

中图法分类号: TP391.9 文献标识码: A 文章编号: 1006-8961(2026)05-1350-22

论文引用格式: Fang H, Yu Z J, Chen Z Y and Cong R M. 2026. Underwater image segmentation methods: a survey. Journal of Image and Graphics, 31(5): 1350-1371(方豪, 于宗吉, 陈志杨, 丛润民. 2026. 水下图像分割方法综述. 中国图象图形学报, 31(5): 1350-1371)[DOI: 10. 11834/jig. 250481]

水下图像分割方法综述

方豪, 于宗吉, 陈志杨, 丛润民*

1. 山东大学控制科学与工程学院, 济南 250061; 2. 机器智能与系统控制教育部重点实验室, 济南 250061

摘要: 水下图像分割作为海洋工程、水下机器人导航、海洋生物监测及水下资源勘探等领域的核心技术, 旨在从复杂且退化的水下图像中精准提取目标区域(如生物体、机器、海底地形等)。然而, 水下环境特有的光照衰减、水体散射、颜色失真及低对比度等问题, 为分割任务带来了严峻挑战。本文系统综述了水下图像分割领域的研究进展, 将现有方法按照任务目标划分为水下显著性目标检测、水下语义分割和水下实例分割3类。阐述了水下图像分割的研究背景、核心挑战及应用价值; 分别详细梳理了3类分割任务的技术演进, 从传统方法到基于深度学习的现代方法, 深入分析了各类方法的核心思想、网络架构与优势; 总结了主流的水下图像分割数据集及对应的评价指标; 通过基准实验结果对比, 直观展示了不同方法在典型数据集上的性能差异; 对该领域面临的挑战及未来的发展趋势进行总结与展望。本文旨在为相关领域的研究人员提供全面的技术参考, 推动水下图像分割技术的进一步发展与应用落地。

关键词: 水下图像分割; 显著性目标检测(SOD); 语义分割; 实例分割; 深度学习

Underwater image segmentation methods: a survey

Fang Hao, Yu Zongji, Chen Zhiyang, Cong Runmin*

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

2. Key Laboratory of Machine Intelligence and System Control, Ministry of Education, Jinan 250061, China

Abstract: Underwater image segmentation serves as a foundational technology for numerous marine-related fields, including ocean engineering, underwater robot navigation, marine biodiversity monitoring, underwater resource exploration, and underwater archaeology. Its primary objective is to accurately extract target regions (such as marine organisms, underwater equipment, and seabed terrain) from underwater images that are inherently degraded owing to the unique underwater environment. Unlike terrestrial image segmentation, underwater scene segmentation poses unprecedented challenges, including light attenuation, water scattering, color distortion (predominantly blue or green tones caused by differential wavelength attenuation), complex target-background interactions, varying water depth effects, and the scarcity of high-quality annotated datasets. These challenges have driven the evolution of underwater image segmentation methods from traditional handcrafted feature-based approaches to modern deep learning-driven paradigms, with task granularity advancing from coarse to fine grained. This review comprehensively summarizes the state-of-the-art progress in underwater image segmentation, categorizing existing methods into three core tasks: underwater salient object detection (USOD), underwater image semantic segmentation (UISS), and underwater image instance segmentation (UIIS). USOD is the most fundamental seg-

收稿日期: 2025-10-13; 修回日期: 2026-01-12; 预印本日期: 2026-01-19

* 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(62471278)

Supported by: National Natural Science Foundation of China(62471278)

mentation task, aiming to rapidly locate visually prominent targets (e. g. , fish and divers) from complex backgrounds and output binary foreground–background maps. Early USOD methods, including quaternion-based distance Weber descriptors and improved histogram equalization techniques, relied on handcrafted low-level visual features such as color, texture, contrast, and contour to model target-background differences. However, these methods lack robustness in low-contrast and color-distorted underwater scenes. With the advent of deep learning, convolutional neural networks (CNNs) have become the mainstream given their ability to automatically learn multilevel semantic features. Representative CNN-based models include SVAM-Net, which integrates bottom-up and top-down learning via a dual-branch architecture. To address the limitations of CNNs in capturing global dependencies, scholars have introduced transformer-based methods, such as TC-USOD with a hybrid transformer-convolution architecture and HEHP, which leverage heterogeneous experts and hierarchical perception to achieve state-of-the-art performance on datasets like USOD10K. Recent advancements have also seen the integration of vision foundation models (VFMs) such as the segment anything model (SAM) and diffusion models. Dual-SAM and MAS-SAM adapt SAM to underwater scenes by enhancing marine feature learning and refining fine-grained details, while DiffMSS and FSCDiff utilize conditional diffusion models to mitigate underwater image degradation and improve feature representation. UISS is a more refined task that assigns predefined semantic class labels (e. g. , fish, coral, seabed, and water column) to each pixel, enabling structured scene understanding. Traditional UISS methods include threshold-based segmentation, clustering algorithms (e. g. , k -means), and classifier-based approaches (e. g. , SVM), which are only applicable to simple scenes with distinct interclass differences. Deep learning has revolutionized UISS, with models based on FCN, U-Net, and DeepLab series dominating the field. Key improvements include module optimization (e. g. , UISS-Net with auxiliary feature extraction networks and SEA-Net with severity-aware dual branches), data augmentation techniques (e. g. , multispatial transformation and CutStitch), and loss function innovations (e. g. , combinations of cross-entropy loss, Dice loss, and boundary-aware loss). Transformer-based methods such as Swin Transformer and SegFormer have further enhanced performance by capturing long-range dependencies, with CoralSCOP (a SAM-based coral segmentation foundation model) achieving exceptional results on the large-scale CoralMask dataset. These advancements have significantly improved the ability of UISS models to handle complex underwater scenes with multiple categories and overlapping targets. UIIS is an advanced task that distinguishes individual instances within the same semantic class (e. g. , different fish in a school), requiring category classification and instance differentiation. Because of its high complexity, UIIS primarily relies on deep learning frameworks adapted from terrestrial instance segmentation. Early UIIS models were based on Mask R-CNN, such as WaterMask, which introduces attention modules and boundary refinement to address underwater image degradation. YOLO series models (e. g. , YOLOv9-N and AASNet) have also been applied for efficient instance segmentation. The integration of VFMs has marked a major breakthrough in UIIS: USIS-SAM adapts SAM with underwater domain prompts, MarineInst optimizes mask filtering to reduce oversegmentation, and UWSAM achieves efficient knowledge distillation from large SAM encoders to smaller models. Emerging architectures like Vision Mamba (e. g. , UIS-Mamba) and self-supervised pretrained models (e. g. , DINOv2-based DiveSeg) have further pushed the boundaries of UIIS performance, with DiveSeg achieving the highest mAP on datasets like UIIS and USIS10K. Meanwhile, various benchmark datasets and evaluation metrics have been proposed to support the development of underwater image segmentation methods. For USOD, datasets such as USOD10K (the largest and most widely used), UFO-120, and MAS3K cover diverse underwater scenes and target types, with evaluation metrics including S-measure (structural consistency), F-measure (balance of precision and recall), E-measure (robustness to degradation), and MAE (pixel-level error). UISS datasets like SUIM (the first large-scale semantic segmentation dataset), DeepFish, and CaveSeg provide annotated samples for different application scenarios, with core metrics including mean intersection over union (IoU), mean accuracy, and average accuracy. UIIS datasets such as UIIS10K and USIS16K (with 158 fine-grained categories) enable comprehensive evaluation of instance segmentation performance, using metrics like mAP, AP50 (loose IoU threshold), and AP75 (strict IoU threshold). Benchmark results show that hybrid architectures (combining CNNs/transformers/VFMs) and domain-specific optimizations (e. g. , underwater image enhancement modules and depth-aware fusion) are key to achieving superior performance across datasets. Despite significant progress, underwater image segmentation still faces several critical challenges. The scarcity of high-quality annotated datasets, especially for deep-sea and rare marine species, limits

the generalization of deep learning models. Additionally, the variability of underwater environments (e. g., light conditions and water turbidity) leads to poor cross-domain performance. Future research directions are focused on addressing the following challenges: 1) few-shot and weakly supervised segmentation to reduce reliance on large annotated datasets, leveraging metalearning and transfer learning with underwater physical priors; 2) open-vocabulary segmentation, integrating vision-language models (e. g., CLIP) to handle unseen target categories in real-world applications; 3) referential segmentation, enabling interactive target localization based on text descriptions or other reference information; 4) real-time multimodal segmentation, fusing RGB, depth, polarization, and hyperspectral data to meet the low-latency requirements of underwater robot navigation; and 5) cross-domain generalization, combining domain adaptation techniques and physical models to enhance robustness to environmental variations. In conclusion, underwater image segmentation has evolved significantly from traditional methods to advanced deep learning and foundation model-based approaches, with continuous improvements in accuracy, robustness, and applicability. The comprehensive overview of segmentation tasks, methods, datasets, and evaluation metrics provided in this review offers valuable insights for researchers in the field. Future advancements in low-annotation-cost, high-practicality, multimodal fusion, and cross-domain robust methods will further unlock the potential of underwater image segmentation, enabling efficient and intelligent exploration and utilization of marine resources.

Key words: underwater image segmentation; salient object detection(SOD); semantic segmentation; instance segmentation; deep learning

0 引言

海洋覆盖了地球表面约71%的面积,蕴藏着丰富的生物资源、矿产资源和能源,是人类探索自然的重要领域。水下图像作为获取海洋信息的主要载体之一,其分析与理解技术是实现海洋自动化、智能化开发利用的基础。水下图像分割作为水下图像处理与计算机视觉交叉领域的关键任务,其目标是根据像素的语义信息或实例属性,将水下图像划分为具有特定含义的连续区域,为后续的目标识别、跟踪、行为分析及决策提供精准的区域级支撑。

在实际应用中,水下图像分割技术展现出重要价值。在海洋生物监测(Xu等,2014)中,对珊瑚礁、鱼类、藻类等海洋生物进行分割与计数,助力生态环境评估与生物多样性研究;在水下机器人导航与避障(Lee等,2012)中,实时分割水下障碍物(如岩石、沉船、渔网),保障水下机器人的安全航行;在水下资源勘探(Bailey和Flemming,2008)中,分割海底矿产、油气管道及水下设备,辅助资源开采与工程维护;在水下考古与救援(Bailey和Flemming,2008)中,对水下遗址、失事船只残骸或失踪目标进行分割定位,为考古发掘与应急救援提供技术支持。

与陆地图像相比,水下图像的质量受环境因素影响极大,这些退化现象直接导致分割任务的难度

显著高于陆地场景,主要挑战包括:1)光照衰减与非均匀照明:光线在水下传播时,会因水体吸收而快速衰减,导致图像呈现“近亮远暗”的非均匀光照效果,远处目标区域细节丢失严重,难以与背景区分;2)水体散射效应:水中悬浮的颗粒物会对光线产生散射,造成图像模糊、对比度降低,使得目标边缘模糊,前景与背景的灰度/颜色差异被削弱;3)颜色失真:不同波长的光在水中衰减速率不同(红光衰减最快,蓝光衰减最慢),导致水下图像普遍呈现偏蓝或偏绿的色调,目标的真实颜色信息丢失,基于颜色特征的传统分割方法失效;4)目标形态与背景复杂性:水下目标(如鱼类群、珊瑚)形态多样、姿态多变,且常与复杂的海底地形(如沙质、岩石、水草)相互交织,导致目标与背景的边界模糊,增加了分割的歧义性;5)水域深度:浅水与深水图像特性差异显著,其中浅水光线充足,但易受水面反射、浮游生物散射影响,存在局部失真和纹理噪声;深水光线匮乏,依赖人工光源,强非均匀光照、散射及颜色失真更突出,目标边界模糊;6)数据稀缺性与标注困难:获取高质量的水下图像需要专业的设备和特定的环境条件,成本高昂;同时,像素级的精准标注需要领域专家参与,耗时耗力,导致公开的大规模高质量水下分割数据集相对匮乏。根据分割任务的输出粒度、目标定位精度及应用场景差异,现有水下图像分割方法可系统性划分为水下显著性目标检测、水下图像语义

分割与水下图像实例分割3大类,3类方法的技术演进呈现“从粗到精、从单一到多维”的递进趋势,经历了“传统手工特征驱动”到“深度学习数据驱动”的范式转变。

作为最基础的分割任务,水下显著性目标检测(underwater salient object detection, USOD)的核心目标是从复杂背景中快速定位“视觉显著”的目标区域(如人类、鱼类),输出二值化的前景—背景区分结果,为后续精细分割或目标识别提供粗定位支撑,如图1(b)所示。早期依赖手工设计的底层视觉特征建模目标与背景的差异,主流思路包括基于颜色、纹理、对比度和轮廓的分割方法。这些方法计算复杂度较低,但对水下低对比度、颜色失真场景鲁棒性差,难以应对复杂背景。随着深度学习在计算机视觉领域的爆发,基于深度神经网络的方法逐渐成为水下显著性目标检测的主流。以卷积神经网络(convolutional neural network, CNN)和Transformer为核心,通过端到端训练自动学习深层语义特征。例如,CNN通过多尺度特征融合提升目标细节捕捉能力,Transformer通过自注意力机制建模全局像素依赖关系。深度学习方法显著提升了对水下退化图像的适应能力。

相较于显著性目标检测(salient object detection, SOD)的“前景—背景二元划分”,水下图像语义分割(underwater image semantic segmentation, UISS)的目标更精细,其为每个像素分配预定义的语义类别标签(如鱼、珊瑚、人、水体等),实现对水下场景的结构化理解,是海洋生物监测、水下地形分析的核心技术,如图1(c)所示。早期方法基于阈值分割、聚类或分类器的方法,依赖手工特征实现类别区分。此类方法仅适用于类别差异显著的简单场景,无法处理多类别、目标形态复杂的水下环境(如不同鱼类与珊瑚的混淆)。深度学习阶段以全卷积网络(fully convolutional network, FCN)(Long等,2015)、U-Net(U-shaped network)(Ronneberger等,2015)、DeepLab系列(Chen等,2018a)等网络为基础,通过编码器—解码器结构、空洞卷积、多尺度特征融合、数据增强以及损失函数等技术,实现像素级语义标注。

水下图像实例分割(underwater image instance segmentation, UIIS)是当前领域的前沿任务,其目标是在语义分割基础上,进一步区分同一类别中的不同个体实例(如同一种鱼群中的不同个体),需同时实

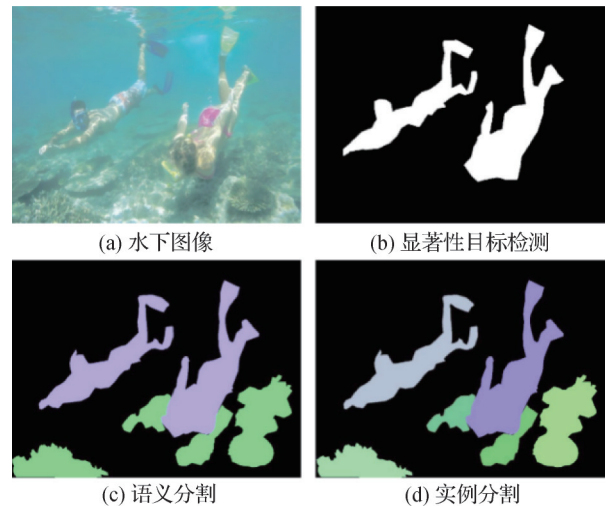


图1 水下图像分割任务示意图

Fig. 1 Schematic diagram for three tasks of underwater image segmentation ((a) underwater image; (b) salient object detection; (c) semantic segmentation; (d) instance segmentation)

现类别判断与实例区分,为水下生物行为分析、个体跟踪提供关键支撑,如图1(d)所示。由于任务复杂度高,其技术演进高度依赖深度学习的发展。主要方法为基于两阶段检测框架Mask R-CNN(mask region-based convolutional neural network)(He等,2017)的迁移应用,针对水下图像退化特性设计针对性模块,通过“候选框生成+掩码预测”实现实例分割,但整体性能依旧受限。为此,研究者将SAM(segment anything model)(Kirillov等,2023)等视觉基础大模型引入水下分割任务,借助强大的通用视觉表征和水下数据集微调,大幅提高了水下实例分割的性能。

近年来,水下视觉综述研究逐渐增多(周玲等,2025a;庄培显等,2026),但仍存在明显局限。Anand等人(2023)仅覆盖水下语义分割,未涉及显著性检测与实例分割;Jian等人(2021)以及González-Sabbagh和Robles-Kelly(2023)涵盖水下感知多任务,仅简要提及水下分割技术,且缺少近几年方法;Li等人(2025d)专注水下珊瑚分割,侧重单一目标类型,应用场景局限;Elmezain等人(2025)作为最新的相关综述,涵盖的水下分割数据集和方法不够全面,缺乏多基准实验对比。本文针对上述局限,从3方面进行补充与创新:1)任务覆盖更完整:系统涵盖显著性检测、语义分割和实例分割3大水下分割任务,尤其是最新的水下实例分割任务,填补该领域空白;

2)技术演进更清晰:呈现从传统手工特征方法到深度学习的技术递进关系,包含最新的基于视觉基础模型的方法,保证技术前沿性;3)数据支撑更全面:构建数据集与评价指标的对比框架,覆盖14个主流数据集与9类核心指标,确保最新基准实验结果。

本文全面系统地阐述水下显著性目标检测、水下图像语义分割以及水下实例分割的研究进展,以便人们能够更好地掌握水下图像分割领域的研究现状;总结当前主流的水下图像分割数据集及对应的评价指标;呈现不同方法在基准数据集上的性能对比结果;最后,对本文进行了总结,并对未来的发展进行了展望。

1 水下图像分割方法

1.1 水下显著性目标检测

水下显著性目标检测旨在模拟人类视觉系统,自动定位并提取水下图像中最“显著”的区域(通常是用户感兴趣的目标,如鱼类、珊瑚等),输出二值化的显著性图。其核心是建模目标与背景在视觉特征上的差异,适用于目标快速定位与粗提取场景,代表性算法的可视化结果如图2所示。

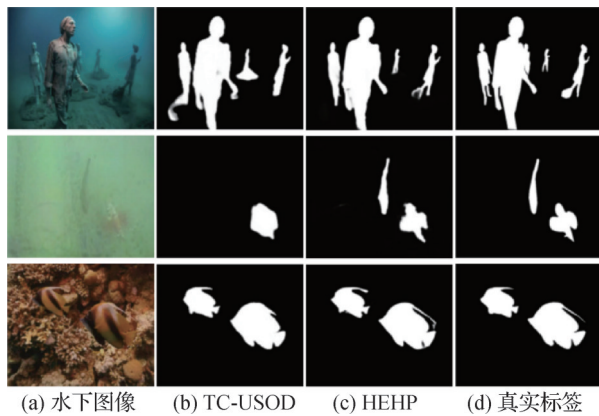


图2 水下显著性目标检测代表性算法的可视化结果
Fig. 2 Visualization results of representative methods for underwater salient object detection ((a) underwater images; (b) TC-USOD; (c) HEHP; (d) ground truths)

1.1.1 基于手工特征的方法

传统显著性目标检测方法主要依赖手工设计的底层视觉特征(如颜色、纹理、对比度和轮廓)来计算像素的显著性。受经典显著性目标检测方法(Itti等,1998)启发,Edgington等人(2003)以及Barat和

Phlypo(2010)通过图像区域的局部特征对比来推断水下场景的视觉显著性。Jian等人(2018)利用基于四元数的距离Weber描述符表征水下图像中的方向线索,采用主成分分析(principal component analysis, PCA)坐标系计算模式差异性,利用局部对比度进一步突出水下的显著区域并抑制背景区域。针对传统直方图均衡化算法存在的问题,Cui等人(2019)提出了一种改进的直方图均衡化方法以提升图像质量,该方法能使显著性区域更平滑和清晰。Chen等人(2020)通过结合手工制作的2D(颜色和强度)和3D(深度图)视觉特征,提出了一种数学模拟水生动物生物视觉机制构建的检测模型。上述方法在高对比度的简单水下场景有一定效果,但面对低对比度、复杂背景时鲁棒性差。

1.1.2 基于CNN的方法

随着算力的持续提升与数据资源的不断丰富,深度学习方法逐步取代传统基于手工设计特征的方法,显著提升了水下显著性检测的任务性能与鲁棒性。CNN因其自动提取多层级语义特征、对水下复杂噪声与光照变化的强适应性,以及高效的局部特征捕捉能力,成为最为主流的水下显著性检测模型架构。Islam等人(2022)提出SVAM-Net(saliency-guided visual attention modeling net),以共享CNN编码层的双分支架构整合自下而上与自上而下学习,设计专用空间注意力模块和残差细化模块进一步增强水下表征。为了追求轻量级,EffiSeaNet(efficient sea network)(Wu等,2025b)通过集成无参数图像增强模块和跨层特征融合模块,实现低参数量和计算复杂度与高检测精度的平衡,适用于资源有限的实际水下场景。Auto-USOD(auto underwater salient object detection)(Liu等,2025b)通过拓扑搜索方案使网络自适应地协调特征提取、特征选择与特征融合之间的配合,从而让网络更好地感知并恢复显著目标特征。

水下场景的专有特性也被引入网络设计中,Peng等人(2024)基于水下场景中“物体距离越远越模糊”的固有特性,提出了融合自衍生模糊线索与输入图像的模糊引导SOD方法,并通过模糊图调整图像以扩大高聚焦与低聚焦区域差异,进而提升模型泛化能力。CSUNet(contour-sensitive underwater salient object detection network)(Wei等,2024)是一种轮廓敏感型水下显著性目标检测网络,通过轮廓

细化模块生成粗轮廓和细轮廓,助力网络有效捕捉并增强显著区域的轮廓。CE³USOD(channel-enhanced, efficient, and effective network for underwater salient object detection)(Wu等,2025a)从邻域视角重新审视长距离依赖关系与特征计算过程,进而进行长距离上下文感知,并利用经水下图像增强技术优化后的较弱颜色通道作为引导,实现多尺度特征融合。Ge和Ouyang(2025)提出双迭代互补增强网络(dual iterative complement enhancement net, DINet),通过迭代交替优化全局与细节特征,并基于傅里叶变换分离振幅分量以缓解水下光传播损失。

1.1.3 基于Transformer的方法

为弥补CNN在全局特征建模与长距离依赖捕捉上的局限,Transformer凭借自注意力机制的核心优势,成为水下显著性检测的重要研究方向。其对水下场景中目标与背景全局关联性的精准刻画,以及多尺度特征的高效融合能力,有效提升了复杂水下环境下的检测精度与边缘细节表现力。此外,随着首个大规模水下显著性目标检测数据集USOD10K(Hong等,2025a)的提出,因其包含深度图信息,研究者同步转向基于RGB-D(RGB-depth)的水下显著性目标检测方法探究。

为了充分利用各种架构的独特优势,Transformer与不同网络结合的混合架构成为首选。Hong等人(2025a)提出的TC-USOD(Transformer-convolution based underwater salient object detection)模型采用Transformer编码器(捕捉水下图像全局依赖)与卷积解码器(捕捉局部细节)的混合架构,结合含跨模态融合的深度图辅助模块及多任务学习策略,为RGBD-USOD任务提供了性能优越的强基线。IF-USOD(interactive feature for underwater salient object detection)(Yuan等,2025)通过融合RGB与深度信息,结合CNN和Transformer构建跨尺度交互嵌入模块提取特征,设计全感知交叉注意力模块增强跨模态互补性。Tang等人(2025b)提出一种无注意力混合网络AFHNet(attention-free hybrid network),整合了无注意力机制和基于Transformer的双编码器,能够在保留关键信息的同时,有效融合RGB特征与深度特征。Ma等人(2025)提出了一种协同Transformer与Mamba的融合网络STAMF(synergistic transformer and Mamba fusion network),通过融合能够表征目标固有物理特性的偏振信息,增强复杂场

景目标与背景之间的区分度。

进一步建模水下数据本身的特性,依旧是基于Transformer的方法提升性能的关键。SODE(salient object depth emphasis)(Jin等,2025)是一种基于迭代双阶段自步学习和显著目标深度强调的USOD框架,通过先聚焦简单样本再以样本难度分数为权重自适应学习所有样本,模拟人类学习过程以应对不同难度水下样本。HDANet(hydro depth aware network)(Liu等,2025c)是一种水深感知网络,集成了一个任务驱动的水下图像增强模块,通过物理模型为USOD任务提供经过优化的增强图像与多模态信息,并利用再增强与再退化的互补效应,提升在多种水下图像退化场景中的泛化能力。Zha等人(2025)分别利用RGB分支与深度分支学习解纠缠表示,构建了异构专家与层级感知网络(heterogeneous experts and hierarchical perception network, HEHP),频率专家混合模块专注于高频与低频建模,并提出设计四路径融合专家模块高效整合多样化的空间与频率信息,以学习跨模态、跨图像的紧致且通用的表示。

1.1.4 基于视觉基础模型的方法

依托大规模数据预训练与通用特征提取能力,视觉基础模型打破了传统模型对特定场景数据的依赖,凭借卓越的跨域泛化性能和复杂环境适配能力,成为水下显著性检测的前沿研究方向。其中,SAM作为典型的通用分割基础模型,以“提示—分割”框架为核心,通过学习图像中目标的通用形态与边界特征,实现对任意类别、任意形态目标的零样本或少量样本分割,其强大的目标定位与边缘捕捉能力为水下复杂场景的显著性检测提供了坚实基础。Zhang等人(2024)提出Dual-SAM(dual-segment anything model)框架,通过双SAM编码器增强海洋特征学习,多尺度耦合提示注入水下先验,以及交叉连通性预测捕捉像素间关联,解决SAM在海洋场景先验不足、单提示引导有限的问题。针对SAM解码器的结构简单性可能导致目标细粒度细节丢失的问题,Yan等人(2024)提出一种用于海洋生物分割的模型MAS-SAM(marine animal segmentation-segment anything model),引入超图提取模块和渐进式预测解码器从全局上下文信息到细粒度局部细节中提取更丰富的海洋生物信息。

Diffusion模型作为基于扩散过程的生成式基础模型,通过模拟从噪声到清晰图像的逐步去噪过程,

具备优异的图像修复、细节增强与条件生成能力,其对数据分布的深层建模特性使其能有效应对水下图像的退化问题,为显著性特征的鲁棒学习提供了新范式。Chang 等人(2025)提出了一种基于条件扩散模型的水下显著性目标检测模型 DiffMSS(diffusion model for marine saliency segmenter),该模型通过语义知识蒸馏引导海洋显著目标的检测过程。Li 等人(2025c)提出了一种基于傅里叶空间纠缠条件扩散模型 FSCDiff,与那些依赖像素级概率的现有空间域感知型 RGB-D USOD 方法不同,FSCDiff 通过利用傅里叶域信息以及扩散模型强大的多步迭代生成能力,有效解决了水下显著物体检测中存在的特征表示不足及边界模糊等问题。

1.2 水下图像语义分割

水下图像语义分割是比显著性目标检测更精细的任务,其目标是为图像中的每个像素分配一个预定义的语义类别标签(如“鱼”、“珊瑚”、“岩石”、“水体”等),实现对水下场景的全面结构化理解。该任务需要模型不仅能区分前景与背景,还能识别不同类别的目标,是水下场景理解的核心技术。代表性算法的可视化结果如图 3 所示。

1.2.1 基于手工特征的方法

传统语义分割方法同样依赖手工特征与传统机器学习模型。首先是基于阈值分割的方法,根据图像中不同类别像素的灰度或颜色分布差异,设定阈值将图像划分为不同区域。例如,针对偏蓝的水下图像,通过蓝色通道的灰度阈值区分水体与海底;或使用 Otsu 算法自动计算最优阈值。此类方法仅适用于类别间灰度差异显著的简单场景(如沙质海底与水体),无法处理多类别、低对比度的复杂水下图像。其次是基于聚类的方法(Li 等,2016),将像素视为数据点,通过聚类算法(如 K-Means、Mean-Shift)将特征相似的像素聚为一类,实现类别划分。例如,利用 K-Means 对水下图像的 HSV 颜色空间特征进行聚类,区分珊瑚(多为彩色)与岩石(多为灰度)。但聚类结果缺乏语义信息,需要人工映射到实际类别,且对噪声和光照变化敏感。最后是基于分类器的方法(Rajasekar 等,2015),提取每个聚类的像素块的手工特征,使用传统分类器(如支持向量机、随机森林)对像素块进行类别标注,实现语义分割。此类方法将像素级分类转化为区域级分类,降低了计算复杂度,但聚类分割的精度直接影响最终结果,且

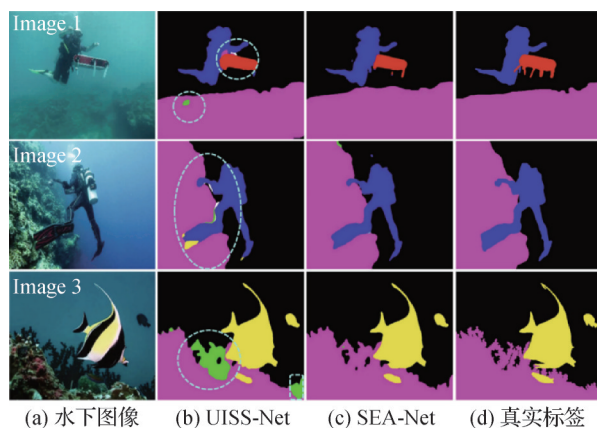


图3 水下语义分割代表性算法的可视化结果

Fig. 3 Visualization results of representative methods for underwater semantic segmentation ((a) underwater images; (b) UISS-Net; (c) SEA-Net; (d) ground truth)

手工特征的表达能力有限,难以区分相似类别。

1.2.2 基于CNN的方法

与水下显著性目标检测一样,深度学习方法也大幅提高了水下语义分割性能,其中,对基于CNN的经典语义分割网络进行模块改进成为主流。O'Byrne 等人(2018)提出利用合成水下图像训练 SegNet(segment network)模型,再通过基于迭代支持向量机的后处理算法优化初始分割结果,实现对真实水下图像的语义分割。Liu 和 Fang(2020)在 DeepLabV3+(Chen 等,2018a)的编码器结构中引入无监督颜色校正方法,以提升图像质量,在解码器结构中增加两个上采样层,从而保留更多目标特征与物体边界信息。Kim 和 Park(2022)提出并行语义分割网络(parallel semantic segmentation network, PSS-Net),以 U-Net 为骨干、VGG-16(Visual Geometry Group)为编码器,通过并行设计前景与背景处理模型,并对两模型提取的特征图进行基于注意力的融合以完成最终分割。苑永起等人(2024)提出一种利用三支注意力模块改进 U-Net 结构的水下语义分割方法,利用注意力机制实现跨通道交互信息,在实现多维交互的同时不降低维度。He 等人(2024)提出水下图像语义分割网络(underwater image semantic segmentation network, UISS-Net),通过引入辅助特征提取网络改善水下场景特征提取难题,设计反向多尺度特征融合解决网络高低层语义信息差异问题。Liu 等人(2025a)提出 DeepLab-FusionNet 模型,以 DeepLabV3+为基础,通过设计基于 HRNet(high

resolution network) (Wang等, 2021)的多尺度特征融合模块增强多尺寸目标分割能力。Zhou和Li (2025b)提出基于HRNetV2的水下高分辨率语义分割网络UHRS-Net,通过集成小波动态注意力模块实现频率域特征分解,多尺度补丁注意力分割头提升小目标分割与边界精度。

除了网络模块设计改进,研究者也在数据增强和损失函数上进行创新。Islam等人(2020a)提出了首个专为水下图像语义分割构建的大规模数据集SUIM (semantic segmentation of underwater imagery)和一种全卷积编码器—解码器模型SUIM-Net。Wang等人(2022)先通过多空间变换进行图像增强以改善原始水下图像质量,再通过基于编码器—解码器架构进行水下图像语义分割。马志伟等人(2022)提出了一个关注边界的半监督水下语义分割网络US-Net (underwater segmentation network),通过设计伪标签生成器和边界检测子网络,实现了对水下物体与背景之间边界的精细学习。Chicchon等人(2023)提出结合交叉熵损失、Dice损失与基于Chan-Vese函数的主动轮廓损失的组合损失函数,用于训练改进的U-Net和DeepLabV3+模型。王金康等人(2023)先通过加权融合的多空间转换增强水下图像,再借助级联空洞卷积空间金字塔池化模块丰富目标细节。郭浩然等人(2023)采用轻量高效的非对称编解码器结构,结合由OHEMCELoss (online hard example mining cross-entropy loss)和BCELoss (binary cross-entropy loss)组成的总损失函数以优化边缘分割。Nunes和Matos(2023)在SUIM数据集上结合图像增强技术(对比有限自适应均衡化(contrast limited adaptive histogram equalization, CLAHE)等)与数据增强策略(针对代表性不足类别差异化扩充样本等)优化SegNet模型。Zhu和Zhang(2025)提出SEA-Net (severity-aware network),通过引入严重程度指标将水下图像分为高/低严重程度,双分支交替训练以增强不同严重程度图像特征学习。

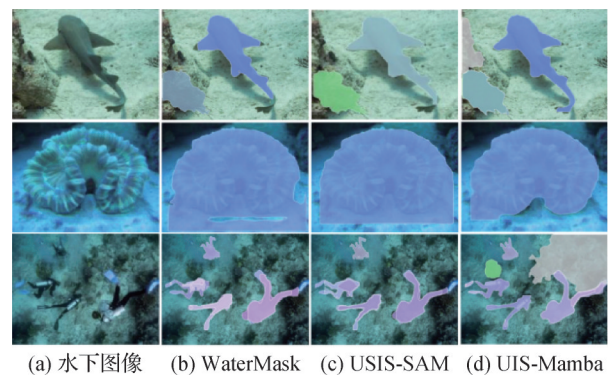
1.2.3 基于Transformer的方法

在基于Transformer的水下语义分割方法中,Abdullah等人(2024)采用Swin Transformer(Liu等, 2021)骨干网络、多尺度金字塔池化头与分层特征聚合模块,相比现有SOTA(state of the art)模型,内存效率和推理速度大幅提高且性能相当。Chen等人(2025a)对标准SegFormer(Xie等, 2021)模型进行

改进,通过将其Mix Transformer骨干网络替换为Swin Transformer以增强特征提取能力,并在骨干下采样阶段和解码器中引入高效多尺度注意力机制以更好捕捉多尺度特征。Zuo等人(2025)提出UWSegFormer (underwater segment Transformer)模型,通过设计水下图像质量注意力模块增强高质量语义信息通道表征,提出多尺度聚合注意力模块高效融合多尺度特征,以及补偿细节语义损失和边缘学习损失提升目标边缘学习能力。Li等人(2025a)提出UISFormer (underwater image segmentation Transformer),通过CutStitch数据增强技术以丰富有限数据集,在编码器中构建残差特征增强模块实现局部与全局特征有效融合,在编码器与解码器间设计多尺度特征残差融合模块解决语义鸿沟。CoralSCOP (segment any coral image on this planet)是首个珊瑚分割基础模型,以SAM为骨干并在解码器处设计新颖并行语义分支,结合“负样本训练”策略解决语义无关与过分割问题,依托大规模CoralMask数据集预训练后具备强零样本泛化能力,为珊瑚礁监测与保护提供高效工具。

1.3 水下图像实例分割

水下图像实例分割是当前水下图像分割领域的前沿任务,其目标是在语义分割的基础上,进一步区分同一语义类别中的不同个体实例。由于实例分割任务的复杂性,其研究主要基于深度学习方法,且大多借鉴陆地实例分割的成功框架,并针对水下场景进行适配优化。代表性算法的可视化结果如图4所示。



(a) 水下图像 (b) WaterMask (c) USIS-SAM (d) UIS-Mamba

图4 水下实例分割代表性算法的可视化结果

Fig. 4 Visualization results of representative methods for underwater instance segmentation ((a) underwater images; (b) WaterMask; (c) USIS-SAM; (d) UIS-Mamba)

1.3.1 基于CNN的方法

Mask R-CNN(He等,2017)是基于Faster R-CNN(faster region-based convolutional neural network)(Ren等,2017)的经典实例分割框架,通过在Faster R-CNN的基础上增加一个掩码预测分支,实现对每个候选框的像素级掩码生成。Lian等人(2023)提出了首个水下图像实例分割模型WaterMask,相比通用场景的Mask R-CNN,一是提出了差异相似图注意力模块,通过计算图像块间注意力聚合相似视觉信息,恢复水下图像因质量退化和下采样丢失的细节;二是设计多级特征细化模块融合特征金字塔粗粒度特征与重建的细粒度特征,分别预测前景掩码与边界掩码;三是引入边界掩码策略及边界学习损失,优化水下聚类实例的边界预测。Chen等人(2024)改进Mask R-CNN的特征金字塔网络(feature pyramid network, FPN),并通过生成对抗网络(generative adversarial network, GAN)扩充水下图像数据集,设计结合色彩偏移去除、去模糊与对比度增强的组合优化算法预处理图像以提升质量。Yi等人(2024)提出CAM-RCNN(coordinate-aware mask R-CNN)模型,通过融合坐标卷积(CoordConv)与分组归一化(group normalization, GN)提升泛化性与稳定性。Pan和Pei(2025)在Mask-RCNN的基础上提出BARIS-ERA框架,包含面向实例分割的边界感知细化解码器(boundary-aware refinement decoder for instance segmentation, BARIS)替换原有的FPN网络以进一步优化多尺度特征融合,环境鲁棒适配器(environmental robust adapter, ERA)用于高效微调参数。

除了Mask R-CNN模型,YOLO(you only look once)系列模型也因高效率广泛应用于实例分割任务。Tang等人(2025a)提出基于YOLOv9(Wang等,2025)的水下图像实例分割算法YOLOv9-N,通过将设计的非局部模块(non-local-s)集成到YOLOv9网络p3层后,结合余弦相似度与欧氏距离以增强模型捕捉图像长距离依赖、处理遮挡问题及保留小目标信息的能力。He等人(2025)提出通道空间交叉协同注意力机制与通道先验注意力融合模块,有效捕捉局部和全局信息并去除冗余信息,嵌入YOLOv8(Varghese和Sambath,2024)网络的性能提升证明了有效性。Kong等人(2025)提出基于YOLOv9的水下实例分割模型AASNet(agricultural aqua segmentation network),通过引入基于皮尔逊相关系数捕捉特

征间线性相关性的线性相关注意力机制,解决水下场景光照变化与颜色差异导致的特征不一致问题,同时设计动态自适应焦点损失应对极端数据不平衡以优化分类性能。

1.3.2 基于视觉基础模型的方法

为了进一步探索视觉基础模型的潜力,研究者也将SAM(Kirillov等,2023)、VMamba(visual state space model)(Liu等,2024)、DINO(distillation with no labels)(Oquab等,2023)等视觉基础模型引入水下实例分割任务。Lian等人(2024)首次将SAM应用于水下显著实例分割任务,提出了USIS-SAM(underwater salient instance segmentation-segment anything model)模型,通过设计水下自适应视觉Transformer编码器,借助适配器融入水下领域视觉提示,提升模型对复杂水下场景的适应性;设计显著特征提示生成器SFPG(salient feature prompt generator),自动生成显著提示以实现端到端分割,无需人工提供前景点或框。MarineInst(marine instance)(Zheng等,2025)在SAM的掩码解码器中加入二元实例过滤,通过构建正负掩码样本训练轻量化二分类器,筛除无实例语义的低质量掩码,并在自己提出的MarineInst20M数据集上持续预训练,有效解决了现有模型在海洋图像上的过分割和部分分割问题,且支持自动与交互式两种推理模式,为后续实例描述及下游海洋视觉分析任务奠定了高质量掩码基础。为了解决SAM较高计算量的问题,Li等人(2025b)提出UWSAM(underwater segment anything model),通过基于掩码图注意力网络的水下知识蒸馏方法,将SAM的ViT-Huge图像编码器中的知识高效蒸馏到更小的ViT-Small图像编码器中,以实现有效的视觉表征学习;此外,也设计了端到端水下提示生成器自动生成水下提示,从而使网络能够自动定位水下实例,实现高效分割。

Vision Mamba作为基于状态空间模型(state-space model, SSM)的新兴视觉架构,以其线性序列建模复杂度和动态路由机制为核心优势,在长距离特征依赖捕捉与计算效率间实现了高效平衡,为视觉实例分割等任务提供了超越Transformer的新范式。Cong等人(2025)提出首个基于Mamba的水下实例分割模型UIS-Mamba(underwater image segmentation Mamba),通过允许补丁动态偏移和缩放以维持实例内部特征连续性,并基于Ncut的隐藏状态弱

化机制抑制复杂背景干扰和聚焦实例状态传播信息流,同时保持较低的数量和计算复杂度。DINOv2作为自监督预训练视觉Transformer的典型代表,凭借无标注数据学习的通用视觉表征和强大的跨域迁移能力,无需额外微调即可在多种下游视觉任务中展现优异性能。Chen等人(2025b)提出首个基于DINOv2的水下实例分割模型DiveSeg(DINO vision enhanced segmenter),将水下色彩风格特征嵌入到DINO的微调过程中,并通过二进制分割提示引入对象级先验,为既需对象又需实例级推理的实例分割任务提供必要的指导,实现了最先进的水下实例分割性能。

1.4 水下图像分割方法关联性分析

尽管水下显著性目标检测、语义分割与实例分割在任务粒度上存在差异,但3类方法在技术演进路径、网络架构设计与优化策略上呈现出清晰的内在关联。从任务粒度看,三者构成从粗到细的完整分割谱系:显著性目标检测聚焦二值化前景提取,为后续任务提供区域提议;语义分割在像素级别引入类别信息;实例分割则进一步区分同类目标的不同个体,三者形成递进式技术栈。从架构演进看,3类方法均经历了从手工特征到CNN、Transformer,再到视觉基础模型的技术跃迁,且创新模块常跨任务迁移复用,如注意力机制、多尺度特征融合与跨模态交互设计在3类任务中均有适配应用。从优化策略看,针对水下特有挑战(光照衰减、颜色失真等),3类方法共享数据增强策略(如水下模拟退化)、损失函数设计(边界感知损失)及后处理技术(轮廓细化)。值得注意的是,随着视觉基础模型(如SAM)的发展,3类任务呈现融合趋势:统一架构通过提示工程适配不同粒度的分割需求,例如DiveSeg框架通过引入对象级先验,使同一模型可灵活支持显著性检测到实例分割的多级任务。这种技术协同与架构统一,正推动水下图像分割向“一模型多任务”的范式演进,为复杂水下视觉理解提供系统化解决方案。

2 数据集和评价指标

2.1 水下图像分割数据集

2.1.1 水下显著性目标检测数据集

UFO-120(underwater fine-grained objects-120)数

据集是Islam等人(2020b)提出的支持“图像增强、超分辨率、显著性预测”多任务的综合性水下图像数据集,共1620幅图像,1500幅用于训练,120幅用于测试。该数据集聚焦“低分辨率+颜色失真”特性,覆盖不同光学特性的水体,可训练模型应对复杂水下光学条件。图像来自不同地点的海洋探索任务,如珊瑚礁、沉船、生物栖息地等。

RUWI(robust marine animal segmentation)数据集(Drews-Jr等,2021)是用于验证算法泛化性的中小型水下数据集,共700幅图像,525幅用于训练,175幅用于测试。目标类别以海洋动物(如章鱼、海星、鱼类)为主。该数据集突出“动态水下场景”特性,图像包含动物运动模糊、水体浑浊度变化,有效评估算法对动态目标的检测能力。

MAS3K(marine animal segmentation 3K)数据集是Li等人(2020)提出的用于精细海洋动物显著性检测的数据集,共3103幅图像,1769幅图像用于训练,1141幅图像用于测试。目标以“形态复杂的海洋动物”为主(如章鱼触手、海星棘刺),包含不同水体透明度(从清澈到浑浊)。

USOD(underwater salient object detection)数据集(Islam等,2022)包含300幅测试图像,整合了USR-248(underwater super resolution-248)、UIEB(underwater image enhancement benchmark)、EUVP(enhancing underwater visual perception)3个水下数据集中未用于训练的子集和实地海洋采集图像,旨在覆盖多样的水下场景与失真类型。图像涵盖鱼类、珊瑚和水下设备等目标类别,水体类型包括淡水与海水,且包含光照衰减、散射和颜色失真等典型水下光学干扰,能有效评估算法在复杂背景下的检测能力。

RMAS(robust marine animal segmentation)数据集由Fu等人(2024)提出,用于海洋动物特异性显著性检测的数据集,共3014幅图像,2514幅用于训练,500幅用于测试。目标类别限定为常见海洋生物(如珊瑚、海胆、扇贝),图像涵盖不同光照条件(表层强光、深层弱光),背景以水底沙质、岩石为主,减少复杂干扰,便于评估算法对特定目标的检测精度。

USOD10K(underwater salient object detection 10K)数据集是Hong等人(2025a)提出的首个大规模水下显著性检测基准数据集,也是当前该领域应用最广泛的数据集。该数据集共10255幅图像,7178幅

用于训练, 2 051 幅用于验证, 1 026 幅用于测试。涵盖 12 类水下场景(如珊瑚礁、深海、极地冰下、港口)与 70 类显著性目标(如鱼类、水下机器人、沉船残骸、潜水员等)。图像包含光照衰减、散射、运动模糊以及颜色偏色等全面的水下失真类型, 并提供所有图像的像素级标注、深度图及目标边界信息。

2.1.2 水下语义分割数据集

SUIM(semantic segmentation of underwater imagery)数据集(Islam 等, 2020a)是专为水下图像语义分割构建的首个大规模数据集, 数据来源于海洋探测、人机协作实验及现有数据集, 能够反映真实且多样的水下环境状况。该数据集共包含 1 635 幅图像, 1 525 幅用于训练和验证, 110 幅作为测试集, 标注类别细致且贴合水下场景实际需求, 覆盖水体背景、人类潜水员、水生植物、沉船/废墟、机器人/仪器、珊瑚礁/无脊椎动物、鱼类/脊椎动物和海床/岩石共 8 类, 为水下语义分割模型的训练、验证与评估提供了丰富且多样化的样本支撑, 成为该领域最重要的基准数据集。

DeepFish(deep learning fish)数据集(Saleh 等, 2020)是面向水下鱼类语义分割任务构建的数据集, 数据来源于 20 个澳大利亚热带海洋环境栖息地, 具有鲜明的地域生态特征。该数据集初始包含约 40 000 幅仅含分类标签的水下图像, 后补充标注了 300 幅具备语义分割标签的样本图像, 标注类别分为背景和前景(鱼类)两类, 聚焦于鱼类目标的分割任务。在实际使用中, 常按照 9:1 的比例将其划分为训练集与测试集, 为水下鱼类语义分割模型的性能验证与优化提供了针对性的数据支持, 尤其适用于鱼类检测、计数及生态监测相关的算法研发。

CaveSeg(cave segmentation)数据集(Abdullah 等, 2024)是首个针对水下洞穴场景语义分割构建的大规模数据集, 数据通过遥控潜水器(remotely operated vehicle, ROV)和潜水员在多个不同地理区域的洞穴系统中采集获取, 包含针对水下洞穴自主导航关键的 13 个目标类别。该数据集共包含 3 350 幅像素级标注样本图像, 分为 2 848 幅训练集、168 幅验证集和 334 幅测试集。该数据集为水下洞穴自主探索提供了关键数据支撑, 助力推动水下洞穴自主导航技术发展。

CoralMask(coral mask dataset)数据集(Zheng 等, 2024)是当前规模最大的珊瑚礁数据集, 数据来

源于互联网、公开数据集、开源贡献及珊瑚生物学家提供的水下勘测图像, 涵盖全球多个珊瑚礁站点, 具有图像多样性高、标注密集精细的特点, 共包含 41 297 幅珊瑚礁图像及 330 144 个珊瑚掩码, 支持从珊瑚/非珊瑚、6 类生长形态到 14 类属级等多粒度语义类别标注, 可满足珊瑚密集分割及多下游珊瑚分析任务需求。

2.1.3 水下实例分割数据集

UIIS(underwater image instance segmentation)数据集是 Lian 等人(2023)提出的首个水下实例分割基准数据集, 其整合了多个水下图像来源, 筛选出 4 628 幅高质量水下图像, 其中, 训练集 3 937 幅、验证集 691 幅。图像涵盖鱼类、珊瑚礁、水生植物、遗迹残骸、水下机器人、人类潜水员、海底礁岩 7 个核心类别, 覆盖浅海、深海等场景, 包含光照不均、水体散射等典型水下失真, 采用 COCO(common objects in context)格式存储像素级实例分割掩码和边界框标注, 能有效评估算法在复杂水下场景下的实例分割能力。

USIS10K(underwater salient instance segmentation 10K)数据集是 Lian 等人(2024)提出的首个水下显著实例分割数据集。数据来源于互联网及开源水下数据集, 筛选含“视觉显著实例”的图像, 标注有像素级显著实例掩码、类别标签和边界框, 支持“类别无关”(仅分割显著实例)和“类别相关”(同时预测类别)两种任务。包含 10 632 幅图像, 划分为训练集 7 442 幅、验证集 1 595 幅和测试集 1 595 幅。涵盖鱼类、珊瑚礁、水生植物、水下废墟、人类潜水员、机器人和海底礁岩 7 个类别。聚焦“低对比度 + 小目标”场景, 50% 图像显著目标占比 < 10%。

UIIS10K 数据集是 Li 等人(2025b)构建的大规模水下实例分割数据集, 是 UIIS 数据集的扩展。该数据集共 10 048 幅图像, 训练集 8 083 幅, 验证集 2 010 幅, 41 862 个像素级实例掩码, 较 UIIS 新增 5 420 幅图像及 13 452 个实例, 在 UIIS 基础上新增爬行动物(海龟、海蛇)、软体动物(章鱼、海星)和垃圾(生活垃圾、工业废弃物)3 个类别, 是目前规模最大的水下实例分割数据集。

USIS16K 数据集是 Hong 等人(2025b)构建的类别丰富的水下显著实例分割数据集。该数据集共 16 151 幅高分辨率水下图像, 分为训练集 11 306 幅、验证集 3 230 幅和测试集 1 615 幅, 包含 158 个细分

类别, 归为 11 个超类别(人类参与者、鱼类、贝类、海洋动物、水下垃圾、水下交通工具、遗失物品、水下设施、水下遗迹、水下残骸及其他)。数据来源于互联网搜索、USOD10K 数据集以及北海原位拍摄(水下工程

机器人采集), 标注有注视点(眼动追踪数据)、边界框、像素级显著掩码和实例级掩码, 分辨率范围广。

各水下图像分割数据集简介及基本情况如表 1 所示。

表 1 水下图像分割数据集
Table 1 Underwater image segmentation datasets

名称	分割任务	发布年份	图像数量	类别数量	数据集简介
UFO-120		2020	1 620	2	早期水下显著性检测基准, 强调颜色退化与对比度损失场景。
MAS3K		2021	3 103	2	首个多类海洋动物的大规模显著性数据集, 强调形态多样与弱边界分割。
RUWI	水下显著性目标检测	2021	700	2	首个真实水下机器人视角采集数据集, 突出真实噪声影响。
USOD		2022	300	2	经典小规模显著性数据集, 突出低光、浑浊等极端环境。
RMAS		2023	3 014	2	混合海洋生物与人工遗迹场景, 强调复杂背景下的鲁棒显著性检测。
USOD10K		2023	10 255	2	首个大规模水下显著性检测数据集, 涵盖多种目标, 使用最广泛。
SUIM		2020	1 635	8	早期多语义水下分割基准, 专注多语义共存的水下场景解析。
DeepFish	水下语义分割	2020	300	2	首个专注鱼类分割的数据集, 强调弱纹理与伪装生物的认识难度。
CaveSeg		2024	3 350	13	首个水下洞穴语义分割数据集, 突出低光、狭窄结构场景。
CoralMask		2024	41 297	14	首个大规模珊瑚生态语义分割数据集, 强调细粒度结构和复杂背景。
UIIS		2023	4 628	7	首个水下实例分割数据集, 涵盖典型生物实例, 侧重大量目标场景。
USIS10K	水下实例分割	2024	10 632	7	首个大规模显著实例分割基准, 支持类别无关与多类实例检测与分割。
UIIS10K		2025	10 048	10	UIIS 数据集的扩展版本, 图像数量和标注掩码进一步增加。
USIS16K		2025	16 151	158	最大规模显著实例分割基准, 水下目标类别丰富, 强调跨类别泛化能力。

2.2 评价指标

2.2.1 显著性目标检测评价指标

1) 结构一致性测度(S-measure, S_m)。由 Fan 等人(2017)提出, 是结合“目标级结构相似性”与“区域级结构相似性”、衡量显著性图与人类视觉感知一致性的指标。具体定义为

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

式中, α 为权重系数; S_o 为目标级相似度, 衡量前景区域的结构一致性; S_r 为区域级相似度, 基于结构相似性(SSIM)衡量局部图像块的结构匹配度。 S_m 越高, 说明显著性图的结构与人类视觉感知的契合度越好。

2) F 测度(F-measure, F_β)。由 Zhao 等人(2019)提出, 是用于平衡水下显著性检测中“查准率(precision)”与“查全率(recall)”的加权调和均值指标。其核心是通过引入权重系数 β (通常取 0.3 以侧重抑制误检), 量化算法对低对比度目标(如浑浊水体中

的鱼类)和小目标(如浮游生物)的检测精度, 定义为

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

式中, $P = TP/(TP + FP)$ (预测为前景的像素中真实目标占比), $R = TP/(TP + FN)$ (真实目标像素中被正确检测的比例), TP 为真阳性像素, FP 为假阳性像素, FN 为假阴性像素。该指标通过滑动阈值(0~255)计算并取最优值, 适用于评估算法在复杂水下场景中的综合检测性能。

3) E 测度(E-measure, E_ϕ)。由 Fan 等人(2018)提出, 是兼顾水下图像“像素级误差”与“结构感知一致性”的鲁棒性指标, 尤其适用于评价存在颜色失真、散射干扰场景的检测效果。其定义通过动态权重 ϕ 平衡像素误差与结构相似性, 计算为

$$E_\phi = \frac{2 \times (1 - \xi) \times \tau + \xi}{(1 - \xi) + \tau + \xi} \quad (3)$$

式中, ξ 为像素级误差(衡量预测图与真实标注的逐

像素偏差), τ 为结构相似性(评估目标轮廓,如珊瑚边缘的保留度), ϕ 为自适应调节因子(在高浑浊度场景中增大结构相似性权重)。该指标通过多阈值计算捕捉最优结构一致性,有效反映算法对水下非均匀光照和模糊干扰的抵抗能力。

4) 平均绝对误差(mean absolute error, MAE)。是衡量预测显著性图与真实图(ground truth)像素级平均偏差的指标,反映算法在像素维度的检测准确性。具体定义为

$$MAE = \frac{1}{H \times W} \sum_{x=1}^W \sum_{y=1}^H |S(x, y) - G(x, y)| \quad (4)$$

式中, $S(x, y)$ 为预测显著性图像素值, $G(x, y)$ 为真实图像素值, H 、 W 分别为图像的高度与宽度。MAE 越小,表明预测图与真实图的像素级偏差越小,尤其适用于低对比度、模糊水下图像的检测精度评估。

2.2.2 语义分割评价指标

1) 平均交并比(mean intersection over union, mIoU)。由 Chen 等人(2018b)提出,在水下语义分割中是衡量模型分割精度的核心指标,广泛应用于各类相关研究。对于每个类别 i , 其交并比计算为

$$IoU_i = \frac{TP_i}{TP_i + FP_i + FN_i} \quad (5)$$

平均交并比则是对所有类别 IoU_i 的平均值,即

$$mIoU = \frac{1}{n} \sum_{i=1}^n IoU_i \quad (6)$$

式中, n 为类别总数。mIoU 取值范围为 $[0, 1]$, 值越接近 1, 表明模型对各类别分割的准确性越高,能综合反映模型在不同类别上的分割性能,尤其适用于评估类别分布较为均衡的水下语义分割任务。

2) 平均准确率(mean accuracy, mACC)。用于衡量模型对所有类别像素预测的平均准确程度。其计算式为

$$mACC = \frac{1}{n} \sum_{i=1}^n \frac{TP_i}{TP_i + FN_i} \quad (7)$$

式中, n 为类别总数, TP_i 和 FN_i 含义同前。mACC 体现了模型在每个类别内部正确预测像素占该类别实际像素的平均比例,值越高说明模型对各类别像素的整体识别准确程度越好。

3) 平均类别准确率(average accuracy, aACC)。计算所有类别预测正确的像素数占总像素数的比例,计算式为

$$aACC = \frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n (TP_i + FP_i + TN_i)} \quad (8)$$

aACC 从整体层面反映模型对所有像素的分类准确情况,是从全局角度评估水下语义分割模型性能的重要指标之一。

2.2.3 实例分割评价指标

1) 平均精度均值(mean average precision, mAP)。由 Lin 等人(2014)提出,在水下实例分割任务中是关键综合评价指标。其计算基于平均精度(average precision, AP), AP 是针对每个类别,通过计算精确率(precision)—召回率(recall)曲线(PR 曲线)下的面积得出,表示模型在该类别上的平均精确率。具体而言,先针对每个类别,依据模型预测结果与真实标签,算出不同置信度阈值下的精确率与召回率,据此绘制 PR 曲线,对该曲线下的面积进行积分即得到 AP。而 mAP 则是将所有类别的 AP 进行算术平均,计算为

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad (9)$$

式中, n 为类别总数, mAP 为第 i 类的平均精度。mAP 的取值范围在 0 到 1 之间,数值越高,表明模型在多个类别上综合的实例分割性能越优,能全面反映模型在整个数据集上对不同类别实例分割的精确率与召回率的平衡情况。

2) AP_{50} 和 AP_{75} 属于特定 IoU 阈值下的平均精度指标。其中, AP_{50} 是在 IoU 阈值设定为 0.5 时计算得到的平均精度,意味着预测实例掩码与真实掩码的重叠部分达到 50% 及以上,便认定为正确检测。该指标相对宽松,更侧重评估模型对实例位置的大致定位能力,在一些对分割边界要求不高、仅需对水下目标进行初步定位与区分的场景中具有重要参考价值。 AP_{75} 则是在 IoU 阈值为 0.75 时计算的平均精度。

2.2.4 评价指标适用边界与局限性

1) 显著性检测指标。S-measure 适用于评估水下低对比度场景下的目标结构保留能力,但对运动模糊图像的敏感度不足;MAE 能直观反映像素级偏差,适合水下弱光场景检测精度评估,但无法区分目标边缘与内部的误差差异;F-measure 通过侧重查准率,有效抑制浑浊水体中的误检问题,是水下场景的

核心评价指标。

2) 语义分割指标。mIoU是综合性能评估的核心指标,但在类别不平衡场景(如深海中少量生物与大量水体)中存在偏差;aACC侧重全局像素分类准确率,适合整体分割效果评估,但对小目标分割精度不敏感;mACC能反映各类别内部的分割准确性,适合水下多类别均衡场景。

3) 实例分割指标。mAP综合反映多类别实例分割性能,适合通用场景评估;AP₅₀适用于对分割边界要求较低的场景(如水下机器人避障),AP₇₅对边界精度要求严苛,适合精细场景(如珊瑚礁个体监测),但在强散射模糊场景中易出现结果偏差。

3 基准结果

3.1 显著性目标检测结果对比

表2列出了主流的水下显著性目标检测方法在USOD10K、USOD、MAS3K、RMAS、UFO-120、RUWI共6个数据集上的基准结果。

HEHP几乎在所有数据集上取得了最优性能,这主要得益于其基于异构专家学习与层级感知的核心设计,通过层级原型引导交互、动态适配频率/空间特征融合、抑制深度噪声干扰,并结合分支差异化监督,在高效建模模态特性与降低干扰的同时,兼顾了性能与泛化能力。HDANet-D是HDANet的增加深度分支版本,在USOD10K数据集上HDANet-D性能略好,在USOD数据集上HDANet性能较为领先,这一方面是因为两个数据集的深度图质量不同,另一方面是HDANet对于深度信息的优势探索尚不够。

两个基于SAM的水下显著性目标检测模型Dual-SAM和MAS-SAM都取得了相对较高的指标,但ViT-H的骨干网络带来了较大的计算量,且相对于ResNet、Swin和PVT等轻量级的骨干网络模型性能提升不明显。这两方面的因素:一是SAM是基于非水下场景的数据集训练而来,与水下场景有较大的领域差距;二是SAM为一个提示分割模型,显著性目标检测是一个无提示分割任务,因此如何为SAM自动生成显著性提示是一个SAM适配的难点。

DiffMSS和FSCDiff在多个数据集上的部分指标上取得次优效果,这说明Diffusion模型作为基于

表2 水下显著性目标检测基准结果

Table 2 Benchmark results for underwater salient object detection

数据集	方法	骨干网络	$S_m \uparrow$	$E_\phi \uparrow$	$F_\beta \uparrow$	MAE \downarrow
USOD-10K	TC-USOD	T2T-ViT-14	0.921	0.968	0.923	0.020
	SOE	Swin-T	0.923	0.968	0.927	0.019
	HDANet	Res2Net-50	0.924	0.968	0.934	0.020
	HDANet-D	Res2Net-50	0.927	0.968	<u>0.937</u>	0.019
	Dual-SAM	ViT-H	0.924	0.968	0.931	0.018
	IF-USOD	-	0.924	0.969	0.931	0.018
	DiffMSS	-	0.922	0.956	0.912	0.020
	DINet	ResNet-50	0.924	0.960	0.917	0.023
	AFHNet	T2T-ViT-14	0.937	0.970	0.915	0.019
USOD	FSCDiff	-	0.932	<u>0.971</u>	0.932	<u>0.017</u>
	HEHP	PVT-v2-S	<u>0.935</u>	0.983	0.940	0.014
	TC-USOD	T2T-ViT-14	0.911	0.948	0.923	0.036
	SOE	Swin-T	0.900	0.934	0.913	0.043
	HDANet	Res2Net-50	<u>0.927</u>	<u>0.958</u>	0.950	<u>0.030</u>
	HDANet-D	Res2Net-50	0.912	0.949	<u>0.939</u>	0.036
	FSCDiff	-	0.906	0.941	0.921	0.040
	HEHP	PVT-v2-S	0.931	0.978	0.950	0.022
	DINet	ResNet-50	0.838	0.879	0.861	0.073
UFO-120	Dual-SAM	ViT-H	0.856	0.914	0.864	0.064
	MAS-SAM	ViT-H	0.861	0.914	0.864	0.063
	DiffMSS	-	<u>0.873</u>	<u>0.927</u>	<u>0.867</u>	<u>0.057</u>
	HEHP	PVT-v2-S	0.875	0.930	0.869	0.052
MAS-3K	DINet	ResNet-50	0.865	0.906	0.788	0.032
	Dual-SAM	ViT-H	0.884	0.933	0.838	<u>0.023</u>
	MAS-SAM	ViT-H	<u>0.887</u>	<u>0.938</u>	<u>0.840</u>	0.025
	HEHP	PVT-v2-S	0.906	0.961	0.862	0.017
RMAS	DINet	ResNet-50	0.862	0.926	0.801	0.025
	Dual-SAM	ViT-H	0.860	0.944	0.812	0.022
	MAS-SAM	ViT-H	<u>0.865</u>	<u>0.948</u>	<u>0.819</u>	<u>0.021</u>
	HEHP	PVT-v2-S	0.885	0.978	0.841	0.018
RUWI	DINet	ResNet-50	<u>0.898</u>	0.944	0.927	0.040
	Dual-SAM	ViT-H	0.903	0.959	0.939	<u>0.035</u>
	MAS-SAM	ViT-H	0.894	<u>0.961</u>	<u>0.941</u>	<u>0.035</u>
	HEHP	PVT-v2-S	0.890	0.972	0.959	0.024

注:加粗、下划线字体表示各数据集上的最优、次优结果。“ \uparrow ”表示值越高越好,“ \downarrow ”表示值越低越好。“-”表示对应文献中未提供相关信息。

扩散过程的生成式基础模型,通过模拟从噪声到清晰图像的逐步去噪过程,即使对于复杂的水下退化

图像,也具备优异的特征表征和重建能力,为水下显著性检测提供了新范式,但训练和测试代价较大也是一个需要重点优化的问题。

3.2 语义分割结果对比

水下语义分割任务在 SUIM、DeepFish、CaveSeg 和 CoralMask 共 4 个数据集上的基准结果如表 3—表 6 所示。

SUIM 数据集覆盖水体背景(BW)、人类潜水员(HD)、水生植物(PF)、沉船/废墟(WR)、机器人/仪器(RO)、珊瑚礁/无脊椎动物(RI)、鱼类/脊椎动物(FV)、海床/岩石(SR)共 8 类,SEA-Net 在 IoU 相关子类别以及 IoU 指标上均取得了最优效果,这得益于 SEA-Net 通过引入严重程度指标将水下图像分为高/低严重程度,双分支交替训练以增强不同严重程度图像特征学习,以及专为水下场景设计的 SEA-Adapter 优化低层次特征。UISS-Net 在多个 IoU 子类别指标上取得了次优效果,仅略逊于 SEA-Net,在对不同水下目标和场景的分割上表现出色。

DeepFish 数据集上有同样的性能表现,但因为只有两个类别,因此数据集的难度较低,所有的方法都取得了 80% 以上的 mIoU。

在 CaveSeg 数据集上, Swin Transformer 在 mIoU、mACC、aACC 指标上均取得了最优的性能,意味着强大的骨干网络依旧是水下语义分割任务的性能关键。

对于数据量较大的珊瑚数据集 CoralMask,基于 Transformer 的 SegFormer 和 Mask2Former 的性能都要强于基于 CNN 的 DeepLabV3 和 BiSeNetV2,这说明 Transformer 结构对于大规模数据集训练的显著优势。CoralSCOP 以 SAM 为骨干并在解码器处设计新颖的并行语义分支,结合“负样本训练”策略解决语义无关与过分割问题,在 CoralMask 上取得了最优的性能。

3.3 实例分割结果对比

水下实例分割方法在 UIIS、USIS10K、UIIS10K 和 USIS16K 这 4 个数据集上的基准结果如表 7 所示。

AASNet 和 BARIS-ERA 模型分别在 UIIS 数据集和 USIS10K 上也取得了次优的表现,这说明基于 CNN 的 YOLO 和 Mask-RCNN 系列模型依旧强劲。UIS-Mamba 在 UIIS 和 UIIS10K 数据集上取得了次优的结果,这表明新的视觉基础架构 Mamba 也能引入水下任务,且在适配水下任务方面依旧有潜在的提升和改进空间。

UWSAM 模型在多个数据集上取得了相对最优的性能指标,这得益于 SAM 的强大泛化能力,但 632 M 的巨大的参数量也带来了显著的训练推理代价。DiveSeg 在 UIIS 和 USIS10K 数据集上都取得了最优指标,这说明 DINOv2 强大的通用视觉表征和跨域迁移能力,DiveSeg 通过将水下色彩风格特征嵌入到微调过程中,并通过二进制分割提示引入对象级

表 3 在 SUIM 数据集的水下语义分割基准结果

Table 3 Benchmark results for underwater semantic segmentation on SUIM dataset

方法	各类别 IoU								mIoU
	BW	HD	PF	WR	RO	RI	FV	SR	
U-Net	79.46	32.25	21.85	33.94	23.65	50.28	38.16	42.16	39.85
SegNet	80.63	45.67	17.45	32.24	55.72	47.62	43.92	51.51	46.85
SUIM-Net	80.64	63.45	23.27	41.25	60.89	53.12	46.02	57.12	53.22
PSPNet	82.51	65.04	<u>28.54</u>	46.56	62.88	55.80	46.78	55.98	55.51
DeepLab	81.82	50.26	17.05	43.33	63.60	57.18	43.56	55.35	51.52
LEDNet	82.96	58.47	18.02	42.86	50.96	58.13	46.13	54.99	51.36
BiSeNetV2	83.67	59.29	18.27	39.58	56.54	58.16	47.33	56.93	52.47
UISS-Net	<u>89.11</u>	<u>80.88</u>	24.37	<u>68.54</u>	<u>79.22</u>	<u>69.78</u>	<u>76.95</u>	<u>65.96</u>	<u>69.35</u>
SEA-Net	89.75	84.69	38.45	71.11	80.49	72.68	79.11	68.92	73.15

注:加粗、下划线字体表示各列最优、次优结果。

表4 在DeepFish数据集的水下语义分割基准结果

Table 4 Benchmark results for underwater semantic segmentation on DeepFish dataset

方法	IoU		mIoU
	背景	前景	
SegNet	98.89	68.94	83.91
SUIM-Net	99.03	78.40	88.71
PSPNet	99.11	71.35	85.23
DeepLabV3	99.21	66.30	82.75
DGCNet	99.21	81.42	90.32
MFAS-Net	99.15	84.86	92.01
UISS-Net	<u>99.49</u>	<u>89.03</u>	<u>94.26</u>
SEA-Net	99.57	90.73	95.15

注:加粗、下划线字体表示各列最优、次优结果。

表5 在CaveSeg数据集的水下语义分割基准结果

Table 5 Benchmark results for underwater semantic segmentation on CaveSeg dataset

方法	mIoU	mACC	aACC
FastFCN	38.86	46.89	72.01
DeepLabV3+	38.46	<u>49.47</u>	71.64
Segmenter	30.81	39.64	69.76
SegFormer	35.36	44.71	70.19
Swin Transformer	48.11	56.69	73.26
CaveSeg	<u>40.22</u>	45.99	<u>72.91</u>

注:加粗、下划线字体表示各列最优、次优结果。

表6 在CoralMask数据集的水下语义分割基准结果

Table 6 Benchmark results for underwater semantic segmentation on CoralMask dataset

方法	mIoU	mACC	MAE
DeepLabV3	59.39	71.76	0.163 4
BiSeNetV2	60.90	78.11	0.112 9
Mask2Former	69.17	77.63	0.084 2
SegFormer	<u>71.51</u>	84.29	<u>0.077 6</u>
CoralSCOP	79.13	<u>81.11</u>	0.076 4

注:加粗、下划线字体表示各列最优、次优结果。

先验,使DINO完美适配水下场景,实现了最先进的水下实例分割性能。

4 结 语

本文系统综述了水下图像分割领域的研究进展,该技术作为海洋工程、水下机器人导航和海洋生物监测等领域的核心支撑,其核心目标是从光照衰减、水体散射、颜色失真等退化问题显著的水下图像中精准提取目标区域。全文首先梳理了水下图像分割的研究背景、核心挑战和应用价值;随后依据任务目标将现有方法划分为水下显著性目标检测、水下语义分割与水下实例分割3大类,系统分析了各类方法从“传统手工特征驱动”到“深度学习数据驱动”的技术演进路径,其中显著性目标检测实现从RGB单模态到RGBD多模态融合的精度提升,语义分割在模块设计、损失函数和数据增强3方面同时改进,实例分割则依托Mask R-CNN、YOLO及SAM等视觉模型实现类别与实例的双重区分;同时,本文总结了主流数据集的场景覆盖与标注特性,明确了不同评价指标的适用场景,并通过基准实验对比直观呈现了代表性模型在不同数据集上的性能差异。本文通过全面梳理技术脉络、数据资源与实验结果,为相关领域研究者提供了系统的技术参考,助力推动水下图像分割技术的理论研究与应用落地。

尽管水下图像分割技术取得了显著进展,但面对复杂多变的水下环境和日益增长的应用需求,仍存在诸多亟待解决的问题。结合当前研究瓶颈与实际应用场景,未来的研究方向可聚焦于以下几个方面:

1)小样本水下图像分割。现有深度学习方法高度依赖大规模标注数据,深海环境图像采集成本极高(需搭载专业深潜设备),且稀有生物、未知沉船等目标标注困难,亟需小样本与弱监督技术突破。最近已有工作(Kabir等,2023;Li等,2025e)结合元学习、迁移学习与水下物理先验,设计面向小样本的特征提取与自适应调优机制。例如,通过元学习挖掘不同水下场景的共性特征,利用预训练视觉基础模型结合少量水下样本微调,减少对标注数据的依赖,满足深海探测、稀有生物监测等小样本场景需求。

2)弱监督水下图像分割。现有高精度分割模型多依赖像素级强监督标注,而水下图像标注耗时耗力,大量实际场景无法提供充足标注。未来需重点探索弱监督(图像级标注、边界框标注、点标注)与无

表7 水下实例分割基准结果

Table 7 Benchmark results for underwater instance segmentation

数据集	方法	骨干网络	参数量/M	mAP/%	AP ₅₀ /%	AP ₇₅ /%
UIIS	Mask R-CNN	ResNet-101	63	23.4	40.9	25.3
	WaterMask	ResNet-101	67	27.2	43.7	29.3
	UW-DETR	ResNet-101	-	29.2	47.1	31.6
	YOLOv9-n	GELAN	-	29.5	48.2	-
	AASNet	GELAN+LCA	28	31.7	<u>49.5</u>	<u>35.1</u>
	USIS-SAM	ViT-H	700	29.4	45.0	32.3
	UIS-Mamba	UIS-Mamba-B	115	31.2	49.1	34.5
	BARIS-ERA	Swin-B	114	31.6	52.0	33.6
	UWSAM	ViT-H	632	<u>31.9</u>	48.2	34.9
	DiveSeg	ViT-L	390	35.6	52.0	38.5
USIS10K	Mask R-CNN	ResNet-101	63	36.9	53.1	41.7
	WaterMask	ResNet-101	67	38.7	54.9	43.2
	USIS-SAM	ViT-H	700	43.1	59.0	48.5
	UIS-Mamba	UIS-Mamba-B	115	46.2	<u>63.2</u>	53.4
	BARIS-ERA	Swin-B	114	<u>47.3</u>	65.1	<u>53.7</u>
	UWSAM	ViT-H	632	46.0	61.7	51.7
	DiveSeg	ViT-L	390	48.4	62.3	54.4
UIIS10K	Mask R-CNN	ResNet-101	63	35.8	53.6	40.2
	WaterMask	ResNet-101	67	37.4	51.6	41.7
	USIS-SAM	ViT-H	700	<u>39.8</u>	<u>52.0</u>	<u>42.6</u>
	UWSAM	ViT-H	632	44.6	58.9	49.2
USIS16K	Mask R-CNN	ResNet-101	63	<u>73.6</u>	<u>90.0</u>	<u>81.7</u>
	WaterMask	ResNet-101	67	72.7	86.8	79.3
	USIS-SAM	ViT-H	700	81.0	90.8	87.1

注:加粗、下划线字体表示各数据集上的最优、次优结果。“-”表示未知。

监督分割方法。例如,弱监督方向可设计适配水下退化特性的损失函数,通过图像级标签引导像素级类别推断,或利用边界框约束目标区域生成;无监督方向可结合生成对抗网络(GAN)、自监督特征学习(如对比学习),挖掘水下图像的颜色、纹理、频率等固有特征,实现无标注情况下的目标与背景区分,降低标注成本。

3)开放词汇水下图像分割。当前语义与实例分割模型均基于预定义类别训练,无法处理实际应用中出现的新类别目标。例如珊瑚礁监测场景中,珊瑚礁种类繁多(全球超过1 000种),现有固定类别

模型无法覆盖所有种类,需融合视觉—语言模型(如CLIP(contrastive language-image pre-training))构建开放词汇分割框架,实现未知珊瑚类别的识别与分割;同时需解决水下颜色失真、模糊等问题对跨模态特征对齐的干扰,或设计模态自适应注意力机制提升文本—视觉匹配精度,满足海洋生物多样性监测中未知物种识别的需求。

4)参考水下图像分割。现有分割模型多为端自动分割,缺乏交互式灵活定位能力,而实际场景中常需基于参考信息(如文本指代)分割用户指定的目标。文本指代分割需融合细粒度文本描述(如“体

长约 30 cm 的黄色鱼类”)与水下目标视觉特征,通过语义解析模块提取文本中的形状、颜色、尺寸等属性,引导模型定位并分割目标。Yao 等人(2025)首次提出语言引导的水下图像推理分割任务,构建了首个水下图像—语言分割数据集,通过多模态大模型(multimodal large language model, MLLM)的方案融入了推理能力,使系统能够解读场景中物体间的复杂关系,并在动态水下环境中实现指定目标的分割。

5)实时多模态水下图像分割。水下机器人导航对分割速度要求严苛,且面临动态障碍物、强散射水体等复杂环境,通常需要融合多模态信息。当前多模态分割多聚焦 RGB 与深度信息融合,而偏振、高光谱、声学等模态考虑较少。未来需研究多模态数据的高效融合策略,一方面需解决模态异质性问题,设计跨模态特征转换与对齐模块;另一方面需结合水下场景特性筛选关键模态,同时考虑多模态设备的小型化与低功耗,适配水下机器人的实际部署。

6)跨域泛化与鲁棒性提升。现有模型在特定数据集(如浅海清澈水体)上表现优异,但迁移至其他场景(如深海弱光、河口高浑浊水体)时性能显著下降,核心原因是水下环境的光照、散射、颜色失真程度差异大,导致模型难以适应域间变化。未来需从两方面提升跨域泛化能力:一是引入域自适应技术,通过域对抗训练、元域自适应等方法,减少源域与目标域的分布差异;二是融合水下物理模型,将光衰减公式、散射系数等物理先验融入模型设计,使模型能根据水体光学参数动态调整特征提取策略。

未来,水下图像分割的研究将朝着“低标注成本”(小样本、弱监督)、“高实用性”(开放词汇、文本指代)、“多信息融合”(实时多模态协同)和“强鲁棒性”(跨域泛化)的方向发展。随着这些关键技术的突破,水下图像分割将为海洋资源勘探、水下机器人智能化以及海洋生态监测等领域提供更强大的技术支持,助力人类更高效、更深入地探索和利用海洋资源。

参考文献(References)

Abdullah A, Barua T, Tibbetts R, Chen Z J, Islam M J and Rekleitis I. 2024. CaveSeg: deep semantic segmentation and scene parsing for autonomous underwater cave exploration//Proceedings of 2024 IEEE International Conference on Robotics and Automation. Yoko-

hama, Japan: IEEE: 3781-3788 [DOI: 10.1109/ICRA57147.2024.10611543]

Anand S K, Kumar P V, Saji R, Gadagkar A V and Chandavarkar B R. 2023. A survey on semantic segmentation models for underwater images//Proceedings of the ICDSAI 2022 on Advances in Data Science and Artificial Intelligence. IIT Patna, India; Springer: 69-86 [DOI: 10.1007/978-3-031-16178-0_7]

Bailey G N and Flemming N C. 2008. Archaeology of the continental shelf: marine resources, submerged landscapes and underwater archaeology. *Quaternary Science Reviews*, 27(23/24): 2153-2165 [DOI: 10.1016/j.quascirev.2008.08.012]

Barat C and Phlypo R. 2010. A fully automated method to detect and segment a manufactured object in an underwater color image. *EURASIP Journal on Advances in Signal Processing*, 2010(1): #568092 [DOI: 10.1155/2010/568092]

Chang L B, Wang Y K, Huang J X, Deng L X, Du B and Xu C. 2025. Marine saliency segmenter: object-focused conditional diffusion with region-level semantic knowledge distillation [EB/OL]. [2025-09-30]. <https://arxiv.org/pdf/2504.02391.pdf>

Chen B W, Zhao W, Zhang Q S, Li M L, Qi M Y and Tang Y. 2025a. Semantic segmentation of underwater images based on the improved SegFormer. *Frontiers in Marine Science*, 12: #1522160 [DOI: 10.3389/fmars.2025.1522160]

Chen J F, Zhu S D and Luo W L. 2024. Instance segmentation of underwater images by using deep learning. *Electronics*, 13(2): #274 [DOI: 10.3390/electronics13020274]

Chen L C, Papandreou G, Kokkinos I, Murphy K and Yuille A L. 2018b. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4): 834-848 [DOI: 10.1109/TPAMI.2017.2699184]

Chen L C, Zhu Y K, Papandreou G, Schroff F and Adam H. 2018a. Encoder-decoder with atrous separable convolution for semantic image segmentation//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 833-851 [DOI: 10.1007/978-3-030-01234-2_49]

Chen Z, Gao H M, Zhang Z, Zhou H L, Wang X and Tian Y. 2020. Underwater salient object detection by combining 2D and 3D visual features. *Neurocomputing*, 391: 249-259 [DOI: 10.1016/j.neucom.2018.10.089]

Chen Z Y, Zhang C, Fang H and Cong R M. 2025b. Empowering DINO representations for underwater instance segmentation via aligner and prompter [EB/OL]. [2025-11-18]. <https://arxiv.org/pdf/2511.08334.pdf>

Chicchon M, Bedon H, Del-Blanco C R and Sipiran I. 2023. Semantic segmentation of fish and underwater environments using deep convolutional neural networks and learned active contours. *IEEE Access*, 11: 33652-33665 [DOI: 10.1109/ACCESS.2023.3262649]

Cong R M, Yu Z J, Fang H, Sun H Y and Kwong S. 2025. UIS-Mamba:

- exploring Mamba for underwater instance segmentation via dynamic tree scan and hidden state weaken [EB/OL]. [2025-09-30]. <https://arxiv.org/pdf/2508.00421.pdf>
- Cui Z, Wu J F, Yu H, Zhou Y Z and Liang L. 2019. Underwater image saliency detection based on improved histogram equalization//Proceedings of 2019 International Conference of Pioneering Computer Scientists, Engineers and Educators. Singapore, Singapore: Springer: 157-165 [DOI: 10.1007/978-981-15-0121-0_12]
- Drewno-Jr P, De Souza I, Maurel 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]
- Edgington D R, Salmay K A, Risi M, Sherlock R E, Walther D and Koch C. 2003. Automated event detection in underwater video//Proceedings of the Oceans 2003. Celebrating the Past Teaming Toward the Future. San Diego, USA: IEEE: 2749-2753 [DOI: 10.1109/OCEANS.2003.178344]
- Elmezain M, Saad Saoud L, Sultan A, Heshmat M, Seneviratne L and Hussain I. 2025. Advancing underwater vision: a survey of deep learning models for underwater object recognition and tracking. *IEEE Access*, 13: 17830-17867 [DOI: 10.1109/ACCESS.2025.3534098]
- Fan D P, Cheng M M, Liu Y, Li T and Borji A. 2017. Structure-measure: a new way to evaluate foreground maps//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 4558-4567 [DOI: 10.1109/ICCV.2017.487]
- Fan D P, Gong C, Cao Y, Ren B, Cheng M M and Borji A. 2018. Enhanced-alignment measure for binary foreground map evaluation//Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm, Sweden: IJCAI: 698-704 [DOI: 10.24963/ijcai.2018/97]
- Fu Z Q, Chen R Z, Huang Y, Cheng E, Ding X H and Ma K K. 2024. MASNet: a robust deep marine animal segmentation network. *IEEE Journal of Oceanic Engineering*, 49(3): 1104-1115 [DOI: 10.1109/JOE.2023.3252760]
- Ge H L and Ouyang J L. 2025. Underwater image segmentation via the progressive network of dual iterative complement enhancement. *Expert Systems with Applications*, 266: #126049 [DOI: 10.1016/j.eswa.2024.126049]
- González-Sabbagh S P and Robles-Kelly A. 2023. A survey on underwater computer vision. *ACM Computing Surveys*, 55(13s): # 268 [DOI: 10.1145/3578516]
- Guo H R, Guo J C and Wang Y D. 2023. Lightweight semantic segmentation network for underwater image. *Journal of Zhejiang University (Engineering Science)*, 57(7): 1278-1286, 1296 (郭浩然, 郭继昌, 汪显东. 面向水下场景的轻量级图像语义分割网络. *浙江大学学报(工学版)*, 57(7): 1278-1286, 1296) [DOI: 10.3785/j.issn.1008-973X.2023.07.002]
- He K M, Gkioxari G, Dollár P and Girshick R. 2017. Mask R-CNN//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 2980-2988 [DOI: 10.1109/ICCV.2017.322]
- He Z Q, Cao L J, Luo J L, Xu X Q, Tang J Y, Xu J H, et al. 2024. UISS-Net: underwater image semantic segmentation network for improving boundary segmentation accuracy of underwater images. *Aquaculture International*, 32(5): 5625-5638 [DOI: 10.1007/s10499-024-01439-x]
- He Z Q, Cao L J, Xu X Q and Xu J H. 2025. Underwater instance segmentation: a method based on channel spatial cross-cooperative attention mechanism and feature prior fusion. *Frontiers in Marine Science*, 12: #1557965 [DOI: 10.3389/fmars.2025.1557965]
- Hong L, Wang X, Li Y H and Wang X. 2025b. USIS16K: high-quality dataset for underwater salient instance segmentation [EB/OL]. [2025-09-30]. <https://arxiv.org/pdf/2506.19472.pdf>
- Hong L, Wang X, Zhang G and Zhao M. 2025a. Usod10k: a new benchmark dataset for underwater salient object detection. *IEEE Transactions on Image Processing*, 34: 1602-1615 [DOI: 10.1109/TIP.2023.3266163]
- 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. Las Vegas, USA: IEEE: 1769-1776 [DOI: 10.1109/IROS45743.2020.9340821]
- Islam M J, Luo P G and Sattar J. 2020b. Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception//Robotics: Science and Systems 2020. Corvallis, USA: MIT Press Journals: 1-11 [DOI: 10.15607/RSS.2020.XVI.018]
- Islam M J, Wang R B and Sattar J. 2022. SVAM: saliency-guided visual attention modeling by autonomous underwater robots [EB/OL]. [2025-09-30]. <https://arxiv.org/pdf/2011.06252.pdf>
- Itti L, Koch C and Niebur E. 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11): 1254-1259 [DOI: 10.1109/34.730558]
- Jian M W, Liu X Y, Luo H J, Lu X W, Yu H and Dong J Y. 2021. Underwater image processing and analysis: a review. *Signal Processing: Image Communication*, 91: #116088 [DOI: 10.1016/j.image.2020.116088]
- Jian M W, Qi Q, Dong J Y, Yin Y L and Lam K M. 2018. Integrating QDWD with pattern distinctness and local contrast for underwater saliency detection. *Journal of Visual Communication and Image Representation*, 53: 31-41 [DOI: 10.1016/j.jvcir.2018.03.008]
- Jin J H, Jiang Q P, Wu Q Y, Xu B W and Cong R M. 2025. Underwater salient object detection via dual-stage self-paced learning and depth emphasis. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(3): 2147-2160 [DOI: 10.1109/TCSVT.2024.3491907]
- Kabir I, Shaurya S, Maigur V, Thakurdesai N, Latnekar M, Raunak

- M, et al. 2023. Few-Shot segmentation and semantic segmentation for underwater imagery//Proceedings of 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems. Detroit, USA: IEEE: 11451-11457 [DOI: 10.1109/IROS55552.2023.10342227]
- Kim Y H and Park K R. 2022. PSS-net: parallel semantic segmentation network for detecting marine animals in underwater scene. *Frontiers in Marine Science*, 9: #1003568 [DOI: 10.3389/fmars.2022.1003568]
- Kirillov A, Mintun E, Ravi N, Mao H Z, Rolland C and Gustafson L. 2023. Segment anything//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 3992-4003 [DOI: 10.1109/ICCV51070.2023.00371]
- Kong J L, Tang S N, Feng J M, Mo L P and Jin X B. 2025. AASNet: a novel image instance segmentation framework for fine-grained fish recognition via linear correlation attention and dynamic adaptive focal loss. *Applied Sciences*, 15 (7) : #3986 [DOI: 10.3390/app15073986]
- Lee D, Kim G, Kim D, Myung H and Choi H T. 2012. Vision-based object detection and tracking for autonomous navigation of underwater robots. *Ocean Engineering*, 48: 59-68 [DOI: 10.1016/j.oceaneng.2012.04.006]
- Li D S, Zhao S W, Hu J Z, Yang Y F and Ding J Q. 2025a. An underwater image segmentation model for complex scenes in aquaculture using vision Transformer. *Computers and Electronics in Agriculture*, 238: #110764 [DOI: 10.1016/j.compag.2025.110764]
- Li H, Lian S J, Li Z Y, Cong R M, Li C Y, Laurence T, et al. 2025b. Advancing marine research: UWSAM framework and UIIS10K dataset for precise underwater instance segmentation [EB/OL]. [2025-09-30]. <https://arxiv.org/pdf/2505.15581.pdf>
- Li H, Lin G W, Li Z Y, Kwong S and Cong R M. 2025c. FSCDiff: frequency-spatial entangled conditional diffusion model for underwater salient object detection//Proceedings of the 33rd ACM International Conference on Multimedia. Dublin, Ireland: ACM: 8379-8388 [DOI: 10.1145/3746027.3755467]
- Li L, Rigall E, Dong J Y and Chen G. 2020. MAS3K: an open dataset for marine animal segmentation//Proceedings of the 3rd BenchCouncil International Symposium on Benchmarking, Measuring, and Optimizing. Virtual Event: Springer: 194-212 [DOI: 10.1007/978-3-030-71058-3_12]
- Li M, Zhang H Q, Gruen A and Li D R. 2025d. A survey on underwater coral image segmentation based on deep learning. *Geo-spatial Information Science*, 28(2): 472-496 [DOI: 10.1080/10095020.2024.2343323]
- Li X, Song J D, Zhang F, Ouyang X G and Khan S U. 2016. MapReduce-based fast fuzzy c-means algorithm for large-scale underwater image segmentation. *Future Generation Computer Systems*, 65: 90-101 [DOI: 10.1016/j.future.2016.03.004]
- Li Z H, Huang Z C, Liu W C, Zhang Z X and Miao J M. 2025e. FSSU-WNet: mitigating the fragility of pre-trained models with feature enhancement for few-shot semantic segmentation in underwater images//Proceedings of 2025 International Joint Conference on Neural Networks (IJCNN). Rome, Italy: IEEE: 1-8 [DOI: 10.1109/IJCNN64981.2025.11229243]
- 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. 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: PMLR: 235: 29545-29559 [DOI: 10.48550/arXiv.2406.06039]
- Lin T Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al. 2014. Microsoft COCO: common objects in context//Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland: Springer: 740-755 [DOI: 10.1007/978-3-319-10602-1_48]
- Liu C X, Yao H X, Qiu W H, Cui H Y, Fang Y B and Xu A Q. 2025a. Multi-scale feature map fusion encoding for underwater object segmentation. *Applied Intelligence*, 55 (2) : #163 [DOI: 10.1007/s10489-024-05971-4]
- Liu F F and Fang M. 2020. Semantic segmentation of underwater images based on improved DeepLab. *Journal of Marine Science and Engineering*, 8(3): #188 [DOI: 10.3390/jmse8030188]
- Liu T W, Wang R Y, Zhang M, Piao Y R and Lu H C. 2025b. Auto-USOD: searching topology for underwater salient object detection//Proceedings of the 7th Chinese Conference on Pattern Recognition and Computer Vision. Urumqi, China: Springer: 3-16 [DOI: 10.1007/978-981-97-8490-5_1]
- Liu Y W, Zhang X Y, Zhu J C, Ma B T, Duan Y T and Tan P L. 2025c. HDANet: enhancing underwater salient object detection with physics-inspired multimodal joint learning. *IEEE Transactions on Geoscience and Remote Sensing*, 63: #5622614 [DOI: 10.1109/TGRS.2025.3565579]
- Liu Y, Tian Y J, Zhao Y Z, Yu H T, Xie L X, Wang Y W, et al. 2024. VMamba: visual state space model//Proceedings of the 38th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: 103031-103063 [DOI: 10.52202/079017-3273]
- Liu Z, Lin Y T, Cao Y, Hu H, Wei Y X, Zhang Z, et al. 2021. Swin transformer: hierarchical vision transformer using shifted windows//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 9992-10002 [DOI: 10.1109/ICCV48922.2021.00986]
- Long J, Shelhamer E and Darrell T. 2015. Fully convolutional networks for semantic segmentation//Proceedings of 2015 IEEE Conference

- on Computer Vision and Pattern Recognition. Boston, USA: IEEE: 3431-3440 [DOI: 10.1109/CVPR.2015.7298965]
- Ma Q W, Li X B, Li B C, Zhu Z, Wu J, Huang F, et al. 2025. STAMF: synergistic transformer and Mamba fusion network for RGB-Polarization based underwater salient object detection. *Information Fusion*, 122: #103182 [DOI: 10.1016/j.inffus.2025.103182]
- Ma Z W, Li H J, Fan X, Luo Z X, Li J J and Wang Z H. 2022. A real scene underwater semantic segmentation method and related dataset. *Journal of Beijing University of Aeronautics and Astronautics*, 48(8): 1515-1524 (马志伟, 李豪杰, 樊鑫, 罗钟铉, 李建军, 王智慧. 2022. 真实场景水下语义分割方法及数据集. *北京航空航天大学学报*, 48(8): 1515-1524) [DOI: 10.13700/j.bh.1001-5965.2021.0527]
- Nunes A and Matos A. 2023. Improving semantic segmentation performance in underwater images. *Journal of Marine Science and Engineering*, 11(12): #2268 [DOI: 10.3390/jmse11122268]
- O'Byrne M, Pakrashi V, Schoefs F and Ghosh B. 2018. Semantic segmentation of underwater imagery using deep networks trained on synthetic imagery. *Journal of Marine Science and Engineering*, 6(3): #93 [DOI: 10.3390/jmse6030093]
- Oquab M, Darcet T, Moutakanni T, Vo H V, Szafraniec M, Khalidov V, et al. 2023. DINOv2: learning robust visual features without supervision [EB/OL]. [2025-09-30]. <https://arxiv.org/pdf/2304.07193.pdf>
- Pan P C and Pei S C. 2025. BARIS: boundary-aware refinement with environmental degradation priors for robust underwater instance segmentation [EB/OL]. [2025-09-30]. <https://arxiv.org/pdf/2504.19643.pdf>
- Peng Y T, Lin Y C, Peng W Y and Chen C Y. 2024. Blurriness-guided underwater salient object detection and data augmentation. *IEEE Journal of Oceanic Engineering*, 49(3): 1089-1103 [DOI: 10.1109/JOE.2023.3344154]
- Rajasekar M, Celine Kavida A and Anto Bennet M. 2015. Underwater K-means clustering segmentation using SVM classification. *Middle-East Journal of Scientific Research*, 23(9): 2166-2172 [DOI: 10.5829/idosi.mejsr.2015.23.09.22582]
- Ren S Q, He K M, Girshick R and Sun J. 2017. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6): 1137-1149 [DOI: 10.1109/TPAMI.2016.2577031]
- Ronneberger O, Fischer P and Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation//*Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention*. Munich, Germany: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Saleh A, Laradji I H, Kononov 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]
- Tang Q, Wang Z, Wang X Q and Zhang S W. 2025b. AFHNet: attention-free hybrid network for salient object detection in underwater images. *IEEE Access*, 13: 89287-89299 [DOI: 10.1109/ACCESS.2025.3567606]
- Tang S N, Jin X B, Bai Y T, Su T L and Kong J L. 2025a. YOLOv9-N: an underwater image instance segmentation algorithm based on YOLOv9 and improved non-local modules//*Proceedings of the 16th International Conference on Modelling, Identification and Control (ICMIC2024)*. Singapore, Singapore: Springer: 339-344 [DOI: 10.1007/978-981-96-1777-7_38]
- Varghese R and Sambath M. 2024. Yolov8: a novel object detection algorithm with enhanced performance and robustness//*Proceedings of 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems*. Chennai, India: IEEE: 1-6 [DOI: 10.1109/ADICSS58448.2024.10533619]
- Wang C Y, Yeh I H and Mark Liao H Y. 2025. Yolov9: learning what you want to learn using programmable gradient information//*Proceedings of the 18th European Conference on Computer Vision*. Milan, Italy: Springer: 1-21 [DOI: 10.1007/978-3-031-72751-1_1]
- Wang J D, Sun K, Cheng T H, Jiang B R, Deng C R, Zhao Y, et al. 2021. Deep high-resolution representation learning for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10): 3349-3364 [DOI: 10.1109/TPAMI.2020.2983686]
- Wang J K, He X H, Shao F M and Lu G L. 2023. Semantic segmentation method of underwater images based on encoder-decoder architecture. *Ship Electronic Engineering*, 43(9): 29-35 (王金康, 何晓晖, 邵发明, 卢冠林. 2023. 基于编解码机制的水下图像语义分割. *舰船电子工程*, 43(9): 29-35) [DOI: 10.3969/j.issn.1672-9730.2023.09.008]
- Wang J K, He X H, Shao F M, Lu G L, Hu R Z and Jiang Q Y. 2022. Semantic segmentation method of underwater images based on encoder-decoder architecture. *PLoS One*, 17(8): #e0272666 [DOI: 10.1371/journal.pone.0272666]
- Wei Y, Wang Y, Yan S J, Wang T Z, Wang Z H, Sun W R, et al. 2024. CSUNet: contour-sensitive underwater salient object detection//*Proceedings of the 6th ACM International Conference on Multimedia in Asia*. Auckland, New Zealand: ACM: #78 [DOI: 10.1145/3696409.3700239]
- Wu Q Y, Fu Z Q, Lin H, Ma C Y, Tu X T and Ding X H. 2025b. EffiSeaNet: pioneering lightweight network for underwater salient object detection//*Proceedings of the 17th Asian Conference on Computer Vision on Computer Vision*. Hanoi, Vietnam: Springer: 89-104 [DOI: 10.1007/978-981-96-0911-6]
- Wu Q Y, Xie J X, Fu Z Q, Tu X T, Huang Y and Ding X H. 2025a. CE³USOD: channel-enhanced, efficient, and effective network for underwater salient object detection. *IEEE Journal of Oceanic Engineering*, 50(2): 941-954 [DOI: 10.1109/JOE.2024.3523356]

- Xie E Z, Wang W H, Yu Z D, Anandkumar A, Alvarez J M and Luo P. 2021. SegFormer: simple and efficient design for semantic segmentation with transformers//Proceedings of the 35th International Conference on Neural Information Processing Systems. Curran Associates Inc.: 34: 12077-12090
- Xu G B, Shen W M and Wang X B. 2014. Applications of wireless sensor networks in marine environment monitoring: a survey. *Sensors*, 14(9): 16932-16954 [DOI: 10.3390/s140916932]
- Yan T Y, Wan Z F, Deng X H, Zhang P P, Liu Y and Lu H C. 2024. MAS-SAM: segment any marine animal with aggregated features//Proceedings of the 33rd International Joint Conference on Artificial Intelligence. Jeju, Korea (South): Curran Associates, Inc.: 6886-6894 [DOI: 10.24963/ijcai.2024/761]
- Yao M D, Tam K M, Wang M L, Li L G and Kawakami R. 2025. Language-guided reasoning segmentation for underwater images. *Information Fusion*, 122: #103177 [DOI: 10.1016/j.inffus.2025.103177]
- Yi D W, Ahmedov H B, Jiang S Y, Li Y R, Flinn S J and Fernandes P G. 2024. Coordinate-aware mask R-CNN with group normalization: a underwater marine animal instance segmentation framework. *Neurocomputing*, 583: #127488 [DOI: 10.1016/j.neucom.2024.127488]
- Yuan G J, Song J T and Li J J. 2025. IF-USOD: multimodal information fusion interactive feature enhancement architecture for underwater salient object detection. *Information Fusion*, 117: #102806 [DOI: 10.1016/j.inffus.2024.102806]
- Yuan Y Q, Tian Y B, Sun Y X and Pan T. 2024. Underwater image segmentation based on attention mechanism and the U-net model. *Journal of Jiangsu University of Science and Technology (Natural Science Edition)*, 38(2): 53-57 (苑永起, 田雨波, 孙颖鑫, 潘婷. 基于注意力机制和U-net模型的水下图像分割方法研究. *江苏科技大学学报(自然科学版)*, 38(2): 53-57) [DOI: 10.20061/j.issn.1673-4807.2024.02.009]
- Zha M F, Wang G Q, Pei Y Q, Li T Y, Tang X X, Li C Y, et al. 2025. Heterogeneous experts and hierarchical perception for underwater salient object detection. *IEEE Transactions on Image Processing*, 34: 3703-3717 [DOI: 10.1109/TIP.2025.3572760]
- Zhang P P, Yan T Y, Liu Y and Lu H C. 2024. Fantastic animals and where to find them: segment any marine animal with dual SAM//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 2578-2587 [DOI: 10.1109/CVPR52733.2024.00249]
- Zhao K, Gao S H, Wang W G and Chen M M. 2019. Optimizing the F-measure for threshold-free salient object detection//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 8848-8856 [DOI: 10.1109/ICCV.2019.00894]
- Zheng Z Q, Chen Y W, Zeng H M, Vu T A, Hua B S and Yeung S K. 2025. Marineinst: a foundation model for marine image analysis with instance visual description//Proceedings of the 18th European Conference on Computer Vision—ECCV 2024. Milan, Italy: Springer: 239-257 [DOI: 10.1007/978-3-031-72627-9_14]
- 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. Seattle, USA: IEEE: 28170-28180 [DOI: 10.1109/CVPR52733.2024.02661]
- Zhuang P X, Wang Y H, Zhang X H, Liu F, Tong J J and Fu Z Q. 2026. A Survey of Underwater Light Field Imaging: Theory and Applications. *Journal of Image and Graphics*, 31(5): 1503-1522. (庄培显, 王一航, 张新恒, 刘飞, 童俊杰, 富振奇. 2026. 水下光场成像综述: 理论与应用. *中国图象图形学报*, 31(5): 1503-1522) [DOI: 10.11834/jig.250255]
- Zhou L, Liu Q M, Jin K J, Zhao W Y and Zhang W D. 2025a. 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]
- Zhou Z Y and Li Z. 2025b. UHRS-Net: underwater high-resolution semantic segmentation network. *Signal, Image and Video Processing*, 19(11): #921 [DOI: 10.1007/s11760-025-04463-3]
- Zhu J Y and Zhang T. 2025. SEA-Net: a severity-aware network with visual prompt tuning for underwater semantic segmentation//Proceedings of the 13th International Conference on Computational Visual Media. Hong Kong, China: Springer: 331-348 [DOI: 10.1007/978-981-96-5809-1_18]
- Zuo X, Jiang J R, Shen J F and Yang W K. 2025. Improving underwater semantic segmentation with underwater image quality attention and multi-scale aggregation attention. *Pattern Analysis and Applications*, 28(2): #80 [DOI: 10.1007/s10044-025-01460-7]

作者简介

方豪,男,博士研究生,主要研究方向为计算机视觉和图像/视频分割。E-mail:fanghaok@mail.sdu.edu.cn

丛润民,通信作者,男,教授,主要研究方向为人工智能、模式识别、机器学习和具身感知。E-mail:rmcong@sdu.edu.cn

于宗吉,男,本科生,主要研究方向为水下图像分割。

E-mail:zjyu@mail.sdu.edu.cn

陈志杨,男,硕士研究生,主要研究方向为受限场景视觉感知。E-mail:zhiyangchen@mail.sdu.edu.cn