

[引用格式] 焦文沛, 李杰, 张春燕, 等. 声呐图像智能感知算法综述 [J]. 水下无人系统学报, 2025, 33(3): 559-572.

声呐图像智能感知算法综述

焦文沛¹, 李杰¹, 张春燕¹, 谢广明², 肖文栋³, 张建磊^{1*}

(1. 南开大学 人工智能学院, 天津, 300350; 2. 北京大学 工学院, 北京, 100071; 3. 北京科技大学 自动化学院, 北京, 100083)

摘要: 声呐图像智能感知算法在海洋探测与水下救援中具有至关重要的作用。近年来, 深度学习技术在声呐图像智能感知任务中取得了显著进展。文中对该领域进行了全面梳理, 从声呐图像数据集与数据增强技术、经典的声呐图像处理算法以及基于深度学习的声呐图像处理方法 3 个方面进行探讨。首先, 归纳了不同任务的开源数据集与常用的数据增强技术, 为后续研究提供数据支撑; 其次, 系统回顾了从经典算法到基于深度学习的先进算法在不同任务中的应用与发展现状, 旨在为研究者提供系统的领域概览; 最后, 展望了未来的研究方向, 指出可通过获取更大规模的声呐图像数据、设计更强健的算法以及开发更适用于真实水下场景的任务设置等方面, 进一步提升声呐图像的解译能力。

关键词: 声呐图像; 智能感知; 海洋探测; 深度学习

中图分类号: TJ630.34; U666.74

文献标识码: R

文章编号: 2096-3920(2025)03-0559-14

DOI: [10.11993/j.issn.2096-3920.2024-0127](https://doi.org/10.11993/j.issn.2096-3920.2024-0127)

Intelligent Perception Algorithms for Sonar Images: A Review

JIAO Wenpei¹, LI Jie¹, ZHANG Chunyan¹, XIE Guangming², XIAO Wendong³, ZHANG Jianlei^{1*}

(1. College of Artificial Intelligence, Nankai University, Tianjin 300350, China; 2. College of Engineering, Peking University, Beijing 100071, China; 3. School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China)

Abstract: Intelligent perception algorithms for sonar images are vital in ocean exploration and underwater rescue. In recent years, deep learning has achieved remarkable advancements in intelligent perception tasks related to sonar images. This paper provided a comprehensive review of the field, focusing on sonar image datasets and data augmentation techniques, classical sonar image processing algorithms, and deep learning-based sonar image processing methods. By summarizing open-source datasets and commonly used data augmentation techniques, the paper provided data support for future research efforts. Additionally, this paper systematically analyzed the application and evolution of both classical and deep learning-based algorithms across various tasks, offering researchers an overview of the current state of the field. Finally, based on the research progress, the paper predicted future research directions. It was pointed out that the interpretation ability of sonar images could be further improved by obtaining larger-scale sonar image data, designing more robust algorithms, and developing task settings that are more suitable for real-world underwater environments.

Keywords: sonar image; intelligent perception; ocean exploration; deep learning

收稿日期: 2024-08-09; 修回日期: 2024-09-29; 录用日期: 2024-10-21.

基金项目: 国家自然科学基金项目(62473211, 62073174, 62073175).

作者简介: 焦文沛(2000-), 男, 在读硕士, 主要研究方向为模式识别及声呐图像处理.

* 通信作者简介: 张建磊(1981-), 男, 博士, 教授, 主要研究方向为深度学习、强化学习及声呐图像处理.

OPEN ACCESS

0 引言

在海洋探测中,海底成像是一种常用的探测手段。常用水下成像方法包括光学成像和声学成像。光学成像受海底光线和海水浑浊度的影响,在水下环境中能见度很差,难以应用于实际探测。而相比于光学成像,声呐成像可以显示浑浊水中的目标,并有较长的探测距离,因而在海上油气勘探、海洋科学、水下搜救和海底管道铺设等许多光信号无法有效传播的场景中广泛应用^[1-3]。

对于长时间的水下作业活动,通常需要声呐操作员持续监控屏幕寻找目标,这样不仅耗费体力、成本高昂,且容易出错^[1]。自动声呐图像感知方法能够将声呐操作员从繁琐费时的目标寻找任务中解脱出来,帮助其安全高效地探索海底世界^[1]。然而由于水下环境十分复杂,经常存在和目标相似的非目标,导致自动目标识别(automatic target recognition, ATR)系统经常出现虚警;有时目标被遮挡、掩埋难以分辨,从而导致 ATR 系统漏警^[2];同时,ATR 系统本身还存在成像噪声较高、图像颜色单一、边缘模糊及图像特征稀疏等现象^[4-6],这些原因给水下声呐图像自动感知任务增加了挑战^[1, 7-8]。

针对这些问题,许多声呐图像自动目标感知方法被提出。早期研究中,研究者们将一些传统的分类算法用于声呐图像的分类,例如决策树^[9]、马尔科夫随机场^[10]和稀疏表示法^[11]等方法。近年来,越来越多的研究者使用深度卷积神经网络(convolutional neural network, CNN)来解决声呐图像自动目标感知问题,取得了显著成果^[12-15]。实验证明,基于深度学习的方法性能优于传统方法^[16]。

不论是传统目标感知算法还是近年来大放异彩的深度学习方法,声呐图像智能感知方法的整体流程是相似的,包括原始图像预处理、特征提取及分类、目标检测或语义分割。图 1 展示了声呐图像目标分类方法的整体流程。

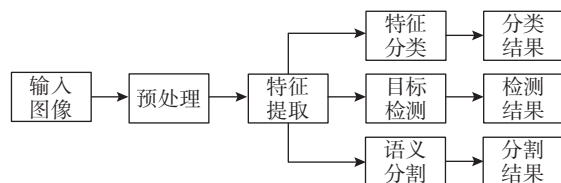


图 1 声呐图像智能感知整体流程

Fig. 1 The whole process of sonar image intelligent perception

文中首先整理和分析了国内外开源的声呐图像分类、检测和分割任务数据集,为声呐图像智能感知算法研究提供便利;其次,对声呐图像智能感知算法中采用的各种机器学习方法进行了全面回顾和分析;此外,还对声呐图像领域的分类、目标检测和语义分割任务所使用的深度学习方法进行了对比分析;最后提出了未来的潜在应用研究方向。

1 数据集与数据增强方法

1.1 数据集

1.1.1 分类任务

公开的声呐图像分类数据集是推动该领域研究进展的基础。现有公开声呐图像分类数据集如表 1 所示。

表 1 现有声呐图像分类数据集信息汇总表

Table 1 Summary of existing sonar image classification datasets information

数据集	类别	数量	总数	种类	不平衡因子
Seabed Objects-KLSG	飞机	62	447	侧扫	6.21
	沉船	385			
FLSMDD	瓶子	449	2 364	前视	6.91
	罐子	367			
	链条	226			
	饮料瓶	349			
	钩子	133			
	螺旋桨	137			
	洗发水瓶	99			
	直立状瓶子	65			
	轮胎	331			
	阀门	208			
NKSID	大螺旋桨	203	2 617	前视	47.55
	圆柱体	288			
	渔网	20			
	浮球	951			
	钢管	112			
	小螺旋桨	94			
	软管	115			
	轮胎	834			

Seabed Objects-KLSG(海底物体数据集)^[1]是收集真实侧扫声呐图像的数据集,目前公开了其中 2 个类别的图像,包含 385 张沉船和 62 张飞机,这些图像是从原始的大幅侧扫声呐图像中裁剪得

到的。前视声呐海洋垃圾数据集(forward-looking sonar marine debris dataset, FLSMDD)^[17-18]在水箱中采集得到, 包含 10 个类别共 2 364 张图片, 内容是一些常见的海洋垃圾, 如瓶子、罐子、钩子及螺旋桨等。南开声呐图像数据集(Nankai university sonar image dataset, NKSID)^[19]在渤海湾(39°N 118°E)采集得到, 由携带多波束前视声呐(具体型号为 Oculus M750d)的设备进行水下数据采集。为了提高数据的丰富程度, 采集时对每个目标都进行了不同视角、量程(2~15 m)和频率(750 kHz, 1.2 MHz)的拍摄, 经筛选、预处理和标注后, 共获得了 8 个类别共 2 617 张图片, NKSID 呈现自然的长尾分布。

1.1.2 目标检测任务

声呐图像目标检测领域也有一些公开数据集, 如表 2 所示。

表 2 现有声呐图像目标检测数据集信息汇总表
Table 2 Summary of existing sonar image target detection datasets information

数据集	类别	数量	总数	种类
FLS-DD	飞机	1 105	3 752	前视
	沉船	884		
	溺水者	1 363		
SCTD	飞机	57	357	侧扫
	沉船	34		
	溺水者	266		
UATD	立方体	2 987	9 200	前视
	球体	3 463		
	圆柱体	657		
	溺水者	1 434		
	轮胎	1 368		
	圆形网箱	860		
	方形网箱	1 318		
	金属桶	487		
	飞机	1 065		
	水下航行器	1 000		

前瞻性声呐检测数据集(forward-looking sonar debris dataset, FLS-DD)^[20]在大型定制实验水箱中采集飞机和沉船模型数据, 并在标准游泳池中进行了模拟人类溺水场景的数据采集。原始实验数据经解析与预处理后转换为二维灰度图像储存, 经过进一步的整理和筛选, 将一些重复帧和相似帧去除后获得了 3 个类别共 3 752 张原始图像, 随后使用 CVAT(computer vision annotation tool)标注工具完成目标检测标注。

声呐通用目标检测数据集(sonar common target detection dataset, SCTD)^[21]公开了 357 张侧扫声呐图像, 其中包含 57 张飞机、34 张沉船和 266 张溺水者, 平均每张图像包含 1.2 个目标。

水下声学目标检测数据集(underwater acoustic target detection dataset, UATD)^[22]使用 Tritech Gemini 1200ik 声呐进行数据采集, 共包含 9 000 多张 BMP 格式的多波束前视声呐(multibeam forward-looking sonar, MFLS)图像。该数据集提供了声呐图像的原始数据, 并标注了 10 类目标物(立方体、圆柱体及轮胎等), 标注文件为 XML 格式, 且与图像文件一一对应, 数据采集于湖泊和浅水区。

1.1.3 语义分割任务

由于标注难度较大, 相比于分类和目标检测任务, 目前公开的声呐图像语义分割数据集更为稀缺。

FLS-DD 数据集^[23]采用 Oculus M1200d 型前视声呐采集, 在可控的实验水箱和游泳池 2 种实验场地分别开展实验, 在数据采集过程中, Oculus M1200d 声呐被安装在 ROBO-ROV 水下机器人下方。水下机器人被远程操控以大约 0.3 m/s 的速度在目标周围多角度探测, 以便捕捉场景中目标的不同视图。此外, 目标的位置和姿态在有效探测范围内被人为多次调整, 以便引入可变性, 避免数据单一和重复。目前已经公开了 400 张前视声呐语义分割数据, 包含飞机和沉船模型 2 个类别的前视声呐图像和对应的分割掩膜。

FLSMDD 数据集^[17]使用 ARIS Explorer 3000 型号前视声呐采集了 11 个类别的目标物, 采集时声呐工作频率 3 MHz。数据集在水箱中采集得到, 目标物为常见的海洋垃圾, 包括瓶子、罐子、链条、饮料瓶、钩子、螺旋桨、洗发水瓶、直立状瓶子、轮胎和阀门。该数据集使用 LabelMe 标注工具生成像素级语义标签。

1.2 数据增强方法

早期识别方法针对声呐图像的高噪声和单通道等特点设计了数据增强的方案, 以提升识别模型的性能。例如: 文献 [16] 使用直方图均衡化调整图像亮度分布, 从而增强图像中物体的可见度; 文献 [7] 为了使迁移学习中源域的训练样本和目标域的声呐图像更接近, 对源域训练样本进行了预处理, 包括移除源域目标细节特征和添加噪声; 文献 [1] 提

出一种半合成数据集生成方法,以扩充数据集中数目较少的类别的样本数,该方法利用光学图像提取目标(如飞机)的主要形状,根据参考声呐图像的概率分布函数匹配来模拟目标和阴影区域的强度分布,并从参考图像中随机选择背景作为半合成图像的背景,将背景和进行强度模拟后的目标及阴影部分组合,得到半合成的声呐图像,有效提升了分类性能。

文献 [2] 提出问题:以往手工设计的数据增强方法是否真的有利于提高分类性能,以及能否放弃选择增强算法的繁重任务,让网络以无监督的方式学习图像增强变换。该研究提出了一种可学习的增强网络来自动学习图像增强变换,以得到对分类网络来说更好的增强效果,实验验证了该方法对性能提升的有效性。

由于声呐图像的小样本特性,在训练过程中对训练数据进行数据增强是基于深度学习的分类方法中的常用做法,通常包括对最小批次的训练样本进行随机水平翻转、90°范围内的随机旋转及裁剪等操作。例如文献 [24] 对声呐图像进行水平翻转、旋转、随机裁剪和拼接来扩充数据集;文献 [25] 和 [19] 对所有的对比方法进行随机水平翻转和 90°范围内的随机旋转,以提高模型的泛化能力,进而对声呐图像数据集进行数据扩增,扩大数据集规模,防止网络过拟合。

2 经典方法

2.1 声呐图像分类任务

在声呐图像目标分类任务中,也有将几种算法相结合的方法,例如将经典的主成分分析法(principal component analysis, PCA)和核极限学习机(kernel extreme learning machine, KELM)相结合^[26],将局部二值模式(local binary pattern, LBP)、方向梯度直方图(histogram of oriented gradient, HOG)和支持向量机(support vector machine, SVM)相结合^[16]。具体来说,文献 [11] 提出了一种基于稀疏表示法的侧扫声呐目标分类方法,利用与每类训练样本最接近的低维子空间来提供辨别能力;文献 [27] 提出了一种基于阴影特征可靠性融合的水下侧扫声呐图像鲁棒目标分类方法,实验验证了该方法在不同数据集上的有效性;文献 [28] 提出了 2 个用

于分类的框架,分别基于单视图内核分类器的 DS 证据理论(Dempster-Shafer's theory, DST)融合概念和多实例分类器概念,并且考虑了数据集不平衡的问题;文献 [26] 先使用 PCA 对声呐图片进行降维,然后利用 KELM 对低维数据进行分类。然而上述方法通常存在较为费时、效率低且难以在不同数据上有效泛化等不足。

2.2 声呐图像目标检测任务

声呐图像目标检测的传统方法主要包括基于模板匹配、基于局部特征符以及基于传统机器学习的方法。

2.2.1 模板匹配方法

基于模板匹配的方法着重于将待识别的声呐图像与一系列预定义模板之间的相似度进行比较,这些模板代表了不同类别的典型特征,通过选取与目标图像相似度最高的模板类别实现对声呐图像的分类。文献 [29] 提出了一种创新的模板匹配技术,专门用于合成孔径声呐(synthetic aperture sonar, SAS)图像中的自动目标识别。该方法首先生成目标的理想图像(模板),再与实际目标图像进行对比匹配。该技术的显著特点是强调了同时利用目标回声和阴影信息的重要性。海上实验表明其优于传统模板匹配方法。文献 [30] 提出了一种结合特征点匹配和相似性计算的方法,以改善侧扫声呐图像的匹配效果。该方法通过对图像进行预处理以消减环境因素的影响,并且利用已知位置信息来提取在仿射变换条件下稳定的特征点,有效提高了匹配速度和准确性。文献 [31] 探讨了一种基于极线模板匹配技术的前视声呐图像拼接方法,在极坐标图像中应用模板匹配解决图像拼接过程中的对齐难题,适用于处理由视角变化所引起的图像扭曲问题。该方法可以有效地对多个前视声呐图像进行精准拼接,从而提供更为连贯和全面的水下视图。文献 [32] 介绍了一种结合先验位置知识的侧扫声呐图像组合成像匹配方法,通过融合特征点匹配和先验位置信息来进行匹配,利用先验位置知识来缩小搜索区域,有效减少错误匹配的概率,降低了计算复杂性。文献 [33] 专注于研究一种适用于前视扫描声呐(forward scanning sonar, FSS)图像的特征匹配和图像配准方法,通过精确提取图像特征并进行有效匹配,提高

声呐图像在复杂水下环境中用于导航和定位的准确性。文献 [34] 探讨了 FSS 成像技术中的特征提取和区域匹配技术, 通过采用改进的特征提取和匹配算法来提升水下目标的识别和图像分类的准确性。

2.2.2 局部特征提取方法

基于局部特征的水下声呐图像检测技术侧重于利用图像中的边缘、角点及纹理等关键视觉特征, 来进行目标的识别和分类, 在处理复杂水下环境中的目标识别和分类时特别有效。文献 [35] 比较了 7 种流行的特征检测算法, 包括 Harris 角点检测、Shi-Tomasi 角点检测、尺度不变特征转换 (scale-invariant feature transform, SIFT)、加速稳健特征 (speeded up robust features, SURF)、中心环绕算法 (center surround extremas, STAR)、加速段测试特征 (features from accelerated segment test, FAST)、定向 FAST 和旋转具有鲁棒性的独立基本特征 (oriented FAST and rotated binary robust independent elementary features, ORB) 在合成声呐图像数据集上的性能表现。利用 SVM 分类器进行训练, 并通过接收者操作特征曲线 (receiver operating characteristic curve, ROC) 来评估各算法在正确识别对象特征 (真正例) 与误将背景噪声识别为特征 (假正例) 之间的表现。文献 [36] 探究了基于 HOG 特征的海洋雷达目标检测技术, 展现了 HOG 特征在提升雷达图像目标检测精度方面的应用潜力。文献 [37] 介绍了 MATLAB 工具箱 Sonar Class, 该工具箱主要应用于侧扫声呐图像分类, 侧重利用本地纹理和回声特征, 强调从侧扫声呐图像中提取的第 2 阶 (即纹理) 统计参数的重要作用。文献 [38] 提出了一种前视声呐图像分割的新方法, 该方法结合了 CNN、Transformer 和 HOG 框架的多特征融合机制, 展示了将深度学习技术与传统特征提取方法相结合在声呐图像处理中的有效性。文献 [39] 探讨了使用基于 HOG 特征提取技术和 SVM 分类器来实现彩色成像声呐图像的目标检测, 研究结果表明该方法具有较高的检测率。文献 [40] 提出了一种基于尺度不变特征变换 (scale-invariant feature transform, SIFT) 的水下图像双向匹配定位方法, 专门用于有效识别特定船只的尾部, 突出了 SIFT 技术在水下环境应用中的潜力, 表明了其在水下无人系统学报

复杂水下环境中的实用性和准确性。

2.2.3 传统机器学习方法

基于传统机器学习方法的水下声呐图像目标检测与分类技术主要依赖于从声呐图像中提取的特征, 结合经典的机器学习算法来有效识别和分类海底目标。文献 [41] 介绍了一种结合 SVM 和 AdaBoost 算法的级联分类器, 用于声呐图像的目标检测和分类。该方法利用 SVM 的高精确度和 AdaBoost 的强大分类性能, 有效实现了对声呐图像中目标的高效识别, 展示了结合多种机器学习技术在声呐图像处理方面的优势。文献 [42] 展示了一种结合通用回归神经网络 (general regression neural network, GRNN) 和 PCA 的技术, 用以提高声呐目标分类的性能。该方法中, PCA 被用于降维和特征提取, 而 GRNN 负责处理复杂的非线性分类问题。二者的结合有效提升了声呐目标分类的准确率, 展现了混合技术在复杂声呐数据分析中的应用潜力。文献 [43] 探讨了利用多波束回波数据和传统机器学习方法进行海底沉积物分类的技术。研究表明, 通过合理选择特征并应用恰当的机器学习算法, 可以有效区分不同类型的海底沉积物。这项研究不仅证实了机器学习在海洋科学领域中的应用潜力, 同时也为海底沉积物的自动化分类提供了有价值的技术参考。文献 [44] 提出了一种基于传统机器学习方法的侧扫声呐图像分割技术。该方法使用将傅里叶谱分解为 3 个谱带: 低频、中频和高频功能对从该过程中提取的数据进行评估, 以确定其潜在的分类性能。文献 [45] 描述了在 2 种监督机器学习分类: SVM 和决策树 (decision tree, DT) 中对侧扫声呐图像的纹理分析和特征提取的使用进行的研究, 探讨了一阶纹理和二阶纹理的组合, 以获得最适合图像分类的纹理特征。

虽然传统算法在前视声呐目标检测领域取得了一定成果, 但是它们更多依赖图像的固有特征, 缺乏一定的普适性, 在噪声异常复杂的水下环境中效果欠佳。目前智能领域快速发展, 基于深度学习的前视声呐目标检测算法表现出显著的优势。

2.3 声呐图像语义分割任务

针对前视声呐图像的语义分割方法主要是对传统图像分割方法的优化和改进, 包括阈值分割法、聚类分割法、马尔可夫随机场分割法、小波变

化分割法和水平集分割法等^[46]。

阈值分割法通过比较图像中像素点的灰度与设定灰度阈值的大小,将每个像素点划分为不同的目标类别。文献^[47]提出一种二维模糊熵阈值分割方法,该方法引入传统帧差运动检测方法中的时间概念,并建立二维时空模糊熵算法确定灰度图像分割阈值,应用于前视声呐图像中运动目标的分割。文献^[48]提出一种基于先验知识的阈值分割方法,并在 36 张具有不同观测角度的包含金属管道目标的声呐图像上开展实验,实验结果验证了该方法在具有旁瓣干扰的前视声呐图像中分割线性目标的适用性。文献^[49]提出一种改进 OTSU 阈值的分割方法以限制分割阈值的搜索范围,并在具有人体目标的前视声呐图像上开展实验,实验结果显示该方法的计算时间只有最大熵阈值分割法的三分之一。

聚类分割法按照特定的距离度量方法度量样本间的相似度,并通过不断迭代更新聚类的质心,直到质心不再变化时完成分割。文献^[50]首先利用二维最大熵分割原理和鸡群优化算法去除前视声呐图像的背景,然后基于二维经验模态分解方法,选择合适的本征模态函数对前视声呐图像进行去噪和增强,最后采用增强模糊 C 均值聚类法对重建的前视声呐图像进行分割。

马尔可夫随机场分割法首先使用贝叶斯分类器基于已知的先验概率计算后验概率,然后利用后验概率分割图像。文献^[51]提出一种基于局部中心编码的马尔可夫随机场模型,有效提取前视声呐图像中的目标局部信息,用于前视声呐图像的目标分割。文献^[52]进一步提出一种基于选择性自编码的马尔可夫随机场模型标签域初始化方法,显著提升了文献^[51]中模型的分割速度。

小波变化分割法将图像信号分解为不同层次的小波系数,通常再结合其他方法进一步细化分割。文献^[53]首先在小波域应用自适应动态随机共振法以增强感兴趣区域,然后采用 Lee 滤波抑制斑点噪声,最后采用基于 C 均值聚类的分割方法计算前视声呐图像中目标的位置、大小和重心。

水平集分割法采用高维零水平集函数表示图像中不同区域的变化曲线方程,从而将复杂的图像曲线演变过程描述成求解偏微分方程近似解的

问题。文献^[54]引入形态学顶帽和底帽变换对前视声呐图像进行预处理,并提出无需重新初始化的改进水平集分割方法。

3 深度学习方法

3.1 表示学习方法

表示学习是声呐图像感知任务的基础。早期的声呐图像识别方法利用了传统的机器学习方法,通过精巧设计的手工特征提取方法对声呐图像的特征进行提取,例如局部二值模式(local binary pattern, LBP)^[55]、方向梯度直方图(histogram of oriented gradients, HOG)^[56]和主成分分析法(principal component analysis, PCA)^[26]等。随着深度学习的发展, CNN 逐渐成为声呐图像特征提取的主流工具。以下从骨干网络演进与声呐图像应用 2 个维度展开分析。

3.1.1 骨干网络的发展

深度学习骨干网络的发展为声呐图像感知任务提供了强大的特征提取能力。一种在深度学习中广泛使用的 CNN 架构视觉几何组(visual geometry group, VGG)网络^[57]由牛津大学开发,通过重复的 3×3 卷积核和 2×2 最大池化层,展示了网络深度在提升图像识别性能中的关键作用。尽管计算量较大,但 VGG 网络(如 VGG-16 和 VGG-19)在各种视觉任务中表现优异。残差神经网络(ResNet)^[58]通过引入残差连接解决了深度网络中的退化问题,使得构建数百层的深度网络成为可能,并迅速成为高性能模型的首选骨干网络。

Inception 网络^[59]通过并行多尺度卷积操作,提升了对多尺度信息的捕捉能力和计算效率,而稠密连接网络(DenseNet)^[60]通过层间密集连接,增强了特征重用和信息流动,同时减少了参数量。MobileNets^[61]通过深度可分离卷积技术,显著减少了模型的计算量和尺寸,非常适合资源受限环境下的设备。EfficientNet^[62]采用复合缩放策略平衡网络宽度、深度和分辨率,进一步提升了模型的精度和效率。

近年来,基于 Transformer 的网络结构在视觉领域也得到了成功应用。Vision Transformer^[63]首次通过将图像拆分为图像块输入 Transformer 网络,证明了视觉 Transformer 的可行性。Swin Trans-

former^[64]通过层级结构和滑动窗口设计,成功处理了图像中的长距离依赖问题,特别是在目标检测和语义分割任务中表现卓越。上述骨干网络的创新推动了深度学习在声呐图像感知任务中的广泛应用。

3.1.2 声呐图像中的表示学习

针对声呐图像的小样本特性,通常使用基于深度迁移学习的训练方法来获得具有较好分辨力的特征表示。具体做法是:在大规模的光学图像数据集上对 CNN 进行预训练,随后在声呐图像数据集上对模型进行微调,如图 2 所示。例如,文献[16]

使用预训练的 AlexNet 作为特征提取器,并在声呐图像数据中微调了最后一层全连接层,该方法的效果被证明显著超过了传统的特征提取方法。文献[12]比较了不同的 CNN 结构在声呐图像中的性能差异,包括 AlexNet 和 VGG-Net 系列,并对 VGG-f 的前 3 个卷积层基础上微调构建模型。文献[65]使用预训练的 ResNet-50 进行特征提取和微调。文献[1]使用半合成的声呐图像,并通过预训练的 VGG-16 和 VGG-19 作为特征提取器,在真实与合成数据上进行微调,有效提升了多类目标的分类性能。

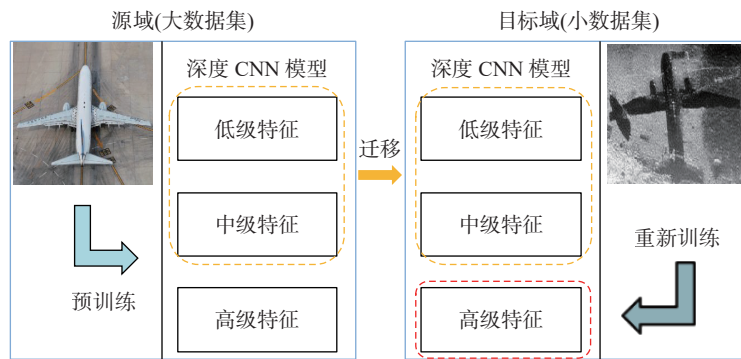


图 2 基于深度 CNN 模型的迁移学习方法示意图

Fig. 2 Schematic diagram of transfer learning method based on deep CNN model

除了直接对 CNN 骨干网络的微调,部分研究在训练方法或损失函数上进行了改进,以获得更加适用于声呐图像的特征提取方法。文献[24]提出了多尺度重复注意力机制和多域协同迁移学习方法,通过多尺度重复注意力机制来有效结合多尺度特征,使模型更加关注目标的形状细节,从而在排除噪声的同时,提取更多有用的特征,并利用了声呐图像与合成孔径雷达图像之间的低级特征相似性,以及声呐图像与光学图像之间的高级表示相似性,共同增强深度学习模型的特征提取能力。文献[25]通过一种两阶段解耦的微调方案来提高模型对尾部类别的特征分辨能力,并提出了平衡集成迁移学习来更充分地利用稀缺的声呐图像信息。文献[19]提出了一种新的损失函数,通过提升正确类别的 Logits 值并抑制错误类别的 Logits 值,从而扩大类别间在特征空间中的间隔,增强特征表示的判别性。

3.2 声呐图像分类算法

3.2.1 分类模型架构

获得特征表示之后,分类任务需要使用分类器

将输入样本特征分类为相应的类别。在基于深度学习的声呐图像分类任务中,通常使用 SVM 或全连接神经网络作为分类器。文献[16]通过使用 AlexNet 得到的图像特征进行分类,证明了 CNN 和 SVM 组合的方法性能优于传统的特征提取方法(CBP、HOG)和 SVM 的组合。文献[12]探讨了不同 CNN 网络和 SVM 结合的方法在声呐图像分类任务中的表现,发现使用 AlexNet 和 VGG19 作为特征提取器并使用 SVM 作为分类器能够取得非常高的识别精度。文献[7]使用全连接神经网络作为分类器,并通过迁移学习对其进行微调,以适应侧扫声呐图像的特定分类任务。文献[2]除了使用全连接层作为分类器外,还在训练中加入目标偏移量预测任务,以训练模型具有粗略定位目标的能力,从而获得更好的分类性能。文献[25]使用多个子训练集训练了多个全连接神经网络作为分类器,并通过权重平均方法使其融合,从而在推理阶段提高了识别模型的分类性能。

3.2.2 任务场景设置

虽然基于深度 CNN 的声呐图像自动目标识

别方法已取得了引人关注的成果^[12-15],但现有研究大都建立在数据集分布平衡且封闭的假设下。在真实的水下应用场景中,声呐图像数据通常存在天然的长尾分布和开放集问题,很大程度上制约了这些方法的应用效果^[18]。近期研究开始关注声呐图像中的长尾识别^[25]和开放集识别问题^[19]。

文献[25]通过实证研究发现,在使用声呐图像时,类别重平衡方案损害了表示器微调的性能,并有条件地促进了迁移学习过程中的分类器微调。为此,将两阶段解耦训练方法引入声呐图像分类任务,提出采用欠拟合头部类或过拟合尾部类的多平衡采样策略来缓解现有重采样方法存在的不足,进而提出平衡集成迁移学习(balanced ensemble transfer learning, BETL)方法:首先在大规模数据集中预训练模型,并使用均匀采样器对表示器进行微调;然后使用多平衡采样器重新训练分类器,并应用集成学习训练多个平衡分类器;再使用集成剪枝方法和融合策略来提高模型性能,减少内存和时间消耗;最后通过大规模实验验证了 BETL 方法的最佳性能。文献[19]构建了声呐图像开放集长尾识别(sonar open-set long-tail recognition, Sonar-OLTR)的任务基准,提出了 2 种新型开放集长尾识别评估指标,通过在训练中合成伪未知类特征来增强模型的未知类检测能力,并设计了新的损失函数来扩大未知类和已知类特征之间的边距。该方法能够在测试阶段获得更具区分性的已知类和未知类特征,从而提升 Sonar-OLTR 任务的性能。

3.3 声呐图像目标检测算法

3.3.1 通用领域

在深度学习领域,目标检测算法主要分为 CNN 和 Transformer 2 类。CNN 基础检测方法通过卷积层提取图像特征,广泛应用于目标检测任务。Transformer 则通过处理图像序列,以不同于传统 CNN 的方式进行图像理解和分析。这 2 种方法各具特色,推动了目标检测技术的发展。

基于 CNN 的目标检测算法分为单阶段和双阶段两类。单阶段算法中,YOLO^[66]系列将目标检测视为回归问题,通过操作同时预测多个边界框和类别,具有较高的处理速度和实时性能;单次多框检测器(single shot multibox detector, SSD)^[67]和

RetinaNet^[68]等改进算法通过多尺度锚框和焦点损失改善了小目标的检测精度及类别不平衡问题。双阶段算法中,如 Faster R-CNN^[69]则更加注重准确性,先通过区域提议网络生成候选区域,再进行精确分类和回归,适用于复杂背景和小目标检测任务;Mask R-CNN^[70]进一步增加实例分割功能,实现了更精细的目标检测。

基于 Transformer 的目标检测算法中,DETR (detection transformer)^[71]通过自注意力机制和编码器-解码器结构,简化了传统目标检测的流程,且无需锚框和非极大值抑制;Sparse R-CNN^[72]引入稀疏注意力机制,选择与目标相关的特征区域,提升了检测效率和精度;Deformable DETR^[73]通过可变形卷积和注意力机制,适应目标形态和尺寸变化,在小目标检测中表现尤为突出。

3.3.2 声呐图像领域

在声呐图像目标检测的深度学习应用领域,CNN 同样被广泛使用,显示出在处理特有的声呐图像特性和挑战时的巨大潜力。文献[74]针对水下声呐图像中诸如海底回波噪声干扰、前景目标区域像素比例低以及成像分辨率低等常见问题,提出了一种名为多级特征融合网络(multi-layer feature fusion module, MLFFNet)的声呐图像目标检测器。该检测器通过不同模块的紧密协作,有效实现了精确的水下目标检测。在真实场景声呐图像数据集上的测试结果表明,MLFFNet 在性能上优于其他现有方法。文献[75]使用 CNN 对前视声呐图像进行目标识别,并发现 CNN 在该类图像上的性能超过了传统的模板匹配方法,显示出更强的泛化能力。文献[76]对比了 CNN 和传统模板匹配方法在水下目标识别任务中所需的可学习参数数量,并证实了 CNN 在识别能力和泛化能力方面的优越性。这一发现为水下目标检测和识别领域的进一步研究和应用提供了有力的理论和技术支持。

近些年的研究中,基于 Transformer 框架的声呐图像目标检测方法也取得了一系列进展,给声呐图像目标检测领域带来了新的启示。文献[77]提出了一种创新的框架,用于基于混合 Transformer 网络的水下小目标检测。这种轻量级的混合 Transformer 网络设计可有效提取全局上下文信息,特别适合于处理水下环境中的细微信号。同时,为

了解决微弱信号在处理过程中可能发生的消失问题,该研究还引入了细粒度特征金字塔网络。此外,为了提高检测效果,还采用了测试时增强方法,且无需引入任何额外的参数。实验证明了该方法在高效检测水下小目标方面的能力,在参数数量和平均精度(mAP)方面,该方法显著优于主流检测算法。文献[78]专注于声呐图像去斑点任务,提出了一种基于Transformer的生成对抗网络(generative adversarial network, GAN)框架,名为声呐图像去噪(sonar image denoising, SID)-TGAN。该框架结合了Transformer和卷积块,以提取声呐图像的全局和局部特征,这些特征随后被融入到生成器和判别器网络中,以实现特征的融合和增强。通过对抗性训练, SID-TGAN能够学习到更全面的声呐图像表示,并在去除斑点方面展现出优异的性能。该方法还引入了一种新颖的对抗性损失函数,该函数综合考虑了图像内容、局部纹理、风格和全局相似性,从而减少训练过程中的图像失真和信息丢失。文献[79]通过结合轻量级YOLO网络和Swin Transformer,对单阶段检测算法YOLO进行了优化,以提高水下目标检测的准确性。研究在特殊的水下环境条件下使用声呐图像进行声学检测,引入深度学习神经网络替代传统检测方法。结果表明,YOLO与Swin Transformer的结合不仅能实现更高的检测精度,而且还满足了实时检测的硬件要求,证明了其在水下声呐图像处理中的应用潜力。

3.4 声呐图像语义分割算法

3.4.1 通用领域

近年来,基于深度学习的图像语义分割方法取得了快速发展,并广泛应用于自动驾驶、医学影像和水下声呐图像分析等领域。经典的语义分割网络包括全卷积网络(fully convolutional networks, FCN)、SegNet、U-Net系列、金字塔场景解析网络(pyramid scene parsing network, PSPNet)以及DeepLab系列等。

FCN^[80]是语义分割领域的开创性工作,其通过用卷积层替代传统分类网络的全连接层,构建端到端的语义分割网络。通过上采样操作恢复原始输入图像的尺寸,并结合浅层和深层特征图以保留更多特征信息。SegNet^[81-82]在FCN的基础

上,采用编码器-解码器结构,并利用最大池化层保存特征位置信息以生成与原始图像尺寸相同的分割结果。

U-Net^[83]最初用于医学图像分割,由于其跳跃连接机制能够更好地融合深层和浅层特征信息,因此在多个领域得到广泛应用。后续的改进结构包括Attention U-Net^[84]、UNet++^[85]和UNet 3+^[86],进一步提高了分割性能。

PSPNet^[87]通过引入空间金字塔池化模块,能够提取图像的多尺度特征,有效解决了尺度不一致和信息不足的问题,在ImageNet场景中表现较好。DeepLab系列则在FCN基础上不断优化:DeepLabV1^[88]引入空洞卷积以增加感受野,并使用全连接条件随机场精细化处理边缘信息;DeepLabV2^[89]通过空洞空间金字塔池化模块进一步提升了分割性能;DeepLabV3^[90]则通过多分辨率策略优化了特征融合,成为语义分割领域的前沿模型。

3.4.2 声呐图像领域

在前视声呐图像语义分割任务中,需要对声呐图像中的目标进行像素级别的分类,其比其他语义分割任务更加具有挑战性,主要表现在以下几个方面。

1) 数据稀少:由于水下环境的复杂性和不可预测性,数据采集过程中存在很大难度和风险,往往需要使用高成本设备,且需要专业人员进行操作、处理和标注。此外,前视声呐通常应用于海洋勘探和探测等领域,其中的数据需要保密处理,难以公开和共享。因此,需要构建可用数据集并采用合适的数据扩增策略,从而提升网络性能。

2) 分辨率较低:受到声波传播质量、能量、探头结构以及信号处理等因素的制约,前视声呐设备的成像分辨率较低,无法提供目标的详细特征^[91]。因此,需要设计合适的结构提升网络的特征提取能力。

3) 目标形态复杂:在前视声呐图像中,不同帧中的同一目标对象存在特征和尺度差异。此外,受到旁瓣干扰的影响,图像中完整的目标可能被割裂为多个相连的亮斑,使得目标的精确分割更为困难。因此,需要设计合适的结构以提升网络的多尺度能力。

4) 复杂噪声干扰:水声信道的复杂多变和声

波在介质中的传播特性会影响前视声呐系统的探测性能,具体表现为信号衰减、失真和混叠等^[92]。因此,当声呐设备通过接收换能器基阵接收目标的反射回波进行探测时,回波信号不仅携带水下目标的主要特征信息,同时也包含大量混杂的噪声干扰,导致目标对比度差、边缘模糊和易与复杂噪声背景混淆等问题,给语义分割任务带来许多挑战。因此,针对性地克服噪声干扰并保留更多目标细节,成为语义分割任务的关键。

针对上述声呐图像语义分割任务所面临的挑战,近年来也有部分研究。文献 [93] 提出了基于残差多级特征提取与融合的语义分割方法。针对前视声呐图像复杂特性,在 U-Net 的基础上设计并提出多级注意力与空洞金字塔嵌套 U 型网络(multi-level attention and atrous pyramid nested U-shaped network, MAANU-Net),实现前视声呐图像的精准语义分割。首先,设计残差 U 型特征提取(residual U-shaped feature extraction, RSU)结构作为该网络主干编码器和解码器结构,以提取上下文特征信息。然后设计注意力与空洞金字塔特征融合(attention and atrous pyramid feature fusion, AA)结构,引入注意力门控(attention gate, AG)模块抑制输入前视声呐图像中噪声区域的特征权重,突出关注局部显著区域的特征信息。同时,设计空洞金字塔池化(atrous pyramid, AP)模块构建具有不同扩张率的卷积核组,进一步扩大感受野,获取多尺度特征信息,解决低分辨率前视声呐图像下采样过程中的信息丢失问题。文献 [23] 针对纯卷积神经网络语义分割方法在前视声呐图像语义分割任务上的局限性,设计并提出融合 Transformer 和卷积神经网络的 U 型网络(mix transformer and convolution layer U-shaped network, MiTU-Net),实现准确语义分割的同时降低了计算开销。首先,设计混合 Transformer 嵌入编码结构(mix transformer embedding encoder block, MiT Block)和混合 Transformer 扩张解码结构(mix transformer expanding decoder block, MiTD Block),改进原始 Transformer 位置编码方式以输入更多特征信息,并通过高效自注意力模块在全局范围获取高分辨率与低分辨率特征图的细节特征信息;其次,在解码器中引入多层感知机(multi-layer perceptron, MLP)层,提取并聚合来自多

个层次的特征信息;最后,采用难例增强(online hard example mining, OHEM)策略改进损失函数,针对性改善网络的学习效果。

4 未来研究展望

水下智能感知算法在实际水下工程中已经得到了良好的应用,但是在工程应用中仍存在诸多不足,无法满足无人水下航行器的任务需求。近年来,计算机视觉技术的蓬勃发展也为声呐图像智能感知任务带来了新的启发,文中从数据集的规模、智能感知算法的性能和实际水下工程应用的任务需求 3 个层面入手对未来声呐图像智能感知算法的发展提出了以下 3 个潜在研究方向。

1) 大规模高质量数据集构建。大规模高质量数据是声呐图像智能感知领域的研究基础。尽管已有部分开源数据集被提出,但与通用图像领域相比,这些数据集的数量和多样性仍然严重不足。因此,未来研究亟需进一步采集和构建更加丰富的声呐图像数据集。首先,图像的规模应从现有的数千张扩展到数十万张,才能充分发挥深度学习的潜力,提升模型的学习效果;其次,目前的声呐图像数据存在较高的重复性,未来的数据采集应注重捕获目标在不同水下环境和条件下的多样性表现,以提高数据集的质量;最后,可以针对不同任务场景构建特定的数据集,如针对小目标、长尾分布、开放环境识别以及多模态数据集等。

2) 提升感知算法性能。未来提升声呐图像感知算法性能的关键在于解决领域泛化问题。当前大多数声呐图像模型通过光学图像数据集进行预训练,然而光学图像和声呐图像之间存在显著领域偏差,限制了模型在声呐图像上的表现。未来研究可以通过领域适应、跨领域特征对齐和对抗训练等方法,减少这种偏差,提高模型在声呐图像上的泛化能力,使其更好地适应复杂多变的海洋环境;其次,数据标注成本是前视声呐图像语义分割任务中的瓶颈,像素级标注耗费大量人力和时间,因此,未来可以通过探索弱监督、半监督和无监督学习等技术,降低对精确标签的依赖,利用伪标签生成、标签传播等方法减少对标注数据的需求,有效降低数据标注成本,同时提升模型训练效率应用范围。

3) 拓展新型任务设置。首先是声呐图像中的开放集长尾目标检测和语义分割任务。声呐图像中的目标检测和语义分割任务在真实环境中同样会面临开放集长尾问题, 定义声呐图像中的开放集长尾目标检测和语义分割任务, 构建系统规范的研究基准, 提出声呐图像中的开放集长尾目标检测和语义分割方法, 能够提高模型在开放环境中的感知能力。其次是声呐图像中的具身智能任务。水下机器人需要在执行任务时整合通信、感知、决策和控制等多重功能, 其中, 感知是决策和控制的基础。未来研究可以重点关注针对水下作业场景的特定感知任务, 使感知结果更好地支持后续的决策和控制任务。

参考文献:

- [1] HUO G, WU Z, LI J. Underwater object classification in sidescan sonar images using deep transfer learning and semisynthetic training data[J]. *IEEE Access*, 2020, 8: 47407-47418.
- [2] GERG I D, MONGA V. Structural prior driven regularized deep learning for sonar image classification[EB/OL]. *IEEE Transactions on Geoscience and Remote Sensing*, (2020-10-26)[2025-05-29]. <http://arXiv:2010.13317> (cs.CV).
- [3] NGUYEN H T, LEE E H, LEE S. Study on the classification performance of underwater sonar image classification based on convolutional neural networks for detecting a submerged human body[J]. *Sensors*, 2020, 20(1): 94.
- [4] HUO G, YANG S X, LI Q, et al. A robust and fast method for sidescan sonar image segmentation using nonlocal despeckling and active contour model[J]. *IEEE Transactions on Cybernetics*, 2016, 47(4): 855-872.
- [5] YUAN F, XIAO F, ZHANG K, et al. Noise reduction for sonar images by statistical analysis and fields of experts [J]. *Journal of Visual Communication and Image Representation*, 2021, 74: 102995.
- [6] MCKAY J, MONGA V, RAJ R G. Robust sonar ATR through Bayesian pose-corrected sparse classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(10): 5563-5576.
- [7] YE X, LI C, ZHANG S, et al. Research on side-scan sonar image target classification method based on transfer learning[C]//OCEANS MTS/IEEE Charleston. Harbin, China: IEEE, 2018.
- [8] OCHAL M, VAZQUEZ J, PETILLOT Y, et al. A comparison of few-shot learning methods for underwater optical and sonar image classification[C]//Global Oceans 2020. Singapore: IEEE, 2020: 1-10.
- [9] NOVAKOVIC J. Using information gain attribute evaluation to classify sonar targets[J]. *Telecommunications Forum Telfor*, 2009: 1351-1354.
- [10] REED S, PETILLOT Y, BELL J. An automatic approach to the detection and extraction of mine features in sidescan sonar[J]. *IEEE Journal of Oceanic Engineering*, 2003, 28(1): 90-105.
- [11] KUMAR N, TAN Q F, NARAYANAN S S. Object classification in sidescan sonar images with sparse representation techniques[C]//2012 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP). Kyoto, Japan: ICASSP, 2012: 1333-1336.
- [12] MCKAY J, GERG I, MONGA V, et al. What's mine is yours: Pretrained CNNs for limited training sonar ATR [C]//Oceans 2017-Anchorage. Anchorage, USA: IEEE, 2017: 1-4.
- [13] WANG X, JIAO J, YIN J, et al. Underwater sonar image classification using adaptive weights convolutional neural network[J]. *Applied Acoustics*, 2019, 146(3): 145-154.
- [14] VALDENEGRO-TORO M. Real-time convolutional networks for sonar image classification in low-power embedded systems[EB/OL]. (2017-09-06)[2025-05-29]. <https://arxiv.org/abs/1709.02153>.
- [15] WILLIAMS D P. On the use of tiny convolutional neural networks for human-expert-level classification performance in sonar imagery[J]. *IEEE Journal of Oceanic Engineering*, 2020, 46(1): 236-260.
- [16] ZHU P, ISAACS J, FU B, et al. Deep learning feature extraction for target recognition and classification in underwater sonar images[C]//2017 IEEE 56th Annual Conference on Decision and Control(CDC). Melbourne, Australia: IEEE, 2017: 2724-2731.
- [17] SINGH D, VALDENEGRO-TORO M. Deep neural networks for marine debris detection in sonar images [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE, 2021: 3741-3749.
- [18] CUI Y, JIA M, LIN T Y, et al. Class-balanced loss based on effective number of samples[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Angeles, USA: CVPR, 2019: 9268-9277.
- [19] JIAO W P, ZHANG J L, ZHANG C Y. Open-set recognition with long-tail sonar images[J]. *Expert Systems with Applications*, 2024, 249: 123495.
- [20] ZHU X, LIANG Y, ZHANG J, et al. STAFNet: Swin transformer based anchor-free network for detection of forward-looking sonar imagery[C]//Proceedings of the 2022 International Conference on Multimedia Retrieval. Newark, USA: ACM, 2022: 443-450.
- [21] ZHANG P, TANG J, ZHONG H, et al. Self-trained target detection of radar and sonar images using automatic deep learning[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 1-10.

- mote Sensing, 2021, 60: 1-14.
- [22] XIE K, YANG J, QIU K. A dataset with multibeam forward-looking sonar for underwater object detection[J]. *Scientific Data*, 2022, 9(1): 739.
- [23] LIANG Y, ZHU X, ZHANG J. MiTU-Net: An efficient mix transformer U-like network for forward-looking sonar image segmentation[C]//2022 IEEE 2nd International Conference on Computer Communication and Artificial Intelligence. Changsha, China: CCAI, 2022: 149-154.
- [24] CHENG Z, HUO G, LI H. A multi-domain collaborative transfer learning method with multi-scale repeated attention mechanism for underwater side-scan sonar image classification[J]. *Remote Sensing*, 2022, 14(2): 355.
- [25] JIAO W, ZHANG J. Sonar images classification while facing long-tail and few-shot[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 1-20.
- [26] ZHU M, SONG Y, GUO J, et al. PCA and kernel-based extreme learning machine for side-scan sonar image classification[C]//2017 IEEE Underwater Technology(UT). Busan, Korea: IEEE, 2017: 1-4.
- [27] KUMAR N, MITRA U, NARAYANAN S S. Robust object classification in underwater sidescan sonar images by using reliability-aware fusion of shadow features[J]. *IEEE Journal of Oceanic Engineering*, 2014, 40(3): 592-606.
- [28] WANG X, LIU X, JAPKOWICZ N, et al. Automated approach to classification of mine-like objects using multiple-aspect sonar images[J]. *Journal of Artificial Intelligence and Soft Computing Research*, 2014, 4(2): 133-148.
- [29] MYERS V, FAWCETT J. A template matching procedure for automatic target recognition in synthetic aperture sonar imagery[J]. *IEEE Signal Processing Letters*, 2010, 17(7): 683-686.
- [30] TAO W, LIU Y. Combined imaging matching method of side scan sonar images with prior position knowledge[J]. *IET Image Processing*, 2018, 12(2): 194-199.
- [31] LIU H, YE X. Forward-looking sonar image stitching based on midline template matching in polar image[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 62: 1-10.
- [32] LANE D M, STONER J P. Automatic interpretation of sonar imagery using qualitative feature matching[J]. *IEEE Journal of Oceanic Engineering*, 1994, 19(3): 391-405.
- [33] AYKIN M D, NEGAHDARIPOUR S. On feature matching and image registration for two-dimensional forward-scan sonar imaging[J]. *Journal of Field Robotics*, 2013, 30(4): 602-623.
- [34] AYKIN M, NEGAHDARIPOUR S. On feature extraction and region matching for forward scan sonar imaging [C]//2012 Oceans. Yeosu, Korea: IEEE, 2012: 1-9.
- [35] TUELLER P, KASTNER R, DIAMANT R. A comparison of feature detectors for underwater sonar imagery [C]//OCEANS 2018 MTS/IEEE Charleston. South Carolina, Charleston, USA: IEEE & MTS, 2018: 1-6.
- [36] LU Z, SHI Y. A Novel target detector of marine radar based on HOG feature[C]//2021 IEEE International Conference on Mechatronics and Automation(ICMA). Takamatsu, Japan: IEEE, 2021: 727-732.
- [37] FAKIRIS E, PAPATHEODOROU G. Sonar class: A MATLAB toolbox for the classification of side scan sonar imagery, using local textural and reverberational characteristics[C]//International Conference on Advanced Concepts for Intelligent Vision Systems(ACIVS 2006). Antwerp, Belgium: ACIVS, 2006: 488-499.
- [38] HE J, CHEN J, XU H, et al. SonarNet: Hybrid CNN-transformer-HOG framework and multifeature fusion mechanism for forward-looking sonar image segmentation[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 1-17.
- [39] WANG X, WANG J, YANG F, et al. Target detection in colorful imaging sonar based on HOG[C]//2018 IEEE International Conference on Signal Processing, Communications and Computing(ICSPCC). Qingdao, China: IEEE, 2018: 1-5.
- [40] LIN Y, LIU B. Underwater image bidirectional matching for localization based on SIFT[J]. *Journal of Marine Science and Application*, 2014, 13(2): 225-229.
- [41] XU H, YUAN H. An SVM-based AdaBoost cascade classifier for sonar image[J]. *IEEE Access*, 2020, 8: 115857-115864.
- [42] ERKMEN B, YLDRM T. Improving classification performance of sonar targets by applying general regression neural network with PCA[J]. *Expert Systems with Applications*, 2008, 35(1-2): 472-475.
- [43] JI X, YANG B, TANG Q. Seabed sediment classification using multibeam backscatter data based on the selecting optimal random forest model[J]. *Applied Acoustics*, 2020, 167: 107387.
- [44] NAIT-CHABANE A, ZERR B, LE CHENADEC G. Sidescan sonar imagery segmentation with a combination of texture and spectral analysis[C]//2013 MTS/IEEE OCEANS. Bergen, Norway: IEEE & MTS, 2013: 1-6.
- [45] FEBRIAWAN H, HELMHOLZ P, PARNUM I. Support vector machine and decision tree based classification of side-scan sonar mosaics using textural features[J]. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2019, 42(1): 27-34.
- [46] WANG J, XIAO L D, ZHOU P. Current situation and review of image segmentation[J]. *Recent Patents on Computer Science*, 2017, 10(1): 70-79.
- [47] LU Z, CHEN Y C, ZHANG T D, et al. A sonar image segmentation algorithm based on two-dimensional spatiotemporal fuzzy entropy[C]//2018 IEEE 8th International Conference on Underwater System Technology: Theory and Applications. Wuhan, China: IEEE, 2018: 1-5.

- [48] LIU L, BIAN H, YAGI S-I, et al. A prior-knowledge-based threshold segmentation method of forward-looking sonar images for underwater linear object detection[J]. *Japanese Journal of Applied Physics*, 2016, 55(7): 07KG06.
- [49] YUAN X, MARTNEZ J-F, ECKERT M, et al. An improved Otsu threshold segmentation method for underwater simultaneous localization and mapping-based navigation[J]. *Sensors*, 2016, 16(7): 1148.
- [50] TAN X, WANG C, CHEN T, et al. Forward looking sonar image segmentation based on empirical mode decomposition[J]. *Journal of Physics: Conference Series*, 2022: 012063.
- [51] SONG S, SI B, FENG X, et al. Prior parameter estimation for Ising-MRF-based sonar image segmentation by local center-encoding[C]//OCEANS 2015-Genova. Genova, Italy: IEEE, 2015: 1-5.
- [52] SONG S, SI B, FENG X, et al. Label field initialization for MRF-based sonar image segmentation by selective autoencoding[C]//OCEANS 2016-Shanghai. Shanghai, China: IEEE, 2016: 1-5.
- [53] BANERJEE S, RAY R, SHOME S N, et al. Noise induced feature enhancement and object segmentation of forward looking SONAR image[J]. *Procedia Technology*, 2014, 14: 125-132.
- [54] LIU G Y, PANG Y J, BIAN H Y, et al. Sonar image segmentation using the level set method without re-initialization[J]. *Advanced Materials Research*, 2014, 981: 368-371.
- [55] OJALA T, PIETIKAINEN M, HARWOOD D. Performance evaluation of texture measures with classification based on Kullback discrimination of distributions[C]//Proceedings of 12th International Conference on Pattern Recognition. Lisbon, Portugal: IEEE, 1994: 582-585.
- [56] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). California, San Diego, USA: IEEE, 2005: 886-893.
- [57] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//International Conference on Learning Representations (ICLR), San Diego, USA: ICRL, 2015: 1-14.
- [58] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: 2016: 770-778.
- [59] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Nevada, Las Vegas, USA: IEEE, 2016: 2818-2826.
- [60] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii, Honolulu, USA: IEEE, 2017: 4700-4708.
- [61] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-01)[2025-05-06]. <http://arxivpreprint arXiv:1704.04861>.
- [62] TAN M, LE Q. Efficientnet: Rethinking model scaling for convolutional neural networks[C]//International Conference on Machine Learning. Vienna, Austria: PMLR, 2019: 6105-6114.
- [63] DOSOVITSKIY A. An image is worth 16×16 words: Transformers for image recognition at scale[EB/OL]. (2020-12-01)[2025-05-06]. <http://arxivpreprint arxiv:2010.11929>.
- [64] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE, 2021: 10012-10022.
- [65] RUTLEDGE J, YUAN W, WU J, et al. Intelligent shipwreck search using autonomous underwater vehicles[C]//2018 IEEE International Conference on Robotics and Automation(ICRA). Brisbane, Australia: IEEE, 2018: 6175-6182.
- [66] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Nevada, Las Vegas, USA: IEEE, 2016: 779-788.
- [67] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]//Computer Vision-ECCV 2016: 14th European Conference. Amsterdam, The Netherlands: ECCV, 2016: 21-37.
- [68] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision. Amsterdam, Netherlands: IEEE, 2017: 2980-2988.
- [69] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2016, 39(6): 1137 - 1149.
- [70] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [C]//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 2961-2969.
- [71] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]//European Conference on Computer Vision. Glasgow, Scotland: ECCV, 2020: 213-229.
- [72] SUN P, ZHANG R, JIANG Y, et al. Sparse R-CNN: End-to-end object detection with learnable proposals[C]//Pro-

- ceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Tennessee, Nashville, USA: IEEE, 2021: 14454-14463.
- [73] ZHU X, SU W, LU L, et al. Deformable DETR: Deformable transformers for end-to-end object detection[C]//International Conference on Learning Representations. Addis Ababa, Ethiopia: ICLR, 2020: 1-12.
- [74] WANG Z, GUO J, ZENG L, et al. MLFFNet: Multilevel feature fusion network for object detection in sonar images[C]//IEEE Transactions on Geoscience and Remote Sensing. New York, USA: IEEE, 2022: 1-19.
- [75] KVASIC I, MIŠKOVIC N, VUKIC Z. Convolutional neural network architectures for sonar-based diver detection and tracking[C]//OCEANS 2019-Marseille. Marseille, France: IEEE, 2019: 1-6.
- [76] SUNG M, CHO H, KIM T, et al. Crosstalk removal in forward scan sonar image using deep learning for object detection[J]. *IEEE Sensors Journal*, 2019, 19(21): 9929-9944.
- [77] CHEN G, MAO Z, WANG K, et al. HTDet: A hybrid transformer-based approach for underwater small object detection[J]. *Remote Sensing*, 2023, 15(4): 1-16.
- [78] ZHOU X, TIAN K, ZHOU Z, et al. SID-TGAN: A transformer-based generative adversarial network for sonar image despeckling[J]. *Remote Sensing*, 2023, 15(20): 5072.
- [79] CHEN R, ZHAN S, CHEN Y. Underwater target detection algorithm based on YOLO and Swin transformer for sonar images[C]//OCEANS 2022. Virginia, Virginia Beach, USA: IEEE, 2022: 1-7.
- [80] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, Massachusetts, USA: IEEE, 2015: 3431-3440.
- [81] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Advances in Neural Information Processing Systems*, 2012, 25: 1097-1105.
- [82] BADRINARAYANAN V, KENDALL A, CIPOLLA R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(12): 2481-2495.
- [83] RONNEBERGER O, FISCHER P, BROX T. U-NET: Convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference. Munich, Germany: Springer, 2015, 18(3): 234-241.
- [84] OKTAY O, SCHLEMPER J, FOLGOC L L, et al. Attention U-Net: Learning where to look for the pancreas[C]//Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2022.
- [85] ZHOU Z, SIDDIQUEE M M R, TAJBAKHS N, et al. UNet++: Redesigning skip connections to exploit multiscale features in image segmentation[J]. *IEEE Transactions on Medical Imaging*, 2019, 39(6): 1856-1867.
- [86] HUANG H, LIN L, TONG R, et al. UNet 3+: A full-scale connected UNET for medical image segmentation[C]//ICASSP 2020 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP). Barcelona, Spain: IEEE, 2020: 1055-1059.
- [87] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii, Honolulu, USA: IEEE, 2017: 2881-2890.
- [88] LIANG-CHIEH C, PAPANDREOU G, KOKKINOS I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFS[C]//International Conference on Learning Representations. San Diego, California, USA: IEEE, 2015.
- [89] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFS[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 40(4): 834-848.
- [90] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation[EB/OL]. (2017-06-01)[2025-05-06]. <http://arXiv preprint arXiv: 1706.05587>.
- [91] ZHANG T, LIU S, HE X, et al. Underwater target tracking using forward-looking sonar for autonomous underwater vehicles[J]. *Sensors*, 2019, 20(1): 1-16.
- [92] GUSSEN C M, DINIZ P S, CAMPOS M L, et al. A survey of underwater wireless communication technologies [J]. *J. Commun. Inf. Sys*, 2016, 31(1): 242-255.
- [93] LIANG Y, ZHU X, ZHANG J. Maanu-Net: Multi-level attention and atrous pyramid nested U-Net for wrecked objects segmentation in forward-looking sonar images [C]//2022 IEEE International Conference on Image Processing (ICIP). Bordeaux, France: IEEE, 2022: 736-740.

(责任编辑: 闫溶蔓)