

中图法分类号: TP24; P71 文献标识码: A 文章编号: 1006-8961(2026)05-1288-12

论文引用格式: Lu H M, Zheng Y C and Li Y J. 2026. Embodied marine perception: a review. Journal of Image and Graphics, 31(5):1288-1299
(陆慧敏, 郑禹超, 李玉洁. 2026. 具身海洋环境感知综述. 中国图象图形学报, 31(5):1288-1299)[DOI:10.11834/jig.250429]

具身海洋环境感知综述

陆慧敏^{1,2*}, 郑禹超^{2,3}, 李玉洁³

1. 东南大学自动化学院, 南京 210096; 2. 东南大学南通海洋高等研究院, 南通 226000;
3. 九州工业大学工学部, 福岡 804-8550, 日本

摘要: 具身感知强调智能体通过物理交互感知和理解环境, 是实现自主智能的重要路径。本文将具身感知理论拓展至复杂的海洋环境, 提出具身海洋环境感知这一战略方向, 该方向面向海洋视觉受限、通信不畅以及信息稀疏等核心挑战, 强调通过主动物理交互机制提升环境建模与认知能力。具身海洋环境感知对于海洋资源开发、国家安全保障以及深海科学探索具有重要的基础性和战略性意义, 是推动我国迈向海洋强国的关键技术基础。本文系统梳理当前研究现状, 构建涵盖视觉、声学、触觉/力觉、流体感知和化学感知的具身海洋五感关键技术体系, 并以深海采矿作业为典型应用场景, 深入分析当前理论研究与技术实践中存在的瓶颈与难点。在此基础上, 凝练出三大核心科学问题: 非完备、非结构化信息下的环境统一表征与理解、任务驱动的主动感知与高效探索, 以及物理交互驱动的感知策略涌现与自适应演化。本文认为, 具身海洋环境感知将朝向与环境深度融合的物理人工智能方向发展, 亟需加快布局具身海洋环境感知基础理论与关键技术攻关, 并推动典型场景的应用示范。

关键词: 具身智能; 具身海洋环境感知; 多模态感知; 海洋机器人; 深海采矿; 人工智能(AI)

Embodied marine perception: a review

Lu Huimin^{1,2*}, Zheng Yuchao^{2,3}, Li Yujie³

1. School of Automation, Southeast University, Nanjing 210096, China; 2. Advanced Ocean Institute of Southeast University, Nantong 226000, China; 3. School of Engineering, Kyushu Institute of Technology, Fukuoka 804-8550, Japan

Abstract: Artificial intelligence (AI) is increasingly permeating marine applications, significantly driving the development of autonomous marine systems. The continuous advancement of AI pushes intelligent agents into progressively complex and challenging physical environments, emphasizing the necessity for effective environmental perception and cognitive understanding within dynamic and uncertain marine contexts. The concept of embodied intelligence provides essential insights for addressing these challenges. Embodied intelligence underscores the importance of ongoing physical interaction between agents and their environments, substantially enhancing perceptual capabilities, environmental understanding, and decision-making through continuous real-time sensing-action loops. Driven by rapid advances in computational capabilities and foundational large-scale models, embodied intelligence has become a prominent research area within terrestrial and aerial domains. Despite notable advancements in these areas, applying embodied intelligence within marine environments introduces fundamentally different and considerably greater challenges due to their inherent complexity and uncer-

收稿日期: 2025-09-08; 修回日期: 2025-12-04; 预印本日期: 2025-12-11

* 通信作者: 陆慧敏 luhuimin@ericlab.org

基金项目: 国家自然科学基金项目(62576090); 中央高校基本科研业务费专项资金资助(4008002403); 江苏省基础研究计划重点项目(SBK20251000047); 江苏省科技厅重大创新平台计划(BM2024020); 江苏省水利厅水利科技项目(2025002)

Supported by: National Natural Science Foundation of China(62576090)

tainties. Marine environments pose significant limitations to traditional perception technologies. Rapid attenuation of light severely restricts underwater visibility, substantially complicating optical sensing methods. Constraints on electromagnetic wave propagation limit wireless communication and navigation techniques underwater, making these conventional methods unreliable. Although acoustic sensing provides possibilities for long-range perception, it is challenged by considerable limitations, including low bandwidth, high latency, and complex multipath effects. Additionally, underwater intelligent agents operate within stringent energy constraints, severely limiting their capacities for sustained movement, prolonged data collection, and processing. These constraints make direct transfers of terrestrial or aerial intelligent solutions into marine environments difficult, substantially hindering the development of foundational marine perception capabilities. In response, this work introduces and systematically elaborates on the concept of embodied marine perception, which emphasizes proactive physical interaction by marine intelligent agents. Embodied marine perception integrates adaptive motion and morphological adjustments to actively acquire and fuse multimodal sensory information, encompassing visual, acoustic, haptic/force, fluid, and chemical perceptions. Such integrated multimodal sensing allows marine intelligent agents to construct comprehensive internal world models that are closely tailored to specific environmental contexts and operational tasks. For instance, underwater robots can autonomously execute optical and acoustic mapping missions, detect chemical concentration variations accurately through fluid traversal, and discern object properties through direct physical interaction. This proactive and multimodal perception approach offers effective solutions to address the inherent sparsity, uncertainty, and variability of marine environmental information. This work systematically reviews the state of embodied marine perception research internationally, highlighting key theoretical gaps and technological limitations. It defines critical scientific challenges, including achieving unified environmental representation from incomplete and unstructured sensory data, developing methods for task-oriented active perception and efficient environmental exploration, and facilitating the emergence and adaptive evolution of perception strategies through continuous physical interaction. This study proposes a comprehensive technological framework termed five embodied marine senses, integrating vision, acoustics, haptic/force perception, fluid perception, and chemical perception. By synthesizing these multimodal sensory capabilities, marine intelligent agents significantly enhance their environmental modeling accuracy, cognitive understanding, and decision-making abilities, thereby effectively addressing the complexities and uncertainties inherent in marine environments. Taking deep-sea mining operations as a representative scenario, this work elucidates the practical implications and transformative potential of embodied marine perception. Deep-sea mining operations require precise multimodal perception to navigate complex seabed terrains, accurately identify and evaluate mineral resources, and effectively manage operational risks and uncertainties. Embodied marine perception's active perception paradigm enables underwater robots to dynamically adjust their sensory strategies and operational behaviors on the basis of real-time environmental feedback, thus significantly enhancing operational efficiency, reliability, and safety. Ultimately, this study envisions embodied marine perception evolving toward physics-integrated AI, highlighting the importance of continuous physical interaction for robust, adaptive, and reliable environmental cognition in marine contexts. It advocates for accelerated foundational research, targeted technological breakthroughs, and practical demonstrations in representative marine scenarios to drive future advancements in this critical area of embodied marine perception.

Key words: embodied intelligence; embodied marine perception; multimodal perception; marine robotic; deep-sea mining; artificial intelligence(AI)

0 引言

“建设海洋强国是实现中华民族伟大复兴的重大战略任务”。党的十八大以来,习近平总书记多次强调要进一步关心海洋、认识海洋、经略海洋,指引我国海洋强国建设不断取得新成就。顺应这一时代

号召,人工智能(artificial intelligence, AI)正以前所未有的速度加速融入海洋领域,成为驱动我国海洋强国建设的重要引擎之一。人工智能的发展不断推动智能系统迈向更复杂、更具挑战性的物理环境,如何实现智能体对复杂动态环境的高效感知与认知理解,已成为当前亟待突破的根本性问题。具身智能(embodied AI)理论的兴起为解决上述问题提供了

关键性启发。具身智能强调智能体通过持续与环境的物理交互,在感知、行动中不断提升对环境的理解与决策能力。近年来,在计算能力、基础大模型的共同推动下,陆空领域的具身智能研究已成为学术界和工业界的前沿热点。例如,在交互式模拟环境与基准方面, AI2-THOR (the house of interactions) (Kolve等, 2022)、ProcTHOR(Deitke等, 2022)等大型平台的建立,为训练能执行复杂日常任务的智能体提供了坚实基础;在机器人操作方面,以 CLIPort (Shridhar等, 2022)为代表的开创性工作,展示了如何融合多模态信息以实现开放世界物体的零样本(zero-shot)灵巧操控。与此同时,面向海洋科学任务的领域大语言模型也开始出现,例如 OceanGPT (generative pre-trained Transformer)展示了大模型在海洋知识获取与推理中的具体应用(Bi等, 2024)。近期也开始出现面向水下机器人的 USIM and UO 等数据集与模型,用于支持通用水下机器人的任务学习与评测(Gu等, 2025)。类似地,水下具身智能体的评测与训练也开始出现专用基准环境,例如 OceanGym 提供了面向水下具身代理的基准环境与任务框架(Xue等, 2025)。

当前,海洋环境下的具身智能发展尚处于起步阶段。海洋感知对传统感知技术的要求极为苛刻:光线迅速衰减导致水下视觉极为受限;电磁波传播受阻严重限制了无线通信与导航方式的可行性;声学手段虽然可实现远程感知,但其传输特性存在低

带宽、高延迟以及复杂多径效应等显著局限(周玲等, 2025)。此外,水下智能体的能量供应极为有限,进一步削弱了其行动与感知能力。这些物理约束决定了陆空环境的智能方案难以直接迁移到海洋环境,进而制约了海洋智能系统基础感知能力的发展。在此背景下,如图1所示,本文提出并系统阐述具身海洋环境感知(embodied marine perception)研究方向。该方向强调通过海洋智能体主动运动与形态调整,实现对视觉、声学、触觉/力觉、流体感知与化学感知等多模态信息的主动获取与融合,构建紧密关联于环境与任务的内部世界模型,引导智能体进行合理的感知决策行为。以水下机器人为例,它通过主动游弋进行有效光学与声学测绘,通过穿越水体精确探测化学浓度变化,通过物理接触感知目标物体属性。该感知模式被视为应对海洋环境信息稀疏与不确定性的有效路径。

具身海洋环境感知的深入发展对国家具有重大的战略意义:在海洋资源开发领域,它是实现深海矿产与油气资源高效绿色开采的基础;在国家安全领域,它对于构建大范围、高效可靠的水下防御系统至关重要;在科学探索领域,将赋予海洋智能体自主探索能力,推动深海科学的重大突破。本文聚焦具身海洋环境感知领域的核心挑战,系统梳理国内外研究现状,明确关键科学问题,提出具身海洋五感关键技术体系,为我国具身海洋环境感知领域的战略布局提供系统性理论指导与技术路径支撑。

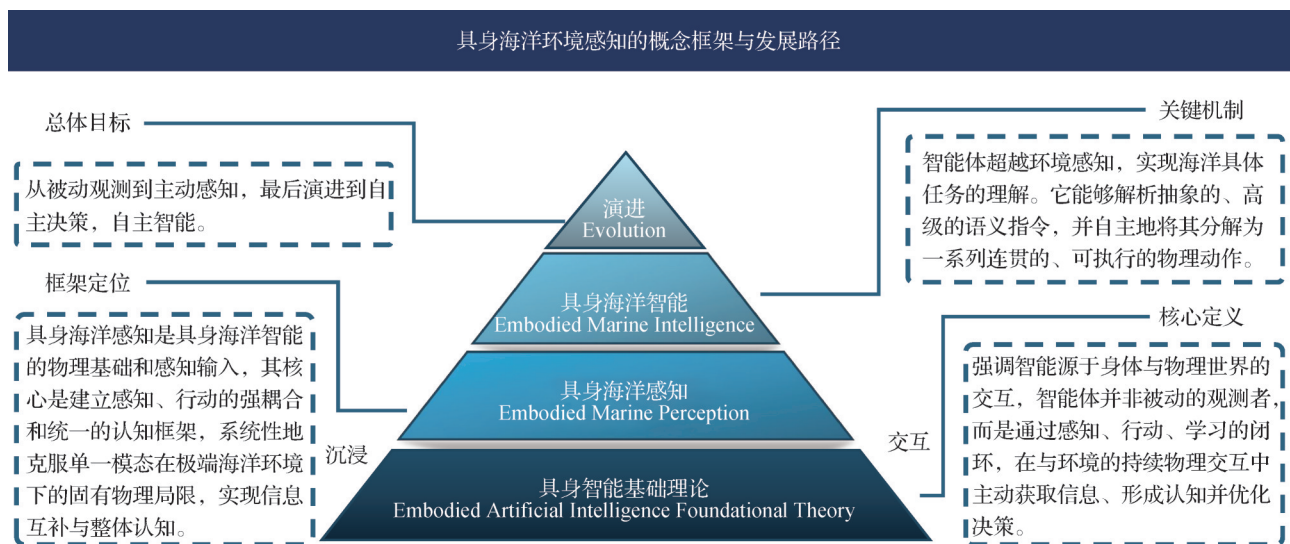


图1 具身海洋环境感知综述内容

Fig. 1 Overview of embodied marine perception

1 具身海洋环境感知的关键技术体系

1.1 具身海洋视觉感知:从看见海洋到看懂海洋

水下视觉感知正在经历从传统图像增强到面向感知任务的场景理解的转变。其核心在于克服水体复杂光学特性,为具身智能体提供高质量、结构清晰的局部三维环境认知基础。首先,水下成像退化是“看见”的根本挑战(郭威等,2025)。光在水中传播时,因波长选择性吸收和悬浮颗粒物散射,导致图像普遍存在色偏、低光照、对比度下降及细节模糊等问题(Jaffe,1990)。早期研究主要依赖物理模型,如基于暗通道先验(dark channel prior, DCP)的方法(He等,2011),通过逆推成像退化过程实现增强。然而,真实海洋环境光学参数的时空快速变化,使得精确物理复原困难,导致模型化能力和鲁棒性不足(Li等,2016)。随着深度学习发展,以卷积神经网络(convolutional neural network, CNN)和生成对抗网络(generative adversarial network, GAN)为代表的纯数据驱动方法成为主流,通过端到端训练展现出显著性能优势。然而,高质量配对水下图像数据集的稀缺性极大限制了这些方法的可训练性与泛化能力(Sun等,2021)。为克服上述局限,当前研究前沿转向物理模型与深度学习相结合的混合方法。例如,利用简化物理模型引导神经网络学习难以估计的动态参数,或将物理约束融入损失函数。物理模型引导神经网络学习的策略可以降低对大规模数据的依赖,同时提升模型的可解释性与跨场景适应能力,被认为是实现稳健、高效水下图像恢复的关键方向。图像增强与恢复技术是后续所有高级视觉任务的基石,是实现高精度感知与自主操作的基础。

在实现清晰“看见”的基础上,具身海洋环境感知更核心的挑战是“看懂”环境,即从像素信息中提取稳健的几何结构与语义信息。视觉同步定位与建图(simultaneous localization and mapping, SLAM)是实现此目标的关键技术(Durrant-Whyte和Bailey,2006)。通过处理来自单目、双目或深度相机的连续图像,视觉SLAM能实时估计智能体运动轨迹,并同步构建周围环境的高精度三维模型。然而,水下环境中的弱纹理、动态光照与视野受限等因素对经典视觉SLAM算法构成显著挑战(Mur-Artal等,2015)。因此,研究前沿集中于利用深度学习提升视觉

SLAM的鲁棒性和精度。深度学习模型能从退化图像中学习更抽象和鲁棒的特征表达,在传统方法失效的场景中保持稳定性能,使得水下智能体能在近距离内构建出精细三维环境模型,为自主导航与交互提供几何先验(庄培显等,2026)。近期也有相关工作探索视觉—语言—动作(vision language action, VLA)架构在水下场景的应用,例如面向水下自主导航的UnderwaterVLA(Wang等,2025)。

1.2 具身海洋声学感知:从听见海洋到听懂海洋

声学感知作为具身海洋智能体的核心支撑技术,凭借其在暗光与高浑浊水体中出色的穿透能力,为水下大范围、长航时自主运行提供了视觉系统无法替代的广域环境感知能力。声学的首要作用在于构建精确可靠的导航与定位基准。在电磁波迅速衰减的海洋环境中,声学传感也是实现长距离定位与测绘的最常用有效手段(Urick,1975)。该技术体系主要依赖于多普勒测速仪(Doppler velocity log, DVL)获取载体对底速度,长基线(long baseline, LBL)或超短基线(ultra short baseline, USBL)定位系统提供绝对或相对位置校正(Paull等,2014),以及利用多波束测深声呐(multibeam echosounder, MBES)进行大规模海底地形测绘(Lurton,2010)。此类传感器的信息融合构成了声学同步定位与建图(声学SLAM)的基础。相较于视觉SLAM,声学SLAM具备数公里级的有效作用距离,且不受光照与水体浑浊度影响,能够为水下智能体构建一个全局一致的稀疏地形或特征地图。然而,声学SLAM由于声学特性也面临很多挑战,包括声速剖面的时空变化、严重的多径干扰以及复杂环境下的噪声污染等问题。为应对这些问题,部分研究正从传统滤波算法向更高鲁棒性的方法改进。例如,利用CNN和点云处理网络从原始声呐数据中提取更鲁棒的地形特征,或使用图神经网络(graph neural network, GNN)优化数据关联与回环检测,提升弱特征区域或动态环境下的定位与建图精度与稳定性。

声学的第2个作用是构建多智能体交互网络。声波是目前已知唯一能在水下实现有效远程通信的媒介。水声通信技术使得分散在广阔海域中的多个水下航行器能够组成一个分布式、协同工作的感知系统。尽管水声信道固有的低带宽(通常为十几kbps级)、高时延和高误码率限制了信息传输的速率和可靠性,但它为多智能体系统的协同感知、任务分

配与数据融合提供了关键的通信链路(Sun等, 2021)。近几年的研究焦点在于如何在极端受限的信道条件下实现高效数据传输和协同作业。例如,通过开发基于深度学习的信源编码和信道编码与译码技术来压缩需交换的信息,确保在低信噪比下也能传输位置估计、目标探测结果或局部地图的紧凑表示等关键数据。海洋环境多智能体系统在受限通信条件下的协同能力,是解决复杂海洋现象观测等科学问题的关键。在协同流场估计或自适应环境采样等任务中,智能体集群可以通过水声网络交换各自的观测数据与状态信息,动态调整各自的感知策略,实现对大尺度、时变海洋环境的涌现式感知与建模。

1.3 具身海洋流体感知:从多点取样到流场重构

在传统水下导航与控制框架中,海洋流场常被视为需实时估计并补偿的外部扰动。传统的惯性导航系统(inertial navigation system, INS)与多普勒测速仪(DVL)组合,通过测量载体与水体的相对运动来滤除流场影响,从而滤除流场干扰以获得精确的对地速度信息(Fossen, 2011)。然而,在具身海洋环境感知的理论体系下,流场是蕴含丰富信息的动态环境介质。主动感知、建模并利用流场,是提升水下智能体自主性、效率与安全性的关键。具身海洋流体感知的核心在于将智能体的本体运动作为一种流场测量的传感模态。

单个智能体仅能提供其航迹上一维、稀疏的流场信息,难以形成全局认知。因此,研究前沿聚焦于多智能体的协同流场估计机制。通过水声通信网络,集群中的每个智能体共享其位置、航迹或运动积分误差等信息。这些分布于广阔空间、时间上连续的稀疏数据,形成覆盖广域空间和连续时间尺度的稀疏观测体系,构建出一个可解的数学逆问题。最新的研究采用高斯过程对流场进行非参数化建模,以更好地表征其时空相关性与不确定性;或利用GNN融合多智能体间的复杂交互,使系统能够在无需部署昂贵的流速计阵列的情况下,仅通过集群运动模式重建高分辨率二维或三维时变流场地图。同时,将物理约束显式纳入学习过程的物理信息神经网络(physics-informed neural network, PINN)也可在稀疏传感器观测下实现流场重建,为解决稀疏观测的逆问题提供了另一条数据-物理融合路径(Hosseini和Shiri, 2024)。

掌握流场动力学模型带来了多重优势。首先,最直接的应用是能源最优路径规划。考虑到水下航行器有限的板载能源,基于实时流场地图所进行的顺流航线优化可显著降低能耗,从而延长任务续航或扩大作业覆盖范围。其次,精确的流场模型能提升科学数据采集质量,例如在海洋溢油、赤潮扩散等研究中,预测物质在动态海洋环境中的分布路径。此外,流体感知还可作用于规避风险。除大尺度洋流外,海洋中还存在涡流、湍流以及海底结构物附近的局部强流等复杂流动现象,可能引发智能体失稳、偏航甚至碰撞风险。因此,通过部署高频压力传感器阵列、仿生侧线系统等局部流体传感器,实时感知近场流速梯度与微尺度涡旋结构(Yang等, 2006; Asadnia等, 2015)。将这些局部感知数据与全局流场模型相融合,可实现对恶劣海洋环境的提前预警和主动规避,增强智能体在未知复杂海域的生存能力与任务执行可靠性。因此,在具身感知体系中,流体感知赋予了智能体理解并利用环境内在动力学规律的能力,使被动适应环境的执行器转变为可主动利用并顺应流场规律的协同智能体。

1.4 具身海洋化学感知:从闻到海洋到听懂海洋

嗅觉感知使水下智能体能够探测和解析由视觉、声学等传感器无法感知的化学信息。该领域的核心科学问题在于:在动态流体环境中,如何依据高度稀疏、间歇性的化学信号,实现对释放源的有效定位与追踪(Farrell等, 2005)。相较于其他感知模态,嗅觉感知的根本挑战源于化学信号在流体介质中传输的物理特性。在湍流主导的海洋环境中,来自热液喷口、海底渗漏点或生物体的化学物质,并不会形成平滑、稳定的浓度梯度场。相反,在平流运输和湍流扩散的共同作用下,会形成时空上极不连续、形态不规则的化学羽流,其结构由一系列离散的化学斑块构成(Webster和Weissburg, 2001)。智能体搭载的化学传感器在长时间内可能探测不到任何信号,随后突然捕获一个短暂的浓度峰值得到的有或无的二元信号,使得传统的基于梯度的搜索方法常常失效。

因此,化学源定位的成功,除了依赖于传感器本身的灵敏度,也取决于智能体所采用的具身化学探索策略。由于无法直接测量梯度,智能体必须通过主动的物理移动,在环境中进行探索性采样,以间接推断源头的时空分布位置。该领域的前沿研究正从

预设的启发式算法向自适应的智能策略演进。早期研究主要借鉴生物的追踪行为,为提升搜索效率,后续研究引入了以信息熵最大化(infotaxis)为代表的信息论驱动策略,将搜索建模为贝叶斯推理过程,智能体的每一步移动都在最大化期望信息增益(Vergassola等,2007)。近期研究趋势则转向强化学习驱动的自主策略学习。通过在模拟的计算流体力学(computational fluid dynamics, CFD)环境中进行大规模训练,强化学习智能体能够自主学习出在复杂流场中进行化学源追踪的最优策略,在多智能体框架下,通过协同感知,一个海洋智能体集群可以分布式地对化学羽流进行采样,构建羽流的边界轮廓与内部结构特征图。嗅觉感知是海洋场景下发现导向型科学探索任务的核心,它在诸多关键场景中发挥着不可替代的作用,包括极地冰下湖探测、远洋海域油气资源勘探、海底热液喷口定位、污染源泄漏快速响应,以及管道微量化学泄漏监测等。

1.5 具身海洋触觉/力觉感知:与海洋目标灵巧交互

触觉与力觉感知是具身智能体与环境进行物理交互的最终环节,标志着智能体从被动环境观测向主动物理作业的转变,是实现具身智能的另一大关键要素。非接触式的声学与视觉感知主要解决导航、避障与目标识别等问题,而触觉/力觉感知则直接为灵巧操作提供关键信息(Cutkosky, 1989)。当智能体的执行器与目标物体接触时,部署在其上的传感器能提供一系列远距离感知无法获取的关键物理信息,包括接触力/力矩、几何轮廓及物体刚度等。

水下触觉传感技术正处于快速发展阶段,其前沿研究集中于在恶劣海洋环境中实现高分辨率与高鲁棒性感知能力的突破。当前技术发展主要体现在以下方面:首先是柔性、贴附式电子皮肤器件的材料革新。该领域自表皮电子(epidermal electronics)等开创性研究(Kim等,2011)起步,已取得长足进展。当前研究正逐渐突破传统的基于压电效应的刚性传感器局限,转而开发基于水凝胶、离子导电聚合物和液态金属等软材料的新型传感器(Sun等,2025)。这类传感器因其与水体声阻抗和机械柔顺性高度匹配,能更好地贴合物体表面,实现微小接触力与压力分布的高灵敏、高精度测量。其次,在仿生设计与结构创新中,研究人员从海洋生物的感知机制中汲取灵感,设计出能解耦压力、剪切力和扭矩的多维力传

感器阵列(Li等,2024)。基于视觉的触觉传感技术在水下场景中的应用探索也正日益活跃。以Gel-Sight为代表的视觉触觉传感器通过内嵌摄像系统记录弹性体表面在接触过程中的微小形变,进而重建高分辨率的三维几何结构与纹理信息(Yuan等,2017),其水下应用已成为前沿方向。在水下操作中,相关研究将视觉提供目标的初始定位,例如将触觉传感器集成于柔顺水下机械臂,并结合驱动系统实现非线性力反馈控制,以提升水下操作的稳定性与力控精度(Lin等,2020)。还有部分研究通过初始接触和主动探索,实现目标的精确位姿估计、表面轮廓重建和操作力的闭环控制(Xu等,2022)。在具身感知体系中,触觉/力觉感知闭合了感知到规划再到行动的完整回路,使智能体具有通过物理接触来改变或探索环境的能力。

具身海洋五大感知模态对比如表1所示,具身海洋环境感知的技术发展时间轴和关键技术体系分别如图2和图3所示。

表1 具身海洋五大感知模态对比

Table 1 Comparison of the embodied marine perception

感知模态	探测范围	分辨率	核心局限
海洋视觉	近	极高	浑浊、光照、距离受限
海洋声学	远	低	多径效应、低带宽、高延迟
海洋流体	局部/广域	中	需多传感器多点协同估计
海洋化学	弥散/点状	极低	信号稀疏、非连续
海洋触觉	非常近	高	需物理接触、易磨损

2 典型应用场景:深海采矿作业中的具身海洋环境感知

为阐明具身海洋环境感知的五大技术体系如何在一个统一框架下协同运作,本节以深海采矿作为典型应用场景进行剖析。深海采矿作业涵盖从广域资源勘探到精准开采与环境监测,对环境感知能力提出了贯穿始终的复杂需求,因此构成了检验具身海洋环境感知系统整体效能的一个代表性范例,如图4所示。

作业流程首先始于广域资源勘探与环境基线调查阶段,核心目标是在数千平方公里的目标海域内,识别出具有开采潜力的矿区。在此宏观尺度下,声学

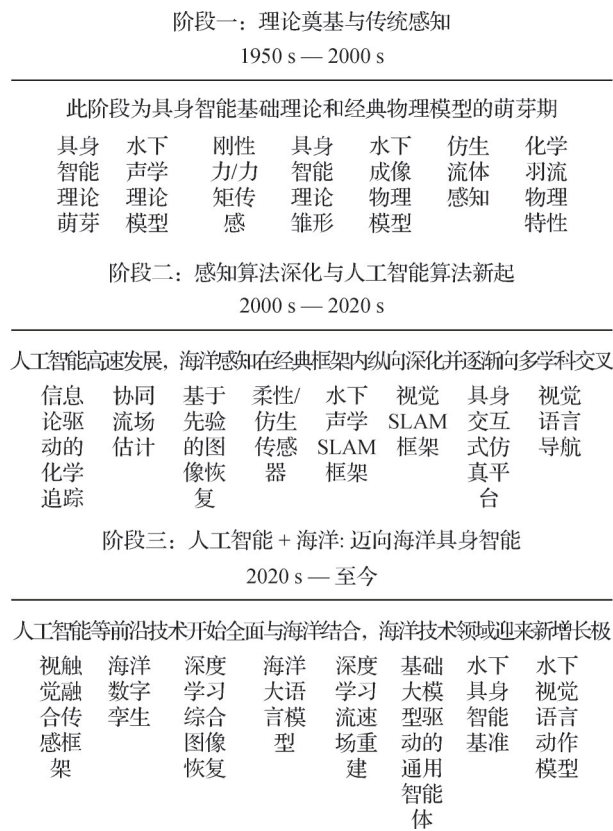


图2 具身海洋环境感知:技术发展时间轴

Fig. 2 Embodied marine perception: technology development timeline

感知承担了主要的远距离、大范围测绘任务。自主水下航行器 (autonomous underwater vehicle, AUV) 集群搭载多波束测深声呐与侧扫声呐, 对海底进行扫描,

生成高精度的地形地貌图。通过声学同步定位与建图技术, 结合水面母船的超短基线 (USBL) 系统进行全局位置校正, 形成一幅广域的资源潜力地质图。若目标为多金属硫化物矿床, 化学感知 (搭载氧化还原电位或溶解甲烷传感器) 通过追踪热液活动释放的化学羽流, 为矿点的间接定位提供关键线索。

在声学勘探圈定的重点矿区内, 作业进入高分辨率精细调查阶段, 任务目标转为 cm 级的资源储量评估与底栖生物多样性编目, 此时视觉感知成为核心模态。AUV 将航行高度降低至近海底数米, 高频率采集高清图像序列。通过高分辨率的视觉数据, 对多金属结核的覆盖率、尺寸分布与赋存状态进行大致统计, 实现可靠的资源储量估算。同时, 对这些影像资料进行底栖生物的分类与计数, 建立采矿生态环境基线, 进行后续环境影响评估。

开采作业阶段是具身感知系统从被动观测转向主动物理交互的核心环节。重型海底采矿车需在崎岖复杂的海底地貌中执行长时程连续作业。此阶段, 触觉/力觉感知成为核心感知模态。采矿车首先依赖声学与视觉感知进行导航与局部避障, 沿规划路径机动至作业点。当其铣挖头与海底接触时, 部署在铣挖头驱动与液压回路中的压力传感器可实时感知作业过程中产生的液压载荷变化, 间接表征铣挖头与结壳、沉积物之间的相互作用力, 从而使控制系统能够调节转矩与推进力, 施加近似最优接触力,

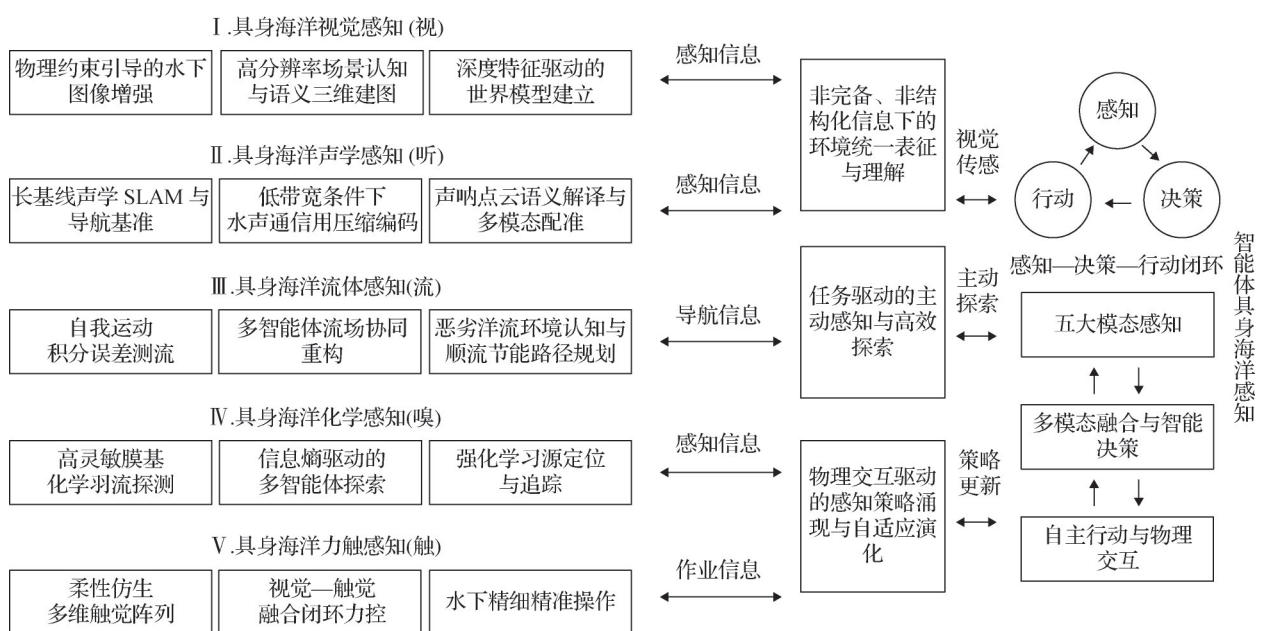


图3 具身海洋环境感知关键技术体系

Fig. 3 Key technology system of embodied marine environment perception

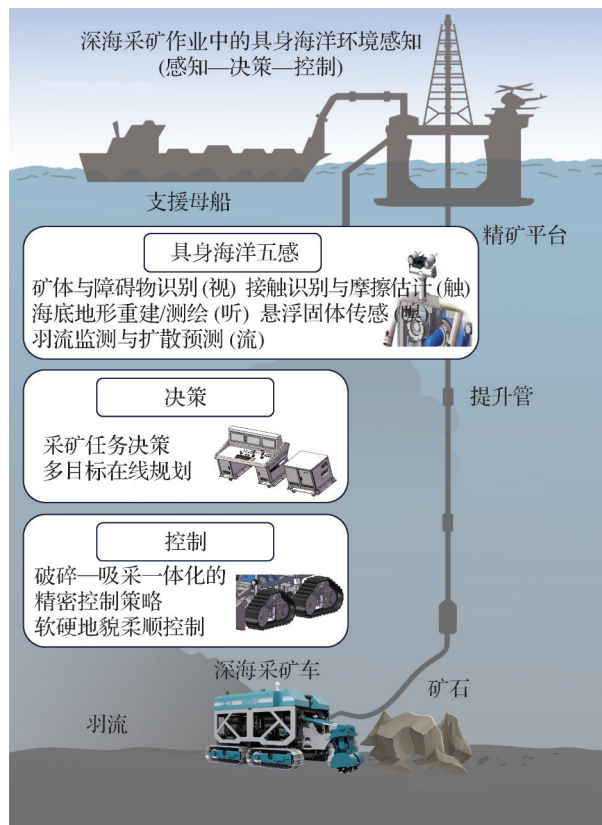


图4 深海采矿作业中的具身海洋感知

Fig. 4 Embodied marine perception in deep-sea mining

实现高效开采并降低对底层沉积物的扰动。

最后,深海采矿面临的核心环境风险之一,是采矿车在作业过程中对海底沉积物的强烈扰动所引发的大尺度悬浮泥沙羽流。该羽流具有扩散范围广、沉降时间长、空间与时间尺度高度不确定等特征,可能对底栖生态系统、滤食性生物及局部生境结构造成持续性影响,因此被认为是当前深海采矿活动中最具争议和最受监管关注的环境问题之一。在贯穿整个作业周期的环境监测与影响评估阶段,具身感知系统成为实现深海采矿环境风险可控与影响可评估的关键支撑。为此,多艘监测AUV在采矿区域协同工作,形成一个移动的分布式感知网络。通过综合运用多种感知模式:利用视觉感知中的光学后向散射传感器(optical back scatter, OBS)原位测量羽流的浊度(悬浮物浓度);利用化学感知分析羽流中可能溶出的金属离子浓度变化;同时结合流体感知获取的实时高分辨率流场数据,驱动羽流扩散的数值模型,实现对羽流运输路径与影响范围的精准追踪与预测。

综上所述,具身海洋环境感知在深海采矿的应用中,是一个动态、协同、贯穿任务执行的智能过程。依

据任务需求进行的多模态切换与深度融合,使得从资源勘探到采矿作业这一复杂的深海工程任务成为可能,也成为智能化、绿色化深海采矿的关键技术支撑。具身海洋环境感知典型应用场景如图5所示。

3 具身海洋环境感知面临的核心挑战与科学问题

回顾已有的海洋感知技术路线,发展具身海洋环境感知的首要任务是正视其所处的极端环境条件。深海中的高压、腐蚀以及剧烈温度梯度等物理因素,对装备的结构强度、密封与材料耐久性提出了极为严苛的挑战;而在信息层面,洋流、浑浊度及海洋生物活动等共同塑造了一个高度非结构化、迅速变化的开放世界,使感知面临显著不确定性。同时,受能源与通信带宽限制,水下智能体必须在极低功耗和有限协同的条件下自主运行。在此情形下,智能体需要在信息稀疏、残缺且噪声显著的环境中,实时构建对复杂海域的可靠认知图景,并做出高效的决策。

然而,与陆地及空中机器人系统相比,海洋环境在信息获取、传输与解析方面具有更高难度,具身海洋环境感知的发展仍处于初期阶段。当前技术尚无法充分满足实际应用中对环境理解的鲁棒性、资源利用的高效性与长期自主适应能力的多重要求,尤其是在信息极端稀疏、动态多变和资源受限的海洋环境下,核心技术瓶颈亟待实质性突破。

3.1 具身海洋环境感知面临的主要挑战

具身海洋环境感知的发展受到海洋环境本身所固有的物理特性的深层制约,构成海洋智能体实现自主感知的系统性难题。具体包括以下3个方面:

首先,极端环境下的信息获取与表征是具身海洋环境感知的根本挑战。水体对光线的选择性吸收与散射显著削弱了视觉感知能力,光学图像清晰度与颜色信息锐减,严重影响了视觉导航与目标识别等核心算法的可靠性。声学感知虽可远程传输,但因传播速度慢、多径效应明显、信号分辨率低等问题,导致水下定位与通信的准确性明显下降。此外,触觉/力觉传感和化学场感知也受到海洋高压、腐蚀环境以及复杂湍流动力学过程的制约,造成了信息稀疏、事件驱动且难以持续稳定获取的困境。因此,海洋智能体必须建立多模态稀疏与异质信息的融合

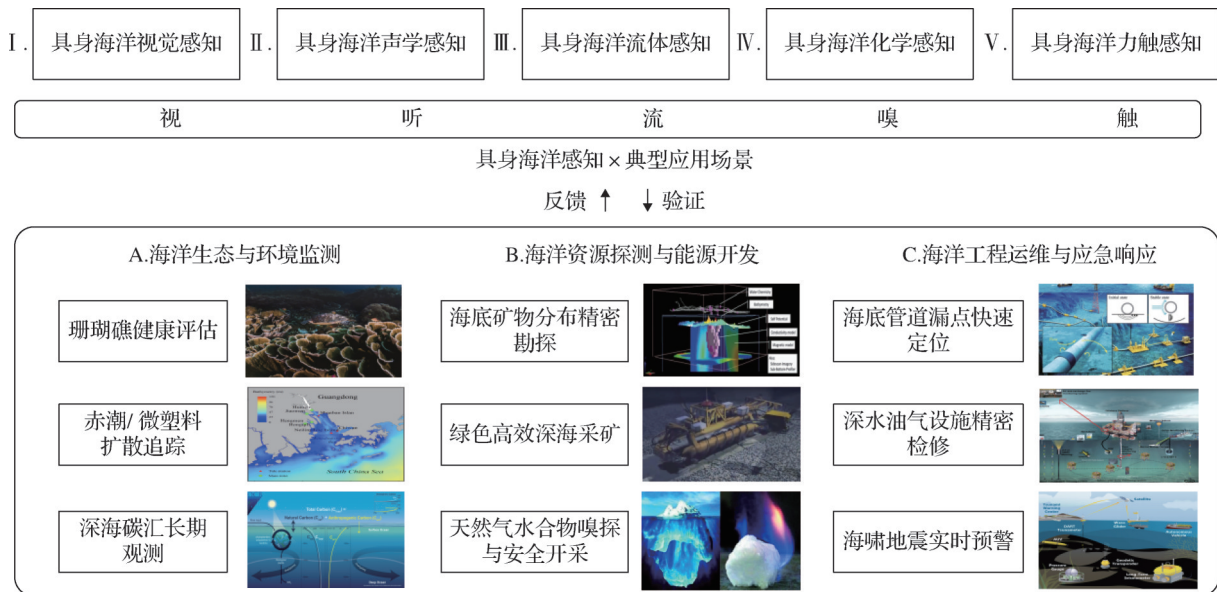


图5 具身海洋环境感知典型应用场景

Fig. 5 Typical application scenarios of embodied marine perception

机制,构建对环境的整体认知。

其次,资源强约束下的主动感知与决策是另一个关键挑战。海洋智能体受限于能源供应、计算资源以及通信带宽等约束,使其感知与决策策略面临高度压缩的选择空间。智能体需要具备主动的信息获取与寻求策略,在保证感知效果的同时尽可能降低能耗开销。此外,低带宽高延迟的水声通信进一步限制了多智能体之间高效的信息交换,促使智能体集群倾向于采用去中心化的协同机制,通过本地感知信息的高效整合实现面向全局的自组织行为。

最后,开放动态环境下的长期自主与适应能力是对海洋智能体鲁棒性的根本考验。长期部署在海洋环境中的智能体不仅面临传感器性能随时间退化的挑战,而且还需应对洋流变化、海底地貌动态演变及水质变化等非平稳环境因素的持续干扰。这些动态特性要求水下智能体具备在线学习、自适应更新内部模型以及实时传感器自标定的能力,确保长期任务执行的稳定性与感知可靠性。

3.2 具身海洋环境感知面临的科学问题

具身海洋环境感知面临的主要挑战可进一步凝练为3个核心科学问题:非完备、非结构化信息下的环境统一表征与理解,任务驱动的主动感知与高效探索,以及物理交互驱动的感知策略涌现与自适应演化。

1)非完备、非结构化信息下的环境统一表征与

理解。智能体所依赖的视觉、声学、触觉/力觉、流体和化学感知模态信息,在物理特性、时空尺度及数据形态上存在明显差异。视觉信息具有高分辨率,但受限于传播距离及水体浑浊度,在水下环境中易失真且感知范围受限;声学信息具有远程传播能力,但面临分辨率低、噪声高和复杂多径效应等问题;触觉/力觉信息在高价值信息提取上有显著优势,但其事件驱动特征导致信息稀疏;流体和化学场信息分布弥散、动态变化强烈,受到复杂流体动力过程影响而表现出高度不确定性。针对上述问题,第1个核心科学问题是:如何有效融合这些异构、稀疏、不确定性强且存在时空错位的信息,建立统一、多尺度且具备明确语义的环境认知模型?这一问题要求精密的数据融合模型,更要求建立容错缺失、时空校正与不确定性量化的系统认知框架。框架需涵盖几何感知、语义识别、定位建图及动态预测,并具备任务相关性与情境自适应能力。例如,在广域搜索场景下以声学模态主导,而精细作业场景则需动态切换为视觉与触觉主导。解决该问题的目标,是实现从零散信息到完整环境理解的转化,为智能体的任务决策提供可靠的认知基础。

2)任务驱动的主动感知与高效探索。海洋环境的开放性与高度不确定性,使基于预设路径与脚本的感知策略难以胜任复杂任务场景,智能体亟需具备主动感知与自适应探索的能力。这引出了第2个

核心科学问题:在能源与通信高度受限的约束条件下,智能体如何通过持续的主动探索与动态决策,发展主动感知与信息获取策略,以优化任务效率与感知质量?这一问题强调智能体在任务执行过程中能够自主优化感知路径和信息获取模式,实现与环境交互的动态适应与高效探索。例如,智能体追踪水下化学羽流时,需要结合间歇性、碎片化的化学信息及局部流场特征,动态优化搜索路径。该类策略可在感知质量与能耗之间实现优化平衡,提升智能体在未知海域中的持续自主探索能力。

3)物理交互驱动的感知策略涌现与自适应演化。在具备全域感知支撑的异构海洋智能体集群中,多样化的任务、庞大的智能体数量与异质性显著增加了群体决策和信息交互的复杂性,亟需构建高效且鲁棒的协同决策与反馈机制。因此,第3个核心科学问题聚焦如何在智能体个体—分组—集群三级结构中,通过物理交互与感知驱动,实现感知策略与协同行为的自主涌现与演化。这一问题涉及全局态势信息的高效分发机制和局部感知信息的汇聚方式,目标在于揭示异构的海洋智能体集群在环境不确定性与资源受限条件下形成集体智能的涌现规律。尤其是在水声通信带宽极为有限的条件下,智能体无法频繁、密集地交换信息,必须依赖环境感知驱动的去中心化协同方式展开群体行为建构。例如,通过交换高度凝练的局部状态信息,实现智能体之间的高效协同定位与全局环境感知。

4 未来发展方向:迈向与环境共融的物理人工智能

具身海洋环境感知的终极目标是实现智能体与物理环境的深度共融,逐步演化为更高级的物理海洋人工智能。物理海洋人工智能的核心理念在于智能体的海洋感知、控制与学习能力同其物理形态以及海洋物理环境的共同演化。针对海洋这一极端复杂的环境,具身海洋环境感知技术的发展是通向物理海洋人工智能的关键途径。具体而言,未来的发展将集中于以下几个前沿方向:

1)感知—形态—控制一体的仿生海洋智能。未来的海洋机器人将不再是搭载传统传感器的刚性平台,而将从设计之初便融入仿生学理念,将感知能力深度嵌入到机器人的物理结构与形态之中。类似于

鱼类的侧线系统对流场的敏感性以及章鱼皮肤对环境纹理的感知能力,仿生软体机器人将通过自身柔性结构的形变以及材料电学或光学特性的变化来实时感知环境状态。西北工业大学潘光教授团队开发的仿蝠鲼机器人,外表面整体可能成为一个高度集成化的触觉、流体与化学感知网络。该类形态即感知的设计,除了显著提升智能体在复杂环境中的感知交互能力,也使其能够以更低功耗、更少干扰、更高隐蔽的方式融入海洋生态系统。

2)基于物理海洋约束的具身感知学习。面对数据稀缺的深海环境,赋予智能体自主学习和泛化能力成为重要的发展方向。未来的具身感知系统将减少对人工标注数据的依赖,而更多利用海洋环境固有的物理规律(流体动力学、声学传播特性及化学扩散原理)进行自监督学习。智能体通过不断将自身的多模态感知结果与物理模型的理论预测进行对比和纠正,动态修正感知模型与环境认知。基于物理规律约束的具身感知学习使智能体在与环境持续交互中逐步理解并内化物理规律,提高其在复杂海洋条件下的感知和预测鲁棒性。

3)能源驱动的持久化海洋具身感知。实现长期、持续性的海洋观测是未来海洋科学与技术的关键目标,而能源供应的限制是实现该目标的主要瓶颈之一。未来的具身感知系统将与环境能源采集技术实现深度融合。智能体能够主动感知并利用环境能源场实现能量补充。能源感知能力可以使智能体摆脱对传统能源供应的依赖,形成覆盖全球海洋的持续、长期观测网络。

5 结 语

具身感知是智能系统发展的核心,具身海洋环境感知是该理念在复杂、动态的海洋环境中实现落地应用的必然演进路径。它决定了未来海洋探索深度、资源开发效率,对国家海洋安全和科技自主权具有战略意义。本研究综述系统地论述了具身海洋环境感知。其核心思想是,智能体需通过视觉、声学、触觉/力觉、流体、化学五大模态的有机协同与深度融合,构建对高度非结构化海洋环境的整体认知模型,并以此支撑与环境的主动物理交互与任务执行。本综述提出物理海洋人工智能,构建具有自主学习、自我进化能力的下一代海洋具身智能体。当前,具

身海洋环境感知作为一个整体性概念,其基础理论和关键技术体系尚处于发展的早期阶段。随着人工智能、新材料等技术的不断突破,系统性地开展具身海洋环境感知研究的条件日趋成熟。需加快布局具身海洋环境感知基础理论研究与关键技术攻关,并推动典型场景的应用示范。

参考文献 (References)

- Asadnia M, Kottapalli A G P, Miao J M, Warkiani M E and Triantafyllou M S. 2015. Artificial fish skin of self-powered micro-electromechanical systems hair cells for sensing hydrodynamic flow phenomena. *Journal of the Royal Society Interface*, 12(111): #20150322 [DOI: 10.1098/rsif.2015.0322]
- Bi Z, Zhang N Y, Xue Y D, Ou Y X, Ji D Z, Zheng G Z, et al. 2024. OceanGPT: a large language model for ocean science tasks//Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics. Bangkok, Thailand: Association for Computational Linguistics: 3357-3372 [DOI: 10.18653/v1/2024.acl-long.184]
- Cutkosky M R. 1989. On grasp choice, grasp models, and the design of hands for manufacturing tasks. *IEEE Transactions on Robotics and Automation*, 5(3): 269-279 [DOI: 10.1109/70.34763]
- Deitke M, Vander Bilt E, Herrasti A, Weihs L, Salvador J, Ehsani K, et al. 2022. ProcTHOR: large-scale embodied AI using procedural generation//Proceedings of the 36th International Conference on Neural Information Processing Systems. New Orleans, USA: Curran Associates Inc.: #433
- Durrant-Whyte H and Bailey T. 2006. Simultaneous localization and mapping: part I. *IEEE Robotics and Automation Magazine*, 13(2): 99-110 [DOI: 10.1109/MRA.2006.1638022]
- Farrell J A, Pang S and Li W. 2005. Chemical plume tracing via an autonomous underwater vehicle. *IEEE Journal of Oceanic Engineering*, 30(2): 428-442 [DOI: 10.1109/JOE.2004.838066]
- Fossen T I. 2011. *Handbook of Marine Craft Hydrodynamics and Motion Control*. Chichester, UK: Wiley
- Gu J W, Wu Z H, Si P X, Qiu S, Feng Y K, Sun L Y, et al. 2025. USIM and U0: a vision-language-action dataset and model for general underwater robots [EB/OL]. [2025-12-26]. <https://arxiv.org/pdf/2510.07869.pdf>
- Guo W, Hua X, Li D N, Cui X P and Deng L. 2025. Research progress on polarization image restoration technology for complex underwater scenes [J/OL]. *Journal of Image and Graphics*, 1-21 (2025-12-05). <https://www.cjig.cn/zh/article/doi/10.11834/jig.250368/> (郭威, 华夏, 李德南, 崔小鹏, 邓璐. 2025. 复杂水下场景偏振图像恢复技术研究进展[J/OL]. *中国图象图形学报*, 1-21 (2025-12-05). <https://www.cjig.cn/zh/article/doi/10.11834/jig.250368/> [DOI: 10.11834/jig.250368])
- He K M, Sun J and Tang X O. 2011. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12): 2341-2353 [DOI: 10.1109/TPAMI.2010.168]
- Hosseini M Y and Shiri Y. 2024. Flow field reconstruction from sparse sensor measurements with physics-informed neural networks. *Physics of Fluids*, 36(7): #073606 [DOI: 10.1063/5.0211680]
- Jaffe J S. 1990. Computer modeling and the design of optimal underwater imaging systems. *IEEE Journal of Oceanic Engineering*, 15(2): 101-111 [DOI: 10.1109/48.50695]
- Kim D H, Lu N S, Ma R, Kim Y S, Kim R H, Wang S D, et al. 2011. Epidermal electronics. *Science*, 333(6044): 838-843 [DOI: 10.1126/science.1206157]
- Kolve E, Mottaghi R, Han W, VanderBilt E, Weihs L, Herrasti A, et al. 2022. AI2-THOR: an interactive 3D environment for visual AI [EB/OL]. [2025-12-26]. <https://arxiv.org/pdf/1712.05474.pdf>
- Li C Y, Guo J C, Cong R M, Pang Y W and Wang B. 2016. Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior. *IEEE Transactions on Image Processing*, 25(12): 5664-5677 [DOI: 10.1109/TIP.2016.2612882]
- Li Y Z, Liu B, Xu P, Liu J H, Dai X R, Yu A Q, et al. 2024. A palm-like 3D tactile sensor based on liquid-metal triboelectric nanogenerator for underwater robot gripper. *Nano Research*, 17(11): 10008-10016 [DOI: 10.1007/s12274-024-6903-3]
- Lin M H, Vatani M, Choi J W, Dilibal S and Engeberg E D. 2020. Compliant underwater manipulator with integrated tactile sensor for nonlinear force feedback control of an SMA actuation system. *Sensors and Actuators A: Physical*, 315: #112221 [DOI: 10.1016/j.sna.2020.112221]
- Lurton X. 2010. *An Introduction to Underwater Acoustics: Principles and Applications*. 2nd ed. Berlin, Germany: Springer
- Mur-Artal R, Montiel J M M and Tardós J D. 2015. ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE Transactions on Robotics*, 31(5): 1147-1163 [DOI: 10.1109/TRO.2015.2463671]
- Paull L, Saeedi S, Seto M and Li H. 2014. AUV navigation and localization: a review. *IEEE Journal of Oceanic Engineering*, 39(1): 131-149 [DOI: 10.1109/JOE.2013.2278891]
- Shridhar M, Manuelli L and Fox D. 2022. CLIPort: what and where pathways for robotic manipulation//Proceedings of the 5th Conference on Robot Learning. London, UK: PMLR: 894-906
- Sun J, Zhang Q F, Lu Y, Huang B D and Li Q. 2025. A review of touching-based underwater robotic perception and manipulation. *Machines*, 13(1): #41 [DOI: 10.3390/machines13010041]
- Sun K, Cui W C and Chen C. 2021. Review of underwater sensing technologies and applications. *Sensors*, 21(23): #7849 [DOI: 10.3390/s21237849]
- Urick R J. 1975. *Principles of Underwater Sound*. 2nd ed. New York, USA: McGraw-Hill

- Vergassola M, Villermaux E and Shraiman B I. 2007. 'Infotaxis' as a strategy for searching without gradients. *Nature*, 445(7126): 406-409 [DOI: 10.1038/nature05464]
- Wang Z Y, Zhu Y P, Yan Y Q, Tian X Y, Shao X H, Li M X, et al. 2025. UnderwaterVLA: dual-brain vision-language-action architecture for autonomous underwater navigation [EB/OL]. [2025-12-26]. <https://arxiv.org/pdf/2509.22441.pdf>
- Webster D R and Weissburg M J. 2001. Chemosensory guidance cues in a turbulent chemical odor plume. *Limnology and Oceanography*, 46(5): 1034-1047 [DOI: 10.4319/lo.2001.46.5.1034]
- Xu P, Liu J H, Liu X Y, Wang X Y, Zheng J X, Wang S Y, et al. 2022. A bio-inspired and self-powered triboelectric tactile sensor for underwater vehicle perception. *NPJ Flexible Electronics*, 6(1): #25 [DOI: 10.1038/s41528-022-00160-0]
- Xue Y D, Mao M J, Ru X Y, Zhu Y Q, Ren B C, Qiao S F, et al. 2025. OceanGym: a benchmark environment for underwater embodied agents [EB/OL]. [2025-12-26]. <https://arxiv.org/pdf/2509.26536.pdf>
- Yang Y C, Chen J, Engel J, Pandya S, Chen N N, Tucker C, et al. 2006. Distant touch hydrodynamic imaging with an artificial lateral line. *Proceedings of the National Academy of Sciences of the United States of America*, 103(50): 18891-18895 [DOI: 10.1073/pnas.0609274103]
- Yuan W Z, Dong S Y and Adelson E H. 2017. GelSight: high-resolution robot tactile sensors for estimating geometry and force. *Sensors*, 17(12): #2762 [DOI: 10.3390/s17122762]
- 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 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]

作者简介

陆慧敏,男,教授,主要研究方向为深海采矿、机器人、人工智能和海洋观测网。E-mail:luhuimin@ericlab.org

郑禹超,男,博士研究生,主要研究方向为复杂场景水下成像系统、水声通信用压缩编解码。E-mail:zyc4718@outlook.com

李玉洁,女,助理教授,主要研究方向为水下视频超分和 underwater image enhancement。E-mail:zyjli@gmail.com