

中图法分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2026)05-1478-25

论文引用格式: Li H, Li Z Y, Liu J W and Cong R M. 2026. A survey of underwater visual datasets for ocean intelligence. Journal of Image and Graphics, 31(5):1478-1502(李华, 李志远, 刘家伟, 丛润民. 2026. 面向海洋智能的水下视觉数据集综述. 中国图象图形学报, 31(5):1478-1502)[DOI:10.11834/jig.250483]

面向海洋智能的水下视觉数据集综述

李华¹, 李志远¹, 刘家伟¹, 丛润民^{2*}

1. 海南大学计算机科学与技术学院, 海口 570228; 2. 山东大学控制科学与工程学院, 济南 250061

摘要: 水下计算机视觉作为海洋科学研究和海洋工程应用的重要技术手段, 受到广泛关注。然而, 与陆地环境相比, 水下成像受光线衰减、颜色偏移、散射效应及设备限制等因素影响, 相关任务面临显著挑战。高质量数据集的构建是推动水下计算机视觉技术发展的关键资源, 其质量与多样性直接影响模型的训练效果和性能表现。为全面了解水下数据集的发展脉络、优势与局限性, 本文对当前主要数据集进行了系统综述, 涵盖水下视觉增强、水下场景理解以及水下三维重建等多个关键任务。在水下视觉增强方面, 分析了图像/视频增强、颜色校正与复原以及超分辨率重建等数据集; 在水下场景理解方面, 系统梳理了目标分类、目标检测、语义分割、实例分割、显著性检测、伪装目标检测以及目标跟踪等任务的代表性数据集; 在水下三维重建方面, 探讨了同时定位与建图(simultaneous localization and mapping, SLAM)、神经辐射场(neural radiance fields, NeRF)和三维高斯散射(three-dimensional Gaussian splatting, 3DGS)等技术相关的数据集发展状况。基于对现有数据集的构建方法、规模特征及应用场景的综合分析, 本文总结了当前面临的主要挑战, 并探讨了未来可能的发展方向, 为相关研究人员提供了全面的数据集资源参考, 有助于推动水下计算机视觉技术的深入研究与广泛应用。本文涉及数据集下载链接: <https://cstr.cn/31253.11.sciencedb.j00240.00173> 或 <https://github.com/Linzy0227/UVd>。

关键词: 水下数据集; 水下场景理解; 图像增强; 目标识别; 三维重构

A survey of underwater visual datasets for ocean intelligence

Li Hua¹, Li Zhiyuan¹, Liu Jiawei¹, Cong Runmin^{2*}

1. School of Computer Science and Technology, Hainan University, Haikou 570228, China;

2. School of Control Science and Engineering, Shandong University, Jinan 250061, China

Abstract: Underwater computer vision has emerged as a critical enabling technology for advancing marine scientific research and ocean engineering applications, attracting considerable attention from academic and industrial communities in recent years. However, compared with terrestrial imaging, underwater imaging presents unique and formidable challenges, including light attenuation, color distortion, scattering effects, and equipment limitations, which together make underwater computer vision tasks significantly more complex and demanding than their land-based counterparts. The construction of large-scale, high-quality, and diverse datasets serves as the fundamental resource for advancing underwater computer vision technologies, with their quality, quantity, and diversity directly influencing the training effectiveness, generalization ability, and overall performance of deep learning models. For a comprehensive understanding of the develop-

收稿日期: 2025-10-13; 修回日期: 2025-12-08; 预印本日期: 2025-12-15

* 通信作者: 丛润民 rmcong@sdu.edu.cn

基金项目: 国家自然科学基金项目(62461018); 海南省自然科学基金项目(625YXQN594); 海南省“南海新星”创新平台项目(NHXXRCXM202306)

Supported by: National Natural Science Foundation of China (62461018); Natural Science Foundation of Hainan Province, China (625YXQN594); Innovation Platform for “New Star of South China Sea” of Hainan Province (NHXXRCXM202306)

ment trajectory, advantages, and limitations of underwater datasets, this paper presents a systematic survey of representative underwater datasets currently available in the research community. The survey encompasses three major domains of underwater computer vision, namely, underwater visual enhancement, scene understanding, and 3D reconstruction, providing a holistic and detailed view of the dataset landscape in this rapidly evolving field. In the underwater visual enhancement domain, we carefully analyze datasets specifically designed for image and video enhancement, color correction and restoration, and super-resolution reconstruction. These datasets directly address the fundamental and long-standing challenge of improving underwater image quality, which is a crucial prerequisite for enabling reliable performance in subsequent high-level vision tasks. For underwater scene understanding, we systematically examine a broad set of representative datasets across multiple core tasks, including object classification, object detection, semantic segmentation, instance segmentation, salient object detection, camouflaged object detection, and long-term object tracking. This comprehensive analysis clearly reveals the gradual progression from basic classification tasks to complex and integrated multimodal understanding challenges. The classification datasets have evolved from image-level classification problems to fine-grained species recognition involving hundreds of distinct categories. Detection datasets have expanded from relatively small-scale image collections to extremely large-scale benchmarks containing millions of bounding box annotations. Segmentation datasets have progressed from coarse pixel-level annotations to precise instance-level masks enriched with multiattribute labeling and contextual information. In the underwater 3D reconstruction domain, we explore datasets related to simultaneous localization and mapping, neural radiance fields, and 3D Gaussian splatting technologies. These emerging datasets reflect the growing research interest in comprehensive 3D scene understanding and reconstruction within underwater environments, explicitly addressing challenges unique to marine settings, such as severely limited visibility and unpredictable lighting. Through an in-depth analysis of existing datasets, this survey highlights and reveals several significant development trends in underwater computer vision dataset construction. First, dataset scales have expanded dramatically from hundreds of images in early collections to current million-scale datasets, demonstrating the research community's recognition of the critical importance of large-scale data for training robust and generalizable models. Second, annotation precision has gradually evolved from coarse-grained image-level labels to fine pixel-level segmentation and multiattribute annotations, enabling sophisticated and reliable algorithmic developments. Third, task complexity has progressed from single-modal classification to multimodal fusion and even general foundation models, reflecting the increasing technical sophistication of modern underwater vision applications. The survey also identifies the adoption of automated annotation techniques as a crucial development in significantly reducing dataset construction costs and human labor requirements. The successful application of advanced models, such as the segment anything model, across multiple representative underwater datasets clearly demonstrates the potential of using powerful pretrained models for efficient, scalable, and semiautomated annotation generation. Moreover, multimodal data fusion has rapidly emerged as an important trend, with new datasets increasingly combining RGB imagery, sonar data, textual descriptions, and additional sensor information to provide rich and comprehensive scene understanding capabilities. Despite these encouraging advances, our analysis reveals several critical and persistent limitations in existing underwater datasets. Geographic coverage remains heavily biased toward developed regions and easily accessible coastlines, with insufficient representation of tropical deep-sea ecosystems or polar marine environments. Species and scenario diversity is still limited, with most datasets focusing predominantly on common marine organisms and typical water scenes, while coverage of rare, endangered, or cryptic species and extreme environmental conditions remains inadequate. The lack of long-term temporal data severely hampers the ability to monitor dynamic ecosystem changes, as most existing datasets represent short-term snapshots rather than continuous longitudinal observations. Annotation standards also vary significantly across datasets, with inconsistent quality control mechanisms negatively affecting interdataset comparability. On the basis of this analysis, we propose several promising future development directions for the underwater computer vision dataset community. The construction of large-scale pretraining datasets will become increasingly important for developing universal underwater foundation models and accelerating advances in self-supervised and weakly supervised learning approaches. Synthetic data generation technologies, supported by generative artificial intelligence techniques, hold great promise for addressing data scarcity challenges, particularly in the representation of rare species. Cross-domain generalization capabilities require the design of enhanced evaluation frameworks to rigorously assess algorithm robustness

across different geographic regions, temporal periods, environmental conditions, and equipment configurations. This survey ultimately provides underwater computer vision researchers with comprehensive dataset resource references and development insights, thereby facilitating deep technological advances and broad applications in marine scientific research, ocean resource development, sustainable fisheries, and marine environmental protection. It will contribute to the gradual construction of comprehensive, high-quality, open-access, and standardized underwater dataset ecosystems that will provide robust and enduring data support for advancing human understanding, exploration, and long-term stewardship of fragile marine environments. The link is <https://cstr.cn/31253.11.sciencedb.j00240.00173> or <https://github.com/Linzy0227/UVD>.

Key words: underwater datasets; underwater scene understanding; image enhancement; object recognition; 3D reconstruction

0 引言

海洋作为地球上最大的生态系统,蕴含着丰富的生物与能源资源,对人类社会的可持续发展具有重要意义(Kittinger等,2024)。无论是海洋生态监测、能源开发、水下考古,还是国防需求,海洋探索与环境理解都发挥着不可替代的作用(周玲等,2025)。在这一背景下,水下计算机视觉逐渐成为推动海洋科学研究与工程应用的重要技术手段。通过对图像和视频的自动化分析,水下计算机视觉为水下作业提供了更高的效率与安全保障,是遥控潜水器(remotely operated vehicle, ROV)、自主式水下航

行器(autonomous underwater vehicle, AUV)以及智能化海洋监测系统的核心功能之一,图1展示了水下数据采集、处理与应用的整个流程。

然而,与陆地环境相比,水下成像受物理规律的影响,使得水下计算机视觉的发展和应用面临诸多挑战(李学龙,2024a)。主要包括:1)水体介质对光的吸收和散射效应,水下图像普遍呈现低对比度、颜色偏移和清晰度下降等问题;2)悬浮颗粒、海水浑浊度以及水面波动引起的动态光照变化,使得图像模糊、亮度分布不均匀、细节丢失;3)水下生物目标具有形态相似、伪装性强等特点,不同类别间差异微小,容易导致特征混淆,从而增加识别难度;4)水下设备和硬件资源受限,数据采集成本高、效率低。这些因素共同加剧

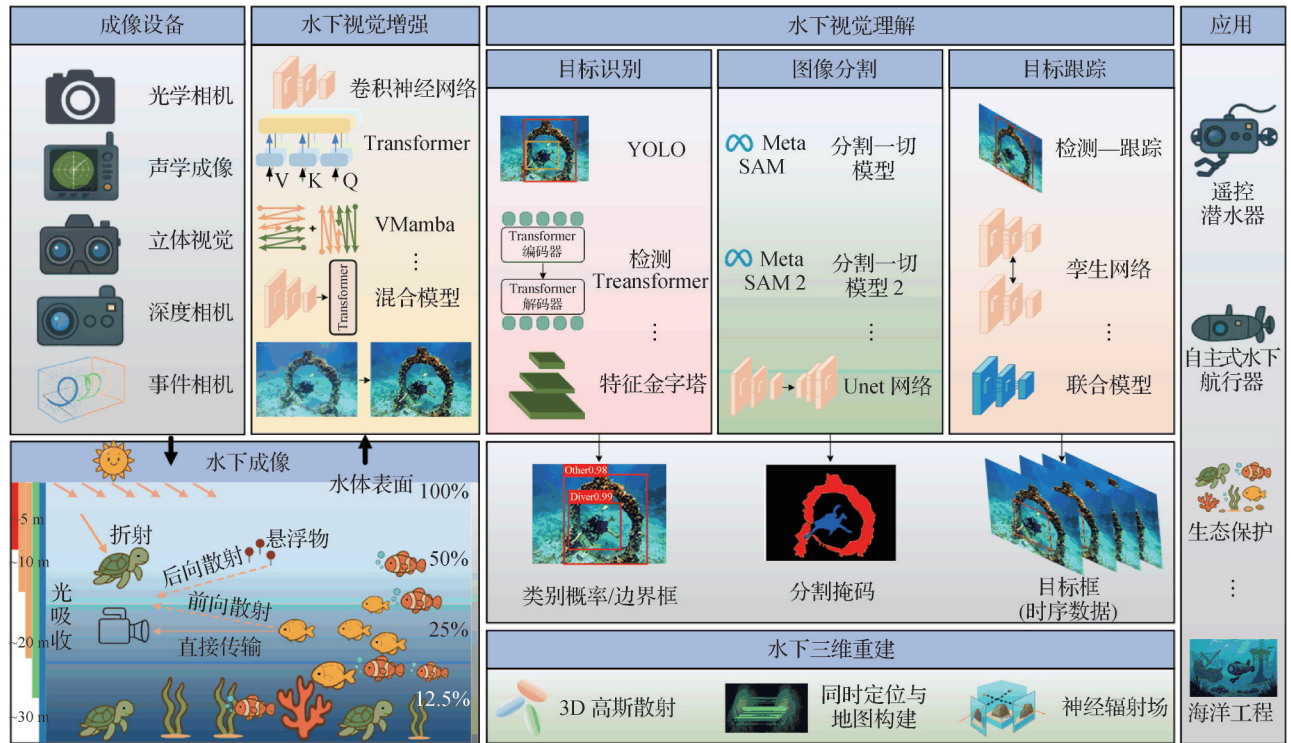


图1 水下视觉数据采集、处理与应用流程概览

Fig. 1 Overview of underwater vision data acquisition, processing, and application workflow

了水下视觉任务的复杂性(李学龙, 2024b), 导致计算机视觉方法在水下环境中难以保持理想性能。

深度学习方法的成功应用使得水下计算机视觉技术快速发展。典型的水下计算机视觉系统通常包括图像采集、特征提取和任务执行等环节: 首先由光学相机或其他传感器捕获图像, 随后通过图像增强或颜色校正改善图像质量(崔峥等, 2025), 再执行分类、检测、分割或跟踪等高级视觉任务。然而, 深度学习的成功高度依赖大规模、高质量数据。在陆地视觉领域, 像 ImageNet (Deng 等, 2009)、SA-1B 等大规模数据集 (Kirillov 等, 2023) 的出现, 推动了多个重要模型的发展和应用, 例如 ResNet (residual network) (He 等, 2016)、视觉 Transformer (Dosovitskiy 等, 2021) 以及分割一切模型 (segment anything model, SAM) 等。为推动水下计算机视觉的发展, 学术界与工业界构建了多种水下视觉数据集, 涵盖不同任务层次和应用场景。例如, 针对图像增强与颜色校正的退化—参考图像配对数据集; 针对水下场景理解的目标检测、语义分割等数据集; 以及水下场景三维重建等专门化数据集。

然而, 在水下视觉中, 由于数据采集难度大、标注成本高和专业知识依赖性强, 现有数据集往往存在规模有限、场景覆盖不足、类别分布不均衡以及标注精度有限等问题(聂婕等, 2022), 不仅限制了水下视觉算法的性能, 也削弱了其跨场景的泛化能力, 成为制约水下视觉技术发展的主要瓶颈。因此, 对现有水下数据集进行系统梳理与全面总结, 不仅有助于厘清不同研究方向的数据资源现状, 也能为未来更大规模、更高质量数据集的构建与应用提供参考, 进而促进水下视觉的发展与应用。尽管已有部分研究对特定类型的水下数据集进行了介绍和分析 (Chen 等, 2025), 但缺乏对整体数据集生态的全面综述, 大多数相关综述集中于某一类任务或应用领域, 如海洋生物识别或图像增强, 难以为研究人员提供全面的数据资源指引。

在此背景下, 本文对现有水下计算机视觉数据集进行系统性综述, 包括水下视觉增强、水下场景理解与水下三维重构等。通过对数据集的规模特征、构建方式、适用任务及应用场景进行全面梳理和深入分析, 并对未来水下数据集的发展趋势进行展望, 旨在为水下计算机视觉研究人员提供全面的数据资源指引, 促进该领域的技术发展与

实际应用。本文所涉及数据集的下载链接整理于 <https://cstr.cn/31253.11.sciencedb.j00240.00173> 或 <https://github.com/Linzy0227/UVd>。

1 水下视觉增强

水下视觉增强任务, 如图像增强与恢复、颜色校正、图像去噪与超分辨率, 旨在改善水下环境对图像质量的退化影响, 是实现后续视觉理解任务的基础。由于水体介质的影响, 水下成像普遍存在对比度低、颜色失真、细节模糊和噪声干扰等问题, 严重影响了水下目标识别任务的性能。通过图像修复, 恢复图像的颜色保真度、去除噪声以及增强细节信息等, 可以改善图像的视觉质量和信息含量。

1.1 任务定义

给定水下退化图像 X , 水下图像修复过程可以通用地表示为

$$\hat{Y} = f(X; \theta) \quad (1)$$

式中, $f(\cdot)$ 表示具有可学习参数 θ 的神经网络, $X, \hat{Y} \in \mathbb{R}^{H \times W \times 3}$ 。 H 和 W 分别表示图像的高度与宽度, $\hat{Y}$ 表示输出的增强图像。对于图像修复模型, 常见的优化目标是通过最小化损失函数求得最优参数, 表示为

$$\hat{\theta} = \arg \min \mathcal{L}(\hat{Y}, Y) \quad (2)$$

式中, $\mathcal{L}(\hat{Y}, Y)$ 表示用于获取最优参数 $\hat{\theta}$ 的损失函数, Y 表参考图像, $\mathcal{L}(\hat{Y}, Y)$ 的具体形式取决于是否存在参考图像, 可以是任何形式的监督或非监督损失函数。

近年来, 研究人员开始系统性地构建水下图像修复数据集。根据数据集构建方法 and 应用目标的不同, 当前的水下图像修复数据集通常分为两类: 一是退化—参考配对样本数据集, 通过在受控条件下采集或模拟生成, 使得退化图像与理想参考图像一一对应, 便于监督学习方法的训练和评估; 二是无参考真实场景数据集, 直接采集不同水域的真实退化图像, 虽然缺乏理想参考标准, 但更接近实际应用场景, 有利于无监督和弱监督方法的发展。配对图像与非配对图像如图2所示。

表1对水下视觉增强任务数据集进行总结, 从不同角度和应用需求出发, 为水下图像增强、质量评

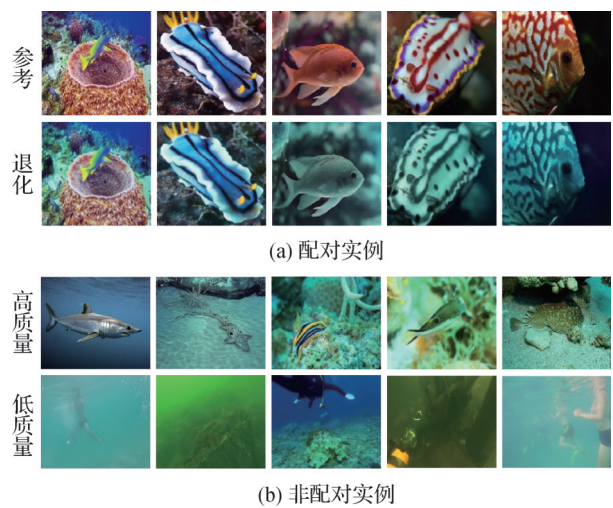


图2 配对实例以及非配对实例

Fig. 2 Paired instances and unpaired instances
((a) paired instances; (b) unpaired instances)

估以及超分辨率重建等任务提供了重要数据支撑。

1.2 评价指标

水下视觉增强任务的评估需要依赖多维度的质量指标,由于不同任务的目标各异,评价指标除了要反映整体图像质量外,还需兼顾可见度、细节保留以及色彩恢复等维度。基于这些特性,现有的水下图像质量评价指标大致可以分为以下几类。

1.2.1 有参考型指标

有参考型指标通常不依赖于特定的图像任务,通过比较增强后的图像与参考图像之间的差异来衡量增强效果。常见指标包括均方误差 (mean squared error, MSE)、峰值信噪比 (peak signal-to-noise ratio, PSNR) 以及结构相似性指数 (structural similarity index, SSIM) 等。MSE 用于衡量增强图像 $\hat{Y}$

表1 水下视觉增强数据集总结

Table 1 Summary of underwater vision enhancement datasets

数据集	类型	数量	主要任务	数据来源	是否成对	特点
TURBID (Duarte 等, 2016)	受控	570 幅	图像复原	受控环境	是	海底场景
OceanDark (Marques 等, 2019)	真实	183 幅	低光照增强	深海环境	否	深海低光环境
U45 (Li 等, 2019)	真实	45 幅	图像增强	多数据源整合	否	3 种退化类型
Sea (Akkaynak 和 Treibitz, 2019)	真实	1 100 幅	水分去除	2 种光学水域	否	5 个不同场景
EUVP (Islam 等, 2020d)	真实	12 000 对 + 8 000 幅	图像增强	多样化场景	部分	生成配对图像
UIEB (Li 等, 2020a)	真实	890 对 + 60 幅	图像增强	珊瑚礁、海洋生物	部分	多样化分辨率
SUID (Hou 等, 2020)	合成	900 + 30 幅参考	图像增强	光学模型合成	是	4 种典型场景
UOQ (Wu 等, 2020)	真实	216 幅	无参考质量评估	算法增强	否	前景区域标注
RUIE (Liu 等, 2020)	真实	约 4 000 幅	多任务基准	黄海实地采集	部分	0.5~8 m 深度变化
USR (Islam 等, 2020c)	真实	2 368 幅	超分辨率	多源场景	是	多分辨率重建
SQUID (Berman 等, 2021)	真实	57 对立体图	颜色复原	红海、地中海	是	多样化水下环境
LNRUD (Ye 等, 2022)	合成	50 000 幅	多类型	神经渲染	是	多样化场景
UIED (Zheng 等, 2022)	真实	1 100 幅	图像增强	增强算法处理	是	多样化场景
UVE-38K (Qi 等, 2022)	真实视频	38 000 帧	视频增强	视频序列	是	多样化水下场景
LSUI (Peng 等, 2023)	真实	4 279 对	多任务	多样化水下场景	是	深海、洞穴场景
UID2021 (Hou 等, 2023)	真实	960 幅	无参考质量评估	增强算法处理	否	6 种退化场景
UIDEF (Chang 等, 2023)	真实	9 200 幅	增强与重建	多地点潜水采集	否	多种水下场景
CSUID (Purnima 和 Kumar, 2024)	合成	15 100 幅	图像增强	地面图像合成	是	自然与人造场景
UVEB (Xie 等, 2024)	真实视频	453 874 帧	视频增强	多水域采集	是	6 种退化类型
SUVE (Du 等, 2024)	合成视频	840 对	视频增强	基于 NYU-Depth 合成	是	3 种水下风格

注:“-”表示数据未知。

与参考图像 Y 之间的误差; PSNR 用于评估图像失真的程度; SSIM 用于综合评价增强图像在亮度、对比度和结构等方面与参考图像的相似程度。

1.2.2 无参考型指标

无参考型指标不依赖参考图像, 可用于评估实际应用中无法获得参考的情况。该类指标能够直接从图像本身提取统计或感知特征, 以衡量图像的视觉质量。常见方法包括: 水下彩色图像质量评价指标 (underwater color image quality evaluation, UCIQE)、水下图像质量评价指标 (underwater image quality measure, UIQM) 等。

1) UCIQE 通过对色度、对比度与饱和度进行加权求和, 对水下图像质量进行定量评价, 具体为

$$f_{\text{UCIQE}} = c_1 \cdot \sigma_c + c_2 \cdot \text{contrast}_L + c_3 \cdot \mu_s \quad (3)$$

式中, σ_c 为色度标准差, contrast_L 为亮度对比度, μ_s 为平均饱和度。系数 c_1 、 c_2 和 c_3 根据经验确定。

2) UIQM 综合考虑色彩丰富度 (underwater image colorfulness measure, UICM)、图像清晰度 (underwater image sharpness measure, UISM) 和对比度 (underwater image contrast measure, UIConM), 具体为

$$f_{\text{UIQM}} = \alpha_1 \cdot f_{\text{UICM}} + \alpha_2 \cdot f_{\text{UISM}} + \alpha_3 \cdot f_{\text{UIConM}} \quad (4)$$

式中, α_1 、 α_2 和 α_3 为经验权重。

1.2.3 人类视觉评估

尽管自动指标在图像质量评价中应用广泛, 但人类主观感知仍不可替代。特别是在低对比度、色彩偏移等复杂情况下, 主观评价能够更准确地反映图像在实际应用中的可视效果与感知质量。因此, 研究中常结合主客观评价, 以全面衡量算法性能。

1.3 分析与讨论

从数据规模来看, 水下图像增强数据集呈现明显的发展趋势。早期数据集 (2016—2019 年) 规模较小, 如 TURBID (turbid dataset) (570 幅)、OceanDark (183 幅)、U45 (45 幅) 等, 主要用于算法验证和概念验证。中期数据集 (2020—2022 年) 规模显著提升, EUVP (enhancement of underwater visual perception) 达到 12 000 对图像, LNRUD (large-scale natural realistic underwater dataset) 达到 50 000 幅, 为深度学习模型训练提供了更充足的数据支撑。近期数据集 (2023—2024 年) 在规模上继续突破, UVEB (underwater video enhancement benchmark) 包含 453 874 帧视频数据, CSUID (color and structure underwater image

dataset) 达到 15 100 幅, 反映了大规模数据集构建的趋势。在数据类型方面, 现有数据集可分为 3 类: 真实数据集 (如 UIEB (underwater image enhancement benchmark)、RUIE (real-world underwater image enhancement)) 直接来源于水下拍摄, 真实性高但采集成本大; 受控数据集 (如 TURBID) 在可控环境中拍摄, 质量稳定但场景多样性有限; 合成数据集 (如 SUID (synthetic underwater image dataset)、LNRUD、CSUID) 通过物理模型或生成模型构建, 规模可扩展但真实性存在差距。当前趋势是将真实数据与合成数据相结合, 如 EUVP 同时包含真实数据和生成式增强数据, 兼顾数据规模与真实性。

成对数据是监督学习的关键, 但不同数据集的成对质量存在显著差异。高质量成对数据集 (如 TURBID、SUID) 提供了准确的参考图像, 通常通过受控环境拍摄或精确的物理建模生成, 适合训练精确的增强模型, 但数据多样性受限; 弱成对数据集 (如 EUVP、UIEB) 的参考图像来自专家手工增强或算法处理, 存在一定主观性, 但场景覆盖更广泛; 非成对数据集 (如 OceanDark、U45、UOQ (underwater open quality)) 仅提供退化图像, 需要依赖无监督或弱监督学习方法, 虽然数据采集成本低, 但训练难度更大。当前研究趋势是开发能够同时利用成对、弱成对和非成对数据的混合学习框架。

水下图像退化类型多样, 包括色彩偏移、对比度下降、细节模糊和噪声干扰等。早期数据集多聚焦于单一或少数退化类型, 如 Sea 主要针对水分散射导致的色彩退化, USR (underwater super-resolution) 专注于分辨率退化。近期数据集更注重多样化退化场景的覆盖, RUIE 包含 0.5~8 m 不同深度的退化变化, LSUI 涵盖深海、洞穴等多种极端环境, UVEB 包含 6 种退化类型的视频序列。然而, 现有数据集对极端环境 (如深海高压、极地低温、高浑浊度) 和混合退化 (多种退化因素同时作用) 的覆盖仍然不足, 限制了算法在复杂实际场景中的泛化能力。

2 水下场景理解

水下场景理解作为计算机视觉在海洋环境中的重要分支, 是实现海洋生物监测、水下机器人导航以及水下生态保护等应用的核心技术。其算法性能与可用性在很大程度上依赖于高质量和多样化的数据

集。按照任务复杂度的不同,水下场景理解通常可以从粗到细进行划分:最基础的是图像分类,仅关注目标类别的判别,不涉及位置信息或几何结构;其次是目标检测,在识别物体类别的同时,还需确定其空间位置,通常通过边界框进行标注;进一步是图像分割任务,其目标是在像素级实现精确划分,包括显著性目标检测、语义分割和实例分割,其中语义分割关注类别识别,而实例分割则需区分同类别下的不同实例。图3展示了从粗粒度到细粒度的3个层次。

1)图像分类层,对潜水员、扇贝、电鳗、海龙等水下目

标进行类别识别;2)目标检测层,通过边界框标注实现目标定位;3)显著性检测和实例分割层,分别用白色掩码和彩色掩码标注目标的精细轮廓。在此之上,目标跟踪任务则面向视频序列,既要求完成识别与定位,还需在时间维度上连续追踪目标的运动轨迹,解决动态场景下的遮挡与交互等问题。为了支撑这些不同层级的任务,研究人员构建并发布了多种水下场景理解数据集,为相关算法的训练与评估提供了重要支撑。针对多任务数据集,本节将结合其主要应用场景进行系统性梳理与总结。

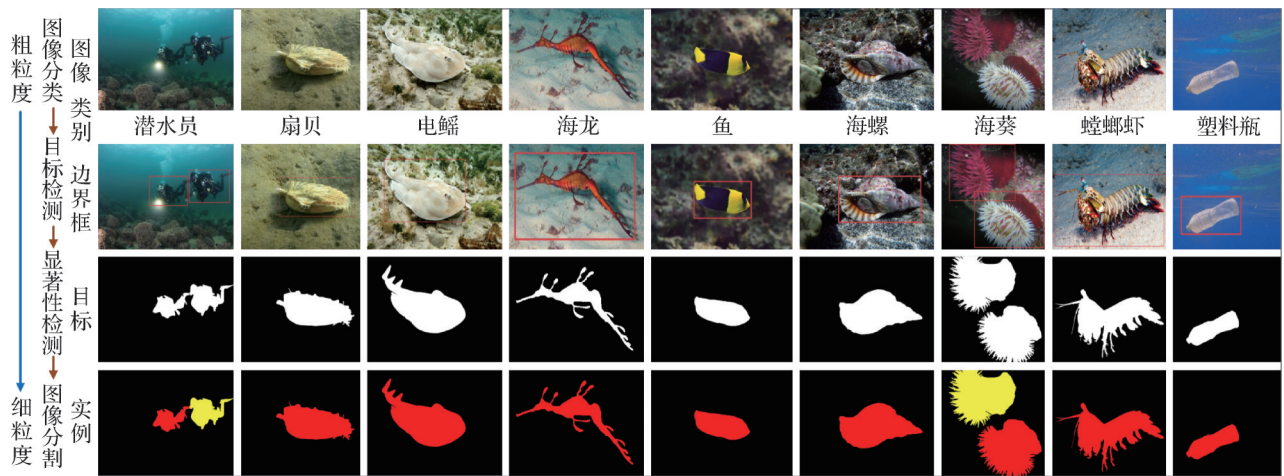


图3 水下视觉场景理解任务的层次结构
Fig. 3 Hierarchical structure of underwater visual scene understanding tasks

2.1 水下目标分类

2.1.1 水下图像分类数据集分类

图像分类任务旨在对水下环境中采集的图像内容进行自动识别和归类,从而预测整幅图像的分类标签。早期研究往往面临数据稀缺与标注不足的挑战,得益于水下摄影设备的普及与数据标注工具的改进,一系列具有代表性的数据集陆续涌现,为算法模型的训练与性能评估提供了坚实基础。根据研究目标和应用需求的不同,现有水下图像分类数据集大致可分为通用类和特定任务类两类,后者涵盖了鱼类识别、珊瑚礁分类、海洋垃圾检测、海洋生物多样性分析、水下环境类型判别以及人机交互等主要方向。这些数据集不仅在类别覆盖上具有较强的代表性,还充分考虑了不同水深、光照条件与水质特征等环境因素对图像质量的影响,从而推动了水下图像分类技术从基于浅层特征的传统方法向深度神经网络驱动的方法转型。表2对水下图像分类数据集进行系统性梳理与总结。

2.1.2 分析与讨论

水下图像分类数据集在物种覆盖度上差异显著。小规模专用数据集(如MLC(marine life classification)9类、NCFM(nature conservancy fisheries monitoring)8类、DIDSON(dual-frequency identification sonar)8类)聚焦于特定应用场景,类别数量少但针对性强,适合特定任务的快速部署;中等规模通用数据集(如WildFish 1 000类)平衡了物种多样性与数据质量,覆盖常见海洋生物,适合通用分类器训练;大规模细粒度数据集(如FishNet 17 357类、Seatizen Atlas 51类)追求物种覆盖的全面性,达到物种级甚至亚种级分类,但面临严重的长尾分布问题:常见物种样本充足,稀有物种样本稀缺,导致模型对少数类别识别率低。在分类粒度方面,早期数据集多采用粗粒度分类(如“鱼类”、“珊瑚”),近期数据集趋向细粒度分类(如具体鱼种、珊瑚种类)。例如,FishNet区分了超过17 000种鱼类,WildFish++提供了2 348种物种的细粒度标注。然而,细粒度标注需要专业

海洋生物学知识,人工成本极高,标注一致性难以保证,且细粒度类别间的视觉差异往往极小,增加了分类难度。

在数据模态与标注类型方面,传统分类数据集主要基于 RGB 图像(如 MLC、WildFish、FishNet),标注形式为图像级类别标签。随着应用需求的发展,部分数据集扩展了多模态数据:FishCLEF(fish task of the Conference and Labs of the Evaluation Forum) 2014 包含视频序列,Marine_PULSE、DIDSON 采用侧扫声呐数据,适合浑浊水体中的识别任务。在标注

丰富度方面,FishCLEF 2015 提供边界框标注支持检测和分类联合训练,WildFish++整合了文本描述支持多模态学习,CADDY(cognitive autonomous diving buddy)系列提供手势和姿态标注支持人机交互研究。然而,多数数据集缺乏属性级标注(如鱼体大小、行为状态、健康状况),限制了细粒度理解和多任务学习的发展。

环境条件的差异显著影响分类任务的难度。可控环境数据集(如 MLC、FishCLEF 2014)在珊瑚礁等相对清澈、光照充足的环境中采集,图像质量高但环

表 2 水下图像分类数据集总结
Table 2 Summary of underwater image classification datasets

数据集	任务类型	数量	物种/类别数	数据模态	标注类型	环境特点
MLC(Beijbom 等,2012)	识别	2 055 幅	9	RGB 图像	点注释	珊瑚礁
FishCLEF 2014(Joly 等,2014)	鱼类识别	约 20 视频	15	视频	标签 + 边界框	珊瑚礁
FishCLEF 2015(Joly 等,2015)	识别 + 计数	>20 000 幅	–	RGB 图像	标签 + 边界框	珊瑚礁
NCFM (FalkSchuetzenmeister 等,2016)	渔业监控	16 915 幅	8	RGB 图像	图像级标签	渔船捕获场景
WildFish(Zhuang 等,2018)	鱼类识别	54 459 幅	1 000	RGB 图像	图像级标签	野生环境
DIDSON(McCann 等,2018)	鱼类识别	524 片段	8	声学图像	物种标签	声呐成像环境
CADDY-Gesture(Gomez 等,2019)	手势识别	16 381 对	15	立体图像	手势标签	开阔海域 + 泳池
CADDY-Pose(Gomez 等,2019)	姿态识别	12 708 对	–	立体图像 + IMU	姿态标注	开阔海域 + 泳池
Fish-gres(Prasetyo 等,2021)	鱼类识别	3 248 幅	8	RGB 图像	图像级标签	多背景条件
SEAMAPD21(Boulais 等,2021)	鱼类细粒度识别	28 328 幅	130	RGB 图像	图像级标签	深度 15~200 m
WildFish++(Zhuang 等,2021)	鱼类识别	103 034 幅	2 348	RGB 图像 + 文本	类别标签	野生环境
CoralNet(Chen 等,2021)	珊瑚礁识别	约 1 700 000 幅	–	RGB 图像	类别标签	全球珊瑚礁
iNaturalistFish ₁₀₀₁ (iNaturalist,2021)	物种识别	54 006 幅	369	RGB 图像	图像级标签	自然环境
FishNet(Khan 等,2023)	多任务物种识别	94 778 幅	17 357	RGB 图像	标签+边界框	全球水域
Marine_PULSE(Du 等,2023)	图像识别	627 幅	4	侧扫声呐	类别标签	海洋工程结构
SeagrassFinder(Elsäßer 等,2024)	海草识别	8 500 幅	2	RGB 图像	二分类标签	海底环境
NKSID(Jiao 等,2024)	开放集长尾识别	2 617 幅	8	前视声呐	类别标签	渤海湾采集
AQUA20(Fuad 等,2025)	挑战性物种识别	8 171 幅	20	RGB 图像	类别标签	多样化退化条件
UVSRD(Saquin 等,2025)	水下信号识别	33 800 幅	12	水下信号图像	信号类别	水下通信场景
Seatizen Atlas(Contini 等,2025)	大规模生物识别	>1 600 000 幅	51	RGB 图像	多标签	印度洋西南部浅水区

注:“–”表示数据未知。

境多样性有限。野生环境数据集(如 WildFish、iNaturalistFish₁₀₀₁)覆盖全球不同水域,包含多种环境条件,但图像质量参差不齐,存在光照不足、色彩偏移和目标模糊等问题。极端环境数据集(如 DIDSON 声呐数据、SEAMAPD21 深度 15~200 m 数据)关注特殊场景,声呐适用于零可见度环境,深度数据覆盖不同水层的光照和压力条件。在场景复杂度方面,复杂背景数据集(如 Fish-gres、iNaturalistFish₁₀₀₁)包含岩石、水草、其他生物等背景干扰,以及目标遮挡、多目标共存等情况,更接近实际应用场景,但分类难度显著增加。当前数据集对动态场景(如鱼群快速游动、水流扰动)和混合物种场景的覆盖仍然不足。

2.2 水下目标检测

2.2.1 水下目标检测数据集总结

水下目标检测不仅要求识别图像中的目标类别,还需精确定位其在图像中的空间位置。与分类任务相比,检测任务对数据集标注的精度与完整性提出了更高要求,需要针对每个目标实例提供准确的边界框标注。随着深度学习目标检测算法的发展,研究人员逐渐开展了针对水下场景的系统性数据集构建工作。根据检测对象与应用需求的不同,现有水下目标检测数据集主要涵盖海洋生物检测、水下工程设施监测、海洋垃圾识别以及水下机器人导航等多个领域。此类数据集的建立不仅为水下目标检测算法的训练、验证与性能评估提供了关键数据支撑,也推动了相关技术在海洋科学研究、生态环境保护与海洋工程应用等方面的深入发展与广泛应用。相关数据集的统计与比较汇总见表3。

2.2.2 分析与讨论

现有水下目标检测数据集在物种覆盖度上呈现两极分化。大规模多物种数据集如 FathomNet (2 244 类)、OBSEA (observatory of the sea) (30 类)、MKUO (marine knowledge-driven underwater organism) (84 类)覆盖了丰富的海洋生物种类,但存在严重的长尾分布问题:常见鱼类样本量充足,而珍稀物种样本极少,导致模型对少数类别检测性能不佳。小规模专用数据集如 UNO (underwater non-living object) (1 类)聚焦于特定任务,类别数量少但样本分布相对均衡,适合专用检测器训练。从应用场景来看,渔业相关数据集(如 Brackish、WildFish++)侧重于商业价值较高的鱼类,生态保护数据集(如 OBSEA、RUOD (real-world underwater object detec-

tion))更关注生物多样性,工程应用数据集(如 SWDD (sea waste detection dataset))针对海洋垃圾、水下结构等非生物目标。当前的主要挑战是如何在保持物种多样性的同时平衡类别分布,未来需要通过数据增强、生成式方法或主动学习策略重点补充稀有物种样本。

在标注精度与细粒度方面,边界框标注数据集(如 OUC-VISION (ocean university of China vision dataset)、Brackish、URPC (underwater robot picking contest)系列)标注效率高但缺乏形状细节,实例分割标注数据集(如 UATD (underwater acoustic target detection)、SCTD (sea creature target detection))提供像素级掩码,标注精度更高但成本显著增加,适合需要精确轮廓的任务。多模态标注数据集(如 FathomNet 的图像+视频+声呐、UXO (underwater explosive ordnance)的声学+光学、Aqua-Eye 的 RGB+事件帧)整合了不同传感器的标注,支持多模态融合检测。标注细粒度方面,早期数据集多采用粗粒度类别(如“鱼类”、“海洋生物”),近期数据集趋向细粒度分类。例如,FathomNet 支持物种级检测,MKUO 区分 84 种水下生物。然而,细粒度标注需要专业海洋生物学知识,人工成本极高,标注一致性也更难保证。部分数据集还提供了丰富的元数据,支持环境因素与检测性能的关联分析。

环境复杂度的差异显著影响检测任务的难度。多样化环境数据集(如 DUO (diverse underwater object)、RUOD)覆盖不同水域、深度和光照条件,训练的模型泛化能力更强;特定环境数据集(如 VDD-C (visual detection dataset-controlled environment)、URPC 系列)场景单一,模型在其他环境中性能下降明显。部分数据集特别关注困难场景:OUC-VISION 针对受控低照环境,Brackish 关注不同浊度条件,DUO 和 UDD (underwater detection dataset)包含大量小尺度目标。然而,现有的数据集对极端深海环境(> 200 m)、极地水域、强湍流条件的覆盖仍然严重不足,对高动态场景(如快速游动的鱼群、水下机器人高速巡航)的覆盖也不足,这些场景的数据采集技术难度大、成本高,是未来需要重点突破的方向。

2.3 水下目标分割

2.3.1 水下语义分割数据集

水下语义分割旨在为图像中的每个像素分配精确的类别标签,从而实现像素级场景解析。在复杂

表 3 水下目标检测数据集总结
Table 3 Summary of underwater object detection datasets

数据集	任务类型	数量	物种/ 类别数	数据模态	标注类型	环境特点
Fish4K-Complex(Kavasidis 等,2014)	复杂场景检测	14 视频	—	水下视频	边界框	复杂水下环境
Labeled Fishes in the Wild (Cutter 等,2015)	鱼类检测跟踪	4 307 幅	—	图像 + 视频	边界框	岩石海底
OUC-VISION(Jian 等,2017)	显著性检测	4 400 幅	220	RGB 图像	边界框	受控水池环境
Brackish(Pedersen 等,2019)	海洋生物检测	14 674 幅	6	RGB 图像	边界框	不同可见度
Trash-ICRA19(Fulton 等,2020)	水下垃圾检测	7 684 幅	3	RGB 图像	边界框	不同水下环境
VDD-C(De Langis 等,2020)	潜水员检测	>100 000 幅	—	RGB 视频	边界框	泳池 + 海域
DUO(Liu 等,2021)	小目标检测	7 782 幅	4	RGB 图像	边界框	多样化水下场景
UODD(Jiang 等,2021)	复杂目标检测	3 194 幅	3	RGB 图像	边界框	复杂/低对比度场景
URPC2017(URPC,2022)	抓取目标检测	18 640 幅	3	RGB 图像	边界框	中国大连幅子岛
URPC2018(URPC,2022)	抓取目标检测	3 701 幅	4	RGB 图像	边界框	大连海域,多分辨率
URPC2019(URPC,2022)	抓取目标检测	5 786 幅	4	RGB 图像	边界框	大连海域,高分辨率
URPC2020 _{ZJ} (URPC,2022)	抓取目标检测	7 543 幅	4	RGB 图像	边界框	大连海域
URPC2020 _{DL} (URPC,2022)	抓取目标检测	8 975 幅	4	RGB 图像	边界框	大连海域,高分辨率
URPC2021(URPC,2022)	抓取目标检测	8 200 幅	4	RGB 图像	边界框	自然环境 ROV 拍摄
UDD(Liu 等,2022)	水下物体检测	2 227 幅	3	RGB 图像	边界框	大连海上农场
FathomNet(Katija 等,2022)	海洋生物检测	84 454 幅	2 244	图像 + 视频 帧	边界框 + 类标签	全球海洋,多深度
UATD(Xie 等,2022)	声呐目标检测	>9 200 幅	10	多波束前视 声呐	边界框	多样化水域
SCTD(Zhang 等,2022)	目标检测	497 幅	3	多种声呐	边界框	多样化场景
UNO(Barrelet 等,2023)	水下垃圾检测	5 902 幅	1	RGB 图像	边界框	多视频序列
RUOD(Fu 等,2023)	水下目标检测	14 000 幅	10	RGB 图像	边界框	多样化挑战场景
OBSEA(Francescangeli 等,2023)	鱼类检测分类	33 805 幅	30	RGB 图像	边界框	西班牙海底缆线
Aqua-Eye(Luo 等,2024)	透明目标检测	13 578 对	5	RGB + 事 件帧	边界框	多光照条件
MKUO(Huang 等,2024)	水下生物监测	10 043 幅	84	RGB 图像	边界框	复杂多样水下环境
SSSforMineDetection (Santos 等,2024)	目标检测	1 170 幅	2	侧扫声呐	边界框	真实水雷探测场景
UXO(Dahn 等,2024)	多模态检测	74 437 幅	3	声学 + 光学	边界框 + 位姿	未爆弹药场景
SWDD(Aubard 等,2025)	目标检测	864 幅 (含增强)	—	侧扫声呐	边界框	水下结构物探测

注:“—”表示数据未知。

的珊瑚礁生态系统、多层次的海洋生物群落以及动态变化的水体环境中,传统分割方法往往难以准确区分不同区域的语义边界。因此,构建高质量的水下语义分割数据集成为推动该领域发展的关键。水下语义分割数据集标注过程成本高昂,不仅需要提供密集的像素级标注,还依赖专业的海洋生物学知

识。近年来,随着深度学习语义分割方法的快速发展,研究人员逐渐系统化地构建面向水下场景的分割数据集。根据分割目标与应用需求的差异,现有水下语义分割数据集主要涵盖珊瑚礁生态系统解析、海底地貌分类、水质环境分析以及海洋生物栖息地划分等多个方向,相关汇总信息见表4。

表 4 水下图像语义分割数据集总结
Table 4 Summary of underwater image semantic segmentation datasets

数据集	任务类型	数量	类别数	数据模态	标注类型	环境特点
DeepFish(Saleh 等, 2020)	多任务分析	43 586 幅	1	RGB 图像	掩码 + 点 + 边界框	20 种海洋栖息地
SUIM(Islam 等, 2020a)	水下场景分割	1 635 幅	8	RGB 图像	掩码	多样化水下场景
WaterDataset(Liang 等, 2020)	水体分割	-	1	图像 + 视频	掩码	多类型水域环境
DUT-USEG(Ma 等, 2021)	生物语义分割	6 617 幅	4	RGB 图像	掩码 + 边界框	多分辨率水下环境
NAUTEC UWI Real (Drews-Jr 等, 2021)	前景背景分割	700 幅	2	RGB 图像	掩码	野外多样水下环境
NAUTEC UWI Sim200 (Drews-Jr 等, 2021)	仿真分割	200 幅	2	RGB 图像	掩码	低浑浊度仿真环境
NAUTEC UWI Sim1000 (Drews-Jr 等, 2021)	仿真分割	1 000 幅	2	RGB 图像	掩码	五级递增浑浊度
MAS3K(Li 等, 2020b)	海洋动物分割	3 103 幅	37 子类	RGB 图像	掩码 + 属性	伪装与普通场景
Marine Debris (Singh 和 Valdenegro-Toro, 2021)	语义分割	1 868 幅	11+背景	前视声呐	掩码	人工水槽垃圾场景
CleanSea(Sánchez-Ferrer 等, 2022)	垃圾分割	1 223 幅	19	RGB 图像	掩码 + 边界框	真实海洋垃圾场景
e-CleanSea(Sánchez-Ferrer 等, 2022)	合成垃圾分割	990 幅	19	合成 RGB 图像	掩码 + 边界框	合成垃圾场景
SASSED(Cobb, 2022)	纹理分割	129 幅	5	合成 孔径声呐	掩码	海底纹理场景
CoralVOS(Zheng 等, 2023)	珊瑚视频分割	60 456 帧	1	RGB 视频	掩码	-
LIACI(Waszak 等, 2023)	船舶检查分割	1 893 幅	10	RGB 图像	掩码	水下船舶检查环境
DeOlhoNosCorais(Furtado 等, 2023)	珊瑚物种分割	1 411 幅	21	RGB 图像	掩码 + 图像级标签	珊瑚礁
UTD(Laptrinhsc, 2023)	垃圾分割	2 900 幅	4	RGB 图像	掩码	水下垃圾环境
Seafloor Sediments(Valles 等, 2023)	海底分割	434 164 幅	-	侧扫声呐	掩码	多种海底类型
CoralSCOP(Zheng 等, 2024)	珊瑚分割	41 297 幅	1	RGB 图像	掩码	全球珊瑚礁站点
S1S2-Water(Wieland 等, 2024)	卫星水体分割	65 组	1	多光谱图像	掩码	全球地表水体
Coralscapes(Sauder 等, 2025)	珊瑚礁分割	2 075 幅	39	RGB 图像	掩码	多状态珊瑚礁环境
UTV(Truong 等, 2025)	视频语义分割	2 000 视频	-	视频 + 文本	掩码 + 文本 描述	真实水下视频场景
SUTV(Truong 等, 2025)	合成视频分割	10 000 视频	-	合成视频	掩码	生成水下场景
AI4Shipwrecks(Sethuraman 等, 2025)	沉船分割	286 幅	2	侧扫声呐	掩码 + 边界框	沉船场景

注：“-”表示数据未知。

2.3.2 水下实例分割数据集

实例分割旨在识别图像中的所有目标,并为每个目标实例生成精确的像素级分割掩码。与传统的图像分类和目标检测相比,实例分割不仅需要判断目标的类别,还须区分同一类别下的不同个体,并对其进行逐像素的分割。在水下场景中,该任务面临更为复杂的挑战。例如,许多海洋生物具有半透明特性,容易导致边界模糊;群体聚集现象与复杂背景

中的目标重叠进一步增加了分割难度。此外,水下目标往往呈现不规则形状和动态姿态变化,对精确提取实例边界提出了更高要求。虽然大规模数据集的建立极大推动了陆地实例分割的研究进展,但水下实例分割仍处于起步阶段。为了推动这一方向的发展,研究人员陆续构建了多个水下实例分割数据集。本小节将对典型的实例分割数据集进行系统介绍,相关信息汇总见表5。

表 5 水下图像实例分割与显著性/伪装目标检测数据集总结

Table 5 Summary of underwater image instance segmentation and saliency/camouflage object detection datasets

数据集	任务类型	数量/幅	类别数	数据模态	标注类型	环境特点
MUED(Jian 等,2019)	显著性目标检测	8 600	430 组	RGB 图像	掩码	多变化因子控制
UFO-120(Islam 等,2020b)	多任务	1 620	—	RGB 图像	掩码	多水域风格
TrashCan-Material(Hong 等,2020)	海洋垃圾实例分割	7 212	16	RGB 图像	掩码 + 边界框	多源视频帧提取
TrashCan-Instance(Hong 等,2020)	海洋垃圾实例分割	7 212	22	RGB 图像	掩码 + 边界框	多源视频帧提取
USOD(Islam 等,2022)	显著性目标检测	300	—	RGB 图像	掩码	加勒比海实地
UHS(Lian 等,2023)	通用水下实例分割	4 628	7	RGB 图像	掩码	质量筛选的复杂场景
Aqua-Test(Liu,2023)	伪装目标检测	455	—	RGB 图像	掩码	强伪装效果场景
MARIS(Đuraš 等,2024)	多域垃圾检测	8 610	40	RGB 图像	掩码 + 标签	ROV 多设备多环境
USIS10K(Lian 等,2024)	显著性实例分割	10 632	7	RGB 图像	掩码 + 边界框	多样化显著性场景
COU(Mukherjee 等,2025)	人造物体实例分割	9 757	24	RGB 图像	掩码 + 边界框	多水环境
UHS10K(Li 等,2025a)	通用实例分割	10 048	10	RGB 图像	掩码 + 边界框	多样化水下环境
USIS16K(Hong 等,2025a)	显著性目标检测	16 151	158	RGB 图像	掩码 + 边界框 + 类别	大规模多层次标注
USOD10K(Hong 等,2025b)	显著性目标检测	10 255	70	RGB 图像	掩码	多样化水下环境

注:“—”表示数据未知。

2.3.3 水下显著性/伪装目标检测

水下显著性检测旨在自动识别和定位水下图像中最能吸引人类视觉注意的区域或目标,从而实现高效的视觉感知。在这一任务中,需要综合考虑海洋生物的运动特征、颜色对比度以及形状特异性等因素。在勘探与测量、船体检查、事件检测、位置识别以及目标定位等任务中,都依赖视觉显著性估计视野中各类目标的重要性,从而做出导航与操作决策。与此相关的水下伪装目标检测则聚焦于识别那些与背景高度融合、难以被常规方法发现的目标。这一任务在海洋生物学研究中具有特殊意义,因为许多海洋生物依靠伪装来适应环境或躲避捕食者。伪装目标检测不仅要应对目标与背景之间的高相似性,还需要克服由水下环境引起的各种视觉干扰。图4展示了4个典型的水下伪装场景:贝类隐藏在沙砾中、比目鱼伪装在海底、章鱼融入礁石环境以及细长的管状生物,这些伪装目标与背景在颜色、纹理上具有高度相似性。第2行对应的二值掩码突出显示了目标的真实轮廓。

随着深度学习技术在这些细分任务中的广泛应用,研究人员们陆续构建了专门面向水下显著性检测与伪装目标检测的数据集,为相关算法的发展提

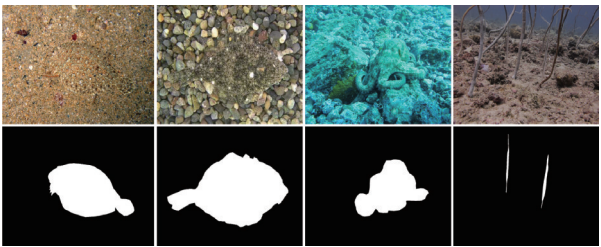


图4 伪装目标示例

Fig. 4 Examples of camouflaged objects

供了关键的数据支撑,相关数据集汇总信息见表5。

2.3.4 分析与讨论

语义分割数据集的类别设计直接影响模型的应用价值。粗粒度分类数据集(如DeepFish 1类)仅区分前景与背景或单一目标类别,标注效率高、一致性好,但语义信息有限;中等粒度分类数据集(如SUIM(segmented underwater image)8类、DUT-USEG(Dalian university of technology underwater segmentation dataset)4类、LIACI(underwater ship lifecycle inspection, analysis and condition information)10类)在常见海洋元素层面进行分类,平衡了标注成本与语义丰富度;细粒度分类数据集(如MAS3K(marine animal segmentation 3K dataset)37类、Coralscapes 39类)对海洋生态系统进行详细分类,支持精细化生态

分析,但标注难度大、主观性强,一致性难以保证。在标注精度方面,掩码标注数据集(占主流)提供像素级类别标签,掩码+边界标注数据集(如DUT-USEG、CleanSea)强化边界信息,掩码+属性标注数据集(如MAS3K)提供多层次语义信息。然而,多数数据集在复杂边界处(如珊瑚分支、鱼鳍边缘)的标注精度不一,细粒度数据集标注主观性问题突出。

在场景多样性方面,单一场景数据集(如DeepFish)虽然在特定场景内数据丰富,但跨场景泛化能力受限。多样化场景数据集(如SUIM、DUT-USEG)覆盖浅海、深海、珊瑚礁和岩石底质等多种环境,训练的模型适应性更强。珊瑚礁数据集(如CoralIVOS、Coralscapes、CoralSCOP)占据重要地位,类别细粒度高、标注精细,但局限于热带和亚热带浅海区域。

综合生态系统数据集(如CleanSea)覆盖更广泛的海洋环境,人工结构数据集(如Marine Debris、LIACI)关注人造目标,支持水下工程和环境监测应用。在数据模态方面,传统数据集主要基于静态RGB图像,部分数据集扩展了新模态:Seafloor Sediments采用侧扫声呐图像,SASSED(synthetic aperture sonar seafloor segmentation dataset)使用合成孔径声呐,AHShipwrecks整合了侧扫声呐和边界框标注,这些声呐数据集填补了视觉方法在恶劣环境下的空白。视频数据集包含时序信息,支持视频语义分割和时序一致性建模,但视频标注成本极高,现有数据集的时序分辨率和标注密度仍然有限。

实例分割和显著性/伪装目标检测任务在标注类型上有所区别。实例分割数据集(如TrashCan系列、UIIS10K(underwater image instance segmentation 10k dataset)、USIS10K、COU、USOD10K)需要为每个目标实例提供独立的像素级掩码和类别标签,标注复杂度高于语义分割。TrashCan系列区分了材质(16类)和实例(22类)两种标注粒度,USIS16K(158类)提供了最细粒度的物种级分类,但标注成本极高。显著性目标检测数据集(如MUED(marine underwater enhancement and detection dataset)、UFO-120(underwater foreground object)、USOD(underwater salient object detection)、Aqua-Test)标注视觉显著的目标区域,不区分具体类别。这类任务通常只需掩码标注,标注效率相对较高,但对标注员的主观判断要求较高。

在应用场景方面,海洋垃圾相关数据集(如

TrashCan系列)标注垃圾的材质和类型,支持自动化海洋清理和污染监测,特点是目标形态多样、姿态不规则、常有遮挡和聚集。生物实例分割数据集(如UIIS10K、USIS10K、USIS16K、USOD10K)关注海洋生物个体的精确分割,支持生物计数、尺寸测量和行为分析,挑战包括生物个体间的遮挡、生物与背景的相似性、透明或半透明特性。多场景通用数据集(如MUED、UIIS10K)同时包含生物和非生物目标,适合通用实例分割模型训练。复杂环境数据集(如TrashCan系列、UIIS10K)包含光照变化、水体浑浊和目标遮挡等多种挑战因素,更接近实际应用场景。在数据规模与标注质量方面,大规模数据集(如USIS16K 16 151幅、USOD10K 10 255幅、UIIS10K 10 632幅)为深度学习模型提供了充足的训练数据,但标注成本极高;中小规模数据集(如MUED 8 600幅、TrashCan系列 7 212幅)虽然规模有限,但往往具有更高的标注精度或更强的任务针对性。高质量标注数据集(如USIS16K)标注精细、边界准确,适合训练精确的分割模型。

2.4 水下目标跟踪

水下目标跟踪旨在在连续的视频序列中持续定位和追踪特定目标的运动轨迹,如图5所示。在海洋生物行为分析、渔业资源监测、水下机器人导航以及海洋环境监控等领域具有广泛的应用前景。与静态图像分析任务不同,目标跟踪需要处理时序信息,在保持跟踪精度的同时确保算法的实时性和鲁棒性。水下环境为目标具有一定的挑战性:1)水体的动态特性导致光照条件快速变化,目标的外观特征在不同帧间可能发生显著差异;2)水流扰动和摄像设备的运动使得背景场景持续变化,增加了跟踪算法的复杂度;3)海洋生物的不规则运动模式、形变以及相互遮挡等现象进一步加剧了跟踪任务的难度。水下跟踪数据集的构建面临着数据采集困难、标注成本高昂以及场景复杂多变等诸多挑战。针对水下场景目标跟踪相关数据集的汇总信息如表6所示。

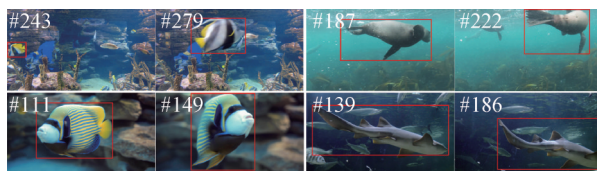


图5 目标跟踪示例

Fig. 5 Examples of object tracking

表 6 水下目标跟踪数据集总结
Table 6 Summary of underwater object tracking datasets

数据集	任务类型	数量	目标类别数	数据模态	环境特点
Fish4K-Trajectory(Beyan 和 Fisher,2013)	鱼类轨迹跟踪	93 序列	–	边界框	珊瑚礁
UOT32(Kezebou 等,2019)	水下目标跟踪	32 序列	12	RGB 视频	高速摄像机拍摄
3D-ZeF(Pedersen 等,2020)	三维鱼类跟踪	8 序列	1	双视角 RGB	商用硬件水族箱
UOT100(Panetta 等,2022)	单目标跟踪	104 序列	–	RGB 视频	真实视频 + 游戏仿真
UTB180(Alawode 等,2022)	水下目标跟踪	180 序列	–	RGB 视频	多种海洋生物
CFC(Kay 等,2022)	声呐鱼类跟踪	1 567 序列	–	声呐视频	5 个河流部署点
BrackishMOT(Pedersen 等,2023)	多目标跟踪	98 序列	6	RGB 视频	咸水环境 9 m 深度
UVOT400(Alawode 等,2023)	水下目标跟踪	400 序列	50	RGB 视频	高分辨率多样化场景
WebUOT-1M(Zhang 等,2024)	目标跟踪	1 500 序列	408	RGB 视频 + 文本	多平台多场景采集
FishTrack23(Dawkins 等,2024)	鱼类单目标跟踪	–	–	RGB 视频	多样化环境
MFT22(Li 等,2024)	鱼类多目标跟踪	10 序列	2	RGB 视频	4 种养殖场景
RGBS50(Li 等,2025c)	多模态目标跟踪	43 702 帧	RGB + 声呐	边界框 + 属性	水池环境多目标
MFT25(Li 等,2025b)	鱼类多目标跟踪	15 序列	2	RGB 视频	水产养殖环境
UW-COT220(Zhang 等,2025)	伪装目标跟踪	220 序列	–	掩码 + 边界框 + 文本	复杂水下伪装环境

注:“–”表示数据未知。

目标跟踪任务依赖于视频序列的时序信息。短序列数据集序列数量有限,如 Fish4K-Trajectory 93 个序列、3D-Zef 8 个序列,主要用于算法验证和概念测试,难以支撑大规模模型训练;中等规模数据集提供了足够的训练样本,如 UTB180(underwater tracking benchmark) 180 个序列、UVOT400 (underwater visual object tracking)400 个序列、UOT100(underwater object tracking) 104 个序列、BrackishMOT(brackish multi-object tracking) 98 个序列、WebUOT-1M(web-scale underwater object tracking)1 500 个序列、UW-COT220 220 个序列,适合主流跟踪算法开发;大规模数据集为深度学习跟踪器提供了充足的数据支撑,如 CFC(continuous fish counting sonar) 1 567 个序列、RGBS50(RGB-sonar)43 702 帧。在序列复杂度方面,简单跟踪场景(如 UOT100 的真实视频 + 游戏仿真、3D-Zef 的商用硬件水族箱)目标运动平滑、遮挡少、背景变化小,跟踪难度较低;复杂跟踪场景(如 UTB180 的多种海洋生物、UVOT400 的高分辨率多样化场景、WebUOT-1M 的多平台多场景采集、UW-COT220 的复杂水下伪装环境)包含快速运动、频

繁遮挡、目标变形、光照剧变以及相似目标干扰等挑战因素。BrackishMOT 和 FishTrack23 专注于咸水环境和多样化环境的多目标跟踪,难度进一步提升。

不同数据集关注的目标类型差异显著。单类目标数据集(如 3D-Zef 的鱼类、UOT32 的 12 类但按单类处理)聚焦特定目标类型,跟踪器可以针对性优化;多类目标数据集(如 UOT32 实际 12 类、UVOT400 的 50 类、WebUOT-1M 的 408 类、RGBS50 的 RGB + 声呐多目标)覆盖多种海洋生物和水下物体,要求跟踪器具备更强的泛化能力。MFT22 和 MFT25 专注于鱼类多目标跟踪,分别在 4 种养殖场和水产养殖环境中采集,具有明显的应用导向。在跟踪粒度方面,单目标跟踪数据集(如 Fish4K-Trajectory、UOT32、UVOT400)为每个视频序列标注一个主要目标的运动轨迹,适合单目标追踪算法开发;多目标跟踪数据集(如 BrackishMOT、RGBS50、MFT22、MFT25)同时标注多个目标的轨迹和身份,支持多目标追踪和目标重识别算法研究,任务难度更高。CFC 提供了声呐鱼类跟踪的 5 个河流部署点数据,结合声呐特性进行目标跟踪。

传统跟踪数据集主要基于RGB视频(如Fish4K-Trajectory、UOT32、UTB180、UVOT400、UOT100、WebUOT-1M、FishTrack23、MFT25和UW-COT220),标注形式为逐帧边界框和目标身份。随着传感器技术的发展,部分数据集扩展了新的模态:双视角RGB数据(如3D-Zef)支持三维跟踪和深度估计;RGB+声呐数据(如RGBS50)融合视觉和声学信息,提升浑浊水体中的跟踪鲁棒性;声呐数据(如CFC)适用于零可见度环境,但目标特征提取难度大。在标注类型方面,边界框标注(占主流)提供目标位置和尺寸,适合大多数跟踪算法;掩码+边界框标注(如UW-COT220提供掩码+边界框+文本)不仅标注目标位置,还提供精确轮廓,支持实例级跟踪和形变分析;属性标注(部分数据集提供目标属性如大小、颜色、行为状态)支持属性驱动的跟踪和检索。此外,WebUOT-1M提供了RGB视频+文本描述的多模态标注,支持语言引导的目标跟踪研究。

水下跟踪面临多种环境挑战。光照变化:从浅海充足光照到深海弱光甚至零光照,光照变化导致目标外观剧烈变化;水体浑浊:悬浮颗粒、藻类和沉积物等导致图像模糊、对比度下降;快速运动:鱼类等海洋生物运动速度快、轨迹不规则,容易运动模糊和跟踪丢失;遮挡:水草、岩石以及其他生物造成的遮挡,特别是密集鱼群中的个体遮挡;目标变形:柔性生物(如水母、章鱼)形态变化大,刚性模板难以适应;相似目标干扰:同种生物的外观高度相似,容易发生身份切换。部分数据集针对特定挑战设计:UW-COT220专注于伪装目标跟踪,目标与背景相似度高;UVOT400强调高分辨率和多样化场景;BrackishMOT关注咸水环境的9 m深度;RGBS50通过RGB+声呐融合应对浑浊水体。然而,现有数据集对极端挑战(如深海高压环境、极地冰下水域、强湍流条件)的覆盖仍然不足,且大多数数据集缺乏系统化的挑战属性标注(如每个序列的光照水平、浑浊度、遮挡程度等),限制了算法性能的细粒度分析。

2.5 评价指标

本节全面介绍了分类、检测、语义分割、实例分割、显著性目标检测和目标跟踪等不同任务的评估指标体系。

2.5.1 基础定义

首先明确以下基本概念:

1) 真正例(true positives, TP): 正确分类的正

样本。

2) 真反例(true negatives, TN): 正确分类的负样本。

3) 假正例(false positives, FP): 错误分类的正样本。

4) 假反例(false negatives, FN): 错误分类的负样本。

2.5.2 分类任务评估指标

1) 准确度(accuracy)。通过正确分类的样本与数据集中样本总数的比值进行计算,衡量整体预测的正确性。

2) 精度(precision, P)。表示正确分类的正样本与预测为正的样本总数之比,量化模型的选择性。

3) 召回率(recall, R)表示所有正样本中被正确预测的比例,量化模型的包容性。

4) F1分数(F1-score)表示精度与召回率的调和平均,平衡模型的性能。

2.5.3 目标检测评估指标

1) 交并比(intersection over union, IoU)。用于评估预测的边界框与真实边界框之间的重叠情况。

2) 平均精度(average precision, AP)表示精度—召回率曲线下的面积,衡量模型在特定类别上排序目标的能力。

3) 平均精度均值(mean average precision, mAP)。表示数据集中所有类别的平均精度。mAP@[0.5:0.95]表示在IoU阈值从0.5到0.95(步长0.05)范围内计算的平均mAP,提供更严格的评估标准。

2.5.4 语义分割评估指标

1) 平均交并比(mean intersection over union, mIoU)。表示数据集中所有类别的平均交并比,具体为

$$f_{\text{mIoU}} = \frac{1}{N} \sum_{i=1}^N \frac{\mathbf{M}_{p,i} \cap \mathbf{M}_{g,i}}{\mathbf{M}_{p,i} \cup \mathbf{M}_{g,i}} \quad (5)$$

式中, $\mathbf{M}_{p,i}$ 和 $\mathbf{M}_{g,i}$ 分别为预测掩码与真实掩码。

2) 平均Dice系数(mean dice coefficient, mDice)。衡量预测与真实掩膜的重叠程度,具体为

$$f_{\text{mDice}} = \frac{1}{N} \sum_{i=1}^N \frac{2|\mathbf{M}_{p,i} \cap \mathbf{M}_{g,i}|}{|\mathbf{M}_{p,i}| + |\mathbf{M}_{g,i}|} \quad (6)$$

3) 平均像素准确度(mean pixel accuracy, mPA)。像素准确度计算正确分类的像素数与图像中像素总数之比,平均像素准确度计算所有类别上平均的像

素准确度,具体为

$$f_{\text{mPA}} = \frac{1}{N} \sum_{i=1}^N \frac{TP_i}{TP_i + FP_i + FN_i} \quad (7)$$

2.5.5 实例分割评估指标

1) 掩膜平均精度 (average precision based on mask, AP^m)。类似于检测任务的AP,但IoU基于分割掩码。

2) 平均召回率 (average recall, AR)。表示在固定数量的检测结果下,不同IoU阈值的平均召回率。

2.5.6 显著性目标检测评估指标

1) 平均绝对误差 (mean absolute error, MAE)。衡量预测显著性图与真实显著性图之间的平均绝对差异,具体为

$$f_{\text{MAE}} = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W |Y(i,j) - \hat{Y}(i,j)| \quad (8)$$

式中, Y 和 $\hat{Y}$ 分别表示真实和预测显著性图, H 和 W 分别为图像的高度和宽度。

2) 最大F-measure。表示在所有可能阈值下的最大F-measure值,具体为

$$F_{\beta}^{\max} = \max_t \frac{(1 + \beta^2) \times P_t \times R_t}{\beta^2 \times P_t + R_t} \quad (9)$$

式中, β^2 通常设置为0.3,强调精度的重要性, t 表示二值化阈值,用于将连续的显著性图转换为二值掩膜。

3) 结构相似性度量 (s-measure, S_m)。用于评估预测与真实显著性图之间的结构相似性,具体为

$$S_m = \alpha \times S_o + (1 - \alpha) \times S_r \quad (10)$$

式中, S_o 和 S_r 分别表示对象级和区域级的结构相似性, α 为权重参数。

4) 增强对齐度量 (enhanced-alignment measure, E_m)。结合了局部像素级和全局图像级的结构信息,具体为

$$E_m = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W \varphi(Y(i,j), \hat{Y}(i,j)) \quad (11)$$

式中, $\varphi(\cdot)$ 表示像素 (i,j) 的增强对齐值。

5) 加权F-measure (F_{β}^w)。表示考虑不同区域重要性的加权F-measure,具体为

$$F_{\beta}^w = \frac{(1 + \beta^2) \times P^w \times R^w}{\beta^2 \times P^w + R^w} \quad (12)$$

式中, P^w 和 R^w 考虑了区域权重。

2.5.7 目标跟踪评估指标

目标跟踪任务通常采用一次性评估 (one-pass

evaluation, OPE) 方法,主要包括以下指标:

1) 精度 (precision, Pre)。用于衡量中心位置误差在预定义阈值内的帧数百分比,具体为

$$Pre = \frac{1}{N_f} \sum_{i=1}^{N_f} \mathbb{I}[d_i \leq \tau] \quad (13)$$

式中, N_f 是总帧数, d_i 是第 i 帧的中心位置误差, τ 是阈值(通常设为20像素), $\mathbb{I}[\cdot]$ 是指示函数。

2) 成功图曲线下面积 (area under curve, AUC)。表示重叠分数高于给定阈值的帧数百分比,定义为成功图曲线下的面积,具体为

$$AUC = \int_0^1 S(t) dt \quad (14)$$

式中, $S(t) = \frac{1}{N_f} \sum_{i=1}^{N_f} \mathbb{I}[IoU_i \geq t]$ 表示IoU阈值为 t 时的成功率。

3) 完整成功图曲线下面积 (complete area under curve, cAUC)。考虑目标的长宽比,使用完整重叠分数,具体为

$$cAUC = \int_0^1 C(t) dt \quad (15)$$

式中, $C(t) = \frac{1}{N_f} \sum_{i=1}^{N_f} \mathbb{I}[cIoU_i \geq t]$,cIoU是考虑长宽比的完整交并比。

4) 平均准确度 (mean accuracy, mACC)。衡量跟踪器提供可靠预测的能力,即使目标消失时也能给出合理估计,具体为

$$mACC = \frac{1}{N_{\text{valid}}} \sum_{i=1}^{N_f} \mathbb{I}[valid_i] \times IoU_i \quad (16)$$

式中, N_{valid} 是有效跟踪帧数, $valid_i$ 表示第 i 帧的跟踪是否有效。

3 水下场景三维重建

水下三维场景重建技术在海底考古、海洋生物研究以及海上基础设施巡检等领域得到了广泛应用。通过水下三维测绘,可以有效监测珊瑚礁的健康状况、追踪入侵物种、沉船检测以及水下建设项目的规划和实施。为了推动水下三维重建技术的发展并对算法性能进行客观评估,研究人员提出了多个针对水下场景三维重建的专用数据集,这些数据集主要服务于不同的技术路线和应用场景。例如,面向实时定位需求的视觉同步定位与建图 (simultaneous localization

and mapping, SLAM)数据集、用于高保真渲染的神经辐射场(neural radiance fields, NeRF)数据集,以及支持快速重建的3D高斯散射(3D Gaussian splatting, 3DGS)数据集等。本节将全面调研现有的水下三维重建数据集,系统分析其数据构成、标注方式、评估指标以及适用的技术方法,为研究人员选择合适的数据集和评估基准提供参考。相关汇总信息如表7所示。

3.1 评价指标

对于基于神经网络的三维重建方法,如3DGS和NeRF,通常采用图像质量评价指标衡量渲染效果,包括峰值信噪比(PSNR)、结构相似性指数(SSIM)和学习感知图像块相似性(learned perceptual image patch similarity, LPIPS)。在水下SLAM系统中,性能评估主要分为地图质量和定位精度两方面。地图质量衡量SLAM生成的环境地图与真实环境之间的匹配程度,常用指标包括SSIM;定位精度衡量机器人在环境中估计位置与姿态的准确性,常用指标包括相对位姿误差(relative pose error, RPE)、绝对位姿误差(absolute pose error, APE)、均方根误差(root mean square error, RMSE)以及绝对轨迹误差(absolute trajectory error, ATE)。

1)SSIM。量化生成图像与真实图像之间的相似性,综合考虑了结构信息、对比度和亮度3个维度的差异。不仅可用于评估三维重建的渲染质量,也可应用于SLAM系统中地图质量的评估。

2)PSNR。衡量图像重建质量的经典指标,数值越高表示重建图像与原图像越相似。

3)LPIPS。基于深度学习特征衡量图像的感知相似性,更能反映人眼对图像质量的主观评价,数值越小表示感知质量越高。

4)RPE。用于评估相邻时刻间位姿估计的相对精度,计算为

$$f_{RPE} = \frac{\|\hat{P} - P\|}{\|P\|} \tag{17}$$

式中, $\hat{P}$ 和 P 分别是预测位置向量和真实位置向量, $\|\cdot\|$ 表示欧几里得范数。

5)APE。用于衡量各时刻位姿估计的绝对精度,反映系统的全局定位能力。具体为

$$f_{APE} = \hat{P} - P \tag{18}$$

6)RMSE。用于衡量位置估计误差的整体水平,定义为

$$f_{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N \|\hat{P}_i - P_i\|^2} \tag{19}$$

式中, $\hat{P}_i$ 和 P_i 分别为时刻 i 的预测位置和真实位置, N 为样本总数。

7)ATE。用于评估系统估计轨迹与真值轨迹的全局一致性,其数值越小表示性能越好,定义为

$$ATE_{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N \|\hat{T}_i - T_i\|^2} \tag{20}$$

式中, $\hat{T}_i$ 和 T_i 分别为时刻 i 的估计位姿和实际位姿, N 是轨迹点数。

3.2 分析与讨论

三维重建数据集的规模通常以图像数量或场景数量衡量。小规模单场景数据集(如UWBundle

表7 水下三维重建数据集总结
Table 7 Summary of underwater 3D reconstruction datasets

数据集	任务类型	数量	场景数	数据模态
UWBundle(Skinner等,2017)	颜色一致性重建	32幅	1	RGB图像+颜色基准
Cave Sonar(Mallios等,2017)	受限环境SLAM	500 m轨迹	2洞	声呐+IMU+DVL+相机
Mbes-slam(Krasnosky等,2022)	测深SLAM	4区域数据	4	多波束+IMU+DVL+GPS
AURORA(Bernardi等,2022)	多传感器SLAM	45 320幅	-	多波束+侧扫+相机+导航
DRUVA(Varghese等,2023)	水下NeRF	20序列	-	水下视频
Seathru(Levy等,2023)	散射介质NeRF	-	4	RGB图像+ROV数据
NUSR(Tang等,2024)	水下NeRF	82幅	4	RGB图像
BVI-Coral(Anantrasirichai,2024)	珊瑚礁3D重建	2 000多幅	23	RGB+激光雷达
S-UW(Wang等,2025)	水下高斯散射	96幅	4	高分辨率RGB

注:“-”表示数据未知。

(underwater bundle adjustment reconstruction) 1个场景 32 幅图像、Cave Sonar 2个洞穴场景、3D-Zef 1个场景、Mbes-slam 4个区域)聚焦特定场景的深度重建,适合算法验证,但泛化能力受限;中等规模多场景数据集(如 Seatru 4个场景、NUSR 4个场景 82 幅图像、S-UW 4个场景 96 幅图像)提供了多个独立场景的数据,支持跨场景性能评估;大规模数据集(如 AURORA 45 (aurora multimodal underwater reconstruction) 320 幅图像、BVI-Coral (british virgin islands coral reef) 2 000+ 幅图像 23 个场景、DRUVA (deep reconstruction underwater video) 20 个视频)为深度学习三维重建方法提供了充足的训练数据。

在场景类型方面,人工环境数据集(如 UWBundle 的颜色卡基准)场景结构规则、纹理丰富,重建相对容易;自然珊瑚礁数据集(如 Seatru、BVI-Coral)场景复杂、纹理变化大、光照不均,更具挑战性但应用价值高;高反射表面数据集(如 S-UW 的高分辨率 RGB)针对金属、玻璃等高反射材料,是三维重建的难点。

多模态融合数据集为三维重建提供了更丰富的信息源。AURORA 融合多波束+侧扫声呐+相机+导航信息,支持复杂环境的多传感器 SLAM。这些多模态数据集的优势在于传感器互补(视觉提供纹理、声呐穿透浑浊水体、IMU 提供姿态、GPS 提供全局定位),但数据采集和标定复杂度高。

在重建方法方面,传统数据集(如 UWBundle)适合离线重建,精度高但计算耗时;SLAM 数据集(如 Cave Sonar、Mbes-slam、AURORA)关注实时定位与建图,适合水下机器人导航;NeRF (neural radiance fields) 数据集(如 DRUVA、Seatru、NUSR、S-UW)采用神经辐射场方法,能够处理光照变化和视角稀疏问题,是当前研究热点。

三维重建的评估需要高质量的标签。精确标定数据集(如 UWBundle)通过受控环境采集,提供准确的相机参数和场景几何,适合定量评估;GPS/导航辅助数据集(如 Mbes-slam、AURORA)提供全局坐标参考,支持大范围重建的精度评估;人工标注数据集(如 Seatru 的 4 个散射质标注场景)通过专家标注或多传感器融合生成参考模型,但精度和完整性有限。

然而,水下三维重建的标签获取极其困难。无标签数据集(如部分 SLAM 数据集)仅提供传感器数据,无法进行绝对精度评估,只能通过相对指标(如

轨迹一致性、闭环误差)间接评价;合成数据集可以提供完美标签,但与真实场景存在差距。未来需要发展高精度水下定位技术(如声学定位网络、光学标记)和标准化测试场景,为三维重建算法提供可靠的评估基准。

4 未来发展方向与挑战

4.1 面向通用基础模型的数据需求

以 GPT (generative pre-trained Transformer)、SAM 等为代表的通用基础模型在自然语言处理和计算机视觉领域取得了革命性进展,其成功经验为水下计算机视觉的发展提供了重要启示。然而,水下领域仍缺乏支撑此类模型训练所需的海量高质量数据资源。目前普遍使用的数据集规模多在数千至数万量级,远不足以满足基础模型对大规模、多样化数据的需求。

因此,构建百万级乃至千万级的水下大规模数据集将成为未来发展的关键方向。这类数据集应具备以下特征:1)具有广泛的地理与环境覆盖范围,囊括全球不同海域、深度与季节的水下场景;2)支持多任务学习,覆盖图像增强、目标检测以及语义分割等多类视觉任务;3)包含多层级标注,从图像级标签到像素级精细标注均具备;4)融合大量弱监督与无监督数据,以支持自监督学习与广义预训练模型的构建。

通用水下基础模型的成功将推动领域范式发生转变。传统依赖特定任务和特定场景的定制化模型开发模式,将逐步演化为“大规模预训练+任务微调”的新范式。相比之下,该范式不仅显著降低特定任务的数据需求与开发成本,还能有效提升模型的泛化能力与环境鲁棒性。同时,基础模型具备强大的迁移学习能力,可将数据丰富海域中学习到的知识迁移至数据匮乏区域,从而缓解水下数据分布不均带来的问题。

4.2 数据标注

复杂的水下环境以及海洋生物的高度多样性增加了水下场景数据标注的复杂性和成本。除了光线折射导致的边界模糊、生物群体聚集造成的个体分割困难等常见技术挑战外,海洋生物种类的精准识别和细分标注更需要依赖专业的海洋生物学知识和丰富的实践经验。传统的人工标注方式面临诸多局

限性:标注过程耗时冗长,难以满足大规模数据处理需求;标注人员的主观判断差异容易引入系统性误差;标注效率相对低下,难以适应快速增长的数据量需求。这些问题严重制约了大规模、高质量水下数据集的构建进程。

以SAM为代表的基础模型的出现为水下数据标注带来了有效的解决方案。通过引入SAM辅助标注技术,可以实现基于用户提示的快速高质量掩码生成,大幅提升标注效率。采用“自动标注+人工校验”的混合标注策略,既体现了基础模型在自动化处理方面的优势,又能确保标注质量的准确性和一致性,从而在保证数据质量的前提下提高标注效率,为大规模水下数据集的构建提供了可行的发展方向。

4.3 泛化能力

水下环境的高度复杂性和多样性严重影响了模型的泛化能力,主要表现在空间维度差异、时间维度变化以及设备局限性等方面。首先,不同水域之间在光照条件、水体透明度、浑浊程度以及生物群落结构等方面存在显著差异,即使是同一水域,在不同深度、不同区域也会呈现出不同的环境特征。其次,同一水域在不同时间节点会因光照强度变化、季节性水温波动以及生物迁徙模式等因素产生明显的环境差异。最后,不同成像传感器、拍摄设备在色彩还原、分辨率以及成像特性等方面存在差异。

当前数据集在构建过程中普遍存在域覆盖不足的问题,多数数据集局限于特定海域或水体环境,并且缺乏长时间跨度和季节性变化的数据采集,同时成像传感器类型相对单一。基于单一场景训练的模型往往难以直接迁移至其他环境,严重限制了模型的泛化能力。

未来的数据集构建应当充分考虑时空完整性和环境多样性,通过跨地域、跨时间和跨设备的全方位数据采集,构建更具代表性的大规模水下数据集,为开发具有强泛化能力的水下视觉识别算法提供数据支撑。

4.4 多模态数据融合

早期水下数据集主要依赖单一RGB视觉模态,在弱光环境、生物遮挡及纹理相似物种识别等场景中存在明显局限。随着深度相机、声呐等设备的发展,多模态融合成为提升系统性能的重要方向。RGB图像提供颜色纹理信息,深度数据补充三维空

间特征,声呐信号在浑浊水体中具备更强穿透能力,融合不同模态可显著提高精度和鲁棒性。然而,当前仍面临传感器时间同步困难、几何配准复杂、数据集规模小且场景单一、融合算法处理模态缺失和噪声能力不足等挑战。

未来多模态水下数据集需在以下方面突破:1)构建标准化采集平台,集成多种传感器并实现高精度时空配准,建立统一的数据采集协议;2)丰富场景覆盖,涵盖浅海、深海、浑浊与清澈水域等环境,支持目标检测、语义分割、三维重建和SLAM等任务,特别关注深海高压、极地低温等极端场景;3)发展通用融合算法框架,支持灵活的模态组合配置和端到端训练,研究模态缺失下的鲁棒学习方法,探索跨模态知识蒸馏技术;4)建立统一标注规范与评估标准,设计综合指标体系评估融合增益和模态缺失时的性能退化。

多模态融合将为水下计算机视觉带来新的技术。视觉与声呐融合可显著提升小目标和遮挡目标的检测精度,多源数据融合能构建更精确的三维场景模型,多模态感知可增强水下机器人的环境理解和作业能力。更重要的是,多模态融合为应对水下环境的不确定性提供了冗余和互补机制,大幅提升系统鲁棒性,为水下智能感知技术的实际应用奠定基础。

4.5 数据共享机制与标准化建设

当前,水下计算机视觉领域面临数据资源分散和标准化不足的挑战。现有数据集在数据格式、标注规范和元数据描述等方面缺乏统一标准,导致不同数据集之间难以直接对比或使用,进而限制了研究成果的可重复性和算法性能的客观评估。同时,部分高质量数据因缺乏完善的数据共享机制而仅在内部流通,造成数据资源重复采集与利用效率低下。因而,构建统一的数据标准化框架已成为推动该领域持续发展的关键基础工作,包括制定规范化的数据格式标准(如图像分辨率、色彩空间等)、统一的标注体系(目标类别定义、边界框与分割标注规则等),以及完善的元数据描述规范(采集时间、地点、水深、水质等环境参数),以确保数据的兼容性、一致性和可追溯性。

构建开放的数据资源共享平台和规范化的基准测试体系,对于促进水下视觉研究的协同发展具有重要意义。共享平台应提供统一的数据存储、检索

与下载服务,支持数据的版本管理、增量更新和质量追溯,并通过合理的激励机制推动高质量数据的贡献,同时兼顾知识产权保护与使用规范。在算法评测方面,需要针对不同任务设计标准化的评估指标与测试协议,构建涵盖多水域、多深度和多光照条件的基准测试数据集,以全面衡量算法的性能与泛化能力,并通过公开透明的排行榜与持续更新机制促进技术进步。通过系统性地推进数据共享机制与标准化建设,可以有效解决当前水下计算机视觉领域数据资源分散、标准不统一以及成果难以对比等问题,为构建更加开放、规范和高效的研究生态系统奠定坚实基础,推动水下计算机视觉技术向更高水平发展。

5 结 语

水下数据集的构建推动了水下计算机视觉的发展。本文系统梳理了当前面向水下任务的数据集。尽管近年来数据集在规模和质量上取得了进步,但仍存在诸多局限性。首先,当前数据集存在地理覆盖不均衡的问题,现有数据集多集中于特定海域,对极地海域等特殊区域的覆盖不足;其次,物种和场景多样性以及时间动态性有限,数据集普遍聚焦于常见海洋生物和典型场景,稀有物种和极端环境场景的样本稀缺。同时,海洋环境具有明显的季节性变化特征,但现有数据集多为短期采集,缺乏长期的动态观测数据;最后,数据的标注质量仍需改进,不同数据集的标注缺乏统一标准,如边界标注精度、显著性目标判断等方面存在差异。这些局限性严重限制了算法的适用性和泛化能力。未来的发展趋势主要体现在3个方面。首先,大规模数据集的构建将成为重要方向,尤其是支撑通用水下基础模型、自监督学习和预训练模型的开发。其次,合成数据将在缓解数据稀缺问题上发挥重要作用,借助生成式技术生成水下合成图像和视频,缓解类不平衡和数据稀缺问题。最后,整合视觉、声呐和深度等多源信息,构建更加丰富和互补的多模态数据集,提升水下视觉任务精度。

总体而言,随着海洋探索需求的不断增长和人工智能技术的快速发展,水下数据集将在推动海洋科学研究、海洋资源开发和海洋环境保护方面发挥越来越重要的作用。未来将有望构建更加全面、高

质量和标准化的水下数据集生态系统,为海洋事业发展提供强有力的数据支撑。

参考文献(References)

- Akkaynak D and Treibitz T. 2019. Sea-Thru: a method for removing water from underwater images//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE: 1682-1691 [DOI: 10.1109/CVPR.2019.00178]
- Alawode B, Dharejo F A, Ummar M, Guo Y H, Mahmood A, Werghi N, et al. 2023. Improving underwater visual tracking with a large scale dataset and image enhancement [EB/OL]. [2025-09-01]. <https://arxiv.org/pdf/2308.15816.pdf>
- Alawode B, Guo Y H, Ummar M, Werghi N, Dias J, Mian A, et al. 2022. UTB180: a high-quality benchmark for underwater tracking//Proceedings of the 16th Asian Conference on Computer Vision on Computer Vision. Macao, China: Springer: 442-458 [DOI: 10.1007/978-3-031-26348-4_26]
- Anantrasrichai N. 2024. BVI-Coral: underwater scenes for 3D reconstruction [EB/OL]. [2025-09-01]. <https://doi.org/10.5281/zenodo.11093417>
- Aubard M, Antal L, Madureira A, Teixeira L F and Ábrahám E. 2025. ROSAR: an adversarial re-training framework for robust side-scan sonar object detection//Proceedings of 2025 Symposium on Maritime Informatics and Robotics. Syros, Greece: IEEE: 1-8 [DOI: 10.1109/MARIS64137.2025.11139627]
- Barrelet C, Chaumont M, Subsol G, Creuze V and Gouttefarde M. 2023. From TrashCan to UNO: deriving an underwater image dataset to get a more consistent and balanced version//Proceedings of the Pattern Recognition, Computer Vision, and Image Processing. ICPR 2022 International Workshops and Challenges. Montreal, Canada: Springer: 403-414 [DOI: 10.1007/978-3-031-37731-0_30]
- Bejbom O, Edmunds P J, Kline D I, Mitchell B G and Kriegman D. 2012. Automated annotation of coral reef survey images//Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE: 1170-1177 [DOI: 10.1109/CVPR.2012.6247798]
- Berman D, Levy D, Avidan S and Treibitz T. 2021. Underwater single image color restoration using haze-lines and a new quantitative dataset. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(8): 2822-2837 [DOI: 10.1109/TPAMI.2020.2977624]
- Bernardi M, Hosking B, Petrioli C, Bett B J, Jones D, Al Huvenne V, et al. 2022. AURORA, a multi-sensor dataset for robotic ocean exploration. The International Journal of Robotics Research, 41(5): 461-469 [DOI: 10.1177/02783649221078612]
- Beyan C and Fisher R B. 2013. Detecting abnormal fish trajectories

- using clustered and labeled data//Proceedings of 2013 IEEE International Conference on Image Processing. Melbourne, Australia; IEEE: 1476-1480 [DOI: 10.1109/ICIP.2013.6738303]
- Boulais O, Alaba S Y, Ball J E, Campbell M, Iftekhar A T, Moorehead R, et al. 2021. SEAMAPD21: a large-scale reef fish dataset for fine-grained categorization//Proceedings of the FGVC8: the Eight Workshop on Fine-Grained Visual Categorization. [s.l.]: [s.n.]: 2-3
- Chang L B, Song H J, Li M J and Xiang M. 2023. UIDEF: a real-world underwater image dataset and a color-contrast complementary image enhancement framework. *ISPRS Journal of Photogrammetry and Remote Sensing*, 196: 415-428 [DOI: 10.1016/j.isprsjprs.2023.01.007]
- Chen L, Huang Y Z, Dong J Y, Xu Q, Kwong S, Lu H M, et al. 2025. Underwater optical object detection in the era of artificial intelligence: current, challenge, and future. *ACM Computing Surveys*, 58(3): #62 [DOI: 10.1145/3759243]
- Chen Q M, Beijbom O, Chan S, Bouwmeester J and Kriegman D. 2021. A new deep learning engine for CoralNet//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Montreal, Canada; IEEE: 3686-3695 [DOI: 10.1109/ICCVW54120.2021.00412]
- Cobb J T. 2022. Synthetic aperture sonar seabed environment dataset (SASSED)[EB/OL]. Mendeley Data, V4 [2025-09-01]. <https://doi.org/10.17632/s5j5gzzr2vc.4>
- Contini M, Illien V, Julien M, Ravitchandiran M, Russias V, Lazenec A, et al. 2025. Seatizen Atlas: a collaborative dataset of underwater and aerial marine imagery. *Scientific Data*, 12(1): #67 [DOI: 10.1038/s41597-024-04267-z]
- Cui Z, Wang S, Wang X, Duan H, Yang C X and Na J. 2025. Hybrid underwater image enhancement based on color transfer and adaptive gain control. *Acta Automatica Sinica*, 51(2): 376-390 (崔峥, 王森, 王嫻, 段浩, 杨春曦, 那靖. 2025. 基于颜色转移和自适应增益控制的混合水下图像增强. *自动化学报*, 51(2): 376-390) [DOI: 10.16383/j.aas.c240491]
- Cutter G, Stierhoff K and Zeng J M. 2015. Automated detection of rockfish in unconstrained underwater videos using Haar cascades and a new image dataset: labeled fishes in the wild//Proceedings of 2015 IEEE Winter Applications and Computer Vision Workshops. Waikoloa, USA; IEEE: 57-62 [DOI: 10.1109/WACVW.2015.11]
- Dahn N, Firvida M B, Sharma P, Christensen L, Geisler O, Mohrmann J, et al. 2024. An acoustic and optical dataset for the perception of underwater unexploded ordnance (UXO)//Proceedings of OCEANS 2024 — Halifax. Halifax, Canada; IEEE: 1-6 [DOI: 10.1109/OCEANS55160.2024.10754316]
- Dawkins M, Prior J, Lewis B, Faillettaz R, Banez T, Salvi M, et al. 2024. FishTrack23: an ensemble underwater dataset for multi-object tracking//Proceedings of 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa, USA; IEEE: 7152-7161 [DOI: 10.1109/WACV57701.2024.00701]
- De Langis K, Fulton M and Sattar J. 2020. An analysis of deep object detectors for diver detection [EB/OL]. [2025-09-01]. <https://arxiv.org/pdf/2012.05701.pdf>
- Deng J, Dong W, Socher R, Li L J, Li K and Li F F. 2009. ImageNet: a large-scale hierarchical image database//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA; IEEE: 248-255 [DOI: 10.1109/CVPR.2009.5206848]
- Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X H, Unterthiner T, et al. 2021. An image is worth 16 × 16 words: transformers for image recognition at scale//Proceedings of the 9th International Conference on Learning Representations. [s.l.]: OpenReview.net: 1-12
- Drewe-Jr P, De Souza I, Maurell I P, Protas E V and Botelho S S C. 2021. Underwater image segmentation in the wild using deep learning. *Journal of the Brazilian Computer Society*, 27(1): #12 [DOI: 10.1186/s13173-021-00117-7]
- Du D Z, Li E H, Si L Y, Xu F J and Niu J W. 2024. End-to-end underwater video enhancement: dataset and model [EB/OL]. [2025-09-01]. <https://arxiv.org/pdf/2403.11506.pdf>
- Du X, Sun Y F, Song Y P, Dong L F and Zhao X L. 2023. Revealing the potential of deep learning for detecting submarine pipelines in side-scan sonar images: an investigation of pre-training datasets. *Remote Sensing*, 15(19): #4873 [DOI: 10.3390/rs15194873]
- Duarte A, Codevilla F, De O Gaya J and Botelho S S C. 2016. A dataset to evaluate underwater image restoration methods//Proceedings of the OCEANS 2016 — Shanghai. Shanghai, China; IEEE: 1-6 [DOI: 10.1109/OCEANSAP.2016.7485524]
- Đuraš A, Wolf B J, Ilioudi A, Palunko I and De Schutter B. 2024. A dataset for detection and segmentation of underwater marine debris in shallow waters. *Scientific Data*, 11(1): #921 [DOI: 10.1038/s41597-024-03759-2]
- Elsäßer J, Wehl L, Cheplygina V and Nielsen L T. 2024. Seagrass-Finder: an underwater eelgrass image classification dataset [EB/OL]. [2025-09-01]. <https://doi.org/10.5281/zenodo.13904604>
- FalkSchuetzenmeister, Matt M, Meg Risdal, suepollock and Wendy Kan. 2016. The nature conservancy fisheries monitoring [EB/OL]. [2025-09-01]. <https://www.kaggle.com/competitions/the-nature-conservancy-fisheries-monitoring/data>
- Francescangeli M, Marini S, Martínez E, Del Río J, Toma D M, Nogueras M, et al. 2023. Image dataset for benchmarking automated fish detection and classification algorithms. *Scientific Data*, 10(1): #5 [DOI: 10.1038/s41597-022-01906-1]
- Fu C P, Liu R S, Fan X, Chen P Y, Fu H, Yuan W Q, et al. 2023. Rethinking general underwater object detection: datasets, challenges, and solutions. *Neurocomputing*, 517: 243-256 [DOI: 10.1016/j.neucom.2022.10.039]
- Fuad T R, Ahmed S and Ivan S. 2025. AQUA20: a benchmark dataset for underwater species classification under challenging conditions

- [EB/OL]. [2025-09-01]. <https://arxiv.org/pdf/2506.17455.pdf>
- Fulton M S, Hong J and Saar J. 2020. Trash-ICRA19: a bounding box labeled dataset of underwater trash [EB/OL]. [2025-09-01]. <https://doi.org/10.13020/x0qn-y082>
- Furtado D P, Vieira E A, Nascimento W F, Inagaki K Y, Bleuel J, Alves M A Z, et al. 2023. #DeOlhoNosCorais: a polygonal annotated dataset to optimize coral monitoring. *PeerJ*, 11: #e16219 [DOI: 10.7717/peerj.16219]
- Gomez A G, Ranieri A, Chiarella D, Zereik E, Babić A and Birk A. 2019. CADDY underwater stereo-vision dataset for human-robot interaction (HRI) in the context of diver activities. *Journal of Marine Science and Engineering*, 7(1): #16 [DOI: 10.3390/jmse7010016]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- Hong J, Fulton M and Sattar J. 2020. TrashCan: a semantically-segmented dataset towards visual detection of marine debris [EB/OL]. [2025-09-01]. <https://arxiv.org/pdf/2007.08097.pdf>
- Hong L, Wang X, Li Y H and Wang X. 2025a. USIS16K: high-quality dataset for underwater salient instance segmentation [EB/OL]. [2025-09-01]. <https://arxiv.org/pdf/2506.19472.pdf>
- Hong L, Wang X, Zhang G and Zhao M. 2025b. USOD10K: a new benchmark dataset for underwater salient object detection. *IEEE Transactions on Image Processing*, 34: 1602-1615 [DOI: 10.1109/TIP.2023.3266163]
- Hou G J, Li Y X, Yang H, Li K Q and Pan Z K. 2023. UID2021: an underwater image dataset for evaluation of no-reference quality assessment metrics. *ACM Transactions on Multimedia Computing, Communications and Applications*, 19(4): #151 [DOI: 10.1145/3578584]
- Hou G J, Zhao X, Pan Z K, Yang H, Tan L and Li J M. 2020. Benchmarking underwater image enhancement and restoration, and beyond. *IEEE Access*, 8: 122078-122091 [DOI: 10.1109/ACCESS.2020.3006359]
- Huang J F, Zhang T J, Zhao S J, Zhang L and Zhou Y C. 2024. An underwater organism image dataset and a lightweight module designed for object detection networks. *ACM Transactions on Multimedia Computing, Communications and Applications*, 20(5): #147 [DOI: 10.1145/3640465]
- iNaturalist. 2021. iNaturalist [EB/OL]. [2025-09-01]. <https://www.inaturalist.org/>
- Islam M J, Edge C, Xiao Y Y, Luo P G, Mehtaz M, Morse C, et al. 2020a. Semantic segmentation of underwater imagery: dataset and benchmark//*Proceedings of 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Las Vegas, USA: IEEE: 1769-1776 [DOI: 10.1109/IROS45743.2020.9340821]
- Islam M J, Enan D S, Luo P G and Sattar J. 2020b. Underwater image super-resolution using deep residual multipliers//*Proceedings of 2020 IEEE International Conference on Robotics and Automation (ICRA)*. Paris, France: IEEE: 900-906 [DOI: 10.1109/ICRA40945.2020.9197213]
- Islam M J, Luo P G and Sattar J. 2020c. Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception [EB/OL]. [2025-09-01]. <https://arxiv.org/pdf/2002.01155.pdf>
- Islam M J, Wang R B and Sattar J. 2022. SVAM: saliency-guided visual attention modeling by autonomous underwater robots//Huang S D, Hauser K, Shell D, eds. *Robotics: Science and Systems 2022*. New York: [s.n.] #48
- Islam M J, Xia Y Y and Sattar J. 2020d. Fast underwater image enhancement for improved visual perception. *IEEE Robotics and Automation Letters*, 5(2): 3227-3234 [DOI: 10.1109/LRA.2020.2974710]
- Jian M W, Qi Q, Dong J Y, Yin Y L, Zhang W Y and Lam K M. 2017. The OUC-vision large-scale underwater image database//*Proceedings of 2017 IEEE International Conference on Multimedia and Expo (ICME)*. Hong Kong, China: IEEE: 1297-1302 [DOI: 10.1109/ICME.2017.8019324]
- Jian M W, Qi Q, Yu H, Dong J Y, Cui C R, Nie X S, et al. 2019. The extended marine underwater environment database and baseline evaluations. *Applied Soft Computing*, 80: 425-437 [DOI: 10.1016/j.asoc.2019.04.025]
- Jiang L H, Wang Y, Jia Q, Xu S W, Liu Y, Fan X, et al. 2021. Underwater species detection using channel sharpening attention//*Proceedings of the 29th ACM International Conference on Multimedia*. [s.l.]: ACM: 4259-4267 [DOI: 10.1145/3474085.3475563]
- Jiao W P, Zhang J L and Zhang C Y. 2024. Open-set recognition with long-tail sonar images. *Expert Systems with Applications*, 249: #123495 [DOI: 10.1016/j.eswa.2024.123495]
- Joly A, Goëau H, Glotin H, Spampinato C, Bonnet P, Vellinga W P, et al. 2014. LifeCLEF 2014: multimedia life species identification challenges//*Proceedings of the 5th International Conference of the CLEF Initiative on Information Access Evaluation—Multilinguality, Multimodality, and Interaction*. Sheffield, UK: Springer: 229-249 [DOI: 10.1007/978-3-319-11382-1_20]
- Joly A, Müller H, Goëau H, Glotin H, Spampinato C, Rauber A, et al. 2015. LifeCLEF 2015: multimedia life species identification//*Proceedings of the 6th International Conference on Experimental IR Meets Multilinguality*. Toulouse, France: Springer: 1-15 [DOI: 10.1007/978-3-319-24027-5_46]
- Katija K, Orenstein E, Schlöning B, Lundsten L, Barnard K, Sainz G, et al. 2022. FathomNet: a global image database for enabling artificial intelligence in the ocean. *Scientific Reports*, 12(1): #15914 [DOI: 10.1038/s41598-022-19939-2]
- Kavasidis I, Palazzo S, Di Salvo R, Giordano D and Spampinato C. 2014. An innovative web-based collaborative platform for video annotation. *Multimedia Tools and Applications*, 70(1): 413-432 [DOI: 10.1007/s11042-013-1419-7]
- Kay J, Kulits P, Stathatos S, Deng S Q, Young E, Beery S, et al.

2022. The Caltech Fish Counting Dataset: a benchmark for multiple-object tracking and counting//Proceedings of the 17th European Conference on Computer Vision (ECCV 2022). Tel Aviv, Israel: Springer: 290-311 [DOI: 10.1007/978-3-031-20074-8_17]
- Kezebou L, Oludare V, Panetta K and Agaian S S. 2019. Underwater object tracking benchmark and dataset//Proceedings of 2019 IEEE International Symposium on Technologies for Homeland Security (HST). Woburn, USA: IEEE: 1-6 [DOI: 10.1109/HST47167.2019.9032954]
- Khan F F, Li X, Temple A J and Elhoseiny M. 2023. FishNet: a large-scale dataset and benchmark for fish recognition, detection, and functional trait prediction//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France: IEEE: 20439-20449 [DOI: 10.1109/ICCV51070.2023.01874]
- Kirillov A, Mintun E, Ravi N, Mao H Z, Rolland C, Gustafson L, et al. 2023. Segment anything//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France: IEEE: 3992-4003 [DOI: 10.1109/ICCV51070.2023.00371]
- Kittinger J N, Rotjan R D, Hanich Q, Pasisi B and Rambourg C. 2024. Balancing protection and production in ocean conservation. npj Ocean Sustainability, 3 (1) : #24 [DOI: 10.1038/s44183-024-00062-w]
- Krasnosky K, Roman C and Casagrande D. 2022. A bathymetric mapping and SLAM dataset with high-precision ground truth for marine robotics. The International Journal of Robotics Research, 41 (1) : 12-19 [DOI: 10.1177/02783649211044749]
- Laptrinhsc. 2023. Underwater trash detection dataset [EB/OL]. Roboflow Universe [2025-09-01].
<https://universe.roboflow.com/laptrinhsc/underwatertrashdetection>
- Levy D, Peleg A, Pearl N, Rosenbaum D, Akkaynak D, Korman S, et al. 2023. SeaThru-NeRF: neural radiance fields in scattering media//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE: 56-65 [DOI: 10.1109/CVPR52729.2023.00014]
- Li C Y, Guo C L, Ren W Q, Cong R M, Hou J H, Kwong S, et al. 2020a. An underwater image enhancement benchmark dataset and beyond. IEEE Transactions on Image Processing, 29: 4376-4389 [DOI: 10.1109/TIP.2019.2955241]
- Li H, Lian S J, Li Z Y, Cong R M, Li C Y, Yang L T, et al. 2025a. Advancing marine research: UWSAM framework and UIIS10K dataset for precise underwater instance segmentation [EB/OL]. [2025-09-01]. <https://arxiv.org/pdf/2505.15581.pdf>
- Li H Y, Li J J and Wang W. 2019. A fusion adversarial underwater image enhancement network with a public test dataset [EB/OL]. [2025-09-01]. <https://arxiv.org/pdf/1906.06819.pdf>
- Li L, Rigall E, Dong J Y and Chen G. 2020b. MAS3K: an open dataset for marine animal segmentation//The 3rd BenchCouncil International Symposium on Benchmarking, Measuring, and Optimizing [s.l.]: Springer: 194-212 [DOI: 10.1007/978-3-030-71058-3_12]
- Li W R, Liu Y Q, Guo Q N, Wei Y J, Leo H L and Li Z B. 2025b. When trackers date fish: a benchmark and framework for underwater multiple fish tracking [EB/OL]. [2025-09-01].
<https://arxiv.org/pdf/2507.06400.pdf>
- Li W R, Liu Y Q, Wang W X, Li Z B and Yue J. 2024. TFMFT: transformer-based multiple fish tracking. Computers and Electronics in Agriculture, 217: #108600 [DOI: 10.1016/j.compag.2023.108600]
- Li X L. 2024a. Water-related optics. Scientia Sinica Informations, 54(2): 227-280 (李学龙. 2024a. 涉水光学. 中国科学: 信息科学), 54(2): 227-280 [DOI: 10.1360/SSI-2022-0398]
- Li X L. 2024b. Water-related vision. Acta Electronica Sinica, 52(4): 1041-1082 (李学龙. 2024b. 涉水视觉. 电子学报, 52(4): 1041-1082) [DOI: 10.12263/DZXB.20230698]
- Li Y F, Wang B, Sun J R, Wu X Y and Li Y. 2025c. RGB-sonar tracking benchmark and spatial cross-attention transformer tracker. IEEE Transactions on Circuits and Systems for Video Technology, 35(3): 2260-2275 [DOI: 10.1109/TCSVT.2024.3497214]
- Lian S J, Li H, Cong R M, Li S Q, Zhang W and Kwong S. 2023. WaterMask: instance segmentation for underwater imagery//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France: IEEE: 1305-1315 [DOI: 10.1109/ICCV51070.2023.00126]
- Lian S J, Zhang Z Y, Li H, Li W J, Yang L T, Kwong S, et al. 2024. Diving into underwater: segment anything model guided underwater salient instance segmentation and a large-scale dataset//Proceedings of the 41st International Conference on Machine Learning. Vienna, Austria: JMLR.org: #1190 [DOI: 10.48550/arXiv.2406.06039]
- Liang Y Q, Jafari N, Luo X, Chen Q, Cao Y P and Li X. 2020. WaterNet: an adaptive matching pipeline for segmenting water with volatile appearance. Computational Visual Media, 6(1): 65-78 [DOI: 10.1007/s41095-020-0156-x]
- Liu C W, Li H J, Wang S C, Zhu M, Wang D, Fan X, et al. 2021. A dataset and benchmark of underwater object detection for robot picking//Proceedings of 2021 IEEE International Conference on Multimedia and Expo Workshops (ICMEW). Shenzhen, China: IEEE: 1-6 [DOI: 10.1109/ICMEW53276.2021.9455997]
- Liu C W, Wang Z H, Wang S J, Tang T, Tao Y L, Yang C F, et al. 2022. A new dataset, poisson GAN and AquaNet for underwater object grabbing. IEEE Transactions on Circuits and Systems for Video Technology, 32(5): 2831-2844 [DOI: 10.1109/TCSVT.2021.3100059]
- Liu R S, Fan X, Zhu M, Hou M J and Luo Z X. 2020. Real-world underwater enhancement: challenges, benchmarks, and solutions under natural light. IEEE Transactions on Circuits and Systems for Video Technology, 30(12): 4861-4875 [DOI: 10.1109/TCSVT.2019.2963772]

- Liu Y W. 2023. Aqua-test [EB/OL]. [2025-09-01].
<https://www.kaggle.com/datasets/mikurules/aqua-test>
- Luo C, Wu J H, Sun S X and Ren P. 2024. TransCODNet: underwater transparently camouflaged object detection via RGB and event frames collaboration. *IEEE Robotics and Automation Letters*, 9(2): 1444-1451 [DOI: 10.1109/LRA.2023.3346754]
- Ma Z W, Li H J, Wang Z H, Yu D, Wang T Y, Gu Y S, et al. 2021. An underwater image semantic segmentation method focusing on boundaries and a real underwater scene semantic segmentation dataset [EB/OL]. [2025-09-01].
<https://arxiv.org/pdf/2108.11727.pdf>
- Mallios A, Vidal E, Campos R and Carreras M. 2017. Underwater caves sonar data set. *The International Journal of Robotics Research*, 36(12): 1247-1251 [DOI: 10.1177/0278364917732838]
- Marques T P, Albu A B and Hoeberechts M. 2019. A contrast-guided approach for the enhancement of low-lighting underwater images. *Journal of Imaging*, 5(10): #79 [DOI: 10.3390/jimaging5100079]
- McCann E, Li L L, Pangle K, Johnson N and Eickholt J. 2018. An underwater observation dataset for fish classification and fishery assessment. *Scientific Data*, 5(1): #180190 [DOI: 10.1038/sdata.2018.190]
- Mukherjee R, Singh S, McWilliams J and Sattar J. 2025. The common objects underwater (COU) dataset for robust underwater object detection//*Proceedings of 2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Hangzhou, China: IEEE: 18597-18603 [DOI: 10.1109/IROS60139.2025.11246089]
- Nie J, Zuo Z J, Huang L, Wang Z G, Sun Z Y, Zhong G Q, et al. 2022. Marine oriented multimodal intelligent computing: challenges, progress and prospects. *Journal of Image and Graphics*, 27(9): 2589-2610 (聂婕, 左子杰, 黄磊, 王志刚, 孙正雅, 仲国强, 等. 2022. 面向海洋的多模态智能计算: 挑战、进展和展望. *中国图象图形学报* 27(9): 2589-2610 [DOI: 10.11834/jig.211267])
- Panetta K, Kezebou L, Oludare V and Agaian S. 2022. Comprehensive underwater object tracking benchmark dataset and underwater image enhancement with GAN. *IEEE Journal of Oceanic Engineering*, 47(1): 59-75 [DOI: 10.1109/JOE.2021.3086907]
- Pedersen M, Haurum J B, Bengtson S H and Moeslund T B. 2020. 3D-ZeF: a 3D zebrafish tracking benchmark dataset//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, USA: IEEE: 2423-2433 [DOI: 10.1109/CVPR42600.2020.00250]
- Pedersen M, Haurum J B, Gade R and Moeslund T B. 2019. Detection of marine animals in a new underwater dataset with varying visibility//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. Long Beach, USA: IEEE: 18-26
- Pedersen M, Lehotský D, Nikolov I and Moeslund T B. 2023. Brackish-MOT: the Brackish multi-object tracking dataset//*Proceedings of the 22nd Scandinavian Conference on Image Analysis (SCIA 2023)*. Sirkka, Finland: Springer: 17-33 [DOI: 10.1007/978-3-031-31435-3_2]
- Peng L T, Zhu C L and Bian L H. 2023. U-shape transformer for underwater image enhancement. *IEEE Transactions on Image Processing*, 32: 3066-3079 [DOI: 10.1109/TIP.2023.3276332]
- Prasetyo E, Suciati N and Fatichah C. 2021. Fish-gres dataset for fish species classification [EB/OL]. Mendeley Data, V1 [2025-09-01].
<https://data.mendeley.com/datasets/76cr3wfhff/1>
- Purnima K and Kumar C S. 2024. CSUID — comprehensive synthetic underwater image dataset. *Data in Brief*, 55: #110723 [DOI: 10.1016/j.dib.2024.110723]
- Qi Q, Zhang Y C, Tian F, Wu Q M J, Li K Q, Luan X, et al. 2022. Underwater image co-enhancement with correlation feature matching and joint learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(3): 1133-1147 [DOI: 10.1109/TCSVT.2021.3074197]
- Saleh A, Laradji I H, Konovalov D A, Bradley M, Vazquez D and Sheaves M. 2020. A realistic fish-habitat dataset to evaluate algorithms for underwater visual analysis. *Scientific Reports*, 10(1): #14671 [DOI: 10.1038/s41598-020-71639-x]
- Sánchez-Ferrer A, Gallego A J, Valero-Mas J J and Calvo-Zaragoza J. 2022. The CleanSea Set: a benchmark corpus for underwater debris detection and recognition//*Proceedings of the 10th Iberian Conference on Pattern Recognition and Image Analysis*. Aveiro, Portugal: Springer: 616-628 [DOI: 10.1007/978-3-031-04881-4_49]
- Santos N P, Moura R, Torgal G S, Lobo V and De Castro Neto M. 2024. Side-scan sonar imaging data of underwater vehicles for mine detection. *Data in Brief*, 53: #110132 [DOI: 10.1016/j.dib.2024.110132]
- Sauder J, Domazetoski V, Banc-Prandi G, Perna G, Meibom A and Tuia D. 2025. The coralscapes dataset: semantic scene understanding in coral reefs [EB/OL]. [2025-09-01].
<https://arxiv.org/pdf/2503.20000.pdf>
- Saquín H. 2025. Underwater visual signals recognition dataset (UVSRD). *IEEE Dataport* [DOI: 10.21227/ew07-nd78]
- Sethuraman A V, Sheppard A, Bagoren O, Pinnow C, Anderson J, Havens T C, et al. 2025. Machine learning for shipwreck segmentation from side scan sonar imagery: dataset and benchmark. *The International Journal of Robotics Research*, 44(3): 341-354 [DOI: 10.1177/02783649241266853]
- Skinner K A, Iscar E and Johnson-Roberson M. 2017. Automatic color correction for 3D reconstruction of underwater scenes//*Proceedings of 2017 IEEE International Conference on Robotics and Automation (ICRA)*. Singapore, Singapore: IEEE: 5140-5147 [DOI: 10.1109/ICRA.2017.7989601]
- Singh D and Valdenegro-Toro M. 2021. The marine debris dataset for forward-looking sonar semantic segmentation [EB/OL]. [2025-09-01].
<https://arxiv.org/pdf/2108.06800.pdf>
- Tang Y K, Zhu C X, Wan R J, Xu C and Shi B X. 2024. Neural underwater scene representation//*Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

- Seattle, USA: IEEE: 11780-11789 [DOI: 10.1109/CVPR52733.2024.01119]
- Truong Q T, Kwan W Y, Nguyen D T, Hua B S and Yeung S K. 2025. AUTV: creating underwater video datasets with pixel-wise annotations [EB/OL]. [2025-09-01]. <https://arxiv.org/pdf/2503.12828.pdf>
- URPC. 2022. Underwater robot professional contest [EB/OL]. [2025-09-01]. <http://www.urpc.org.cn/index.html>
- Valles M C, De Borja F, Rajani H H, Gracias N and Garcia R. 2023. A large scale side-scan sonar dataset of seafloor sediments for self-supervised pretraining [EB/OL]. [2025-09-01]. <https://doi.org/10.5281/zenodo.10209445>
- Varghese N, Kumar A and Rajagopalan A N. 2023. Self-supervised monocular underwater depth recovery, image restoration, and a real-sea video dataset//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France: IEEE: 12214-12224 [DOI: 10.1109/ICCV51070.2023.01125]
- Wang H R, Anantrasirichai N, Zhang F and Bull D. 2025. UW-GS: distractor-aware 3D Gaussian splatting for Enhanced underwater scene reconstruction//Proceedings of 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Tucson, USA: IEEE: 3280-3289 [DOI: 10.1109/WACV61041.2025.00324]
- Waszak M, Cardaillac A, Elvæsæter B, Rødølen F and Ludvigsen M. 2023. Semantic segmentation in underwater ship inspections: benchmark and dataset. IEEE Journal of Oceanic Engineering, 48(2): 462-473 [DOI: 10.1109/JOE.2022.3219129]
- Wieland M, Fichtner F, Martinis S, Groth S, Krullikowski C, Plank S, et al. 2024. SIS2-water: a global dataset for semantic segmentation of water bodies from sentinel-1 and sentinel-2 satellite images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17: 1084-1099 [DOI: 10.1109/JSTARS.2023.3333969]
- Wu D, Yuan F and Cheng E. 2020. Underwater no-reference image quality assessment for display module of ROV. Scientific Programming, 2020(1): #8856640 [DOI: 10.1155/2020/8856640]
- Xie K B, Yang J and Qiu K. 2022. A dataset with multibeam forward-looking sonar for underwater object detection. Scientific Data, 9(1): #739 [DOI: 10.1038/s41597-022-01854-w]
- Xie Y F, Kong L W, Chen K, Zheng Z Q, Yu X, Yu Z B, et al. 2024. UVEB: a large-scale benchmark and baseline towards real-world underwater video enhancement//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 22358-22367 [DOI: 10.1109/CVPR52733.2024.02110]
- Ye T, Chen S X, Liu Y, Ye Y, Chen E K and Li Y C. 2022. Underwater light field retention: neural rendering for underwater imaging//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. New Orleans, USA: IEEE: 487-496 [DOI: 10.1109/CVPRW56347.2022.00064]
- Zhang C H, Liu L, Huang G J, Wen H, Zhou X and Wang Y F. 2024. WebUOT-1M: advancing deep underwater object tracking with a million-scale benchmark//Proceedings of the 38th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc
- Zhang C H, Liu L, Huang G J, Zhang Z P, Wen H, Zhou X, et al. 2025. Underwater camouflaged object tracking meets vision-language SAM2 [EB/OL]. [2025-09-01]. <https://arxiv.org/pdf/2409.16902.pdf>
- Zhang P, Tang J S, Zhong H P, Ning M Q, Liu D D and Wu K. 2022. Self-trained target detection of radar and sonar images using automatic deep learning. IEEE Transactions on Geoscience and Remote Sensing, 60: #4701914 [DOI: 10.1109/TGRS.2021.3096011]
- Zheng Y N, Chen W L, Lin R F, Zhao T S and Le Callet P. 2022. UIF: an objective quality assessment for underwater image enhancement. IEEE Transactions on Image Processing, 31: 5456-5468 [DOI: 10.1109/TIP.2022.3196815]
- Zheng Z Q, Liang H X, Hua B S, Wong Y H, Ang P, Chui A P Y, et al. 2024. CoralSCOP: segment any coral image on this planet//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 28170-28180 [DOI: 10.1109/CVPR52733.2024.02661]
- Zheng Z Q, Xie Y F, Liang H X, Yu Z B and Yeung S K. 2023. CoralVOS: dataset and benchmark for coral video segmentation [EB/OL]. [2025-09-01]. <https://arxiv.org/pdf/2310.01946.pdf>
- Zhou L, Liu Q M, Jin K J, Zhao W Y and Zhang W D. 2025. Research progress of underwater image restoration and enhancement methods. Journal of Image and Graphics, 30(1): 51-65 (周玲, 刘庆敏, 金凯杰, 赵文义, 张卫东. 2025. 水下图像复原和增强方法研究进展. 中国图象图形学报, 30(1): 51-65) [DOI: 10.11834/jig.240050]
- Zhuang P Q, Wang Y L and Qiao Y. 2018. WildFish: a large benchmark for fish recognition in the wild//Proceedings of the 26th ACM International Conference on Multimedia. Seoul, Korea (South): ACM: 1301-1309 [DOI: 10.1145/3240508.3240616]
- Zhuang P Q, Wang Y L and Qiao Y. 2021. Wildfish++: a comprehensive fish benchmark for multimedia research. IEEE Transactions on Multimedia, 23: 3603-3617 [DOI: 10.1109/TMM.2020.3028482]

作者简介

李华,女,副教授,主要研究方向为计算机视觉、水下图像处理、机器学习和深度学习。E-mail:lihua@hainanu.edu.cn

丛润民,通信作者,男,教授,主要研究方向为计算机视觉、水下环境感知和显著性检测。E-mail:rmcong@sdu.edu.cn

李志远,男,博士研究生,主要研究方向为计算机视觉和图像处理。E-mail:lizhiyuan@hainanu.edu.cn

刘家伟,男,硕士研究生,主要研究方向为图像处理和语义分割。E-mail:liujiawei@hainanu.edu