用 CLIP 相似性来选择在上下文上合适的模型。尽管它的准确性 (mAP 0.444) 和速度 (80 FPS)

有超过最好的 YOLO 模型，但它超越了许多传统和基于变压器的模型，验证了语言引导的模型路由在可变条件下的可行性。

定性结果显示了一些失败案例，包括在伪装场景中的定位不准确、在遮挡情况下的漏检，以及对小或模糊物体处理不佳。在珊瑚和复杂结构周围的误报也持续存在，强调了在水下条件下进行细粒度判别和上下文感知检测的持续挑战。

E. 数据集和伦理

FishDet-M 是一个用于在各种水下环境中检测鱼类的大规模基准测试 [14]。它将 13 个公开可用并获得许可的数据集组合为一个鱼类检测数据集，包括海洋、淡水及水族馆环境。此整合解决了之前数据集结构和评估中的不一致性问题 [20]。

该数据集包括 105,556 张图像和 296,885 个鱼的标注，具有用于训练、验证和测试的分层拆分，以保持栖息地、清晰度和密度的变化。图像大小范围从 78×53 到 4608×3456 像素，平均为 711×465 ，每张图像中的鱼类数量从 1 到 256 不等。标注被归为不区分物种的鱼类类别，以优先考虑一般检测，同时允许后续分类。一个彻底的质量保证过程确保了格式一致性、标签标准化以及去除了损坏或模糊的数据。

所有数据均来自于符合伦理的存储库 [25], [89], [70], [90], [85]。发布旨在推动海洋科学、水产养殖和保护领域的可重复性研究 [2], [94]。

VI. 限制和未来方向

尽管 FishDet-M 非常全面，它仍然有一些局限性。实际的水下环境仍然高度多变，极端混浊、动态光照和杂乱的背景对当前的检测模型提出了显著的挑战 [1], [29]。未来的数据集可能会通过合成增强或多传感器数据的结合来提高在这些条件下的鲁棒性。

目前，FishDet-M 专注于目标检测和定位。然而，下游应用如行为分析和特定物种跟踪需要更丰富的标注，包括姿态、个体身份和活动标签 [95]。扩展标注范围将能够提供更深入的生物学见解。

虽然基于 YOLO 的模型在处理小型或遮挡的鱼类时表现优于其他模型，但在复杂场景中仍存在局限性 [13]。多尺度特征融合和上下文感知检测架构可以帮助缓解这些问题，并在遮挡和尺度变化下提高鲁棒性。

泛化分析进一步强调了多样化训练数据的必要性。尽管在 FishDet-M 上训练的模型在外部基准测试中表现良好，广泛地包括地理区域、水体类型和深度将增强其全球适用性。

在自适应推理方面，我们的 CLIP 引导的模型选择器通过动态将输入引导到合适的检测器展示了潜力。然而，这一策略引入了计算开销，与顶尖的固定模型相比，精度略有下降。探索优化的提示、置信度加权投票或集成策略可能会为实时系统提供更好的速度与精度平衡。

可解释性仍然是一个开放的问题。整合 XAI 技术可以阐明复杂水下环境中的模型决策和失败模式。这对海洋保护和水产养殖等高风险应用至关重要，在这些领域中，信任和可解释性是必不可少的。

如表 ?? 所示，表现优异的检测器如 YOLOv12x 在未见过的数据上取得了超过 0.62 的 mAP 值，表现优于它们在 FishDet-M 测试集上的得分。这表明，通过在像 FishDet-M`push0 g 0 G`pop

这样的汇总数据集上训练，具有很强的泛化能力。这种趋势不仅限于 YOLO，其他检测器如 Faster R-CNN (0.45 mAP)、RetinaNet (0.578 mAP) 和 DETR (0.540 mAP) 也表现出更高的得分。

FishDet-M 为推动鱼类检测研究提供了坚实的基础。持续的进展将依赖于拓展注释深度、提高跨领域泛化能力以及优化适应性和可解释性推理策略，以适应动态的水下环境。

VII. 结论

FishDet-M 是一个统一的基准数据集，旨在解决水下鱼类检测中的持久挑战。通过将 13 个异构来源合并到一个标准化的仓库中，它解决了碎片化和不一致的评估协议相关的问题。我们对 28 种目标检测模型进行了基准测试，包括 YOLOv8 至 YOLOv12、基于 Transformer 的架构和 R-CNN 变体。结果表明，YOLO 模型在检测准确性、在各种物体尺度上的鲁棒性和推理效率方面始终表现优越，特别是 YOLO12x 和 YOLO11m。在检测小型或遮挡鱼类时，Transformer 模型表现下降，并且计算需求更大。对多种水域条件的评估揭示了检测难度的显著差异。泛化测试表明，基于 FishDet-M 训练的模型在以前未见数据上保持可靠的表现，强调了该数据集在支持适应性和现场可部署检测系统方面的实用性。通过发布 FishDet-M 及其注释脚本和训练模型，我们旨在促进可重复的研究，并加速海洋科学、生态监测和智能水下系统中的创新。

VIII.

代码可用性

所有 FishDet-M 资源，包括数据集、基准测试脚本、预训练模型和文档，都可以通过官方项目页面公开获取 <https://lyessaadsaoud.github.io/FishDet-M/>。

IX.

致谢 部分由 Khalifa 大学自主机器人系统中心 (KUCARS) 根据 RC1-2018-KUCARS 奖项资助，部分由 CIRAC 根据奖项 8474000419 和 8434000534 资助。

X.

声明 在撰写这项工作时，我们使用了 Grammarly 和 AI 工具来改善论文的英语语法和流畅性。

REFERENCES

- [1] L. Liu, J. Wu, H. Zhao, H. Kong, T. Zheng, B. Qu, and H. Yu, "Fish-finder: A robust small target detection method for aquaculture fish in low-quality underwater images," *Journal of Fish Biology*, vol. 106, no. 3, pp. 908–920, 2025.
- [2] Y. Jiang, Y. Wang, Y. Zhang, Q. Guo, M. Zhao, and H. Qin, "A feature-enhanced and adaptive routing framework for fish school detection on auvs for degraded underwater imaging environments," *IEEE Internet of Things Journal*, vol. 11, no. 10, pp. 18335–18350, May 2024.
- [3] A. Sasithradevi, R. Suganya, P. Prakash, S. Mohamed Mansoor Roomi, M. Vijayalakshmi, S. Nathan, P. Kasthuri, J. Persiya, L. Brighty Ebenezer, S. Jain, S. Verma, S. Balasubramanian, M. Sai Subramaniam, T. Sai Sriram, M. Pranav Phanindra Sai, C. Raj, A. Yadav, R. Payal, S. Paul Choudhury, and R. Singh, "Depondfi '23 challenge on real-time pond environment: Methods and results," *IEEE Access*, vol. 12, pp. 157975–157987, 2024.

- `push0 g 0 G[4]`pop C. Yang, J. Xiang, X. Li, and Y. Xie, "Fishdet-yolo: Enhanced underwater fish detection with richer gradient flow and long-range dependency capture through mamba-c2f," *Electronics*, vol. 13, no. 18, 2024.
- `push0 g 0 G[5]`pop C. Shah, M. M. Nabi, S. Y. Alaba, I. A. Ebu, J. Prior, M. D. Campbell, R. Caillouet, M. D. Grossi, T. Rowell, F. Wallace, J. E. Ball, and R. Moorhead, "Yolov8-tf: Transformer-enhanced yolov8 for underwater fish species recognition with class imbalance handling," *Sensors*, vol. 25, no. 6, 2025.
- `push0 g 0 G[6]`pop M. A. Duhayyim, H. M. Alshahrani, F. N. Al-Wesabi, M. Alamgeer, A. M. Hilal, and M. A. Hamza, "Intelligent deep learning based automated fish detection model for uwsn," *Computers, Materials & Continua*, vol. 70, no. 3, pp. 5871–5887, 2022.
- `push0 g 0 G[7]`pop S. Fayaz, S. A. Parah, and G. J. Qureshi, "Underwater object detection: architectures and algorithms – a comprehensive review," *Multimedia Tools and Applications*, vol. 81, no. 15, pp. 20871–20916, June 2022.
- `push0 g 0 G[8]`pop V. Pagire and A. Phadke, "Underwater fish detection and classification using deep learning," in *2022 International Conference on Intelligent Controller and Computing for Smart Power (ICICSP)*, 2022, pp. 1–4.
- `push0 g 0 G[9]`pop P. J. Reddy, M. Malathi, and A. N. Julaiha, "Deep fish: An approach to fish species identification through deep learning techniques," in *Emerging Trends in Expert Applications and Security*, V. S. Rathore, V. Piuri, R. Babo, and V. Tiwari, Eds. Singapore: Springer Nature Singapore, 2024, pp. 261–272.
- `push0 g 0 G[10]`pop M. Jia-Fa, X. Gang, S. Wei-Guo, and X. Liu, "A 3d occlusion tracking model of the underwater fish targets," in *2015 IEEE International Conference on Electro/Information Technology (EIT)*, May 2015, pp. 082–086.
- `push0 g 0 G[11]`pop Q. Liu, X. Gong, J. Li, H. Wang, R. Liu, D. Liu, R. Zhou, T. Xie, R. Fu, and X. Duan, "A multitask model for realtime fish detection and segmentation based on YOLOv5," *PeerJ Computer Science*, vol. 9, p. e1262, 2023.
- `push0 g 0 G[12]`pop M. Dawkins, J. Prior, B. Lewis, R. Faillietaz, T. Banez, M. Salvi, A. Rollo, J. Simon, M. Campbell, M. Lucero, A. Chaudhary, B. Richards, and A. Hoogs, "Fishtrack23: An ensemble underwater dataset for multi-object tracking," in *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Jan 2024, pp. 7152–7161.
- `push0 g 0 G[13]`pop E. Li, Q. Wang, J. Zhang, W. Zhang, H. Mo, and Y. Wu, "Fish detection under occlusion using modified you only look once v8 integrating real-time detection transformer features," *Applied Sciences*, vol. 13, no. 23, 2023.
- `push0 g 0 G[14]`pop G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics yolov8," 2023. [Online]. Available: <https://github.com/ultralytics/ultralytics>
- `push0 g 0 G[15]`pop Z. Cai and N. Vasconcelos, "Cascade r-cnn: High quality object detection and instance segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, p. 1–1, 2019.
- `push0 g 0 G[16]`pop N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I*. Berlin, Heidelberg: Springer-Verlag, 2020, p. 213–229.
- `push0 g 0 G[17]`pop K. Chieza, D. Brown, J. Connan, and D. Salie, "Automated fish detection in underwater environments: Performance analysis of yolov8 and yolo-nas," in *Artificial Intelligence Research*, A. Gerber, J. Maritz, and A. W. Pillay, Eds. Cham: Springer Nature Switzerland, 2025, pp. 334–351.
- `push0 g 0 G[18]`pop J. Fabric, I. Turla, J. Capacillo, L. David, and P. C. Naval, "Fish population estimation and species classification from underwater video sequences using blob counting and shape analysis," in *2013 IEEE International Underwater Technology Symposium (UT)*, March 2013, pp. 1–6.
- `push0 g 0 G[19]`pop L. Saad Saoud, Z. Niu, L. Seneviratne, and I. Hussain, "Real-time and resource-efficient multi-scale adaptive robotics vision for underwater object detection and domain generalization," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 2024, pp. 3917–3923.
- `push0 g 0 G[20]`pop V. Mohankumar and S. Anbalagan, "A benchmark dataset and ensemble yolo method for enhanced underwater fish detection," *ETRI Journal*. [Online]. Available: <https://doi.org/10.4218/etrij.2024-0383>
- `push0 g 0 G[21]`pop M. Iqtait, M. H. Alqaryouti, A. E. Sadeq, A. Aburomman, M. Baniata, Z. Mustafa, and H. Y. Char, "Enhanced fish species detection and classification using a novel deep learning approach," *International Journal of Advances Computer Science and Applications*, vol. 15, no. 10, 2024.
- `push0 g 0 G[22]`pop C. Spampinato, S. Palazzo, B. Boom, and R. B. Fisher, "Overview of the lifeclef 2014 fish task," in *Working Notes for CLEF 2014 Conference, Sheffield, UK, September 15–18, 2014*, ser. CEUR Workshop Proceedings, L. Cappellato, N. Ferro, M. Halvey, and W. Kraaij, Eds., vol. 1180. CEUR-WS.org, 2014, pp. 616–624. [Online]. Available: <https://ceur-ws.org/Vol-1180/CLEF2014wn-Life-SpampinatoEt2014.pdf>
- `push0 g 0 G[23]`pop Y. Wei, Y. Wang, B. Zhu, C. Lin, D. Wu, X. Xue, and R. Wang, "Underwater detection: A brief survey and a new multitask dataset," *International Journal of Network Dynamics and Intelligence*, vol. 3, no. 4, p. 100025, 2024, published: 25 December 2024.
- `push0 g 0 G[24]`pop S. Fan, C. Song, H. Feng, and Z. Yu, "Take good care of your fish: fish re-identification with synchronized multi-view camera system," *Frontiers in Marine Science*, vol. 11 - 2024, 2024.
- `push0 g 0 G[25]`pop A. Saleh, I. H. Laradji, D. A. Kononov, M. Bradley, D. Vazquez, and M. Sheaves, "A realistic fish habitat dataset to evaluate algorithms for underwater visual analysis," *Scientific Reports*, vol. 10, no. 1, p. 14671, 9 2020.
- `push0 g 0 G[26]`pop Australian Institute of Marine Science (AIMS), University of Western Australia (UWA), and Curtin University, "Ozfish dataset - machine learning dataset for baited remote underwater video stations," 2019, accessed: 22-Jul-2025.
- `push0 g 0 G[27]`pop S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: towards real-time object detection with region proposal networks," p. 91–99, 2015.
- `push0 g 0 G[28]`pop L. Saad Saoud, M. Elmezzain, A. Sultan, M. Heshmat, L. Seneviratne, and I. Hussain, "Seeing through the haze: A comprehensive review of underwater image enhancement techniques," *IEEE Access*, vol. 12, pp. 145 206–145 233, 2024.
- `push0 g 0 G[29]`pop M. Pedersen, D. Lehotský, I. Nikolov, and T. B. Moeslund, "Brackishmot: The brackish multi-object tracking dataset," in *Image Analysis*, R. Gade, M. Felsberg, and J.-F. Kämäräinen, Eds. Cham: Springer Nature Switzerland, 2023, pp. 17–33.
- `push0 g 0 G[30]`pop J. Hong, M. Fulton, and J. Sattar, "Trashcan: A semantically-segmented dataset towards visual detection of marine debris," *arXiv preprint arXiv:2007.08097*, 2020. [Online]. Available: <https://arxiv.org/abs/2007.08097>
- `push0 g 0 G[31]`pop S. Lian, H. Li, R. Cong, S. Li, W. Zhang, and S. Kwong, "Watermask: Instance segmentation for underwater imagery," in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct 2023, pp. 1305–1315.
- `push0 g 0 G[32]`pop P. Zhang, L. Wang, G. Wang, and D. L. "Eornet: An improved rotating box detection model for counting juvenile fish under occlusion and overlap," *Engineering Applications of Artificial Intelligence*, vol. 124, p. 106528, 2023.
- `push0 g 0 G[33]`pop J. Mao, G. Xiao, W. Sheng, Z. Qu, and Y. Liu, "Research on realizing the 3d occlusion tracking location method of fish's school target," *Neurocomputing*, vol. 214, pp. 61–79, 2016.
- `push0 g 0 G[34]`pop M. Sudhakara, M. J. Meena, K. R. Madhav, P. Anjaiah, and L. P. K, "Fish classification using deep learning on small scale and low-quality images," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 10, no. 1s, pp. 279–288, October 2022, research Article. [Online]. Available: <https://www.ijisae.org/index.php/IJISAE/article/view/2292>
- `push0 g 0 G[35]`pop C. Qiu, S. Zhang, C. Wang, Z. Yu, H. Zheng, and B. Zheng, "Improving transfer learning and squeeze-and-excitation networks for small-scale fine-grained fish image classification," *IEEE Access*, vol. 6, pp. 78 503–78 512, 2018.
- `push0 g 0 G[36]`pop R. vanäEssen, A. Mencarelli, A. vanäHelmond, L. Nguyen, J. Batsleer, J.-J. Poos, and G. Kootstra, "Automatic discard registration in cluttered environments using deep learning and object tracking: class imbalance, occlusion, and a comparison to human review,

- ICES *Journal of Marine Science*, vol. 78, no. 10, pp. 3834–3846, 11 2021.
- `push0 g 0 G[37]`pop R. J. M. Veiga and J. M. F. Rodrigues, “Fine-grained fish classification from small to large datasets with vision transformers,” *IEEE Access*, vol. 12, pp. 113 642–113 660, 2024.
- `push0 g 0 G[38]`pop X. Geng, J. Gao, Y. Zhang, and R. Wang, “A dual-branch feature fusion neural network for fish image fine-grained recognition,” *The Visual Computer*, vol. 40, no. 10, pp. 6883–6896, October 2024.
- `push0 g 0 G[39]`pop P. C. Brady, A. A. Gilerson, G. W. Kattawar, J. M. Sullivan, M. S. Twardowski, H. M. Dierssen, M. Gao, K. Travis, R. I. Etheredge, A. Tonizzo, A. Ibrahim, C. Carrizo, Y. Gu, B. J. Russell, K. Mislinski, S. Zhao, and M. E. Cummings, “Open-ocean fish reveal an omnidirectional solution to camouflage in polarized environments,” *Science*, vol. 350, no. 6263, pp. 965–969, 2015.
- `push0 g 0 G[40]`pop C. Shah, M. M. Nabi, S. Y. Alaba, R. Cailloet, J. Prior, M. Campbell, M. D. Grossi, F. Wallace, J. E. Ball, and R. Moorhead, “Active detection for fish species recognition in underwater environments,” in *Ocean Sensing and Monitoring XVI*, W. Hou and L. J. Mullen, Eds., vol. 13061, International Society for Optics and Photonics. SPIE, 2024, p. 130610D.
- `push0 g 0 G[41]`pop S. Berlia, V. K. Singh, M. Kumar, R. Mahato, and M. Mishra, “Enhanced fish species identification using transfer learning on balanced datasets,” in *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, June 2024, pp. 1–5.
- `push0 g 0 G[42]`pop J. Zhai, L. Han, Y. Xiao, M. Yan, Y. Wang, and X. Wang, “Few-shot fine-grained fish species classification via sandwich attention covamnet,” *Frontiers in Marine Science*, vol. Volume 10 - 2023, 2023.
- `push0 g 0 G[43]`pop S. T. K R, K. S. Ananda Kumar, S. P. R, and V. L, “Comparative analysis of neural architectures for underwater object detection,” in *2024 Second International Conference on Advances in Information Technology (ICAIT)*, vol. 1, July 2024, pp. 1–7.
- `push0 g 0 G[44]`pop M. Abdalhafez, I. M. H. AbdelDaam, M. E. H. Eltaib, and M. Abdelrahim, “Enhanced detection and classification of underwater objects using rov and computer vision,” *JES. Journal of Engineering Sciences*, vol. 52, no. 2, pp. 73–86, 2024.
- `push0 g 0 G[45]`pop M. Shen, M. Yang, J. Zhong, H. Liu, and C. Pan, “Underwater image quality evaluation: A comprehensive review,” *IET Image Processing*, vol. 19, no. 1, p. e70068, 2025.
- `push0 g 0 G[46]`pop H. Ma, Y. Zhang, S. Sun, W. Zhang, M. Fei, and H. Zhou, “Weighted multi-error information entropy based you only look once network for underwater object detection,” *Engineering Applications of Artificial Intelligence*, vol. 130, p. 107766, 2024.
- `push0 g 0 G[47]`pop S. Bhalla, A. Kumar, and R. Kushwaha, “A novel underwater marine dataset with diverse scenarios for robust object detection,” in *Proceedings of the 2024 Sixteenth International Conference on Contemporary Computing*, ser. IC3-2024. New York, NY, USA: Association for Computing Machinery, 2024, p. 1–11.
- `push0 g 0 G[48]`pop W. Liu, Q. Ma, P. Liu, and H. Zhao, “A performance evaluation method for distant early warning sonar mobile area search,” in *2023 3rd International Conference on Electronic Information Engineering and Computer Science (EIECS)*, Sep. 2023, pp. 360–364.
- `push0 g 0 G[49]`pop P. Guo, H. Liu, D. Zeng, T. Xiang, L. Li, and K. Gu, “An underwater image quality assessment metric,” *IEEE Transactions on Multimedia*, vol. 25, pp. 5093–5106, 2023.
- `push0 g 0 G[50]`pop Q. Jiang, X. Yi, L. Ouyang, J. Zhou, and Z. Wang, “Toward dimension-enriched underwater image quality assessment,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 35, no. 2, pp. 1385–1398, Feb 2025.
- `push0 g 0 G[51]`pop M. Elmezain, L. Saad Saoud, A. Sultan, M. Heshmat, L. Seneviratne, and I. Hussain, “Advancing underwater vision: A survey of deep learning models for underwater object recognition and tracking,” *IEEE Access*, vol. 12, 2025, early Access.
- `push0 g 0 G[52]`pop W. Akram, A. Casavola, N. Kapetanović, and N. Miškovic, “A visual servoing scheme for autonomous aquaculture net pens inspection using rov,” *Sensors*, vol. 22, no. 9, p. 3525, 2022.
- `push0 g 0 G[53]`pop W. Akram, T. Hassan, H. Toubar, M. Ahmed, N. Miškovic, L. Seneviratne, and I. Hussain, “Aquaculture defects recognition via multi-scale semantic segmentation,” *Expert systems with applications*, vol. 237, p. 121197, 2024.
- `push0 g 0 G[54]`pop M. Vijayalakshmi and A. Sasithradevi, “Aquayolo: Advanced yolo-based fish detection for optimized aquaculture pond monitoring,” *Scientific Reports*, vol. 15, no. 1, p. 6151, February 2025.
- `push0 g 0 G[55]`pop Q. Zhang and S. Chen, “Research on improved lightweight fish detection algorithm based on yolov8n,” *Journal of Marine Science and Engineering*, vol. 12, no. 10, 2024.
- `push0 g 0 G[56]`pop B. Teixeira, A. P. Lima, C. Pinho, D. Viegas, N. Dias, H. Silva, and J. Almeida, “Feedfirst: Intelligent monitoring system for indoor aquaculture tanks,” in *OCEANS 2023 Hampton Roads*, Oct 2022, pp. 1–7.
- `push0 g 0 G[57]`pop Y.-S. Ryuh, G.-H. Yang, J. Liu, and H. Hu, “A school of robotic fish for mariculture monitoring in the sea coast,” *Journal of Bionic Engineering*, vol. 12, no. 1, pp. 37–46, 2015.
- `push0 g 0 G[58]`pop H. Wang, Z. Gao, B. Li, and N. Gao, “Research on robotic fish swarm network technology based on underwater acoustic communication,” in *2023 IEEE International Conference on Image Processing and Computer Applications (ICIPCA)*, Aug 2023, pp. 475–479.
- `push0 g 0 G[59]`pop M. Ahmed, A. B. Bakht, T. Hassan, W. Akram, A. Humais, L. Seneviratne, S. He, D. Lin, and I. Hussain, “Vision-based autonomous navigation for unmanned surface vessel in extreme marine conditions,” in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2023, pp. 7097–7103.
- `push0 g 0 G[60]`pop M. U. Din, A. Humais, W. Akram, M. Alblooshi, L. Saad Saoud, A. Alblooshi, L. Seneviratne, and I. Hussain, “Marine X: Design and implementation of unmanned surface vessel for vision guided navigation,” in *2023 21st International Conference on Advanced Robotics (ICAR)*. IEEE, 2023, pp. 226–231.
- `push0 g 0 G[61]`pop W. Akram, A. B. Bakht, M. U. Din, L. Seneviratne, and I. Hussain, “Enhancing aquaculture net pen inspection: A benchmark study on detection and semantic segmentation,” *IEEE Access*, 2024.
- `push0 g 0 G[62]`pop W. Zhang, L. Zhang, Y. Zhong, P. Lin, and F. Zhang, “Recognition and calculation of fish rafts in mariculture on the basis of artificial intelligence,” in *Proceedings of the 2023 International Conference on Wireless Communications, Networking and Applications*, P. Siarry, M. A. Jabbar, S. K. S. Cheung, and X. Li, Eds. Singapore: Springer Nature Singapore, 2025, pp. 203–210.
- `push0 g 0 G[63]`pop L. Falconer, S. Halstensen, S. F. Rinø, C. Norblø, T. Dale, R. Alvestad, and E. Ytteborg, “Marine aquaculture sites have huge potential as data providers for climate change assessments,” *Aquaculture*, vol. 595, p. 741519, 2025.
- `push0 g 0 G[64]`pop L. Saad Saoud, A. Sultan, M. Elmezain, M. Heshmat, L. Seneviratne, and I. Hussain, “Beyond observation: Deep learning for animal behavior and ecological conservation,” *Ecological Informatics*, vol. 85, p. 102893, 2024.
- `push0 g 0 G[65]`pop J. Kong, S. Tang, J. Feng, L. Mo, and X. Jin, “Aasnet: A novel image instance segmentation framework for fine-grained fish recognition via linear correlation attention and dynamic adaptive focal loss,” *Applied Sciences*, vol. 15, no. 7, 2025.
- `push0 g 0 G[66]`pop W. Wang, B. He, and L. Zhang, “High accuracy real-time fish detection based on self-build dataset and rird-yolov3,” *Complexity*, vol. 2021, no. 1, p. 4761670, 2021.
- `push0 g 0 G[67]`pop M. U. Din, A. B. Bakht, W. Akram, Y. Dong, L. Seneviratne, and I. Hussain, “Benchmarking vision-based object tracking for usvs in complex maritime environments,” *IEEE Access*, 2025.
- `push0 g 0 G[68]`pop L. Hong, X. Wang, G. Zhang, and M. Zhao, “Usod10k: A new benchmark dataset for underwater salient object detection,” *IEEE Transactions on Image Processing*, vol. 34, pp. 1602–1615, 2025.
- `push0 g 0 G[69]`pop COCO Consortium, “Coco - common objects in context,” 2014, accessed: 2025-07-15. [Online]. Available: <https://cocodataset.org>

- `push0 g 0 G[70]`pop F. F. Khan, X. Li, A. J. Temple, and M. El-hoseiny, "Fishnet: A large-scale dataset and benchmark for fish recognition, detection, and functional trait prediction," in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 20 439–20 449.
- `push0 g 0 G[71]`pop A. Wang, H. Chen, L. Liu, Y. Wang, Y. Zhong, C. Shan, Z. Guo, C. Xu, and S. Chen, "Yolov10: Real-time end-to-end object detection," *arXiv preprint arXiv:2405.14458*, 2024, <https://arxiv.org/abs/2405.14458>.
- `push0 g 0 G[72]`pop G. Jocher and J. Qiu, "Ultralytics yolo11," 2024. [Online]. Available: <https://github.com/ultralytics/ultralytics>
- `push0 g 0 G[73]`pop Y. Tian, Q. Ye, and D. Doermann, "Yolov12: Attention-centric real-time object detectors," 2025. [Online]. Available: <https://github.com/sunsmarterjie/yolov12>
- `push0 g 0 G[74]`pop S. Aharon, Louis-Dupont, Ofri Masad, K. Yurkova, Lotem Fridman, Lkdc, E. Khvedchenya, R. Rubin, N. Bagrov, B. Tymchenko, T. Keren, A. Zhilko, and Eran-Deci, "Super-gradients," 2021.
- `push0 g 0 G[75]`pop P. Sun, R. Zhang, Y. Jiang, T. Kong, C. Xu, W. Zhan, M. Tomizuka, Z. Yuan, and P. Luo, "Sparse r-cnn: An end-to-end framework for object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 12, pp. 15 650–15 664, Dec 2023.
- `push0 g 0 G[76]`pop Y. Zhao, W. Lv, S. Xu, J. Wei, G. Wang, Q. Dang, Y. Liu, and J. Chen, "Detrs beat yolos on real-time object detection," pp. 16 965–16 974, June 2024.
- `push0 g 0 G[77]`pop X. Zhu, W. Su, L. Lu, B. Li, X. Wang, and J. Dai, "Deformable detr: Deformable transformers for end-to-end object detection," 2021. [Online]. Available: <https://arxiv.org/abs/2010.04159>
- `push0 g 0 G[78]`pop T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," vol. 42, no. 2, Feb 2020, pp. 318–327.
- `push0 g 0 G[79]`pop W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," pp. 21–37, 2016.
- `push0 g 0 G[80]`pop A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, *PyTorch: an imperative style, high-performance deep learning library*. Red Hook, NY, USA: Curran Associates Inc., 2019.
- `push0 g 0 G[81]`pop G. Jocher, J. Qiu, and A. Chaurasia, "Ultralytics YOLO," Jan. 2023. [Online]. Available: <https://github.com/ultralytics/ultralytics>
- `push0 g 0 G[82]`pop K. Chen, J. Wang, J. Pang, Y. Cao, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Xu, Z. Zhang, D. Cheng, C. Zhu, T. Cheng, Q. Zhao, B. Li, X. Lu, R. Zhu, Y. Wu, J. Dai, J. Wang, J. Shi, W. Ouyang, C. C. Loy, and D. Lin, "Mmdetection: Open mmlab detection toolbox and benchmark," 2019. [Online]. Available: <https://arxiv.org/abs/1906.07155>
- `push0 g 0 G[83]`pop T.-Y. Lin and M. C. Consortium, "pycoco-tools: Coco api for python," <https://github.com/cocodataset/cocoapi>, 2015, accessed: 2025-07-15.
- `push0 g 0 G[84]`pop g18L5754, "Fish4knowledge dataset dataset," <https://universe.roboflow.com/g18l5754/fish4knowledge-dataset>, oct 2023, visited on 2025-05-21. [Online]. Available: <https://universe.roboflow.com/g18l5754/fish4knowledge-dataset>
- `push0 g 0 G[85]`pop A. MURME, "Fish video dataset," <https://universe.roboflow.com/aarjoo-murme/fish-video-ls42k>, nov 2023, visited on 2025-05-21. [Online]. Available: <https://universe.roboflow.com/aarjoo-murme/fish-video-ls42k>
- `push0 g 0 G[86]`pop seoultech, "Fish-video dataset," <https://universe.roboflow.com/seoultech/fish-video>, may 2022, visited on 2025-05-21. [Online]. Available: <https://universe.roboflow.com/seoultech/fish-video>
- `push0 g 0 G[87]`pop I. A. Catalán, A. Álvarez Ellacuría, J.-L. Lisani, J. Sánchez, G. Vizoso, A. E. Heinrichs-Maquilón, H. Hinz, J. Alós, M. Signarioli, J. Aguzzi, M. Francescangeli, and M. Palmer, "Automatic detection and classification of coastal mediterranean fish from underwater images: Good practices for robust training," *Frontiers in Marine Science*, vol. Volume 10–2023, 2023.
- `push0 g 0 G[88]`pop aquarium, "detect aqurium dataset," <https://universe.roboflow.com/aquarium-lrui2/detect-aqurium>, 2023, visited on 2025-06-30. [Online]. Available: <https://universe.roboflow.com/aquarium-lrui2/detect-aqurium>
- `push0 g 0 G[89]`pop M.-Q. Le, T.-N. Le, T. V. Nguyen, I. Ech, and M.-T. Tran, "Aqatic animal species (aas)," 2023, data.
- `push0 g 0 G[90]`pop PAFD, "Fish clean dataset," <https://universe.roboflow.com/pafd/fish-clean>, feb 2023, visited on 2025-06-30. [Online]. Available: <https://universe.roboflow.com/pafd/fish-clean>
- `push0 g 0 G[91]`pop Z. Tian, C. Shen, H. Chen, and T. He, "F1: Fully convolutional one-stage object detection," pp. 9626–9637, Oct 2019.
- `push0 g 0 G[92]`pop test, "fish_dataset_florence_1 dataset," https://universe.roboflow.com/test-fllhx/fish_dataset_florence_1, Aug. 2024, accessed on 2025-07-20.
- `push0 g 0 G[93]`pop L. Saad Saoud and I. Hussain, "Ebio: Ethics-guided bias-aware ai for efficient underwater image enhancement and coral reef monitoring," 2025. [Online]. Available: <https://arxiv.org/abs/2507.15036>
- `push0 g 0 G[94]`pop A. Salman, S. A. Siddiqui, F. Sha, A. Mian, M. R. Shortis, K. Khurshid, A. Ulges, and U. Schaefer, "Automatic fish detection in underwater videos by a deep neural network-based hybrid motion learning system," *Journal of Marine Science*, vol. 77, no. 4, pp. 1295–1307, 2019.
- `push0 g 0 G[95]`pop A. Jalal, A. Salman, A. Mian, S. Ghafoor, F. Shafait, "Deepfins: Capturing dynamics in underwater videos for fish detection," *Ecological Informatics*, vol. 86, p. 101583, 2025.

`push0 g 0 G`pop

