

## 全球尺度下的海洋鱼类图像智能分类研究进展

周鹏<sup>①②③</sup> 李昌永<sup>③</sup> 步雨馨<sup>①②</sup> 周芷诺<sup>③</sup> 王春生<sup>①②</sup> 沈红斌<sup>③</sup> 潘小勇<sup>\*③</sup>

<sup>①</sup>(自然资源部第二海洋研究所 杭州 310012)

<sup>②</sup>(自然资源部海洋生态系统动力学实验室 杭州 310012)

<sup>③</sup>(上海交通大学图像处理与模式识别研究所 上海 200240)

**摘要:** 在全球尺度上了解鱼类物种组成、丰度及时空分布等,将有助于其生物多样性保护。水下图像采集是获取鱼类物种多样性数据的主要调查手段之一,但图像信息分析工作耗时耗力。2015年以来,海洋鱼类图像数据集更新和深度学习模型算法优化等方面取得了一系列进展,但细粒度分类表现仍显不足,研究成果的生产实践应用相对薄弱。因此,该文首先分析海洋相关行业对鱼类自动化图像分类的需求,然后综合介绍鱼类图像数据集和深度学习算法应用,并分析了所面临的小样本下的细粒度分析等主要挑战及相应解决方法。最后探讨了基于深度学习的海洋鱼类图像自动化分类对相关图像信息处理研究及应用平台对未来在生态环境监测等海洋相关产业领域的重要性及其前景。该文旨在为快速了解基于深度学习的海洋鱼类图像自动化分类的研究背景、进展和未来方向的工作者提供相关信息。

**关键词:** 深度学习; 海洋鱼类; 生物多样性; 小样本; 图像分类

中图分类号: TN081; TN911.7

文献标识码: A

文章编号: 1009-5896(2025)03-0001-12

DOI: [10.11999/JEIT231365](https://doi.org/10.11999/JEIT231365)

## A Review of the Artificial Intelligence-based Image Classification of Fishes in the Global Oceans

ZHOU Peng<sup>①②③</sup> LI Changyong<sup>③</sup> BU Yuxin<sup>①②</sup> ZHOU Zhinuo<sup>③</sup>

WANG Chunsheng<sup>①②</sup> SHEN Hongbin<sup>③</sup> PAN Xiaoyong<sup>③</sup>

<sup>①</sup>(Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou 310012, China)

<sup>②</sup>(Key Laboratory of Marine Ecosystem Dynamics, Ministry of Natural Resources, Hangzhou 310012, China)

<sup>③</sup>(Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, and Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai 200240, China)

**Abstract:** Understanding the species composition, abundance and temporal and spatial distribution of fish on a global scale will help their biodiversity conservation. Underwater image acquisition is one of the main means to survey fish species diversity, but image data analysis is time-consuming and labor-intensive. Since 2015, a series of progress has been made in updating the datasets of marine fish images and optimizing the algorithm of deep learning models, but the performance of fine-grained classification is still insufficient, and the production practice application of research results is relatively weak. Therefore, the need for automated fish image classification in marine investigations is firstly studied. Then a comprehensive introduction to fish image datasets and deep learning algorithm applications is provided, and the main challenges and the corresponding solutions are analyzed. Finally, the importance of automated classification of marine fish images for related image information processing research is discussed, and its prospects in the field of marine monitoring are summarized.

**Key words:** Deep learning; Marine fish; Biodiversity; Few shots; Image classification

收稿日期: 2023-12-11; 改回日期: 2024-03-29; 网络出版: 2024-04-12

\*通信作者: 潘小勇 [2008xypan@sjtu.edu.cn](mailto:2008xypan@sjtu.edu.cn)

基金项目: 国家重点研发计划(2023YFC2811502), 上海交通大学深蓝计划(SL2022ZD108, SL2021MS005)

Foundation Items: The National Key Research and Development Program of China (2023YFC2811502), The Oceanic Interdisciplinary Program of Shanghai Jiao Tong University (SL2022ZD108, SL2021MS005)

## 1 引言

海洋约占地球表面积的71%。约占地球上总水量的97%，是地球最大的连续水体生态系统。海洋生物是其重要组成部分，包括多孔动物、刺胞动物、线虫动物、环节动物、节肢动物、软体动物、棘皮动物、脊索动物等真核生物以及原核生物等，多样的海洋生物发挥着各自的作用，维持着海洋系统正常生态功能。其中，海洋鱼类由于其巨大的经济和生态价值，其物种丰度和多样性深受关注。海洋鱼类超过10 000种，分为圆口纲、软骨鱼纲、硬骨鱼纲。鱼类外观形态特征丰富多样，是其分类的重要依据，人类在认识海洋、利用海洋和保护海洋的过程中已积累了大量的海洋鱼类样本及图像资料，如World Atlas of Marine fishes<sup>[1]</sup>，Field guide to marine fishes of tropical Australia<sup>[2]</sup>，中国海洋鱼类<sup>[3-5]</sup>。

遗憾的是，过度开发、环境污染、等对海洋生物多样性构成了重大威胁。在全球尺度上较全面了解鱼类物种组成、丰度及时空分布等是海洋生态系统健康可持续性发展的基础，可以帮助研究人员更好地了解生态系统变化，进而采取适当行动保护生物多样性，其重要性体现在海洋资源开发、生态保护、防灾减灾、海防安全、科学认知、海洋渔业、滨海旅游等各行业需求中(如表1所示)。

在全球海洋生物多样性保护背景下，基于图像采集的调查方式由于对生态系统的干扰或损伤小，逐渐成为海洋鱼类调查主要技术手段之一，如水下摄像拖体、潜水器、海底观测网等。随着海洋鱼类调查监测范围不断扩大、调查监测能力不断提高、

调查监测强度不断加大，海洋鱼类图像数据获取的频率不断提高，基于图像的海洋生物分类鉴定需求相应得不断增加。然而，海洋鱼类种类数量庞大，基于图像的海洋鱼类识别具有挑战性，鉴定人员需要了解相关类别分类知识，掌握分类方法，根据形态特征进行检索，有些样品鉴定需要大量翻阅分类学资料，效率低、耗时长，准确性受鉴定人员专业知识和经验影响。面对日益增长的海洋生物调查要求，分类鉴定专业人员相对缺乏，限制了全球海洋生物多样性评估。鉴于高效海洋鱼类图像自动化专业分类鉴定技术的重要性和紧迫性，提高海洋鱼类图像专业分类鉴定自动化，实现高效准确分类、降低调查成本，提高生态监测能力，是目前国内外海洋生物信息学研究和计算机应用的热点之一。

由于鱼类图像自动化分析的重要性与紧迫性，国内外工作者开展了较为广泛的研究。鉴于海洋调查方法与技术手段不断提高、调查范围不断扩大，鱼类数据集也在不断更新，而且深度学习模型的层出不穷，因此，与以往综述相比，本文介绍了一些最新鱼类图像数据集和深度学习模型在鱼类自动化分类上的应用进展，并分析了目前所面临的主要问题与挑战，如小样本学习问题、实践应用不足的原因等。为促进鱼类图像自动化分析发展并将研究成果转化为生产力，针对问题与挑战，本文探讨了开发用户友好的应用平台等相应的解决方法和未来发展方向。

## 2 研究现状

由于鱼类的重要性，基于深度学习的自动化分类研究相对较早也较多，2015年Li等人<sup>[6]</sup>将快速区

表1 海洋鱼类多样性调查的主要海洋产业需求及示例场景

海洋产业	行业需求	功能场景示例
海洋资源开发利用	重大工程生态环境影响评价/监测	核电对鱼类多样性影响
	海洋油气、海底矿产等资源开发	采矿对鱼类多样性影响
海洋生态环境保护	海洋污染治理评估	微塑料等污染对鱼类影响
	海洋保护区设立及划区依据	鱼类多样性及地理分布调查
海洋防灾减灾	赤潮、绿潮、海洋酸化	灾害对鱼类影响
	全球气候变化	厄尔尼诺和拉尼娜现象对鱼类多样性影响；鱼类迁移变化
海洋安全保障	海洋环境安全保障	核污水对鱼类影响
	港口航运	外来鱼类物种入侵监测
海洋科学认知	海底深部探测	鱼类新物种发现
	公众普及	海洋鱼类识别
海洋渔业	海洋捕捞	鱼类资源调查与评估
	海水养殖	环境污染对鱼类影响
滨海旅游	海岸带可持续发展	污染影响评估；濒危鱼类识别
	健康与安全	危险物种识别，如有毒虾虎鱼

域卷积网络算法(Faster Regions with Convolution Neural Network features, Faster R-CNN)应用于水下图像, 通过使用鱼类图像对深度学习模型进行训练, 以检测和自动化识别鱼类物种, 获得了81.4%的分类准确率。2016年Villon等人<sup>[7]</sup>的研究结果也显示了相较于过去传统的特征提取方法深度学习在图像分类上有更好的表现。随着海洋生物图像数据不断积累、深度学习算法快速发展和计算能力的提升, 基于深度学习的海洋鱼类智能识别近年来正在不断取得进展, 以下从数据集、模型算法和应用3方面总结如下:

### 2.1 数据集

数据作为深度学习的基础要素, 对训练效果起着重要作用。由于图像数据的采集目的、海域和方法等不同, 用于模型训练的海洋鱼类图像数据集也是多种多样(如表2所示)。例如, 鱼类认知(Fish4-Knowledge, F4K)图像集常用于深度学习模型训练与表现评估, 该数据集来自台湾附近水下观测点收集的影像数据<sup>[8]</sup>。该数据集包括了23种鱼类(如表2所示)应用较为广泛, 跨语言评估论坛图像集(Image of the Cross Language Evaluation Forum, ImageCLEF)亦是源于此数据集(<http://www.imageclef.org/2014/lifeclef/fish>)。Croatian fish dataset数据集中图像主要采集自亚得里亚海<sup>[9]</sup>, DeepFish数据集则主要是针对澳大利亚附近海域鱼类, 并且还利用ResNet-50主干框架和前馈网络(Feed Forward Network, FFN)在该数据集上进行了分类性能的验

证<sup>[10]</sup>, Kay等人<sup>[11]</sup>构建的Fishnet图像数据集含有86 029张, 主要收集自太平洋西部和中部, Lin等人<sup>[12]</sup>通过收集主要栖息在我国东海的海洋鱼类图像, 构建了OceanFish数据集。为了面向全球尺度的海洋鱼类, 满足对能识别大量物种及其功能特征的通用系统的需求, 研究人员不断的完善数据集, 提高机器学习模型的代表能力。Zhuang等人<sup>[13]</sup>整合Fishbase数据库等海洋鱼类图像及标注、环境信息, 形成Wildfish++鱼类数据集, 覆盖了2 348种, 包含了103 034张图片, 还包括了生物分类标签标记(如目、科、属和种等)。该数据集融合多模态数据, 加入了文本描述等更多的特征, 在文本描述的指导下, 提出了一种新颖的细粒度识别任务, 通过利用门控循环单元(Gate Recurrent Unit, GRU)来提取文本知识, 以有效区分易混淆的类别。2023年, Khan等人<sup>[14]</sup>发布了FishNet, 这是一个更大规模的多样化数据集, 包含来自17 357个水生物种的94 532张图像, 可用于训练自动鱼类分类的高容量模型。这些大规模的数据集将促进适用于现实场景的鱼类分类、检测和功能特征预测等通用系统的开发。需要指出的是, 目前常用的图像数据集多为2维图片或视频, 不能完全反映物体空间立体信息。随着更多水下图像采集技术的发展和广泛使用, 如基于双目视觉的3维测量技术的更广泛使用<sup>[15]</sup>, 全画幅摄像<sup>[16]</sup>等, 更多可以提供空间位置信息的数据集也将产生。

表2 深度学习在鱼类图像自动化识别中的应用实例

年份	数据源	物种数	原始图像张数	准确率(%)	参考文献
2015	ImageCLEF	12	24 277	81.4	[6]
2015	Croatian fish dataset	12	794	66.78	[9]
2016	Fish4Knowledge	23	27 370	98.64	[17]
2017	ImageCLEF	12	24 277	89.95	[18]
2017	Fish4Knowledge	23	27 142	96.29	[19]
2018	Fish4Knowledge	23	27 000	99.45	[20]
2018	Croatian fish dataset	12	794	83.92	[21]
2018	自建图集	16	1 647	94.30	[22]
2018	WildFish	1 000	54 459	74.70	[23]
2020	DeepFish	20	39 766	99.00	[10]
2020	WildFish++	2 348	103 034	74.70 *	[13]
2021	自建图集	15 †	23 211	99.23	[24]
2021	QUT dataset	6	1 334	90.48	[25]
2023	OceanFish	136	63 622	97.12	[12]
2023	FishNet	17 357	94 532	61.38 △	[14]

注: \* 仅针对选取出的易混淆的22对鱼类, 训练集含1 668张图片, 测试集含1 320张图片。

† 包括鱼类11种、棘皮动物类3种和甲壳动物类1种。

△ 在科一级分类水平上, 共463科。



## 2.2 传统机器学习中的模型算法

### 2.2.1 支持向量机

支持向量机(Support Vector Machine, SVM)是一种强大的监督学习算法<sup>[26]</sup>,旨在解决分类和回归问题。SVM的主要目标是通过最大化数据点到超平面的距离,找到一个最优的超平面,从而实现数据的分类。在SVM中,最接近超平面的数据点被称为“支持向量”,它们决定了分类决策。同时,SVM可以使用核函数,如线性核、多项式核等,将数据从原始特征空间映射到更高维度的空间,以处理非线性分类问题。高维空间的映射处理使得SVM适用于具有大量特征的数据集的相关任务,如图像识别、文本分类等。在处理大规模数据集时,SVM只关注支持向量的特性提高了模型的计算效率,但SVM的训练时间和内存要求可能很高,因为它需要解决凸优化问题。同时,SVM对噪声和异常值敏感,这可能会影响模型的性能。

这些年来,SVM也结合基于深度学习的表征学习。Huang等人<sup>[27]</sup>提出一种带有拒绝选项的分层分类方法,其中SVM被设计为 $p$ 分类器,对应 $p+1$ 个决策,包括 $p$ 种类别的预测和拒绝选项。基于这一方法的拒绝式平衡强制优化树(Balance-Enforced Optimized Tree with Reject option, BEO-TR)算法在来自南台湾海的鱼类图像数据集上取得了84.8%的平均精确率(Average Precision, AP),91.4%的平均召回率(Average Recall, AR)和97.5%的准确率(ACcuracy, AC)。SVM在类似的分层分类算法中也有其他的应用,如Chuang等人<sup>[28]</sup>提出的一种由无监督学习和错误弹性分类器组成的水下鱼类识别框架,其中SVM被用作为类层次的分类器,该框架在NOAA Fisheries数据集上表现出97.1%的AP,98.9%的AR和98.4%的AC,在Fish4Knowledge数据集的最高的15个类别上表现出的AP,AR和AC分别为92.1%,91.6%和97.7%,超过了先前的BEO-TR(91.4%,84.8%和97.5%)。一些结合深度学习和SVM的算法,通常先采用深度学习的框架提取特征,然后再用分类器进行分类。Villon等人<sup>[7]</sup>通过比较深度学习和方向梯度直方图(Histogram of Oriented Gradients, HOG)+SVM方法在水下视频中珊瑚礁鱼类检测与识别的效果,发现深度学习可能更为有效,具有更高的准确性和泛化能力。Qin等人<sup>[17]</sup>提出了一种用于水下摄像机捕捉的视频中鱼类识别的框架,使用深度学习架构提取前景鱼类图像的特征,结合主成分分析、二进制哈希、块状直方图、空间金字塔池化(Spatial Pyramid Pooling, SPP)等技术,最终采用线性SVM分类器实现高效

训练,并在实际鱼类识别数据集上达到了98.64%的最先进准确率。Tamou等人<sup>[20]</sup>利用卷积神经网络AlexNet进行迁移学习,提出自动鱼类物种分类方法,通过对水下数据集的前景鱼类图像进行特征提取,并使用线性SVM分类器实现,实验结果在鱼类识别基准数据集上表现出99.45%的准确性。SVM也常和其他的算法结合被用于鱼类的分类任务。Islam等人<sup>[29]</sup>将决策树、K最近邻算法(K-Nearest Neighbors, KNN)、支持向量机(SVM)、朴素贝叶斯和人工神经网络同时用于融合特征的分类,最后根据多数投票选择最终的类别,该算法在QUT\_fish\_data数据集上进行了实验(AC 98.64%),算法表现优于许多当时的方法。Ou等人<sup>[30]</sup>利用灰度共生矩阵(Gray-Level Co-occurrence Matrix, GLCM)和(Visual Geometry Group, VGG16)从纹理图像中获得纹理特征指数数据(Texture Feature Index Data, TFD)、深度特征数据(Deep Feature Data, DFD)及其组合特征数据(Combined Feature Data, CFD)。采用不同核函数的SVM对金枪鱼的表型纹理进行自动分类,其中CFD在金枪鱼数据集上的平均AC达到了95%。

### 2.2.2 随机森林

随机森林(Random Forest, RF)是集成学习(Ensemble Learning, EL)的一种方法,通过结合多个决策树来提高模型的性能和鲁棒性。在训练时,每个决策树通过有放回抽样随机选择训练样本和特征,从而减小过拟合的风险,提高模型泛化能力,这使得RF在处理高维数据和数据中存在噪声的情况下表现的很出色。在分类任务中,RF采用投票的方式,每棵决策树都会对预测的样本投票出一个类别,最后的预测结果是所有树的投票结果中获胜的类别;对于回归任务,样本的预测结果则是取所有树预测结果的平均值。RF的一个主要优点是它对于超参数调整的需求相对较小,通常默认参数设置就能取得不错的结果。除此之外,每棵决策树的训练都是相互独立的,所以RF很适合同行计算,可以有效地处理大规模的数据集。在一些需要高性能和鲁棒性的情况下,RF通常是一种强有力的选择。RF的优势使其在海洋生物智能识别与跟踪的领域也有很多的应用和研究。Luan等人<sup>[31]</sup>利用随机森林模型基于拖网调查数据构建了21种底层鱼类的物种分布模型,结果显示在样本大小从10增加到30时,大多数物种的预测性能显著提高,且不同物种特征对模型预测的影响存在差异,为有效预测鱼类丰度分布提供了指导。Mampitiya等人<sup>[32]</sup>探讨了不同鱼类水下图像的最佳机器学习分类器,如SVM,

RF和神经网络。结果表明,使用主成分分析(Principal Component Analysis, PCA)和隐含狄利克雷分布(Latent Dirichlet Allocation, LDA)等降维方法可以提高某些算法的精度,RF能够以99.89%的AC优于其他特征提取方法。

## 2.3 计算机视觉研究中的模型算法

### 2.3.1 卷积神经网络

卷积神经网络(Convolutional Neural Network, CNN)是一种深度学习神经网络,其由Lecun<sup>[33]</sup>在1998年提出,主要用于图像处理任务,如图像分类、目标检测、图像分割等。CNN通常由多个卷积层、池化层和全连接层组成,这些层级的堆叠使网络能够学习更复杂的特征表示。其中,卷积层通过卷积核的滑动捕获图像的特征,池化层用于减小特征图的尺寸,降低计算复杂度。CNN通常在每个卷积层和全连接层后面添加激活函数,从而引入非线性。CNN还有一个比较重要的特征是权重共享,是指在卷积层中使用相同的卷积核来处理图像的不同区域,这样做可以减少参数数量,降低过拟合的风险,并加速计算。随着深度学习的发展,CNN在各种领域都取得了卓越的成就,成为许多复杂任务的首选神经网络架构。

基础的深度卷积神经网络与一些传统机器学习算法的结合模型在海洋生物的智能识别任务上有较好的表现。Sarigül等人<sup>[34]</sup>采用具有不同卷积层数量和不同滤波器大小的深度卷积神经网络来对具有挑战性的鱼类数据集进行分类。结果显示,在测试数据集上,较浅的结构并配以较大的滤波器取得了更好的结果,增大滤波器的尺寸可能带来最高40.73%的性能提升。此外,通过增加成功结构每个卷积层的滤波器数量进行测试,这一操作将当前结构的性能额外提升了14.28%。Deep等人<sup>[35]</sup>提出了一种混合卷积神经网络框架,首先使用CNN进行特征提取,然后使用SVM和kNN进行分类,DeepCNN-SVM和DeepCNN-kNN分别取得了98.32%和98.79%的准确率。Konovalov等人<sup>[36]</sup>提出了一种高效的标记方法,用于在来自20个不同的项目域水下有鱼/无鱼图像(4 000个)和17 000张已知的负面(缺失鱼)通用域(VOC2012)水上图像上训练基于Xception CNN的检测器。通过多域图像集合,经过训练的2元(鱼/非鱼)分类器在测试图像上分别实现了0.17%的假阳性和0.61%的假阴性。AUC为99.94%。Zhuang等人<sup>[13]</sup>构建的Wildfish++是一个涵盖高质量的2 348类,包含1 003 034张图片的基准数据集,他们使用CNN作为多模型鱼类网络(Multi-Modal Fish Net, MMFN)的骨干模型,模

型在Wildfish++上取得比之前的最先进的模型(State-Of-The-Art, SOTA)更高的准确率。Zhao等人<sup>[37]</sup>通过对残差网络的改进,设计了一种新的复合骨干网络来学习水下生物图像视频的场景变化信息,从而减少了水下环境信息对目标特征的干扰,加强了主网络对目标信息的输出。此外,他们还设计了增强路径聚合网络来解决线性上采样导致的语义信息利用率不足的问题。实验结果表明,Composited FishNet的(AP) 0.5:0.95, AP50和(AR)max=10分别为75.2%, 92.8%和81.1%。除此之外,针对目标检测和目标识别的任务特性,越来越多里程碑式的基于CNN的模型被人们提出,如R-CNN、单镜头多盒检测器(Single Shot MultiBox Detector, SSD)和你只看1次(You Only Look Once, YOLO)系列算法等。Li等人<sup>[6]</sup>利用Fast R-CNN(具有卷积神经网络的区域)特性,从复杂的水下环境中高效准确地检测和识别鱼类物种,取得了令人期待的平均精度(mean Average Precision, mAP)为81.4%的成果,并在单张鱼类图像上实现了80倍的检测速度提升,超越了先前R-CNN方法。Zeng等人<sup>[38]</sup>通过向标准Faster R-CNN算法中加入对抗遮挡网络(Adversarial Occlusion Network, AON)构成新的算法Faster R-CNN-AON,通过共同竞争学习,提高算法的鲁棒性,也可以有效防止检测网络对生成的固定特征过拟合,实验结果表明,与标准Faster R-CNN相比,Faster R-CNN-AON在VOC07数据集上的mAP提高了2.6%,在水下数据集上的mAP提高了4.2%。Khai等人<sup>[39]</sup>使用参数校准策略来发现合适的参数,并提供了改进的Mask区域卷积神经网络(Mask R-CNN)模型。结果表明,增强后的Mask R-CNN模型准确率高达97.48%。Song等人<sup>[40]</sup>提出了一种R-CNN的变体为激发式区域卷积网络(Boosting R-CNN),其中包含了一种新的区域建议网络和新提出的样例挖掘算法(boosting re-weighting),还引入了概率推理。该算法在UTDAC-2020数据集上以51.4%的AP优于Faster R-CNN的45.5%,SSD512的40.0%等,证明了算法的有效性和鲁棒性。

不同于R-CNN系列的two-stage方法,SSD和YOLO系列采取one-stage的方法,均匀地在突破的不同位置进行密集抽样,然后利用CNN提取特征值直接进行分类与回归。Xu等人<sup>[41]</sup>在现实世界的水力发电站记录的3个非常不同的数据集上训练Yolov3模型,平均精度(mAP)得分为0.539 2。Hu等人<sup>[42]</sup>过修改特征金字塔网络(Feature Pyramid Networks, FPN)的连接方式,残差连接模式和去



冗余操作,提出了一种改进的YOLOv4网络检测模型。在真实鱼场的实验结果表明,改进后的模型检测精度优于原始YOLOv4网络,平均精度从65.40%提高到92.61%(交并比为0.5时),提高了27.21%。此外,计算量减少了约30%。Muksit等人<sup>[43]</sup>提出了YOLO-Fish模型。其中YOLO-Fish-1通过修复上采样步长问题来增强YOLOv3,以减少对小鱼的误检。YOLO-Fish-2则通过在第1个模型中添加空间FPN,以添加动态环境中检测鱼类外观的功能。两个模型在无约束的真实海洋环境中的鱼类检测平均精度分别为76.56%和75.70%,明显优于YOLOv3。与YOLOv4等最新版本的YOLO相比,这两个模型以轻量级的参数达到了相似的性能。Li等人<sup>[44]</sup>提出了一种基于YOLOv5改进的网络,用于密集鱼群和小目标的检测。该网络将坐标注意力机制(Coordinate Attention, CA)和3卷积层(C3)结构的跨阶段局部网络融合,取代YOLOv5中主干的C3模块,并将3层检测层扩展为4层,最后,采用高效交联损失函数代替广义交联损失函数。基于实际图像数据和在线获取的少量数据集,实验结果表明,算法的平均精度(mAP@0.50)达到了94.9%,比YOLOv5算法提高4.4%。Alaba等人<sup>[45]</sup>所提出的模型由MobileNetv3-large, VGG16主干网络和SSD检测头组成。结合一种类感知损失函数来解决数据集的类不平衡问题。该算法在大规模礁鱼数据集SEAMAPD21上的实验结果表明,类别感知损失使模型损失比原始损失提高了79.7%,在Pascal VOC数据集上的实验结果也表明该模型优于原始SSD目标检测模型。

### 2.3.2 注意力机制

注意力机制(Attention Mechanism, AM)是一种在机器学习和深度学习中广泛应用的技术,用于加强模型对输入数据的关注程度,使模型能够根据输入数据的重要性动态地分配注意力<sup>[46]</sup>。注意力机制的核心思想是,不同的输入部分或特征在不同情境下有不同的重要性。在注意力机制中,每个输入位置或特征都被分配一个相应的注意力权重,这些权重表示了模型对每个输入的关注程度。自注意力机制(Self-Attention, SA)是一种特殊的注意力机制,广泛用于处理序列数据,例如自然语言文本。自注意力机制允许模型在同一序列中动态分配注意力,以更好地捕获序列中不同部分之间的关系。注意力机制是一种强大的工具,已经成为现代深度学习架构的核心组成部分,如Transformer模型。这一模型已经在自然语言处理和计算机视觉领域取得了巨大的成功,为这些领域的进步作出了重要贡

献。注意力机制的不断演进和应用将继续推动深度学习在更多领域中的创新和发展。Zhang等人<sup>[47]</sup>提出了一种双池聚合注意力网络(Depth Potentiality-Aware gated attention Network, DPANet),设计了一种新颖的池化聚合位置注意模块和池化聚合通道注意模块,分别用于聚合空间维度和通道维度中的上下文。这两个模块分别采用沿通道维度和沿空间维度的池化操作来聚合信息。DPANet在鱼类图像数据集DeepFish以及水下图像数据集水下图像的语义分割(Semantic segmentation of Underwater Imagery, SUIM)上实现了新的最先进的分割性能,平均交并比(Intersection over Union, IoU)得分分别为91.08%和85.39%,同时将注意力模块的每秒浮点计算次数(Floating-point Operations per second, FLOPs)显著降低了约93%。Gupta等人<sup>[48]</sup>提出了深度鱼类追踪网络(Deep Fish Tracking Network, DFTNet),该网络结合了孪生网络来编码外观相似性和注意长短期记忆网络来捕获后续帧的运动相似性。该框架可以提供联合优化评分,以保持轨道信息编码的外观、运动和空间相似性线索,在Fish4knowledge数据集中上的竞争技术进行了比较,在ID转换量上显著平均减少了60.9%。Li等人<sup>[49]</sup>提出一种基于GhostNet和注意力机制的改进Ghost-YOLOv5网络,减少模型参数,增强水下鱼类检测性能,相较原算法mAP值提高约5%。Chen等人<sup>[50]</sup>提出了一种融合注意力机制的YOLOv5s海洋鱼类检测方法,缓解了由于海洋鱼类目标的形状、颜色和大小而导致的误检和漏检问题。该网络在数据集上以更小的参数量,获得了高于基础YOLOv5的84.5%的精确率和82.3%的mAP。

### 2.3.3 Vision Transformer

预训练大模型(Pretrained Large Models, PLM)是深度学习领域中的一种重要技术趋势,它们是在大规模数据上进行预训练的神经网络模型,通常具有数亿到数万亿的参数<sup>[51]</sup>。这些模型在自然语言处理、计算机视觉、强化学习和其他领域取得了显著的成功,具有出色的泛化性能。通过大量的数据进行训练,可以使预训练大模型能够捕获数据的广泛分布和丰富的特征。在预训练之后,这些模型通常在特定任务上进行微调,以适应具体的问题。一些著名的预训练大模型包括基于变换器的双向编码表征器(Bidirectional Encoder Representations from Transformers, BERT)、生成式预训练变换器(Generative Pre-trained Transformer, GPT)、视觉变换器(Vision Transformer, ViT)等。

Transformer最初是为自然语言处理(Natural

Language Processing, NLP)任务设计的<sup>[52]</sup>, 它使用了自注意力机制(self-attention mechanism)来处理输入序列的不同位置之间的关系。Transformer包括编码器和解码器两部分, 通常用于序列到序列的任务, 如机器翻译。

ViT是一种将Transformer模型应用于计算机视觉(Computer Vision, CV)任务的模型<sup>[53]</sup>。它最初被提出来处理图像分类任务, 但后来也被用于目标检测、图像分割和其他CV任务。ViT采用了与传统图像处理卷积神经网络(CNN)不同的方法。它将图像分为图像块(patch)并将这些块嵌入到一个序列中, 然后使用Transformer编码器处理该序列。ViT引入了一个位置嵌入(position embeddings)来处理图像块之间的空间关系。在训练阶段, ViT还将分类头部附加到Transformer的输出, 以执行图像分类。

Transformer和Vision Transformer的模块细节不同, 但它们共享相同的核心思想, 即基于自注意力机制的解码编码架构。这些模型已经取得了显著的成功, 并在众多领域内推动了深度学习的发展。在实际应用中, 它们通常需要大量的数据和计算资源来进行预训练和微调, 才能够实现卓越的性能。

Saleh等人<sup>[54]</sup>提出了一种基于Transformer的方法, 使用自监督来进行高质量的鱼类分割。实验结果表明, 当对一个数据集的一组水下视频进行训练时, 该方法超越了之前的自监督方法, 并且在两个新的未见过的水下视频数据集上实现了与监督方法相对接近的性能。此外, 有非常多的基于ViT的优化算法被用于海洋生物领域。Saleh等人<sup>[55]</sup>提出了一种名为移动鱼类地标检测网络(Mobile Fish Landmark Detection network, MFLD-net)的新型模型, 基于ViT进行卷积操作, 在低数据环境下能够轻量且具有竞争性的性能, 实现了与一些最先进的CNNs相媲美甚至更好的鱼类图像关键点估计。Liu等人<sup>[56]</sup>提出了一种名为双路鱼网络(dual-Path Fish Network, P-FishNet)的双路径金字塔ViT特征提取网络, 旨在提高水下复杂环境中鱼类检测的准确性, 通过全局和局部特征提取, 以及内容感知特征重组, 实现了76.0%和95.2%的AP和AP50, 并减少了计算量和参数数量约40%。Gong等人<sup>[57]</sup>提出了一种基于迁移学习和视觉变换的多水鱼类分类方法(Fish-TViT), 使用标签平滑损失函数来解决分类器的过拟合和过置信度问题。实验结果表明, Fish-TViT在低分辨率海鱼数据(94.33%)和高分辨率淡水鱼数据(98.34%)上都取得了较高的分类准确率。

## 2.4 小样本学习中的模型算法

对于某类任务T和性能度量P, 一个计算机程序被认为可以从经验E中学习是指, 通过经验E改进后, 它在任务T上由性能度量P衡量的性能有所提升, 解决这样问题的算法被称为机器学习算法。小样本学习(Few-Shot Learning, FSL)是在经验E在任务T上有少量监督信息的样本的情况下, 旨在使模型能够在只有极少数量的样本可用的情况下进行学习和泛化。在经验E中若只有一个监督信息样本时, 则FSL称之为单样本学习(one-shot learning), 若经验E未包含任何监督信息样本时, 则FSL成为零样本学习<sup>[58]</sup>(Zero-Shot Learning, ZSL)。

由于海洋生物类别众多, 同时图像采集受水下条件限制(如深海环境等), 水下目标和图像数据获取困难, 存在相当一部分的物种只有很少的已知样本。因此, FSL在海洋生物的智能识别与跟踪方面有着很大的应用潜力。Villon等人<sup>[59]</sup>探讨了如何利用少样本学习(FSL)克服深度学习(DL)在水下视频物种分类中的限制, 发现FSL在有限注释图像情况下的分类准确性优于经典DL方法。FSL的研究通常从3个方面提高模型在某一任务下的表现: 数据、模型和算法。在数据层面, 常常使用先验知识来扩充和增强数据集, 在扩充之后的数据集上使用标准的机器学习方法可以得到更精确的结果。具体来说, 在小样本学习的研究中, 当存在无标签或弱标签的大数据集时, 可以通过从更大的数据集中选择带有目标标签的样本来进行对数据进行增强; 相似的数据集也可以提供一些先验知识。除此之外, 当没有额外的相似数据集提供数据支持时, 可以对原训练集中的样本进行多种变换, 进而利用转换过程中产生的先验知识提高算法的表现, 例如, Guo等人<sup>[60]</sup>基于FUNIT设计了一个针对水下图像的生成模型, 通过向原数据集中加入生成的假图像, 提高了同一模型在数据集上的分类性能。Liu等人<sup>[61]</sup>提出了一种基于数据增强和迁移学习的水下目标识别的少样本学习方法, 以2维时频谱为输入, 采用多种数据增强方案, 结合迁移学习方法实现目标分类。实验结果表明该系统可以达到0.82的准确率。在算法和模型层面, FSL的研究把关注点放在缩小假设空间和用先验知识搜索假设空间中模型的最优参数。在模型方面, 小样本学习主要通过多任务学习、嵌入学习、从外部记忆中学习和生成式建模4个方面来优化模型的结构和设计。在算法方面, 小样本学习的研究在过去主要集中在搜索策略的优化上: 如何完善现有参数、如何完善元学习参数以及如何学习优化器。其中, 多任务学习可以通过已有的相

似或是辅助任务来对小样本任务进行约束,在Gong等人<sup>[62]</sup>的工作中,为了解决鱼类根据数据集数据缺乏的问题,提出了几种基于半监督和注意机制的样本水下鱼类跟踪方法,使用自训练方法训练添加了卷积块注意力机制(Convolutional Block Attention Module, CBAM)的Yolov4-tiny模型,并将改进的目标检测器与Sort跟踪器相结合。这种通过目标检测辅助目标跟踪任务的想法成功约束了该小样本任务的假设空间。实验结果表明,这一跟踪方法在跟踪精度和跟踪效率方面具有较好的性能。针对注意力机制的优化能够很好地帮助模型实现特征提取,从而让模型关注到数据中最需要注意的特征,这样的特征能够让模型参数在假设空间中更好地被优化。Zhai等人<sup>[63]</sup>提出了一种新颖的注意力网络,名为三明治注意力协方差度量网络(Sandwich Attention Covariance Metric Network, SACovamNet),该网络在基于度量学习的协方差度量网络(CovamNet)上添加了新的三明治状注意力模块,增强了CNN对少样本细粒度鱼图像进行特征提取的能力。实验表明,该SACovamNet在WildFish 1-shot 5-way上取得了72.68%的准确率,5-shot 5-way上取得了86.12%的准确率。

此外,鱼类图像数据集从高一级向低一级系统分类级别的划分普遍呈现出类别数增加、类别内图像数据减少的现象,属于对同一类别的图像进行更细粒度的子类划分。因此,细粒度图像分析对小样本图像分析也具有参考意义,如Wei等人<sup>[64]</sup>对深度学习在细粒度图像分析方面进行了广泛、系统、深入总结。

上述的算法发展如图1所示。

## 2.5 应用

随着鱼类图像数据集的不断完善,如面向全球尺度的wildfish++和Fishnet(如表2所示)鱼类图像数据集的出现,通过训练深度学习网络模型生成的图像分类器已可对海洋鱼类多种类别进行较准确的快速自动化分类,在某些极易混淆鱼类物种间也能找出其中的细微差别,这些自动化分类的研究成果

对于全球海洋鱼类多样性和地理分布调查等具有重要意义。例如,美国国家海洋和大气管理局(National Oceanic and Atmospheric Administration, NOAA)与Kitware公司合作开发了以鱼类在内的海洋物种自动化检测与鉴定分析为目标的海洋环境视频和图像分析工具(Video Image Analytics for the Marine Environment, VIAME)。VIAME是一个开源计算机视觉软件平台(<https://www.viametoolkit.org>),也可作为桌面或web应用程序使用,同时也是一个不断发展的工具包,包含了许多常见的算法和库,许多用于生成不同对象检测器、全帧分类器、图像镶嵌、快速模型生成、图像和视频搜索以及立体测量方法的工作流。然而,海洋调查工作者实际应用中仍需要自行安装、配置环境、模型部署,甚至下载数据、训练模型等,这大大局限了分类器的实际应用性。国内也有“识鱼”、“南海小渔”等鱼类识别微信小程序发布,即用即走,操作简单,可识别我国常见鱼类等,但微信小程序等手机应用程序仍受限于操作系统或者国家区域,无法满足目前全球海洋生物调查对较全面的图像自动化分类需求,需要有面向全球用户的稳定服务平台,提供易用的专业化自动分析,用于调查实践(如表3所示)。目前一些自动化机器学习(Automated Machine Learning, AutoML)平台(如Google Cloud AutoML和百度EasyDL平台)用户无需输入代码即可实现图像数据上传与标注、训练任务配置及调参以及模型部署,以及支持迁移学习和模型结构搜索。AutoML让深度学习应用变得更容易、更广泛,将加快更多的基于个性化数据集的图像分类工具的出现。

## 3 挑战与机遇

尽管海洋鱼类智能化分类研究取得了一系列进展,受篇幅所限,未一一列出,但是目前在数据集、算法、应用以及相关规范标准等方面仍然存在许多问题与挑战,包括:

(1)鱼类图像数据集仍然不够完善,许多类别

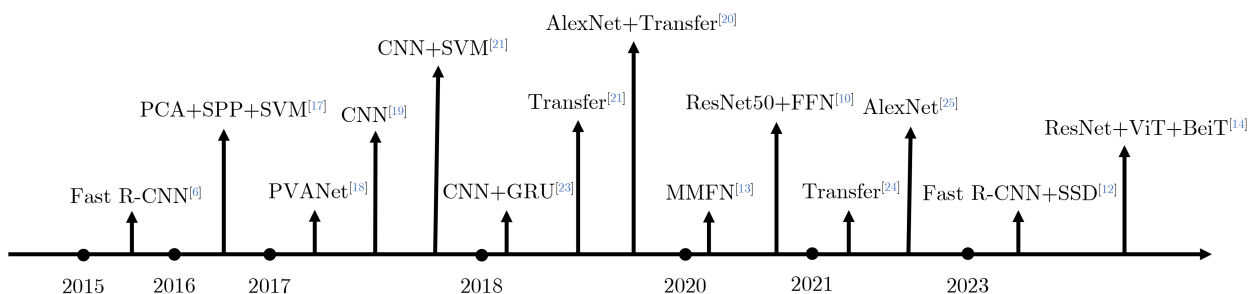


图1 鱼类检测算法发展时序图



表 3 主要问题、原因分析及解决方案

问题	原因	解决方案
识别准确率低	训练数据集小	扩大数据集
	算法需要优化	使用和优化小样本算法
使用流程复杂	需数学、编程等专业知识	利用网络平台提高易用性
应用范围不广	受操作系统、区域限制	

样本数量较少，例如一些生活在深海环境中的鱼类；随着调查技术进步及全球图像数据不断积累，数据集的整合与更新有待加强；

(2)基于小样本学习的模型识别准确率不高；鱼类系统分类学层级信息尚未充分用于提高细粒度分析表现中；

(3)许多基于深度学习的鱼类图像分类研究成果停留在研究成果发表阶段，尚未在海洋产业中得到充分得实际应用；

(4)相关规范标准不完善，如代码共享程度不高，导致研究结果的可重复性和可比较性降低；不利于无计算机视觉或深度学习背景的海洋产业从业人员进行成果的实践应用。

针对目前海洋鱼类图像分类识别上存在的问题与挑战，未来机遇与发展方向包括以下4方面：

(1)不断完善多模态图像数据集，提高数据数量与质量，同时，可纳入数字化海底生境调查产生的3维立体图像。完善的数据集不仅可提高水下环境物种识别准确率，而且有利于水下视频中鱼类的识别与跟踪，加快海洋鱼类基线调查。

(2)水下模糊图像的海洋生物分割与识别，由于水下光照条件较弱，获取的图像比较模糊，开发

人工智能方法针对这些模糊图像进行增强，并提高识别率显得非常重要。

(3)优化小样本深度学习框架，提高物种鉴定(或识别)准确率，开发适用于新型图像数据分类的先进算法，推进海洋生物图像算法自主研发与创新。

(4)提升产学研用深度融合应用，搭建在线鱼类物种智能鉴定网络服务平台，提供友好的访问界面，用户仅需要输入待分析图像即可实现物种智能识别，提供调用接口，供不同的操作系统及版本使用(如图2所示)。基于全球海洋尺度的鱼类图像分类服务平台不仅可以提高海洋调查效率及标准化、有利于环境评估，还能服务社会大众、有利于海洋物种知识普及。

### 4 结束语

海洋是我国经济社会发展的重要依托和载体，海洋生物图像分析智慧化是我国自然资源两统一、优化自然资源规划制定和海洋经济高质量发展需求，服务于国家海洋发展战略和全球海洋生物多样性养护。随着数字化海底生境绘制和精准测量等需求，电子与信息学将推动海洋生物调查向信息化、智能化、现代化转型升级，如人工智能在海洋生物图像分类上的应用将加深海洋认知，在保护海洋生物多样性的同时，推进海洋产业高质量发展，更好地让海洋开发不断向深蓝挺进

当前人工智能已成为战略性科技产业，促进和革新海洋产业在内各行业传统的科学研究方法，需要切实增强海洋生物调查智能化。本文通过总结海洋鱼类图像自动化智能分类技术及未来机遇与挑战，促进包括鱼类在内的海洋生物图像分类，有利

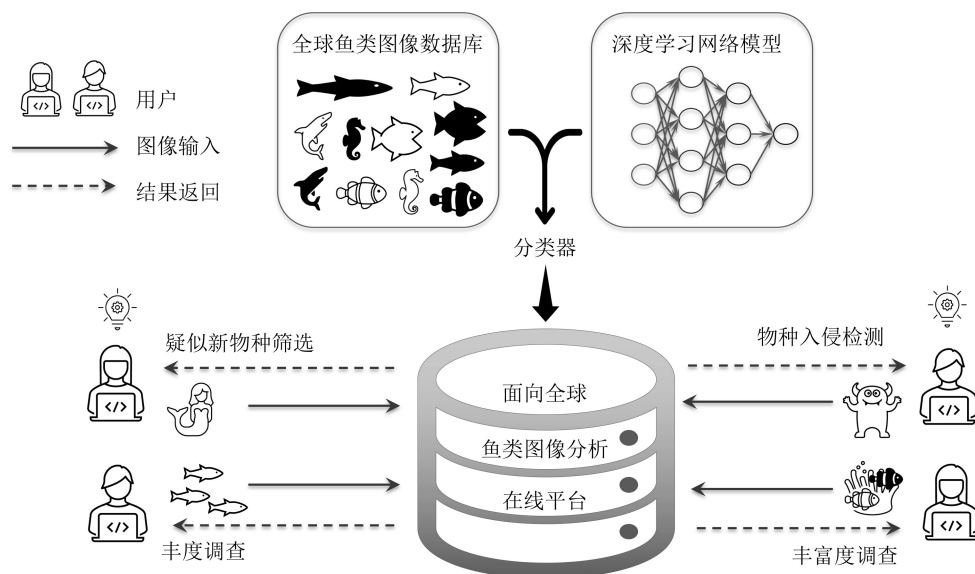


图 2 全球海洋鱼类图像分类实现及应用示意图

于更全面、更有效、更准确的海洋生态监测系统开发,推动全球“智慧海洋生物调查”体系构建,为促进海洋经济高质量发展和全球海洋生物多样性保护做出贡献。

### 参 考 文 献

- [1] KUITER R H and DEBELIUS H. World atlas of marine fishes[M]. 2nd ed. Frankfurt: IKAN-Unterwasserarchiv, 2006.
- [2] ALLEN G R. Field Guide to Marine Fishes of Tropical Australia and South-East Asia[M]. Western Australian Museum, 2009.
- [3] 陈大刚, 张美昭. 中国海洋鱼类 上[M]. 青岛: 中国海洋大学出版社, 2015.  
CHEN Dagang and ZHANG Meizhao. Marine Fishes of China[M]. Qingdao: China Ocean University Press, 2015.
- [4] 陈大刚, 张美昭. 中国海洋鱼类 下[M]. 青岛: 中国海洋大学出版社, 2015.  
CHEN Dagang and ZHANG Meizhao. Marine Fishes of China[M]. Qingdao: China Ocean University Press, 2015.
- [5] 陈大刚, 张美昭. 中国海洋鱼类 中[M]. 青岛: 中国海洋大学出版社, 2015.  
CHEN Dagang and ZHANG Meizhao. Marine Fishes of China[M]. Qingdao: China Ocean University Press, 2015.
- [6] LI Xiu, SHANG Min, QIN Hongwei, *et al.* Fast accurate fish detection and recognition of underwater images with Fast R-CNN[C]. MTS/IEEE Washington, Washington, USA, 2015: 1–5. doi: [10.23919/OCEANS.2015.7404464](https://doi.org/10.23919/OCEANS.2015.7404464).
- [7] VILLON S, CHAUMONT M, SUBSOL G, *et al.* Coral reef fish detection and recognition in underwater videos by supervised machine learning: Comparison between Deep Learning and HOG+SVM methods[C]. The 17th International Conference on Advanced Concepts for Intelligent Vision Systems, Lecce, Italy, 2016: 160–171. doi: [10.1007/978-3-319-48680-2\\_15](https://doi.org/10.1007/978-3-319-48680-2_15).
- [8] BOOM B J, HUANG P X, HE Jiyin, *et al.* Supporting ground-truth annotation of image datasets using clustering[C]. The 21st International Conference on Pattern Recognition, Tsukuba, Japan, 2012: 1542–1545.
- [9] JÄGER J, SIMON M, DENZLER J, *et al.* Croatian fish dataset: Fine-grained classification of fish species in their natural habitat[C]. Machine Vision of Animals and their Behaviour Workshop 2015, 2015: 2. doi: [10.5244/C.29.MVAB.6](https://doi.org/10.5244/C.29.MVAB.6).
- [10] SALEH A, LARADJI I H, KONOVALOV D A, *et al.* A realistic fish-habitat dataset to evaluate algorithms for underwater visual analysis[J]. *Scientific Reports*, 2020, 10(1): 14671. doi: [10.1038/s41598-020-71639-x](https://doi.org/10.1038/s41598-020-71639-x).
- [11] KAY J and MERRIFIELD M. The fishnet open images database: A dataset for fish detection and fine-grained categorization in fisheries[J]. arXiv: 2106.09178, 2021.
- [12] LIN Yuan, CHU Zhaoqi, KORHONEN J, *et al.* Fast accurate fish recognition with deep learning based on a domain-specific large-scale fish dataset[C]. The 29th International Conference on Multimedia Modeling, Bergen, Norway, 2023: 515–526. doi: [10.1007/978-3-031-27077-2\\_40](https://doi.org/10.1007/978-3-031-27077-2_40).
- [13] ZHUANG Peiqin, WANG Yali, and QIAO Yu. Wildfish++: A comprehensive fish benchmark for multimedia research[J]. *IEEE Transactions on Multimedia*, 2021, 23: 3603–3617. doi: [10.1109/TMM.2020.3028482](https://doi.org/10.1109/TMM.2020.3028482).
- [14] KHAN F F, LI Xiang, TEMPLE A J, *et al.* FishNet: A large-scale dataset and benchmark for fish recognition, detection, and functional trait prediction[C]. 2023 IEEE/CVF International Conference on Computer Vision, Paris, France, 2023: 20439–20449. doi: [10.1109/ICCV51070.2023.01874](https://doi.org/10.1109/ICCV51070.2023.01874).
- [15] TORISAWA S, KADOTA M, KOMEYAMA K, *et al.* A digital stereo-video camera system for three-dimensional monitoring of free-swimming Pacific bluefin tuna, *Thunnus orientalis*, cultured in a net cage[J]. *Aquatic Living Resources*, 2011, 24(2): 107–112. doi: [10.1051/alr/2011133](https://doi.org/10.1051/alr/2011133).
- [16] MENG Lin, HIRAYAMA T, and OYANAGI S. Underwater-drone with panoramic camera for automatic fish recognition based on deep learning[J]. *IEEE Access*, 2018, 6: 17880–17886. doi: [10.1109/ACCESS.2018.2820326](https://doi.org/10.1109/ACCESS.2018.2820326).
- [17] QIN Hongwei, LI Xiu, LIANG Jian, *et al.* DeepFish: Accurate underwater live fish recognition with a deep architecture[J]. *Neurocomputing*, 2016, 187: 49–58. doi: [10.1016/j.neucom.2015.10.122](https://doi.org/10.1016/j.neucom.2015.10.122).
- [18] LI Xiu, TANG Youhua, and GAO Tingwei. Deep but lightweight neural networks for fish detection[C]. OCEANS 2017 - Aberdeen, Aberdeen, UK, 2017: 1–5. doi: [10.1109/OCEANSE.2017.8084961](https://doi.org/10.1109/OCEANSE.2017.8084961).
- [19] RATHI D, JAIN S, and INDU S. Underwater fish species classification using convolutional neural network and deep learning[C]. The 2017 Ninth International Conference on Advances in Pattern Recognition, Bangalore, India, 2017: 1–6. doi: [10.1109/ICAPR.2017.8593044](https://doi.org/10.1109/ICAPR.2017.8593044).
- [20] TAMOU A B, BENZINO A, NASREDDINE K, *et al.* Underwater live fish recognition by deep learning[C]. The 8th International Conference on Image and Signal Processing, Cherbourg, France, 2018: 275–283. doi: [10.1007/978-3-319-94211-7\\_30](https://doi.org/10.1007/978-3-319-94211-7_30).
- [21] QIU Chenchen, ZHANG Shaoyong, WANG Chao, *et al.* Improving transfer learning and squeeze- and-excitation networks for small-scale fine-grained fish image classification[J]. *IEEE Access*, 2018, 6: 78503–78512. doi: [10.1109/ACCESS.2018.2885055](https://doi.org/10.1109/ACCESS.2018.2885055).
- [22] SIDDIQUI S A, SALMAN A, MALIK M I, *et al.* Automatic fish species classification in underwater videos: Exploiting pre-trained deep neural network models to compensate for limited labelled data[J]. *ICES Journal of Marine Science*, 2018, 75(1): 374–389. doi: [10.1093/icesjms/fsx109](https://doi.org/10.1093/icesjms/fsx109).
- [23] ZHUANG Peiqin, WANG Yali, and QIAO Yu. WildFish: A large benchmark for fish recognition in the wild[C]. The 26th ACM international conference on Multimedia. Seoul,

- Republic of Korea, 2018: 1301–1039. doi: [10.1145/3240508.3240616](https://doi.org/10.1145/3240508.3240616).
- [24] 孙东洋, 刘辉, 张纪红, 等. 基于深度卷积神经网络的海洋牧场岩礁性生物图像分类[J]. *海洋与湖沼*, 2021, 52(5): 1160–1169. doi: [10.11693/hyh20210100005](https://doi.org/10.11693/hyh20210100005).
- SUN Dongyang, LIU Hui, ZHANG Jihong, *et al.* Classification of reef biological images of marine ranch based on deep convolution neural network[J]. *Oceanologia et Limnologia Sinica*, 2021, 52(5): 1160–1169. doi: [10.11693/hyh20210100005](https://doi.org/10.11693/hyh20210100005).
- [25] IQBAL M A, WANG Zhijie, ALI Z A, *et al.* Automatic fish species classification using deep convolutional neural networks[J]. *Wireless Personal Communications*, 2021, 116(2): 1043–1053. doi: [10.1007/s11277-019-06634-1](https://doi.org/10.1007/s11277-019-06634-1).
- [26] PLATT J C. Sequential minimal optimization: A fast algorithm for training support vector machines[R]. MSR-TR-98-14, 1998.
- [27] HUANG P X, BOOM B J, and FISHER R B. Hierarchical classification with reject option for live fish recognition[J]. *Machine Vision and Applications*, 2015, 26(1): 89–102. doi: [10.1007/s00138-014-0641-2](https://doi.org/10.1007/s00138-014-0641-2).
- [28] CHUANG Mengche, HWANG J N, and WILLIAMS K. A feature learning and object recognition framework for underwater fish images[J]. *IEEE Transactions on Image Processing*, 2016, 25(4): 1862–1872. doi: [10.1109/TIP.2016.2535342](https://doi.org/10.1109/TIP.2016.2535342).
- [29] ISLAM S M M, BANI S I, and GHOSH R. Content-based fish classification using combination of machine learning methods[J]. *International Journal of Information Technology and Computer Science (IJITCS)*, 2021, 13(8): 62–68. doi: [10.5815/ijitcs.2021.01.05](https://doi.org/10.5815/ijitcs.2021.01.05).
- [30] OU Liguo, LIU Bilin, CHEN Xinjun, *et al.* Automatic classification of the phenotype textures of three *Thunnus* species based on the machine learning SVM algorithm[J]. *Canadian Journal of Fisheries and Aquatic Sciences*, 2023, 80(8): 1221–1236. doi: [10.1139/cjfas-2022-0270](https://doi.org/10.1139/cjfas-2022-0270).
- [31] LUAN Jing, ZHANG Chongliang, XU Binduo, *et al.* The predictive performances of random forest models with limited sample size and different species traits[J]. *Fisheries Research*, 2020, 227: 105534. doi: [10.1016/j.fishres.2020.105534](https://doi.org/10.1016/j.fishres.2020.105534).
- [32] MAMPITIYA L I, NALMI R, and RATHNAYAKE N. Performance comparison of sea fish species classification using hybrid and supervised machine learning algorithms[C]. 2022 Moratuwa Engineering Research Conference, Moratuwa, Sri Lanka, 2022: 1–6. doi: [10.1109/MERCon55799.2022.9906206](https://doi.org/10.1109/MERCon55799.2022.9906206).
- [33] LECUN Y, BOTTOU L, BENGIO Y, *et al.* Gradient-based learning applied to document recognition[J]. *Proceedings of the IEEE*, 1998, 86(11): 2278–2324. doi: [10.1109/5.726791](https://doi.org/10.1109/5.726791).
- [34] SARIGÜL M and AVCI M. Comparison of different deep structures for fish classification[J]. *International Journal of Computer Theory and Engineering*, 2017, 9(5): 362–366. doi: [10.7763/IJCTE.2017.V9.1167](https://doi.org/10.7763/IJCTE.2017.V9.1167).
- [35] DEEP B V and DASH R. Underwater fish species recognition using deep learning techniques[C]. The 2019 6th International Conference on Signal Processing and Integrated Networks, Noida, India, 2019: 665–669. doi: [10.1109/SPIN.2019.8711657](https://doi.org/10.1109/SPIN.2019.8711657).
- [36] KONOVALOV D A, SALEH A, BRADLEY M, *et al.* Underwater fish detection with weak multi-domain supervision[C]. 2019 International Joint Conference on Neural Networks, Budapest, Hungary, 2019: 1–8. doi: [10.1109/IJCNN.2019.8851907](https://doi.org/10.1109/IJCNN.2019.8851907).
- [37] ZHAO Zhenxi, LIU Yang, SUN Xudong, *et al.* Composited fishnet: Fish detection and species recognition from low-quality underwater videos[J]. *IEEE Transactions on Image Processing*, 2021, 30: 4719–4734. doi: [10.1109/TIP.2021.3074738](https://doi.org/10.1109/TIP.2021.3074738).
- [38] ZENG Lingcai, SUN Bing, and ZHU Daqi. Underwater target detection based on Faster R-CNN and adversarial occlusion network[J]. *Engineering Applications of Artificial Intelligence*, 2021, 100: 104190. doi: [10.1016/j.engappai.2021.104190](https://doi.org/10.1016/j.engappai.2021.104190).
- [39] HONG KHAI T, ABDULLAH S N H S, HASAN M K, *et al.* Underwater fish detection and counting using mask regional convolutional neural network[J]. *Water*, 2022, 14(2): 222. doi: [10.3390/w14020222](https://doi.org/10.3390/w14020222).
- [40] 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](https://doi.org/10.1016/j.neucom.2023.01.088).
- [41] XU Wenwei and MATZNER S. Underwater fish detection using deep learning for water power applications[C]. 2018 International Conference on Computational Science and Computational Intelligence, Las Vegas, USA, 2018: 313–318. doi: [10.1109/CSCI46756.2018.00067](https://doi.org/10.1109/CSCI46756.2018.00067).
- [42] HU Xuelong, LIU Yang, ZHAO Zhengxi, *et al.* Real-time detection of uneaten feed pellets in underwater images for aquaculture using an improved YOLO-V4 network[J]. *Computers and Electronics in Agriculture*, 2021, 185: 106135. doi: [10.1016/j.compag.2021.106135](https://doi.org/10.1016/j.compag.2021.106135).
- [43] AL MUKSIT A, HASAN F, HASAN BHUIYAN EMON F, *et al.* YOLO-Fish: A robust fish detection model to detect fish in realistic underwater environment[J]. *Ecological Informatics*, 2022, 72: 101847. doi: [10.1016/j.ecoinf.2022.101847](https://doi.org/10.1016/j.ecoinf.2022.101847).
- [44] LI Jianyuan, LIU Chunna, LU Xiaochun, *et al.* CME-YOLOv5: An efficient object detection network for densely spaced fish and small targets[J]. *Water*, 2022, 14(15): 2412. doi: [10.3390/w14152412](https://doi.org/10.3390/w14152412).
- [45] ALABA S Y, NABI M M, SHAH C, *et al.* Class-aware fish species recognition using deep learning for an imbalanced dataset[J]. *Sensors*, 2022, 22(21): 8268. doi: [10.3390/s22218268](https://doi.org/10.3390/s22218268).
- [46] MNIH V, HEES N, GRAVES A, *et al.* Recurrent models



- of visual attention[C]. The 27th International Conference on Neural Information Processing Systems, Cambridge, USA, 2014: 2204–2212. doi: [10.5555/2969033.2969073](https://doi.org/10.5555/2969033.2969073).
- [47] ZHANG Wenbo, WU Chaoyi, and BAO Zhenshan. DPANet: Dual pooling-aggregated attention network for fish segmentation[J]. *IET Computer Vision*, 2022, 16(1): 67–82. doi: [10.1049/cvi2.12065](https://doi.org/10.1049/cvi2.12065).
- [48] GUPTA S, MUKHERJEE P, CHAUDHURY S, *et al.* DFTNet: Deep fish tracker with attention mechanism in unconstrained marine environments[J]. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 5016213. doi: [10.1109/TIM.2021.3109731](https://doi.org/10.1109/TIM.2021.3109731).
- [49] LI Shanmin, PAN Bei, CHENG Yuanshun, *et al.* Underwater fish object detection based on attention mechanism improved ghost-YOLOv5[C]. The 2022 7th International Conference on Intelligent Computing and Signal Processing, Xi'an, China, 2022: 599–603. doi: [10.1109/ICSP54964.2022.9778582](https://doi.org/10.1109/ICSP54964.2022.9778582).
- [50] CHEN Lulu, ZANG Zhaoxiang, HUANG Tianxing, *et al.* Marine fish object detection based on YOLOv5 and attention mechanism[C]. 2022 IEEE Smartworld, Ubiquitous Intelligence & Computing, Scalable Computing & Communications, Digital Twin, Privacy Computing, Metaverse, Autonomous & Trusted Vehicles (SmartWorld/UIC/ScalCom/DigitalTwin/PriComp/Meta), Haikou, China, 2022: 1252–1258. doi: [10.1109/SmartWorld-UIC-ATC-ScalCom-DigitalTwin-PriComp-Metaverse56740.2022.00185](https://doi.org/10.1109/SmartWorld-UIC-ATC-ScalCom-DigitalTwin-PriComp-Metaverse56740.2022.00185).
- [51] HAN Xu, ZHANG Zhengyan, DING Ning, *et al.* Pre-trained models: Past, present and future[J]. *AI Open*, 2021, 2: 225–250. doi: [10.1016/j.aiopen.2021.08.002](https://doi.org/10.1016/j.aiopen.2021.08.002).
- [52] VASWANI A, SHAZEER N, PARMAR N, *et al.* Attention is all you need[C]. The 31st International Conference on Neural Information Processing Systems, Long Beach, USA, 2017: 6000–6010. doi: [10.5555/3295222.3295349](https://doi.org/10.5555/3295222.3295349).
- [53] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, *et al.* An image is worth 16x16 words: Transformers for image recognition at scale[C]. The 9th International Conference on Learning Representations, 2021.
- [54] SALEH A, SHEAVES M, JERRY D, *et al.* Transformer-based Self-Supervised Fish Segmentation in Underwater Videos[J]. arXiv: 2206.05390, 2022.
- [55] SALEH A, JONES D, JERRY D, *et al.* A lightweight Transformer-based model for fish landmark detection[J]. arXiv: 2209.05777, 2022.
- [56] LIU Yang, AN Dong, REN Yinjie, *et al.* DP-FishNet: Dual-path Pyramid Vision Transformer-based underwater fish detection network[J]. *Expert Systems with Applications*, 2024, 238: 122018. doi: [10.1016/j.eswa.2023.122018](https://doi.org/10.1016/j.eswa.2023.122018).
- [57] GONG Bo, DAI Kanyuan, SHAO Ji, *et al.* Fish-TViT: A novel fish species classification method in multi water areas based on transfer learning and vision transformer[J]. *Heliyon*, 2023, 9(6): e16761. doi: [10.1016/j.heliyon.2023.e16761](https://doi.org/10.1016/j.heliyon.2023.e16761).
- [58] WANG Yaqing, YAO Quanming, KWOK J T, *et al.* Generalizing from a few examples: A survey on few-shot learning[J]. *ACM Computing Surveys*, 2020, 53(3): 63. doi: [10.1145/3386252](https://doi.org/10.1145/3386252).
- [59] VILLON S, IOVAN C, MANGEAS M, *et al.* Automatic underwater fish species classification with limited data using few-shot learning[J]. *Ecological Informatics*, 2021, 63: 101320. doi: [10.1016/j.ecoinf.2021.101320](https://doi.org/10.1016/j.ecoinf.2021.101320).
- [60] GUO Zonghui, ZHANG Liqiang, JIANG Yufeng, *et al.* Few-shot fish image generation and classification[C]. Global Oceans 2020: Singapore – U. S. Gulf Coast, Biloxi, USA, 2020: 1–6. doi: [10.1109/IEEECONF38699.2020.9389005](https://doi.org/10.1109/IEEECONF38699.2020.9389005).
- [61] LIU Feng, DING Hao, LI Daihui, *et al.* Few-shot learning with data enhancement and transfer learning for underwater target recognition[C]. 2021 OES China Ocean Acoustics, Harbin, China, 2021: 992–994. doi: [10.1109/COA50123.2021.9519853](https://doi.org/10.1109/COA50123.2021.9519853).
- [62] GONG Longqin, HU Zhuhua, and ZHOU Xiaoyi. A few samples underwater fish tracking method based on semi-supervised and attention mechanism[C]. The 2022 6th International Conference on Robotics, Control and Automation, Xiamen, China, 2022: 18–22. doi: [10.1109/ICRCA55033.2022.9828911](https://doi.org/10.1109/ICRCA55033.2022.9828911).
- [63] ZHAI Jiping, HAN Lu, XIAO Ying, *et al.* Few-shot fine-grained fish species classification via sandwich attention CovamNet[J]. *Frontiers in Marine Science*, 2023, 10: 1149186. doi: [10.3389/fmars.2023.1149186](https://doi.org/10.3389/fmars.2023.1149186).
- [64] WEI Xiushen, SONG Yizhe, MAC AODHA O, *et al.* Fine-grained image analysis with deep learning: A survey[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(12): 8927–8948. doi: [10.1109/TPAMI.2021.3126648](https://doi.org/10.1109/TPAMI.2021.3126648).
- 周 鹏: 男, 高级工程师, 研究方向为海洋生物学与生物信息学研究。  
李昌永: 男, 博士生, 研究方向为模式识别与生物信息学。  
步雨馨: 女, 硕士生, 研究方向为海洋生物学。  
周芷诺: 女, 本科生, 研究方向为模式识别与生物信息学。  
王春生: 男, 研究员, 研究方向为海洋生态学。  
沈红斌: 男, 教授, 研究方向为识别与生物信息学。  
潘小勇: 男, 长聘教轨副教授, 研究方向为图像处理与模式识别研究。

责任编辑: 余 蓉