基于深度学习的水下图像目标检测综述

罗逸豪*[①] 刘奇佩[①] 张 吟[①] 周河宇[①] 张钧陶[②] 曹 翔[③]

[①](宜昌测试技术研究所 宜昌 443003)

[②](军事科学院系统工程研究院 北京 100141)

[③](长沙学院 长沙 410022)

摘 要：水下图像目标检测是水下智能化探测的核心技术之一，广泛应用于工业及军事领域。深度学习相关技术的突破为水下图像目标检测的发展带来了新的机遇，但是目前该领域的综述较为陈旧，并且缺乏一定的系统性和全面性。该文对基于深度学习的水下可见光图像和声呐图像目标检测研究工作进行了详细总结与分析。首先，对基于深度学习的通用目标检测算法框架进行了梳理，包含骨干网络、颈部模块、检测头部、训练算法、推理策略、数据集6项要素，并系统性地总结了每个要素存在的问题及最新研究工作；然后，调研了水下可见光图像目标检测最新进展，分别从数据集发展、模型设计、训练算法进行总结；同时，归纳并分析了水下声呐图像目标检测相关工作，包含前视、侧扫、合成孔径3种声呐。最后，结合深度学习最新研究探讨了该领域的研究趋势。

关键词：水下图像目标检测；深度学习；可见光图像；声呐图像；数据集

中图分类号：TN911.73; TP391.4 **文献标识码**：A **文章编号**：1009-5896(2023)10-3468-15

DOI: 10.11999/JEIT221402

Review of Underwater Image Object Detection Based on Deep Learning

LUO Yihao[①] LIU Qipei[①] ZHANG Yin[①] ZHOU Heyu[①]

ZHANG Juntao[②] CAO Xiang[③]

[①](*Yichang Testing Technique Research Institute*, *Yichang* 443003, *China*)

[②](*Institute of System Engineering*, *AMS*, *PLA*, *Beijing* 100141, *China*)

[③](*Changsha University*, *Changsha* 410022, *China*)

Abstract: Underwater image object detection is one of the core technologies of underwater intelligent exploration, which is widely used in industrial and military fields. The breakthrough of deep learning related technologies has brought new opportunities for the development of underwater image object detection, but the current reviews are relatively old and lack a certain degree of systematicness and comprehensiveness. In this paper, the research of underwater visible and sonar image detection based on deep learning is summarized and analyzed in detail. Firstly, the general object detection algorithm framework based on deep learning is sorted out, including six elements: backbone, neck, head, training algorithm, inference strategy, and evaluation criteria, and the problems of each element and the latest research work are systematically summarized; Then, the latest progresses of underwater visible image object detection are investigated and summarized from three aspects: data set, model design, and training method; Meanwhile, the works related to underwater sonar image detection are summarized and analyzed, including forward-looking sonar, side-scanning sonar and synthetic aperture sonar. Finally, the research trend of underwater image object detection is discussed based on the latest research on deep learning.

Key words: Underwater image object detection; Deep learning; Visible image; Sonar image; Data set

收稿日期：2022-11-08；改回日期：2023-04-10；网络出版：2023-04-24

*通信作者： 罗逸豪 luoyihao@hust.edu.cn

1　引言

随着工业及军事应用中智能化水下探测的需求增多，水下图像目标检测相关研究日益活跃，涉及水生物探测、水环境勘探、海床建模、打捞救助、海底管道探测、反水雷、反潜等众多项任务[1]。由于水下环境复杂多变、信号衰减失真、信号获取传输成本高，水下图像目标检测也是计算机视觉和图像处理领域中最具挑战性的应用研究之一[2]。目前国内水下无人探测尚未进行大规模应用，一个重要的原因就是检测算法性能不足，多数情况需要人工进行干预。如何提高算法精度和速度、丰富水下图像数据集、增强应对复杂环境的鲁棒性、提高算法的泛化性、降低模型计算复杂度，均是该领域中亟需解决的关键问题。

目标检测需要对图像中的目标进行分类和定位，早期依赖人工提取图像特征。然而面对各式各样的应用场景和复杂的环境干扰，传统的人工特征已经无法满足日益增长的需求。随着2012年AlexNet[3]采用卷积神经网络(Convolutional Neural Network, CNN)在ImageNet[4]大规模图像分类数据集上取得的突破性效果，深度学习被逐步应用于计算机视觉领域中的各项应用。深度学习利用大数据对网络模型进行端到端训练，克服了传统方法的诸多缺点。在水下图像目标检测领域，深度学习方法借助数据驱动的优势，已在鱼类图像数据集Fish4-Knowledge、全国水下机器人大赛(Underwater Robot Professional Contest, URPC)等开源可见光图像数据集和一些非公开声呐图像数据集中实现了更优的效果[2,5]。

系统性、模块化地分析通用目标检测算法框架，对水下图像目标检测的应用研究具有十分重要的指导意义，而目前的相关综述较为陈旧。数年前就有文献[6]对早期基于深度学习的通用目标检测(common object detection)研究进行了分类与总结，并与传统方法进行了对比，体现出深度学习的杰出效果。近几年深度学习算法研究呈井喷式增长，克服了模型设计和训练过程中的诸多难题，精度已接近早期深度学习方法的两倍。然而，较新的综述[7,8]依旧沿用早期的模型分类方法(2阶段与1阶段检测)，未对较新的研究进行归纳。针对水下图像应用领域，林森等人[9]对光学图像中目标探测关键技术进行了总结，文献[1,5]对声呐图像目标检测研究进行了总结，但他们梳理的文献较旧，并且对深度学习方法提及过少。Fayaz等人[10]着重介绍了早期通用目标检测算法，未对水下相关应用研究进行详细梳理。

基于此，本文第2节对基于深度学习的通用目标检测算法框架进行了系统性梳理，分类总结了最新研究工作；第3节从数据集构建及方法研究两方面总结了水下可见光图像目标检测最新进展；第4节对前视、侧扫、合成孔径3种声呐图像目标检测研究进行了归纳分析；第5节进行总结与展望。

2　基于深度学习的通用目标检测算法框架

2013年—2019年处于深度学习目标检测算法早期研究阶段，人们主要根据是否存在显式的候选框提取过程，将目标检测模型分为2阶段(two-stage)和1阶段(one-stage)。2阶段检测模型通过候选框提取方法首先筛选感兴趣区域(Region of Interest, RoI)，然后再进行识别与定位，精度更高，代表作是R-CNN家族[11–13]。1阶段检测模型直接使用固定的锚框(anchor)进行识别定位，速度更快，代表作包括SSD(Single Shot Detector)系列[14]和YOLO(You Only Look Once)家族[15–17]。随着研究的深入，人们提出了更多类型的检测模型，比如根据是否需要显式定义先验锚框，可以分为基于锚框(anchor-based)和无锚框(anchor-free)方法，后者可以避免人工预先设置锚框，通用性更强，代表作为CenterNet[18]和FCOS(Fully Convolutional One-Stage object detection)[19]。大部分2阶段模型属于基于锚框的方法，而1阶段模型则两者皆有。最近，Transformer[20]目标检测模型又开辟了基于目标查询和集合预测的新范式，不同于常规CNN。因此，仅以2/1阶段检测模型类别来概括现有方法已不再合适。

借鉴开源项目MMDetection[21]的代码实现方式，本文将深度学习通用目标检测算法框架总结为6个要素：骨干网络、颈部模块、检测头部、训练算法、推理策略、数据集。其中前3项要素属于模型设计过程，以构成目标检测网络模型，如图1所示。本节将总结每个要素的功能、存在问题及最新的算法研究工作，为解决水下图像目标检测应用难题提供支撑。

2.1　骨干网络

骨干网络作为图像特征提取模块，可以提取层次化、模块化、抽象化的特征信息，是深度学习模型最重要的组成部分之一。大多数在图像分类领域中具备良好效果的骨干网络也可在目标检测中获得较高精度。在AlexNet[6]开启CNN研究热潮之后，许多研究致力于对网络模型进行加深加宽，但这会引起计算成本增长与梯度消失问题。2017年ResNet[22]通过残差学习和跳跃连接(skip connection)缓解了梯度消散问题，可以构建上百层甚至更深的网络，

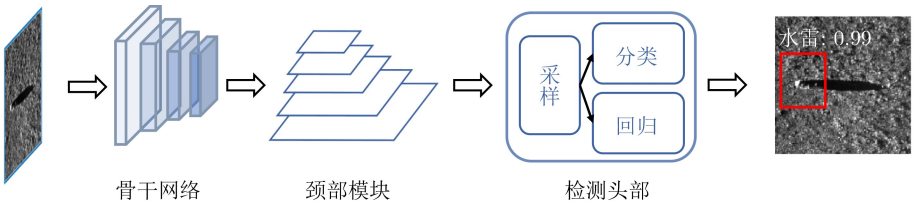

图 1 基于深度学习的目标检测模型

广泛应用于众多视觉任务，并不断被改进优化，比如DenseNet[23]等。近年来，许多不同于常规CNN卷积滤波核的骨干网络被提出，比如可变形卷积(Deformable Convolutional Network, DCN)[24]、多层感知机(MultiLayer Perceptron, MLP)[25]和Transformer[26]，它们的性能不弱于CNN。

随着嵌入式环境中目标检测任务需求日益上升，人们对目标检测算法的实时性要求也水涨船高。由于精度提升往往伴随着模型规模和参数量大幅增长，许多研究工作致力于在保证精度的同时设计轻量化骨干网络。MobileNet系列模型[27]深度可分离卷积，将标准CNN分解成深度(depthwise)卷积和逐点(pointwise)卷积，大幅降低了模型参数量与运算量。轻量化的骨干网络设计可以确保系统运行的实时性，适用于缺陷检测、水下探测等众多工业应用项目。

2.2 颈部模块

颈部模块提取多尺度特征，以提高模型检测精度。深度神经网络理论认为模型中不同的层具备不同的功能，即捕捉不同感受野(receptive field)的信息。通常来说，浅层网络提取的高分辨率特征具有更丰富的空间、边缘等信息，其较小的感受野更适合检测小尺寸的目标；深层网络提取的低分辨率特征具有更丰富的语义信息，其较大的感受野更适合检测大尺寸的目标。为解决单张特征图对大、中、小目标适应性差的问题，特征金字塔网络[28](Feature Pyramid Network, FPN)以自顶向下的方式将不同层级的骨干网络输出特征逐级融合，再对各个尺度执行独立的预测。FPN由于简单的结构设计和优越的性能，成为颈部模块的标准范式。然而FPN结构本身也存在一定的缺陷，比如高层特征通道信息衰减、特征融合过程中的信息稀释和混叠歧义。

为了改善这些问题，PAFPN[29]在FPN原有的自顶向下结构后，又增加了自底向上的连接，使得各层特征都能较好地融合其他层的信息，实现更加丰富的多尺度特征表示。之后以RCNet[30]为代表的诸多研究工作尝试堆叠更多的特征图节点与连接来增强特征，并引入注意力机制优化特征表达。然而复杂的特征堆叠会使FPN的计算复杂度急剧上升。因此，以NAS-FCOS[31]为代表的方法权衡模型精度与推理效率，在保证不引入复杂计算量的情况下设计了更为健壮的FPN结构。

水下图像目标检测应用场景对小目标检测和实时性要求较高，因此颈部模块需要兼顾精度与速度。

2.3 检测头部

检测头部通常包含采样、分类器和回归器，一般也是多尺度的，在提取的各尺度特征图上进行正负样本的采样，然后将其输入到分类器和回归器网络模型(通常为CNN)中进行预测，得到最终的检测结果。

采样过程包含样本生成和类别分配，这也是2阶段、1阶段、无锚框、Transformer检测模型的主要区别所在。2阶段检测模型[11–13]采用区域推荐网络(Region Proposal Network, RPN)提取一定数量的正负样本；1阶段模型[14–17]将特征图上的每一个坐标点都视作具有潜在目标，以固定锚框长宽比和数量生成训练样本；无锚框方法[18,19]不需要人工设定锚框，直接预测目标框的关键点，或是以坐标点是否落入真实框内来区分正负样本，并额外设计适用于无锚框的输出分支；Transformer[20]设计基于目标查询的可学习位置编码，通过解码器生成一定数量的预测框。而随着各类研究的不断深入，不同类型的检测模型在通用目标检测数据集上的效果没有较大差异，性能差异主要体现在精度与速度的权衡。

为了获得更精确的定位结果，以SABL(Side-Aware Boundary Localization)[32]为代表的研究利用语义特征来引导高质量锚框生成，同时适用于1阶段检测器和2阶段检测器的RPN。在此基础上，Wu等人[33]重新思考了分类分支和回归分支模型结构的并行设计，并根据它们的相关性设计了交互相关的模型结构，不再将它们看作为独立的并行结构。延续该思路，TOOD(Task-aligned One-stage Object Detection)[34]等工作进一步挖掘分类任务和回归任务的关联性，共同优化了分类和回归分支。

2.4 训练算法

目标检测训练算法为每个输出分支设计相应的

损失函数对比样本预测值和真实值(标签)以产生损失值，通过迭代最小化损失使得预测结果逼近真实值。在早期研究阶段，分类损失通常采用交叉熵函数进行计算，回归损失通常采用Smooth L1函数或预测框与真实框交并比(Intersection over Union, IoU)计算。目前目标检测模型在训练阶段面临3个主要问题：正负样本不平衡、样本框质量低、任务优化失衡。

在一幅图像中待检测目标面积通常只占小部分，因此在采样过程中会出现大量的背景负样本。目前主流的思想就是依据重要性对训练样本进行加权，以平衡正负样本对梯度的影响。早期难样本挖掘(Hard-example mining)方法认为难样本(产生较大损失值的样本)对训练更加重要，比如Focal Loss[35]和Libra R-CNN[36]巧妙地减弱简单样本的权重并加大难样本的重要程度。最近的研究重新思考了何为重要样本，以IQDet(Instance-wise Quality Distribution sampling detector)[37]为代表的方法引入概率得分作为样本重要性的依据，ATSS(Adaptive Training Sample Selection)[38]等方法则是根据IoU设计自适应的样本选择策略。除此之外，OTA (Optimal Transport Assignment)[39]将样本分配问题看作运输优化问题，以最小运输成本求解最优运输计划；基于AP-Loss[40]的方法利用平均精度设计损失函数，将分类任务替换为排序任务。

样本框质量低可以在检测头部模型设计中得到一定改善(2.3节)。在训练阶段主要体现在优化回归损失函数。Libra R-CNN[36]设计了平衡的L1损失函数来减少异常值和离群值对回归损失的影响；Dynamic R-CNN[41]采用动态训练方法来调整训练过程中损失函数的阈值，逐步提高锚框的质量。由于预测框与真实框的IoU可以直观反映其质量高低，基于IoU的回归损失[42]也被广泛研究。

训练过程同时对分类任务和回归任务进行优化，属于多任务学习(multi-task learning)。梯度较大的任务将会占据训练优化的主导地位，造成任务优化失衡的问题[43]。因此，以SWN(Sample Weighting Network)[44]为代表的方法通过同方差不确定性为分类和回归损失设置权重。后续GFL(Generalized Focal Loss)[45]等方法研究梯度、分布、质量等分类和回归的相关性因素，巧妙地设计了可以进行协同优化的总体损失函数。

除此之外，训练过程中还可以采用数据增强方法来达到提升模型精度的目的。通常而言，训练数据越庞大、样本越丰富，模型泛化能力更好。数据增强方法包括翻转、旋转、裁剪、变形、缩放等几何变换操作，以及颜色和空间变换[17]。在水下图像数据缺乏时，数据增强方法至关重要。

2.5 推理策略

训练结束的目标检测模型用于推理，通常会输出数量繁多的预测框，需要后处理方法删除冗余的预测框以得到精准的结果。目前最常用的方法是非极大值抑制(Non-Maximum Suppression, NMS)，通过迭代算法删除冗余框。当同类目标分布密集且存在遮挡时，NMS极易产生漏检。Soft-NMS[46]在后处理过程中不是粗暴地删除IoU大于阈值的预测框，而是降低其置信度，在密集目标检测任务中效果优越。

由于NMS的独立性，目标检测算法并不是严格意义上的端到端结构。NMS-Loss[47]致力于NMS与检测模型的共同优化训练。最近，Sparse R-CNN[48]和POTO(Prediction-aware OneTo-One)[49]颠覆性地提出了稀疏性目标检测新结构，抛弃了常规的大量候选框提取和NMS过程，取得了较高的检测精度，但尚未在工业界广泛应用。

在水下图像应用场景中，如果需要检测密集的水下生物，比如鱼群、珊瑚群，可以借助Soft-NMS方法得到更精确的结果。如果需要在广阔海域中搜寻个别物体，比如水雷、潜艇，不需要大量的候选框提取过程，可以尝试类似Sparse R-CNN的端到端新架构。此外，图像预处理在推理过程中也至关重要，诸如图像增强、超分辨率[50]等方法能改善水下图像质量，以获得更准确的结果。

2.6 数据集

随着模型规模和参数量越来越大，深度神经网络对于训练数据的依赖也越高，为提升并验证模型精度以及泛化性，建立大规模数据集至关重要。最常用的通用目标检测数据集是PASCAL VOC(Pattern Analysis, Statical modeling and ComputAtional Learning Visual Object Classes)[51]和MS COCO (MicroSoft Common Objects in COntext)[52]，分别包含20类与80类常见目标，共计接近200 000张可见光图像。基于这两个数据集的目标检测模型与训练/推理算法研究常具备较好的泛化性，被广泛应用于医学图像检测、红外目标检测等具体应用场景[7]。不同的目标检测应用任务的主要区别在于图像风格差异，而检测模型与算法类似。当训练集和测试集图像风格差异较大时(比如用可见光图像训练，在声呐图像上测试)，模型检测精度通常很低。

因此在水下可见光、声呐图像目标检测应用中，首要任务是构建各自的大规模数据集，防止深度神经网络训练欠拟合或过拟合，同时便于不同的网络

模型进行精度对比。然后，再基于数据集及图像的特性，对通用目标检测模型与算法进行优化。

3 基于深度学习的水下可见光图像目标检测

水下可见光图像信息量较为丰富，在近距离的水下目标探测任务中具有突出优势。然而，由于受水下特殊成像环境的限制，可见光图像往往存在颜色失真、噪声多、边缘纹理模糊、可见度低等众多问题，比通用目标检测更具挑战性。遵循第2节概括的检测框架，本节从数据集发展、模型设计、训练算法3个方面总结了水下可见光图像目标检测研究进展。

3.1 数据集发展

水下场景环境具备多样性，在不同水域/海域采集的图像具有不同的图像质量与目标种类。为了面对不同类型的探测需求，研究者构建了种类繁多的水下数据集。表1按照3个部分归纳了目前一些可用于水下可见光图像目标检测的公开数据集，其中“*”表示标注信息未公开，“-”表示未专门划分测试集。

针对水下机器人自主抓取所需的感知探测技术，中国连续数年举办了全国水下机器人大赛[53](Underwater Robot Professional Contest, URPC)，采集海参、海胆、扇贝、海星等近海底常见目标构建数据集。URPC2017存在大量相似或重复的图像，精简后的URPC2018常用于算法的对比，后续的版本在前一年的图像库中逐渐增加新图像。美中不足的是，URPC的部分数据缺少海星标签，容易出现错误或标签缺失，并且测试集图像的标注没有公开。RUIE(Real-time Underwater Image Enhancement)[54]数据集构建了目标检测子集，但是图像数量不多。为了解决上述问题，UDD(Underwater open-sea farm object Detection Dataset)[55]收集了高清海底图像并进行了精细的标注；UWD(Under-Water Dataset)[56]收集了URPC及大量互联网图像，构建超过一万张图像的大型数据集进行模型训练。DUO(Detecting Underwater Objects)[57]基于相关数据集进行收集和重新注释，并公平比较了十余种通用目标检测模型的效果，为后续研究提供了重要实验数据支撑。

为了研究海洋生态与动物，Fish4Knowledge[58]构建了目前最大的海底鱼类目标检测数据集，包含23种不同的鱼类以及密集、遮挡、模糊等干扰情况；Brackish数据集[59]扩充了水母、螃蟹等更多的海洋生物。为了研究海洋污染问题，Fulton等人[60]引入了塑料垃圾和人为目标两种大类，与海洋生

表 1 可用于水下可见光图像目标检测的数据集

数据集	训练集图像数	测试集图像数	类别数	类别描述	用途	年份
URPC2017[53]	17655	985*	3	海参、海胆、扇贝	目标检测	2017
URPC2018[53]	2901	800*	4	海参、海胆、扇贝、海星	目标检测	2018
URPC2019[53]	4757	1029*	4	海参、海胆、扇贝、海星	目标检测	2019
URPC2020-ZJ[53]	5543	2000*	4	海参、海胆、扇贝、海星	目标检测	2020
URPC2020-DL[53]	6575	2400*	4	海参、海胆、扇贝、海星	目标检测	2020
URPC2021[53]	7600	2400*	4	海参、海胆、扇贝、海星	目标检测	2021
RUIE-UHTS[54]	300	–	3	海参、海胆、扇贝	目标检测	2020
UDD[55]	1827	400	3	海参、海胆、扇贝	目标检测	2022
UWD[56]	10000	–	4	海参、海胆、扇贝、海星	目标检测	2020
DUO[57]	6671	1111	4	海参、海胆、扇贝、海星	目标检测	2021
Fish4Knowledge[58]	27370	–	23	海底鱼类	目标检测	2013
Brackish[59]	14518	–	6	大鱼、小鱼、水母、螃蟹等	目标检测	2019
Marine Litter[60]	5720	–	3	塑料垃圾、人为目标、生物	目标检测	2019
TrashCan[61]	7212	–	22	海底垃圾、动植物等	目标检测/分割	2020
SUIM[62]	1525	110	8	鱼类、珊瑚、植物、人、残骸等	目标检测/分割	2020
Kyutech10K[63]	10728	–	7	虾、鱿鱼、螃蟹、鲨鱼等	图像分类	2018
UIEB[64]	950	–	8	各类珊瑚与海洋生物等	图像增强	2020
MUED[65]	8600	–	430	430个海底物体	显著性检测	2019
UOT32[66]	24241	–	32	32个海底目标视频	目标跟踪	2019
UOT100[67]	74042	–	104	104个海底目标视频	目标跟踪	2021

物进行区分；TrashCan数据集[61]将海底垃圾和海洋生物进行了更加细致的分类，并且对目标包含的像素点进行了标注。为了应对海底打捞与救助任务，海洋与机器人研究者标注了SUIM(Segmentation of Underwater IMagery)[62]水下图像语义分割数据集。

还有一些用于其他视觉任务的水下图像数据集，经过处理或转换之后可以用于目标检测。日本海洋地球科学技术厅提供了大型深海海洋生物分类数据集Kyutech10K[63]，由于图像中海洋生物清晰可辨，可以对包含的动物进行定位标注。用于水下图像增强算法评估的UIEB(Underwater Image Enhancement Benchmark)数据集[64]，图像可见度及分辨率高，也可以进行目标标注。MUED(Marine Underwater Environment Database)数据集[65]包含8600张图像上430个目标的显著性像素点标注，将目标类别合并之后可以用于目标检测训练。UOT32[66]和UOT100[67]是用于海底目标跟踪的数据集，部分单目标视频标注可以直接用于检测模型训练。

3.2 模型设计

早期的大型水下可见光图像目标检测数据集较少，因此相关研究并不火热。2017年前后，研究者[68-70]分别将当时最受欢迎的3种通用目标检测模型(YOLO[15], SSD[14], Faster R-CNN[12])直接应用到鱼类检测，优于传统算法。这说明当训练数据和计算资源充足时，深度神经网络也可以在水下目标检测任务取得良好效果。随着更多水下数据集的建立，越来越多的研究将通用模型运用到水下目标检测，并不同程度地改进了骨干网络、颈部模块和检测头部。

YOLOv3模型[16]由于运行速度快且易于部署实现深受欢迎。Knausgård等人[71]利用SE(Squeeze-and-Excitation)模块改进了YOLOv3的骨干网络，取得了更高的鱼类检测精度。叶赵兵等人[72]以YOLOv3-SPP骨干网络为基础，增加网络预测尺度以提高URPC数据集[53]中小目标检测性能，同时利用K-Means++聚类算法筛选最佳的锚框。张艳等人[73]基于通道注意力突出不同通道特征图的显著性，提高了骨干网络对水下图像高频信息的提取能力，并且优化了颈部模块多尺度特征融合过程，在URPC上取得了较大提升。

无锚框能够避免人工预先设置锚框，通用性更强，可以改进水下目标漏检问题。王蓉蓉等人[74]改进了CenterNet[34]的骨干网络，降低了模型参数量以提升网络推理速度，同时引入空间注意力和通道注意力，使骨干网络和颈部模块关注重要目标特征信息，在URPC上取得了良好效果。蔡达等人[75]设计了自适应加权融合特征金字塔优化FCOS[19]模型，实现多尺度空间特征选择，同时借鉴了基于空间特征解耦的检测头部，实现了中心和边界区域的特征选择。

两阶段模型由于运行速度较慢，未被广泛应用于水下可见光图像目标检测。为了改进此问题，喻明毫等人[76]设计了一种轻量级检测器，首先使用高效卷积池化来获取不同特征表达，然后在稠密连接结构的基础上增加两路稠密连接以提高网络表征能力，在RUIE[54]和Marine Litter数据集[60]上实现了较高精度和速度的平衡。除此之外，Liang等人[77]借鉴了特征解耦、位置编码和注意力机制优化RoI特征，设计了一种通用的检测头部，在2阶段和1阶段检测器中均实现了较大的精度提升。

3.3 训练算法

尽管最近研究者构建了诸多水下可见光图像数据集，它们离MS COCO的规模大小仍相去甚远。因此许多工作在模型训练过程中引入数据增强方法，充分挖掘模型拟合能力。早期的数据增强方法通常对单个图像进行操作，Lin等人[78]研究了用于模拟重叠、遮挡和模糊对象的增强策略，提出了RoIMix方法，从不同图像中提取的目标混合在一起创建新的训练图像。与此类似，史朋飞等人[79]设计了一种数据增强方法以模拟水下生物重叠、遮挡等显示不完全的情形，增强了网络模型鲁棒性。除了直接针对图像操作的数据增强方法，Li等人[80]提出了一种多任务训练方法，引入自监督去模糊子网络以获得高质量图像，同时设计了基于视角变换的改进空间变换网络，自适应丰富网络内的图像特征。上述方法均在URPC上实现了精度提升。

正负样本不平衡问题在水下应用中也十分严重，因此一些方法延用了重要样本加权的思想。SWIPNET(Sample-WeIghted hyPEr NETwork)模型[81,82]引入了一种噪声鲁棒的训练范式CMA，首先在每个训练迭代中减少未检测到的目标的损失权重，因为它们很可能是噪声数据，然后在模型趋于收敛时增加难例正样本的权重值，直至模型收敛。类似地，Boosting R-CNN模型[83]设计了多级RPN，并引入boosting reweighting难样本挖掘方法，在RPN错误地计算了样本的对象先验概率时，增加样本在检测头部的分类损失值，同时减少具有准确估计先验的简单样本的损失，在水下数据集Brackish[59]和通用目标检测数据集上均取得了性能提升。

4　基于深度学习的水下声呐图像目标检测

可见光图像仅在近距离水下探测时具有较高清晰度，在船舶海洋业应用中限制极大。成像声呐能

够在低可见度条件下可靠运行，是目前最常用的水下探测手段。成像声呐设备主要包括前视声呐、侧扫声呐、合成孔径声呐、干涉合成孔径声呐等，其中前三者最为常用[1]。成像声呐通常安装在水下航行器或水面船只拖曳设备上，在行进过程中不断发射和接收声信号，根据回波信号成像。声呐图像的自主目标识别(Autonomous Target Recognition, ATR)即目标检测，对可疑目标进行定位并确定类别。海水介质的非均匀性会造成声信号的衰减和畸变，同时各种漂浮物和颗粒都会增大声波传输过程中的多径效应[5]，使得声呐图像目标检测难度远大于可见光图像。随着深度学习技术的成熟，越来越多的研究者借助深度神经网络解决声呐图像目标检测难题。本章从声呐图像特点与数据集发展、模型设计和训练算法方面总结了前视、侧扫、合成孔径声呐图像目标检测相关研究。

4.1 声呐图像特点与数据集发展

前视声呐使用1个或多个波束对前方扇形区域进行扫描，需要扩大探测区域时通常转动波束或增加波束数量。其优点是可以使用多个频率的波束进行探测，能耗较低且尺寸较小，因此在民用和军事领域中被广泛应用；缺点是图像分辨率低，扇形图像包含的目标信息量少且对噪声敏感，旁瓣干扰严重。

侧扫声呐基于目标物对入射声波的反向散射原理，将回波数据逐行排列以生成图像，能够直观地反映水下目标物形态。其优点是图像分辨率高，左右声呐生成的矩形图像探测覆盖面大，因此常被用于大面积海域的勘探、搜救、探雷等任务；缺点是图像质量较低，难以从大图上辨识小目标轮廓。

合成孔径声呐是将合成孔径雷达原理推广到水声领域而形成的一种新型高分辨率水下成像声呐[1]，通过小孔基阵移动而在不同位置接受回波信号。其优点是图像分辨率和精度高，相比于侧扫声呐减弱了探测距离对图像质量的影响；缺点是数据处理量极大，对设备要求高，因此合成孔径声呐设备往往价格高昂。

由于前视声呐使用成本较低，研究者多在仿真环境或小型试验场地自采数据集进行算法验证。Image Gallery数据集[84]是由搭载在水下机器人上的前视声呐采集而成的，共有1 500幅图像，包含鱼、鲨鱼、沉船、管道、人等10类目标的位置标注。Singh等人[85]也使用此设备在室内模拟的海洋环境中捕获了1868幅前视声呐图像，包含瓶子、罐头、链条、挂钩等11类海洋垃圾目标的像素点标注，可以用于检测与分割模型训练。尽管如此，1 000多张图像的训练集规模依然很小，易造成训练欠拟合或过拟合问题。

侧扫和合成孔径声呐图像风格类似，因为难以在室内及小型试验场地中进行采集，湖试海试成本高昂，同时涉及军事保密问题，所以目前公开的数据集十分有限，现有的研究工作多在各自采集的非公开数据集上进行训练和测试。Barngrover等人[86]提供了一些真实水雷的侧扫声呐图像和合成图像，用于自主目标识别算法的训练。SeabedObjects-KLSG数据集[87]用于侧扫声呐图像分类任务，含有沉船残骸图像385张、溺水受害者36张、失事飞机62张、水雷129张、海底图像578张，如需用于目标检测模型训练，还需在原始声呐图像上进行矩形位置标注。最近，声呐常见目标检测数据集(Sonar Common Target Detection dataset, SCTD)[88]收集了497张分辨率较高的图像，包含水下沉船、失事飞机残骸、遇难者3类典型目标的位置标注，共计596个样本。这些样本以侧扫声呐图像为主，还包含了一些合成孔径声呐图像、干涉合成孔径声呐图像、前视声呐图像。虽然SCTD1.0图像数量较少，但它填补了开源的侧扫、合成孔径声呐图像目标检测数据集的空白。

4.2 模型设计

基于深度学习的声呐图像目标检测模型设计大致可分为3类：特征提取与分类模型、通用目标检测模型、语义分割模型。

早期训练数据相对匮乏时，研究者通常采用传统方法和深度学习相结合的思路设计目标检测算法，以借鉴深度学习模型提取图像特征的优势。2016年左右CNN被应用到前视声呐目标识别任务中[89]。在此基础上，Valdenegro-Toro[90]使用共享CNN提取的128维图像特征向量进行边界框和类标签的训练，与R-CNN[11]类似采用SVM作为分类器，分为多个步骤实现目标检测。Palomeras等人[91]采用CNN提取声呐图像特征，将检测器、分类器与概率网格图相结合，通过概率图过滤误报信息并与检测结果相组合，极大限度地提高了算法检测精度。Zhou等人[92]首先采用FCM和K-means聚类方法对声呐图像进行全局聚类，以获得更多的RoI，然后使用CNN提取特征，经过非线性变换器和Fisher判别器得到分类结果。该方法的检测精度和实时性较好，不亚于一些深度学习方法。

将CNN用于侧扫、合成孔径声呐图像切片分类也可实现定位效果。Gebhardt等人[93]采用CNN提取侧扫声呐海底图像中水雷的特征并进行分类，Hoang 等人[94]借助DenseNet[23]识别合成孔径声呐图像中的未爆炸弹药。由于侧扫声呐探测面积广，目标只占据图像中极小部分，因此他们将原始图像

切片逐一分类，可以得到粗略的定位结果。虽然此类方法得到的定位框并不能紧密包含目标，但相较于高分辨率的海底声呐图像，此定位误差可以忽略不计。

随着端到端模型训练的成熟，许多研究将通用目标检测模型应用到水下检测任务。相比于较易采集的水下可见光图像，3种声呐图像均缺少大型开源数据集，深度神经网络易在小规模数据集上训练会产生参数冗余和过拟合的问题，这极大限制了声呐图像目标检测的应用研究。因此，许多工作使用轻量化设计的骨干网络缓解此问题，其中YOLO系列模型[15–17]被频繁采用。

对于前视声呐图像，Fan等人[95]利用残差模块构建了32层骨干网络，取代了Mask R-CNN[13]中的Resnet50/101，在保证检测性能的同时大幅减少了网络的训练参数，这对实时性和嵌入式部署具有重要意义。类似地，Fan等人[96]对YOLOv4[17]中的骨干网络进行改进，以减少模型参数和网络深度；同时，他们借鉴了自适应空间特征融合模块(ASFF)优化了颈部模块PAFPN，以获得更好的特征融合效果。Zhang等人[97]也优化了YOLOv5骨干网络以提高检测速度。最近，Zhu等人[98]结合了Swin Transformer[26]和DCN[24]设计了骨干网络和检测头部，构建了一种无锚框检测模型STAFNet，在自采前视声呐数据集上对受害者、船只、飞机3类目标达到了优越的检测性能，领先于YOLOv5,Faster R-CNN,FCOS等经典模型。

对于侧扫、合成孔径声呐图像，Wang和Li等人[99,100]直接将YOLOv3应用于该任务即可实现较好的检测效果。陈禹蒲等人[101]改进了YOLOv3模型检测头部的采样过程，设计了一种超参数锚框映射关系对聚类后的锚框进行拉伸变换，改进了检测精度。虽然他们耗时数月采集了26 689张侧扫声呐图像，但由于大量区域是海底背景，最终符合要求的图像仅有237张，由此可见侧扫声呐数据的采集成本极高。为了应对侧扫图像目标稀疏和特征贫乏的特点，Yu等人[102]将Transformer[20]的自注意力机制引入到YOLOv5s的骨干网络和颈部模块，提高模型在全局图像中检测小目标的能力。Fu等人[103]也采用了空间和通道注意力模块来改善YOLOv5的颈部模块。李宝奇等人[104]设计了一种可扩张、可选择的轻量化CNN，改进了SSD[14]的骨干网络，在中国科学院声学研究所采集的高频合成孔径声呐图像数据上取得了优越的检测效果。Zhang等人[105]提出了一种具有灵活搜索空间和内存高效的可差分结构搜索算法(FL-DARTS)，自动设计轻量CNN处理雷达或声呐图像，在SCTD1.0和合成孔径雷达船舶检测数据集SSDD[106]上实现了良好的性能。

此外，基于语义分割的深度学习模型也可以用于声呐目标检测。Wu 等人[107]设计编码器-解码器结构对侧扫声呐图像进行像素级分类，连续大面积的前景类别像素点构成目标。MB-CEDN模型[108]设计多分支网络，通过级联的方式细化合成孔径声呐图像分割结果。然而此类方法容易受到海底地形的影响，降低小目标检测精度，在实际应用中性能可能不佳。

4.3 训练方法

为缓解声呐图像数据不足、样本不均衡等问题，许多工作采用迁移学习或图像生成等方法，隐式或者显式地扩充训练数据。

Fuchs等人[109]预先使用来自不同领域的数据训练网络模型，以学习通用特征，再通过迁移学习的方法将模型应用到前视声呐数据上。Lee等人[110]采用CNN设计了一种端到端的前视声呐图像合成方法，通过风格转换使仿真数据逼近真实数据，然后使用从水箱和海水中获得的真实水下声呐图像测试仿真图像。Lou等人[111]借鉴了显著性特征可视化方法和生成对抗网络(Generative Adversarial Networks, GAN)[112]学习光学图像和声呐图像之间的转换关系，来解决目标检测CNN欠拟合的问题。然而使用模拟生成的前视声呐图像训练的模型并未在实际场景中得到验证。Jegorova等人[113]基于Pix2Pix[112]引入了一种马尔可夫策略，旨在真实模拟声传感器、物体高度和环境因素的特定伪影，定量评估结果表明生成的图像与真实数据几乎没有区别。凡志邈等人[114]借鉴了合成孔径雷达图像转换思路，基于CycleGAN[115]实现光学图像到合成孔径声呐图像的风格迁移，利用生成图像训练的Mask R-CNN[13]能够在真实环境中良好应用。

迁移学习和图像生成方法也可结合使用。盛子旗等人[116]首先根据侧扫声呐成像机理建立水雷目标的仿真模型进行样本生成，然后采用开源数据集ImageNet[4]对深度卷积神经网络进行预训练，再分别用仿真和真实水雷样本对骨干网络进行微调，最后，将骨干网络嵌入Faster R-CNN, YOLOv3等目标检测模型，使用真实水雷样本进行训练。该方法分3步实现整个训练过程，大幅提高了模型检测精度。

5　总结与展望

水下图像目标检测技术在工业及军事应用中有着巨大的发展前景，受到越来越多研究者的关注。近年来随着深度学习的发展，该领域取得了较大突破，但仍存在一些问题，总结如下：

(1)水下复杂多变的环境使得图像信息易衰减失真，目标检测难度高。近几年基于深度学习的通用目标检测算法在骨干网络、颈部模块、检测头部、训练算法、推理策略方面均取得了众多研究成果。然而水下图像目标检测研究相对滞后，许多工作仅将数年前的深度学习模型稍加改动进行简单应用，对于该领域的特点和困难进行针对性的研究较少，比如在水下可见光图像中目标极易发生密集、遮挡、模糊等情况，在水下声呐图像中目标具有分布稀疏、特征匮乏等特点。

(2)水下可见光图像目标检测数据集众多，然而大部分数据集只包含少量类别的水下目标，一味地扩充图像数量并不能增加深度学习模型在更多类别目标中的通用性。同时，众多的数据集使得不同算法模型之间的公平对比存在困难。

(3)水下声呐图像由于采集成本高昂、涉及军事秘密等原因，公开数据集较少，限制了深度学习模型的应用。迁移学习、数据增强、图像生成等训练方法能够在一定程度上改善数据量不足的问题。现有研究工作的训练集和测试集规模小，虽然取得了良好的推理精度，但模型的泛化能力未能验证。

结合深度学习最新研究，对水下图像目标检测的未来研究做出如下展望：

(1)大规模数据集构建与Transformer模型研究。通过大规模数据训练的深度学习模型具有更好的精度和泛化性，因此构建大规模的水下可见光图像和声呐图像目标检测数据集是未来重要的发展方向。此外，CNN在处理图像数据中更关注局部信息，注重空间上的紧密元素，限制了数据集规模的上限。Transformer模型关注图像全局信息，计算复杂度更低，还能避免深层特征过度平滑，在大规模数据集上表现出更优越的性能。

(2)基于图像修复与目标检测的多任务模型研究。为了应对水下可见光图像模糊、声呐图像失真等问题，一些工作采用图像预处理方法[1,9]，但效果并不突出。深度学习领域中的多任务学习可以令一个网络同时学习多项任务，旨在不同任务之间能够协同优化，实现“1+1>2”。因此，将图像修复和目标检测作为子任务，设计端到端的多任务网络模型，可以更好地应对图像信息受损的问题。

(3)小样本学习相关训练算法研究。水下声呐图像数据量少，用于训练深度学习模型属于小样本学习任务。同时已有数据可能存在标注缺失、类别不明确、标注错误等问题。除了人工扩充数据与标注，研究数据增强、无监督、弱监督、半监督等训练算法，可以充分利用已有的少量声呐图像数据提高目标检测模型精度，也是重要的研究方向之一。

(4)多模态融合算法研究。多模态学习即是从多个模态表达或感知事物，比如通过两种不同成像原理的设备采集的数据进行协同与融合分析。未来在水下探测任务中，可以同时使用可见光相机、前视声呐、侧扫声呐采集图像数据，研究多模态数据特征之间的关联关系，有助于提高数据的利用率，构建鲁棒的算法系统。

参 考 文 献

[1] 郭戈, 王兴凯, 徐慧朴. 基于声呐图像的水下目标检测、识别与跟踪研究综述[J]. 控制与决策, 2018, 33(5): 906–922. doi: 10.13195/j.kzyjc.2017.1678.
GUO Ge, WANG Xingkai, and XU Huipu. Review on underwater target detection, recognition and tracking based on sonar image[J]. *Control and Decision*, 2018, 33(5): 906–922. doi: 10.13195/j.kzyjc.2017.1678.

[2] GOMES D, SAIF A F M S, and NANDI D. Robust underwater object detection with autonomous underwater vehicle: A comprehensive study[C]. 2020 International Conference on Computing Advancements, Dhaka, Bangladesh, 2020, 17. doi: 10.1145/3377049.3377052.

[3] KRIZHEVSKY A, SUTSKEVER I, and HINTON G E. ImageNet classification with deep convolutional neural networks[C]. The 25th International Conference on Neural Information Processing Systems, Lake Tahoe, USA, 2012: 1097–1105. doi: 10.5555/2999134.2999257.

[4] RUSSAKOVSKY O, DENG Jia, SU Hao, *et al*. ImageNet large scale visual recognition challenge[J]. *International Journal of Computer Vision*, 2015, 115(3): 211–252. doi: 10.1007/s11263-015-0816-y.

[5] 檀盼龙, 吴小兵, 张晓宇. 基于声呐图像的水下目标识别研究综述[J]. 数字海洋与水下攻防, 2022, 5(4): 342–353. doi: 10.19838/j.issn.2096-5753.2022.04.010.
TAN Panlong, WU Xiaobing, and ZHANG Xiaoyu. Review on underwater target recognition based on sonar image[J]. *Digital Ocean & Underwater Warfare*, 2022, 5(4): 342–353. doi: 10.19838/j.issn.2096-5753.2022.04.010.

[6] ZOU Zhengxia, CHEN Keyan, SHI Zhenwei, *et al*. Object detection in 20 years: A survey[EB/OL]. https://arxiv.org/pdf/1905.05055.pdf, 2019.

[7] 邵延华, 张铎, 楚红雨, 等. 基于深度学习的YOLO目标检测综述[J]. 电子与信息学报, 2022, 44(10): 3697–3708. doi: 10.11999/JEIT210790.
SHAO Yanhua, ZHANG Duo, CHU Hongyu, *et al*. A review of YOLO object detection based on deep learning[J]. *Journal of Electronics & Information Technology*, 2022, 44(10): 3697–3708. doi: 10.11999/JEIT210790.

[8] ZAIDI S S A, ANSARI M S, ASLAM A, *et al*. A survey of

modern deep learning based object detection models[J]. *Digital Signal Processing*, 2022, 126: 103514. doi: 10.1016/j.dsp.2022.103514.

[9] 林森, 赵颖. 水下光学图像中目标探测关键技术研究综述[J]. 激光与光电子学进展, 2020, 57(6): 060002. doi: 10.3788/LOP57.060002.
LIN Sen and ZHAO Ying. Review on key technologies of target exploration in underwater optical images[J]. *Laser & Optoelectronics Progress*, 2020, 57(6): 060002. doi: 10.3788/LOP57.060002.

[10] FAYAZ S, PARAH S A, and QURESHI G J. Underwater object detection: Architectures and algorithms-a comprehensive review[J]. *Multimedia Tools and Applications*, 2022, 81(15): 20871–20916. doi: 10.1007/s11042-022-12502-1.

[11] GIRSHICK R B, DONAHUE J, DARRELL T, *et al*. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, USA, 2014: 580–587. doi: 10.1109/CVPR.2014.81.

[12] REN Shaoqing, HE Kaiming, GIRSHICK R B, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137–1149. doi: 10.1109/TPAMI.2016.2577031.

[13] HE Kaiming, GKIOXARI G, DOLLÁR P, *et al*. Mask R-CNN[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(2): 386–397. doi: 10.1109/TPAMI.2018.2844175.

[14] LIU Wei, ANGUELOV D, ERHAN D, *et al*. SSD: Single shot MultiBox detector[C]. 14th European Conference on Computer Vision, Amsterdam, The Netherlands, 2016: 21–37. doi: 10.1007/978-3-319-46448-0_2.

[15] REDMON J, DIVVALA S, GIRSHICK R, *et al*. You only look once: Unified, real-time object detection[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 779–788. doi: 10.1109/CVPR.2016.91.

[16] REDMON J and FARHADI A. YOLOv3: An Incremental Improvement[EB/OL].https://arxiv.org/pdf/1804.02767.pdf, 2018.

[17] BOCHKOVSKIY A, WANG C Y, and LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection[EB/OL].https://arxiv.org/pdf/2004.10934.pdf, 2020.

[18] DUAN Kaiwen, BAI Song, XIE Lingxi, *et al*. CenterNet: Keypoint triplets for object detection[C]. 2019 IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 2019: 6568–6577. doi: 10.1109/ICCV.2019.00667.

[19] TIAN Zhi, SHEN Chunhua, CHEN Hao, *et al*. FCOS: Fully convolutional one-stage object detection[C]. 2019 IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 2019: 9626–9635. doi: 10.1109/ICCV.2019.00972.

[20] CARION N, MASSA F, SYNNAEVE G, *et al*. End-to-end object detection with transformers[C]. 16th European Conference on Computer Vision, Glasgow, UK, 2020: 213–229. doi: 10.1007/978-3-030-58452-8_13.

[21] CHEN Kai, WANG Jiaqi, PANG Jiangmiao, *et al*. MMDetection: Open MMLab detection toolbox and benchmark[EB/OL].https://arxiv.org/abs/1906.07155, 2019.

[22] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, *et al*. Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 770–778. doi: 10.1109/CVPR.2016.90.

[23] HUANG Gao, LIU Zhuang, VAN DER MAATEN L, *et al*. Densely connected convolutional networks[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 2261–2269. doi: 10.1109/CVPR.2017.243.

[24] ZHU Xizhou, HU Han, LIN S, *et al*. Deformable ConvNets V2: More deformable, better results[C]. 2019 IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 9300–9308. doi: 10.1109/CVPR.2019.00953.

[25] TOLSTIKHIN I O, HOULSBY N, KOLESNIKOV A, *et al*. MLP-mixer: An all-MLP architecture for vision[C/OL]. Proceedings of the 35th International Conference on Neural Information Processing Systems, 2021: 24261–24272.

[26] LIU Ze, LIN Yutong, CAO Yue, *et al*. Swin transformer: Hierarchical vision transformer using shifted windows[C]. 2021 IEEE/CVF International Conference on Computer Vision, Montreal, Canada, 2021: 9992–10002. doi: 10.1109/ICCV48922.2021.00986.

[27] HOWARD A, SANDLER M, CHEN Bo, *et al*. Searching for MobileNetV3[C]. 2019 IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 2019: 1314–1324. doi: 10.1109/ICCV.2019.00140.

[28] LIN T Y, DOLLÁR P, GIRSHICK R, *et al*. Feature pyramid networks for object detection[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 936–944. doi: 10.1109/CVPR.2017.106.

[29] LIU Shu, QI Lu, QIN Haifang, *et al*. Path aggregation network for instance segmentation[C]. 2018 IEEE/CVF

Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 8759–8768. doi: 10.1109/CVPR.2018.00913.

[30] ZONG Zhuofan, CAO Qianggang, and LENG Biao. RCNet: Reverse feature pyramid and cross-scale shift network for object detection[C]. The 29th ACM International Conference on Multimedia, Chengdu, China, 2021: 5637–5645. doi: 10.1145/3474085.3475708.

[31] WANG Ning, GAO Yang, CHEN Hao, *et al*. NAS-FCOS: Fast neural architecture search for object detection[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 11940–11948. doi: 10.1109/CVPR42600.2020.01196.

[32] WANG Jiaqi, ZHANG Wenwei, CAO Yuhang, *et al*. Side-aware boundary localization for more precise object detection[C]. 16th European Conference on Computer Vision, Glasgow, UK, 2020: 403–419. doi: 10.1007/978-3-030-58548-8_24.

[33] WU Yue, CHEN Yinpeng, YUAN Lu, *et al*. Rethinking classification and localization for object detection[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 10183–10192. doi: 10.1109/CVPR42600.2020.01020.

[34] FENG Chengjian, ZHONG Yujie, GAO Yu, *et al*. TOOD: Task-aligned one-stage object detection[C]. 2021 IEEE/CVF International Conference on Computer Vision, Montreal, Canada, 2021: 3490–3499. doi: 10.1109/ICCV48922.2021.00349.

[35] LIN T Y, GOYAL P, GIRSHICK R B, *et al*. Focal loss for dense object detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(2): 318–327. doi: 10.1109/TPAMI.2018.2858826.

[36] PANG Jiangmiao, CHEN Kai, SHI Jianping, *et al*. Libra R-CNN: Towards balanced learning for object detection[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 821–830. doi: 10.1109/CVPR.2019.00091.

[37] MA Yuchen, LIU Songtao, LI Zeming, *et al*. IQDet: Instance-wise quality distribution sampling for object detection[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, USA, 2021: 1717–1725. doi: 10.1109/CVPR46437.2021.00176.

[38] ZHANG Shifeng, CHI Cheng, YAO Yongqiang, *et al*. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 9756–9765. doi: 10.1109/CVPR42600.2020.00978.

[39] GE Zheng, LIU Songtao, LI Zeming, *et al*. OTA: Optimal transport assignment for object detection[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, USA, 2021: 303–312. doi: 10.1109/CVPR46437.2021.00037.

[40] OKSUZ K, CAM B C, AKBAS E, *et al*. Rank & sort loss for object detection and instance segmentation[C]. 2021 IEEE/CVF International Conference on Computer Vision, Montreal, Canada, 2021: 2989–2998. doi: 10.1109/ICCV48922.2021.00300.

[41] ZHANG Hongkai, CHANG Hong, MA Bingpeng, *et al*. Dynamic R-CNN: Towards high quality object detection via dynamic training[C]. 16th European Conference on Computer Vision, Glasgow, UK, 2020: 260–275. doi: 10.1007/978-3-030-58555-6_16.

[42] GAO Yan, WANG Qimeng, TANG Xu, *et al*. Decoupled IoU regression for object detection[C]. The 29th ACM International Conference on Multimedia, Chengdu, China, 2021: 5628–5636. doi: 10.1145/3474085.3475707.

[43] GUO M, HAQUE A, HUANG Dean, *et al*. Dynamic task prioritization for multitask learning[C]. 15th European Conference on Computer Vision, Munich, Germany, 2018: 282–299. doi: 10.1007/978-3-030-01270-0_17.

[44] CAI Qi, PAN Yingwei, WANG Yu, *et al*. Learning a unified sample weighting network for object detection[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 14161–14170. doi: 10.1109/CVPR42600.2020.01418.

[45] LI Xiang, WANG Wenhai, HU Xiaolin, *et al*. Generalized focal loss V2: Learning reliable localization quality estimation for dense object detection[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, USA, 2021: 11627–11636. doi: 10.1109/CVPR46437.2021.01146.

[46] BODLA N, SINGH B, CHELLAPPA R, *et al*. Soft-NMS - Improving object detection with one line of code[C]. 2017 IEEE International Conference on Computer Vision, Venice, Italy, 2017: 5562–5570. doi: 10.1109/ICCV.2017.593.

[47] LUO Zekun, FANG Zheng, ZHENG Sixiao, *et al*. NMS-loss: Learning with non-maximum suppression for crowded pedestrian detection[C]. 2021 International Conference on Multimedia Retrieval, Taipei, China, 2021: 481–485. doi: 10.1145/3460426.3463588.

[48] SUN Peize, ZHANG Rufeng, JIANG Yi, *et al*. Sparse R-CNN: End-to-end object detection with learnable proposals[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, USA, 2021: 14449–14458. doi: 10.1109/CVPR46437.2021.01422.

[49] WANG Jianfeng, SONG Lin, LI Zeming, *et al*. End-to-end

object detection with fully convolutional network[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, USA, 2021: 15844–15853. doi: 10.1109/CVPR46437.2021.01559.

[50] CAO Xiang, LUO Yihao, XIAO Yi, *et al*. Blind image super-resolution based on prior correction network[J]. *Neurocomputing*, 2021, 463: 525–534. doi: 10.1016/j.neucom.2021.07.070.

[51] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, *et al*. The PASCAL visual object classes (VOC) challenge[J]. *International Journal of Computer Vision*, 2010, 88(2): 303–338. doi: 10.1007/s11263-009-0275-4.

[52] LIN T Y, MAIRE M, BELONGIE S, *et al*. Microsoft COCO: Common objects in context[C]. 13th European Conference on Computer Vision, Zurich, Switzerland, 2014: 740–755. doi: 10.1007/978-3-319-10602-1_48.

[53] URPC. Underwater robot professional contest[EB/OL]. http://www.urpc.org.cn/index.html, 2022.

[54] LIU Risheng, FAN Xin, ZHU Ming, *et al*. Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020, 30(12): 4861–4875. doi: 10.1109/TCSVT.2019.2963772.

[55] LIU Chongwei, WANG Zhihui, WANG Shijie, *et al*. A new dataset, Poisson GAN and AquaNet for underwater object grabbing[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(5): 2831–2844. doi: 10.1109/TCSVT.2021.3100059.

[56] FAN Baojie, CHEN Wei, CONG Yang, *et al*. Dual refinement underwater object detection network[C]. 16th European Conference on Computer Vision, Glasgow, UK, 2020: 275–291. doi: 10.1007/978-3-030-58565-5_17.

[57] LIU Chongwei, LI Haojie, WANG Shuchang, *et al*. A dataset and benchmark of underwater object detection for robot picking[C]. 2021 IEEE International Conference on Multimedia & Expo Workshops, Shenzhen, China, 2021: 1–6. doi: 10.1109/ICMEW53276.2021.9455997.

[58] Fish4Knowledge.https://homepages.inf.ed.ac.uk/rbf/Fish4Knowledge/index.html, 2013.

[59] PEDERSEN M, HAURUM J B, GADE R, *et al*. Detection of marine animals in a new underwater dataset with varying visibility[C]. 2019 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, USA, 2019: 18–26.

[60] FULTON M, HONG J, ISLAM M J, *et al*. Robotic detection of marine litter using deep visual detection models[C]. 2019 International Conference on Robotics and Automation, Montreal, Canada, 2019: 5752–5758. doi: 10.1109/ICRA.2019.8793975.

[61] HONG J, FULTON M, and SATTAR J. TrashCan: A semantically-segmented dataset towards visual detection of marine debris[EB/OL].https://arxiv.org/abs/2007.08097, 2020.

[62] ISLAM M J, EDGE C, and XIAO Yuyang. Semantic segmentation of underwater imagery: Dataset and benchmark[C]. 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems, Las Vegas, USA, 2020: 1769–1776. doi: 10.1109/IROS45743.2020.9340821.

[63] LU Huimin, LI Yujie, UEMURA T, *et al*. FDCNet: Filtering deep convolutional network for marine organism classification[J]. *Multimedia Tools and Applications*, 2018, 77(17): 21847–21860. doi: 10.1007/s11042-017-4585-1.

[64] LI Chongyi, GUO Chunle, REN Wenqi, *et al*. An underwater image enhancement benchmark dataset and beyond[J]. *IEEE Transactions on Image Processing*, 2020, 29: 4376–4389. doi: 10.1109/TIP.2019.2955241.

[65] JIAN Muwei, QI Qiang, YU Hui, *et al*. The extended marine underwater environment database and baseline evaluations[J]. *Applied Soft Computing*, 2019, 80: 425–437. doi: 10.1016/j.asoc.2019.04.025.

[66] KEZEBOU L, OLUDARE V, PANETTA K, *et al*. Underwater object tracking benchmark and dataset[C]. 2019 IEEE International Symposium on Technologies for Homeland Security, Woburn, USA, 2019: 1–6. doi: 10.1109/HST47167.2019.9032954.

[67] PANETTA K, KEZEBOU L, OLUDARE V, *et al*. Comprehensive underwater object tracking benchmark dataset and underwater image enhancement with GAN[J]. *IEEE Journal of Oceanic Engineering*, 2022, 47(1): 59–75. doi: 10.1109/JOE.2021.3086907.

[68] SUNG M, YU S C, and GIRDHAR Y. Vision based real-time fish detection using convolutional neural network[C]. OCEANS 2017, Aberdeen, UK, 2017: 1–6. doi: 10.1109/OCEANSE.2017.8084889.

[69] CHRISTENSEN J H, MOGENSEN L V, GALEAZZI R, *et al*. Detection, localization and classification of fish and fish species in poor conditions using convolutional neural networks[C]. 2018 IEEE/OES Autonomous Underwater Vehicle Workshop, Porto, Portugal, 2018: 1–6. doi: 10.1109/AUV.2018.8729798.

[70] MANDAL R, CONNOLLY R M, SCHLACHER T A, *et al*. Assessing fish abundance from underwater video using deep neural networks[C]. 2018 International Joint Conference on Neural Networks, Rio de Janeiro, Brazil, 2018: 1–6. doi: 10.1109/IJCNN.2018.8489482.

[71] KNAUSGÅRD K M, WIKLUND A, SØRDALEN T K, *et al*. Temperate fish detection and classification: A deep learning based approach[J]. *Applied Intelligence*, 2022,

52(6): 6988–7001. doi: 10.1007/s10489-020-02154-9.

[72] 叶赵兵, 段先华, 赵楚. 改进YOLOv3-SPP水下目标检测研究[J]. 计算机工程与应用, 2023, 59(6): 231–240. doi: 10.3778/j.issn.1002-8331.2204-0264.
YE Zhaobing, DUAN Xianhua, and ZHAO Chu. Research on underwater target detection by improved YOLOv3-SPP[J]. *Computer Engineering and Applications*, 2023, 59(6): 231–240. doi: 10.3778/j.issn.1002-8331.2204-0264.

[73] 张艳, 李星汕, 孙叶美, 等. 基于通道注意力与特征融合的水下目标检测算法[J]. 西北工业大学学报, 2022, 40(2): 433–441. doi: 10.3969/j.issn.1000-2758.2022.02.025.
ZHANG Yan, LI Xingshan, SUN Yemei, *et al*. Underwater object detection algorithm based on channel attention and feature fusion[J]. *Journal of Northwestern Polytechnical University*, 2022, 40(2): 433–441. doi: 10.3969/j.issn.1000-2758.2022.02.025.

[74] 王蓉蓉, 蒋中云. 基于改进CenterNet的水下目标检测算法[J]. 激光与光电子学进展, 2023, 60(2): 0215001.
WANG Rongrong and JIANG Zhongyun. Underwater object detection algorithm based on improved CenterNet[J]. *Laser & Optoelectronics Progress*, 2023, 60(2): 0215001.

[75] 蔡达, 范保杰. 基于空间特征选择的水下目标检测方法[J]. 信息与控制, 2022, 51(2): 214–222. doi: 10.13976/j.cnki.xk.2022.1597.
CAI Da and FAN Baojie. Spatial feature selection for underwater object detection[J]. *Information and Control*, 2022, 51(2): 214–222. doi: 10.13976/j.cnki.xk.2022.1597.

[76] 喻明毫, 高建瓴. 轻量级水下目标检测器LUDet[J]. 计算机工程与科学, 2022, 44(9): 1638–1645. doi: 10.3969/j.issn.1007-130X.2022.09.014.
YU Minghao and GAO Jianling. LUDet: A lightweight underwater object detector[J]. *Computer Engineering & Science*, 2022, 44(9): 1638–1645. doi: 10.3969/j.issn.1007-130X.2022.09.014.

[77] LIANG Xutao and SONG Pinhao. Excavating RoI attention for underwater object detection[C]. 2022 IEEE International Conference on Image Processing, Bordeaux, France, 2022: 2651–2655. doi: 10.1109/ICIP46576.2022.9897515.

[78] LIN Weihong, ZHONG Jiaxing, LIU Shan, *et al*. ROIMIX: Proposal-fusion among multiple images for underwater object detection[C]. 2020 IEEE International Conference on Acoustics, Speech and Signal Processing, Barcelona, Spain, 2022: 2588–2592. doi: 10.1109/ICASSP40776.2020.9053829.

[79] 史朋飞, 韩松, 倪建军, 等. 结合数据增强和改进YOLOv4的水下目标检测算法[J]. 电子测量与仪器学报, 2022, 36(3): 113–121. doi: 10.13382/j.jemi.B2104168.
SHI Pengfei, HAN Song, NI Jianjun, *et al*. Underwater object detection algorithm combining data enhancement and improved YOLOv4[J]. *Journal of Electronic Measurement and Instrumentation*, 2022, 36(3): 113–121. doi: 10.13382/j.jemi.B2104168.

[80] LI Xiuyuan, LI Fengchao, YU Jiangang, *et al*. A high-precision underwater object detection based on joint self-supervised deblurring and improved spatial transformer network[EB/OL].https://arxiv.org/abs/2203.04822, 2022.

[81] CHEN Long, LIU Zhihua, TONG Lei, *et al*. Underwater object detection using Invert Multi-Class Adaboost with deep learning[C]. 2020 International Joint Conference on Neural Networks, Glasgow, UK, 2020: 1–8. doi: 10.1109/IJCNN48605.2020.9207506.

[82] CHEN Long, ZHOU Feixiang, WANG Shengke, *et al*. SWIPENET: Object detection in noisy underwater scenes[J]. *Pattern Recognition*, 2022, 132: 108926. doi: 10.1016/j.patcog.2022.108926.

[83] SONG Pinhao, LI Pengteng, DAI Linhui, *et al*. Boosting R-CNN: Reweighting R-CNN samples by RPN's error for underwater object detection[J]. *Neurocomputing*, 2023, 530: 150–164. doi: 10.1016/j.neucom.2023.01.088.

[84] Sound Metrics. Image Gallery[EB/OL]. http://www.soundmetrics.com/Image-Gallery, 2020.

[85] SINGH D and VALDENEGRO-TORO M. The marine debris dataset for forward-looking sonar semantic segmentation[C]. 2021 IEEE/CVF International Conference on Computer Vision Workshops, Montreal, Canada, 2021: 3734–3742. doi: 10.1109/ICCVW54120.2021.00417.

[86] BARNGROVER C, KASTNER R, and BELONGIE S. Semisynthetic versus real-world sonar training data for the classification of mine-like objects[J]. *IEEE Journal of Oceanic Engineering*, 2015, 40(1): 48–56. doi: 10.1109/JOE.2013.2291634.

[87] HUO Guanying, WU Ziyin, and LI Jiabiao. Underwater object classification in Sidescan sonar images using deep transfer learning and semisynthetic training data[J]. *IEEE Access*, 2020, 8: 47407–47418. doi: 10.1109/ACCESS.2020.2978880.

[88] 周彦, 陈少昌, 吴可, 等. SCTD1.0: 声呐常见目标检测数据集[J]. 计算机科学, 2021, 48(S2): 334–339. doi: 10.11896/jsjkx.210100138.
ZHOU Yan, CHEN Shaochang, WU Ke, *et al*. SCTD1.0: Sonar common target detection dataset[J]. *Computer Science*, 2021, 48(S2): 334–339. doi: 10.11896/jsjkx.210100138.

[89] VALDENEGRO-TORO M. Object recognition in forward-looking sonar images with convolutional neural

networks[C]. OCEANS 2016 MTS/IEEE Monterey, Monterey, USA, 2016: 1–6. doi: 10.1109/OCEANS.2016.7761140.

[90] VALDENEGRO-TORO M. End-to-end object detection and recognition in forward-looking sonar images with convolutional neural networks[C]. 2016 IEEE/OES Autonomous Underwater Vehicles, Tokyo, Japan, 2016: 144–150. doi: 10.1109/AUV.2016.7778662.

[91] PALOMERAS N, FURFARO T, WILLIAMS D P, *et al*. Automatic target recognition for mine countermeasure missions using forward-looking sonar data[J]. *IEEE Journal of Oceanic Engineering*, 2022, 47(1): 141–161. doi: 10.1109/JOE.2021.3103269.

[92] ZHOU Tian, SI Jikun, WANG Luyao, *et al*. Automatic detection of underwater small targets using forward-looking sonar images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 4207912. doi: 10.1109/TGRS.2022.3181417.

[93] GEBHARDT D, PARIKH K, DZIECIUCH I, *et al*. Hunting for naval mines with deep neural networks[C]. OCEANS 2017, Anchorage, UK, 2017: 1–5.

[94] HOANG T, DALTON K S, GERG I D, *et al*. Domain enriched deep networks for munition detection in underwater 3D sonar imagery[C]. 2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 2022: 815–818. doi: 10.1109/IGARSS46834.2022.9884793.

[95] FAN Zhimiao, XIA Weijie, LIU Xue, *et al*. Detection and segmentation of underwater objects from forward-looking sonar based on a modified Mask RCNN[J]. *Signal, Image and Video Processing*, 2021, 15(6): 1135–1143. doi: 10.1007/s11760-020-01841-x.

[96] FAN Xinnan, LU Liang, SHI Pengfei, *et al*. A novel sonar target detection and classification algorithm[J]. *Multimedia Tools and Applications*, 2022, 81(7): 10091–10106. doi: 10.1007/s11042-022-12054-4.

[97] ZHANG Haoting, TIAN Mei, SHAO Gaoping, *et al*. Target detection of forward-looking sonar image based on improved YOLOv5[J]. *IEEE Access*, 2022, 10: 18023–18034. doi: 10.1109/ACCESS.2022.3150339.

[98] ZHU Xingyu, LIANG Yingshuo, ZHANG Jianlei, *et al*. STAFNet: Swin transformer based anchor-free network for detection of forward-looking sonar imagery[C]. The 2022 International Conference on Multimedia Retrieval, Newark, USA, 2022: 443–450. doi: 10.1145/3512527.3531398.

[99] WANG Yanmei, LIU Jiaxin, YU Siquan, *et al*. Underwater object detection based on YOLO-v3 network[C]. 2021 IEEE International Conference on Unmanned Systems, Beijing, China, 2021: 571–575. doi: 10.1109/ICUS52573.2021.9641489.

[100] LI Jiawen and CAO Xiang. Target recognition and detection in side-scan sonar images based on YOLO v3 model[C]. 41st Chinese Control Conference, Hefei, China, 2022: 7186–7190. doi: 10.23919/CCC55666.2022.9902742.

[101] 陈禹蒲, 马晓川, 李璇. 基于YOLOv3锚框优化的侧扫声呐图像目标检测[J]. 信号处理, 2022, 38(11): 2359–2371. doi: 10.16798/j.issn.1003-0530.2022.11.013.
CHEN Yupu, MA Xiaochuan, and LI Xuan. Target detection in side scan sonar images based on YOLOv3 anchor boxes optimization[J]. *Journal of Signal Processing*, 2022, 38(11): 2359–2371. doi: 10.16798/j.issn.1003-0530.2022.11.013.

[102] YU Yongcan, ZHAO Jianhu, GONG Quanhua, *et al*. Real-time underwater maritime object detection in side-scan sonar images based on transformer-YOLOv5[J]. *Remote Sensing*, 2021, 13(18): 3555. doi: 10.3390/rs13183555.

[103] FU Shunan, XU Feng, LIU Jia, *et al*. Underwater small object detection in side-scan sonar images based on improved YOLOv5[C]. 3rd International Conference on Geology, Mapping and Remote Sensing, Zhoushan, China, 2022: 446–453. doi: 10.1109/ICGMRS55602.2022.9849382.

[104] 李宝奇, 黄海宁, 刘纪元, 等. 基于改进SSD的合成孔径声呐图像水下多尺度目标轻量化检测模型[J]. 电子与信息学报, 2021, 43(10): 2854–2862. doi: 10.11999/JEIT201042.
LI Baoqi, HUANG Haining, LIU Jiyuan, *et al*. Synthetic aperture sonar underwater multi-scale target efficient detection model based on improved single shot detector[J]. *Journal of Electronics & Information Technology*, 2021, 43(10): 2854–2862. doi: 10.11999/JEIT201042.

[105] ZHANG Peng, TANG Jinsong, ZHONG Heping, *et al*. Self-trained target detection of radar and sonar images using automatic deep learning[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 4701914. doi: 10.1109/TGRS.2021.3096011.

[106] LI Jianwei, QU Changwen, and SHAO Jiaqi. Ship detection in SAR images based on an improved faster R-CNN[C]. 2017 SAR in Big Data Era: Models, Methods and Applications, Beijing, China, 2017: 1–6. doi: 10.1109/BIGSARDATA.2017.8124934.

[107] WU Meihan, WANG Qi, RIGALL e, *et al*. ECNet: Efficient convolutional networks for side scan sonar image segmentation[J]. *Sensors*, 2019, 19(9): 2009. doi: 10.3390/s19092009.

[108] SLEDGE I J, EMIGH M S, KING J L, *et al*. Target detection and segmentation in circular-scan synthetic aperture sonar images using Semisupervised convolutional encoder-decoders[J]. *IEEE Journal of Oceanic Engineering*,

2022, 47(4): 1099–1128. doi: 10.1109/JOE.2022.3152863.

[109] FUCHS L R, GÄLLSTRÖM A, and FOLKESSON J. Object recognition in forward looking sonar images using transfer learning[C]. 2018 IEEE/OES Autonomous Underwater Vehicle Workshop, Porto, Portugal, 2018, 1–6. doi: 10.1109/AUV.2018.8729686.

[110] LEE S, PARK B, and KIM A. Deep learning from shallow dives: Sonar image generation and training for underwater object detection[EB/OL].https://arxiv.org/abs/1810.07990, 2018.

[111] LOU Guanting, ZHENG Ronghao, LIU Meiqin, *et al*. Automatic target recognition in forward-looking sonar images using transfer learning[C]. Global Oceans 2020: Singapore – U. S. Gulf Coast, Biloxi, USA, 2020: 1–6. doi: 10.1109/IEEECONF38699.2020.9389217.

[112] ISOLA P, ZHU Junyan, ZHOU Tinghui, *et al*. Image-to-image translation with conditional adversarial networks[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 5967–5976. doi: 10.1109/CVPR.2017.632.

[113] JEGOROVA M, KARJALAINEN A I, VAZQUEZ J, *et al*. Full-scale continuous synthetic sonar data generation with markov conditional generative adversarial networks[C]. 2020 IEEE International Conference on Robotics and Automation, Paris, France, 2020: 3168–3174. doi: 10.1109/ICRA40945.2020.9197353.

[114] 凡志邈, 夏伟杰, 刘雪. 基于修正Cycle GAN的声呐图像库构建方法研究[J]. 声学技术, 2021, 40(6): 890–894. doi: 10.16300/j.cnki.1000-3630.2021.06.023.
FAN Zhimiao, XIA Weijie, and LIU Xue. Modified CycleGAN based sonar image library construction[J]. *Technical Acoustics*, 2021, 40(6): 890–894. doi: 10.16300/j.cnki.1000-3630.2021.06.023.

[115] ZHU Junyan, PARK T, ISOLA P, *et al*. Unpaired image-to-image translation using cycle-consistent adversarial networks[C] 2017 IEEE International Conference on Computer Vision, Venice, Italy, 2017: 2242–2251. doi: 10.1109/ICCV.2017.244.

[116] 盛子旗, 霍冠英. 样本仿真结合迁移学习的声呐图像水雷检测[J]. 智能系统学报, 2021, 16(2): 385–392. doi: 10.11992/tis.202101030.
SHENG Ziqi and HUO Guanying. Detection of underwater mine target in sidescan sonar image based on sample simulation and transfer learning[J]. *CAAI Transactions on Intelligent Systems*, 2021, 16(2): 385–392. doi: 10.11992/tis.202101030.

罗逸豪：男，博士，研究方向为深度学习、计算机视觉、声呐图像处理.
刘奇佩：男，博士，研究方向为水声信号处理、声呐图像处理.
张 吟：男，高级工程师，研究方向为声呐图像处理.
周河宇：男，博士，研究方向为深度学习、计算机视觉.
张钧陶：男，工程师，研究方向为深度学习、计算机视觉.
曹 翔：男，博士，研究方向为深度学习、计算机视觉.

责任编辑：余 蓉