

基于多专家动态协作学习的长尾声呐图像识别

崔国恒,周浩*,王超,张汀
(海军工程大学,湖北 武汉 430033)

摘要:声呐图像识别在水下环境探测领域发挥着重要作用。现有基于深度神经网络的声呐图像识别方法一定程度提高了声呐图像识别的准确性,但在实际应用中往往遭受长尾数据分布不平衡的挑战,导致部分稀有高价值目标无法被准确识别。为此,本文提出了一种新颖的多专家动态协作模型来提高模型对稀有类别的识别准确率,实现不平衡声呐图像识别。多专家动态协作模型由多专家网络和动态学习策略两部分组成。多专家网络包含一个用于骨干网络特征学习的传统分支和两个用于学习尾部类别样本的再平衡分支。三个专家共同协作实现不平衡声呐图像识别。动态学习策略用于在模型训练中转移模型对传统分支和再平衡分支之间的注意力来同时提高模型的特征学习能力和分类器识别能力。最后,大量的实验在KLSG、FLSMDD、NKSID等三个声呐图像识别数据集上证明了本文模型的有效性,分别实现了91.51%、99.74%和96.19%的总体准确率。

关键词:声呐图像识别;数据不平衡;长尾分布;多专家协作;动态学习策略

中图分类号:TP391.4 **文献标志码:**A **文章编号:**0253-2395(2026)02-0232-12

Multi-expert Dynamic Collaboration Model for Long-tailed Sonar Image Recognition

CUI Guoheng, ZHOU Hao*, WANG Chao, ZHANG Ting
(Naval University of Engineering, Wuhan 430033, China)

Abstract: Sonar image recognition plays a crucial role in the field of underwater environment detection. While existing sonar image recognition models based on deep neural network have improved classification accuracy, they often face the challenges of long-tailed distribution in practice, leading to insufficient identification of rare yet high-value targets. To remedy this, we propose a novel Multi-expert Dynamic Collaboration model to enhance recognition accuracy for long-tailed sonar image (MEDC-SI). Our model consists of multi-expert network and dynamic learning strategy. The multi-expert network contains a conventional branch for feature representation learning and two re-balancing branches for tail samples learning. And three experts collaborate to achieve imbalanced sonar image recognition. The dynamic learning strategy is designed to shift the focus of model training between the conventional branch and re-balancing branches to improve the feature learning and classifier learning simultaneously. Finally, extensive experimental results on three sonar image recognition datasets, KLSG, FLSMDD, and NKSID, demonstrate the effectiveness of the proposed model, achieving overall accuracies of 91.51%, 99.74%, and 96.19%, respectively.

Key words: sonar image recognition; data imbalance; long-tailed distribution; multi-expert collaboration; dynamic learning strategy

收稿日期:2025-07-11;修回日期:2025-09-19

基金项目:国家自然科学基金(62302516)

作者简介:崔国恒(1981-),男,湖北武汉人,博士,副教授,研究方向为目标识别。E-mail:cgh3899@sina.com

* 通信作者:周浩(ZHOU Hao),E-mail:zhhaohg93@nue.edu.cn

引文格式:崔国恒,周浩,王超,等.基于多专家动态协作学习的长尾声呐图像识别[J].山西大学学报(自然科学版),2026,49(2):232-243. DOI:10.13451/j.sxu.ns.2025100.

0 引言

随着水下探测需求的持续拓展,水下目标识别作为支撑水下作业精准化、智能化的关键技术,在学术研究与工程应用领域均已成为重点关注的前沿课题^[1-2]。在昏暗和低光的水下环境中,声呐成像在水下目标识别和探测中至关重要。相比于自然场景图像和遥感图像,声呐图像的获取更加困难,成像质量差,使得声呐图像识别面临着更多的挑战^[3]。因此,研究声呐图像识别方法能够广泛应用于水环境测深与建模^[1]、海底目标定位与识别^[4]、水下目标物探测^[5]等领域。

当前,声呐图像识别方法主要可以分为两类^[1]:基于传统机器学习的方法和基于深度学习的方法。基于传统机器学习的方法通常先对声呐图像中感兴趣区域进行“切片”,在“切片”的基础上对目标进行分割处理。然后,采用多种传统方法对分割处理后的图像进行特征提取,并利用人工设计的分类器对提取的图像特征进行分类识别^[6-7]。例如,Barngrover^[8]首先提取声呐图像中的Haar特征和尺度不变特征转换(Scale-invariant Feature Transform, SIFT)特征,组合多种特征后设计了改进的Boosting分类器对侧扫声呐图像进行识别。Chhabra等^[9]提取了离散余弦变换特征、Gabor变换特征、Haar变换特征和金字塔定向滤波器组的特征,在字典学习的基础上利用k邻近算法实现了声呐图像的

识别分类。基于深度学习的方法通常利用卷积神经网络进行端对端的特征提取和分类^[10]。由于声呐图像样本数量少,通常基于迁移学习、小样本学习和图像生成等方法来提高声呐图像识别的准确性。例如,Cheng等^[11]采用迁移学习的方法先用大规模数据集对VGG19模型框架进行预训练,然后在引入通道注意力和空间注意力的骨干网络上对模型进行微调,实现小目标的声呐图像识别。Xu等^[12]提出了基于域适应的深度自适应声呐图像分类模型,通过域适配模块弥合源域和目标域之间的差异。

虽然这些方法一定程度上有效地提高了声呐图像识别的准确性,但声呐图像识别模型的训练通常遭受不平衡数据分布的影响,导致其在实际应用中对于样本稀少的重要水下目标识别准确度低^[13-15]。如图1所示,采集的声呐图像数据往往呈现典型的不平衡分布,且尾部类别通常包含具有更高价值的水下目标。例如,声呐图像数据集包含了大量的海底轮胎图像,而对于更具价值的小型螺旋桨样本却十分稀少。在这种情况下,训练出来的声呐图像识别模型仅能较好地识别多数类样本,对少数类样本的识别精度仍有待进一步提高。为了提高模型对尾部类别的识别率,Jiao等^[13]提出了一种平衡集成的迁移学习模型,该模型通过多平衡采样方法来提高模型对尾部类别的采样样本数,但模型需要采用两阶段解耦的训练模式,这一定程度增加了模型训练的困难和限制了其在实际场景中的应用。

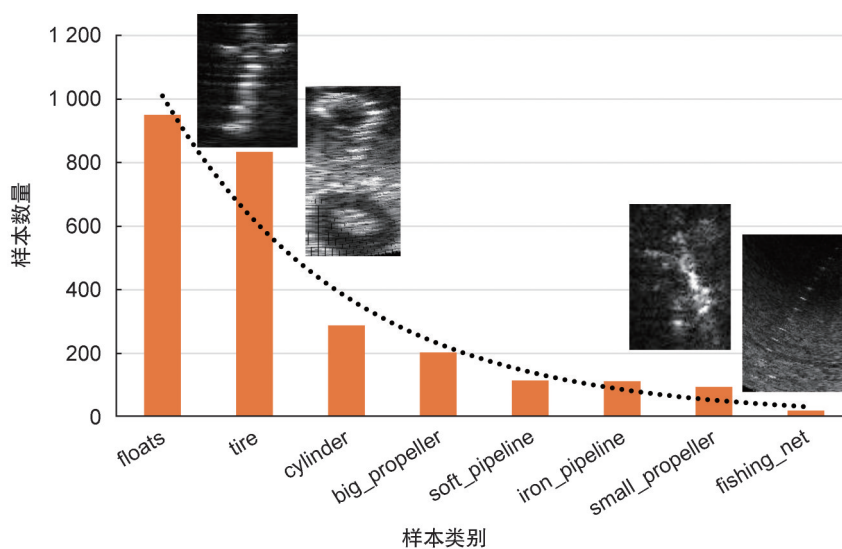


图1 声呐图像数据呈现的长尾不平衡分布

Fig. 1 The long-tailed distribution of sonar image data

为了解决上述问题,本文提出了一种端对端的多专家动态协作学习模型来实现不平衡分布下的声呐图像识别(Multi-expert Dynamic Collaboration Model for long-tailed Sonar Image Recognition, MEDC-SI)。首先, MEDC-SI 模型将在大规模数据集上进行预训练,并在小型声呐图像数据集上进行骨干网络和分类器的微调。其次,在模型微调阶段,我们设计了一个三支网络来提高模型对尾部类别的识别能力,且每一个分支都有相对应的专家分类器进行决策。为了提高专家分类器的多样性,三个分支包括一个传统分支和两个再平衡分支。其中,传统分支配备典型的均匀采样器以进行通用模式的学习;两个再平衡分支分别为类弱平衡分支(Weakly Class-Balance Branch, WCB)和类平衡分支(Class-Balance Branch, RBB)。类平衡分支配备类别平衡采样器以均匀的采样所有类别以进行再平衡学习;为了缓解类平衡分支对头部类别的欠拟合和尾部类别的过拟合,类弱平衡分支采取了轻微增加对尾部类别样本采样和维持头部类别采样相结合的方式,缩小数据分布的不平衡因子并进一步增加不同专家的多样性。再者,在模型训练阶段,我们设计了一种动态学习策略,通过转移模型对不同分支的训练重点来提高模型的特征学习能力和分类器识别能力。在自适应权重的控制下,模型的训练重点将从传统分支逐步转移到类弱平衡分支和类平衡分支以提高模型分类器的学习能力,并最终转移回传统分支以恢复模型的特征表达能力。最后,在三个公开的不平衡声呐图像识别数据集 KLSG、FLSMDD 和 NKSID 上验证了本文 MEDC-SI 模型的有效性,分别实现了 91.51%、99.74% 和 96.19% 的总体准确率。与其他声呐图像识别模型相比,本文模型提高了 0.86%~2.64% 的 F1 精度。本文的主要贡献包括:

1) 本文提出了一个多专家网络协作模型,设置了三种不同的再平衡采样分支来训练三个专家分类器,包括一个传统分支和两个再平衡分支,并通过多专家分类器之间的协作实现模型对不平衡声呐图像的识别,有效提高了模型对稀有样本的识别率。

2) 本文提出了一种动态学习策略来训练多专家网络,通过动态权重转移模型在不同训

练阶段对不同专家的学习注意力,在维持骨干网络特征表达提取能力的同时提高模型分类器对尾部类别的识别能力。

3) 本文在三个不平衡声呐图像识别数据集上进行了实验,在三种不同的测度上都取得了更优的性能表现,证明模型不仅能够准确识别头部类别样本,也有效地提高了稀有样本的识别率。

1 相关工作

声呐图像识别 早期的声呐图像识别模型通常是基于传统机器学习方法的。这些模型主要使用模板匹配、多图像特征组合、多视角多模态联合、字典学习等方法来进行特征提取和目标分类^[1,16]。近年来,基于深度学习的声呐图像识别方法逐步占据了主导地位。由于水下声呐图像数据难以获取而导致其数据规模较小,研究者通常会通过深度学习与迁移学习相结合的方式来提高声呐图像识别的准确性。Zhu 等^[17]在预训练的 AlexNet 模型的基础上采取微调全连接层的方式极大地提高了声呐图像识别的性能。Valdenegro-Toro 等^[3]通过迁移学习的方式专门研究了多种预训练模型在声呐图像识别中的性能,例如 ResNet20^[18], MobileNets^[19], DenseNet121^[20], MiniXception^[21]等。数据合成和样本增强类方法也被广泛使用在声呐图像识别中来解决小样本问题。Yang 等^[22]提出了一种适用于多任务场景的侧扫声呐全类图像样本增强方法,通过扩散模型和迁移学习来生成声呐图像。Xu 等^[23]提出了一种多特征融合自注意力网络通过光学-声学图像对来生成新类别的侧扫声呐图像。陈禹乐等^[24]提出一种带有类别偏好的标签平滑正则化方法,并结合迁移学习的方式构建了声呐图像小样本分类模型。除小样本问题外,不平衡问题也是声呐图像识别所面临的挑战。Jiao 等^[13]发现声呐图像数据呈现出典型的长尾不平衡分布,并提出了一种两阶段解耦的迁移学习模型来提高模型对稀有类别的关注程度。进一步地, Jiao 等^[14]研究了开放集中的声呐图像长尾识别(Open-set Long-Tail Recognition, OLTR)问题,并提出了一种集成方法作为基线模型,结合 PLUD 损

失有效提高了开放集声呐图像识别的性能。

不平衡图像分类 为了缓解图像分类中的不平衡问题,研究者通常采用数据分布再平衡或算法再平衡等方式来提高模型对尾部类别的关注度和识别准确率^[25-26]。数据分布再平衡策略主要包括重采样和数据增强。重采样通过过采样尾部类别样本或欠采样头部类别样本实现类别分布平衡,但容易导致尾部类别的过拟合和头部类别的欠拟合^[27]。此外,类别平衡采样、平方根采样与重采样方法等也被广泛应用^[28]。数据增强通过合成稀有类别样本来增加尾部类别的样本数量^[29-30],常用的数据增强方法包括 Mix-up^[31], ReMix^[32]等。算法分布再平衡策略通常包括损失函数改进和模型的集成学习。重加权是一种被广泛使用的损失函数改进方法,通过给不同频率的类别赋予不同的损失权重来强化模型对尾部类别的关注并平衡不同类别在分类器中的贡献度,例如再平衡损失函数通过每个类别的样本数量来重新为每个类的损失值分配权重,数量越少的样本类别将在损失函数中具有更大的权重,从而促使模型更加关注尾部样本稀有的类别,但重加权也会损害模型的泛化性和增强模型的优化难度^[33]。基于集成学习的模型策略通常将多个网络和分类器组合在一起形成专家组,每个专家组都可以处理特定的领域来解决长尾视觉学习问题^[34]。Zhou 等^[27]提出了将特征学习和

分类器学习解耦的思想,并通过预实验证明了均匀采样有助于模型的特征学习,而再平衡采样有助于分类器学习,但可能会损害模型主干网络的特征表达能力。因此,他们提出了一个双分支网络模型,通过一个均匀采样分支负责特征学习和一个反向采样器分支负责再平衡学习。Zhou 等^[25]将数据集划分为几个相互有重叠的子集,并在每个子集上训练一个领域内的分类专家进行协作决策。

2 多专家动态协作迁移学习模型

2.1 模型框架

多专家动态协作学习模型 MEDC-SI 框架如图 2 所示。模型骨干网络首先在大规模数据集上进行训练,然后在声呐图像数据集上进行微调。在微调阶段,我们分别设计了多专家网络和动态学习策略。多专家网络包含三个输入分支,包括一个传统分支和两个再平衡分支,且每一个分支拥有独立的分类器专家进行决策。为了训练不同专家的多多样性,三个分支分别对声呐图像数据集进行均匀采样(Uniform Sampling, US)、类弱平衡采样(Weakly Class-Balance Sampling, WCBS)和类平衡采样(Class-Balance Sampling, CBS),且三个分支共享骨干网络进行图像特征提取。动态学习策略能够自动地转移模型在不同训练阶段的学习注意力。在模型训练的前半

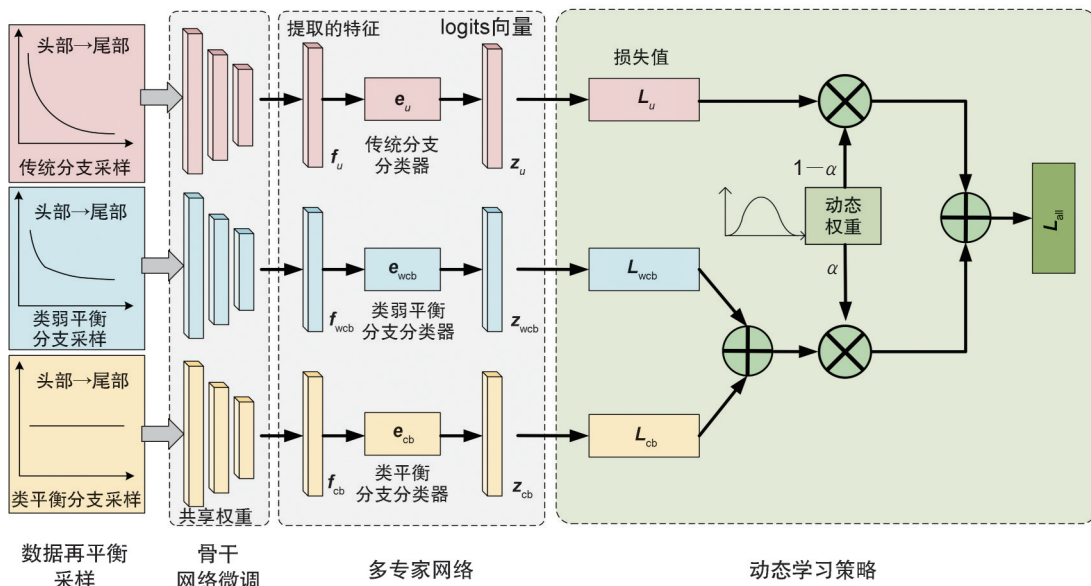


图2 多专家动态协作模型框架

Fig. 2 The framework of the multi-expert dynamic collaboration model

部分,动态权重将控制模型的训练注意力从传统分支向再平衡分支逐步转移以充分训练所有分支的分类器;在模型训练的后半部分,为了恢复由于训练再平衡分支而降低的模型骨干网络的特征表达能力,动态权重将控制模型的训练注意力从再平衡分支逐步转移回传统分支。最后,在推理阶段,MEDC-SI模型将三个专家的预测结果进行集成,实现协同决策。

2.2 迁移学习

声呐图像的获取受水下环境、设备和技术等因素的影响通常费时又昂贵,从而导致声呐图像数据集通常面临着数据量少的问题,这使得从头开始训练卷积神经网络(Convolutional Neural Network, CNN)变得具有挑战性。使用迁移学习的方式来对声呐图像识别模型进行训练是当前的主流做法^[35]。通过在大型数据集上预训练CNN,然后将学习到的知识转移到声呐图像小型数据集进行微调,可以克服数据集不足的影响。在本模型中,我们首先在ImageNet数据集上对模型进行预训练,并将预训练的模型权重设置为初始权重进行微调。在微调中,模型的骨干网络部分会被赋予较低的学习率以更适合声呐图像的特征提取,分类器全连接层部分替换为多专家网络中的分类器并进行重新训练。特别地,在动态学习策略中,模型后四分之一的训练将会冻结再平衡分类器权重以恢复骨干网络的特征提取能力。

2.3 多专家网络

为了提高长尾声呐图像识别问题的性能,多专家网络由传统分支和两个再平衡分支组成。其中传统分支采取均匀采样的方式进行训练,主要负责强化模型的特征表达能力和通用模式的识别;再平衡分支包含类平衡分支和类弱平衡分支,分别采取不同的再平衡采样方式进行训练,主要负责提高模型对尾部类别样本识别的能力。每个分支通过不同的采样方式训练各自所对应的专家分类器:

(1)传统分支。传统分支的数据输入来源于均匀采样器,主要负责从原始数据分布中学习通用模式的识别。在该分支的训练中,通过均匀采样器每一个样本都有均等的机会被选取,且在一个训练周期内每个样本仅会被采样

一次。因此,均匀采样器保留了数据分布的原始特征,在模型训练中能够更好地提高骨干网络的特征表达能力。

(2)类平衡分支。类平衡分支的数据输入来源于类平衡采样的方式。在该分支的训练中,每一类别的样本都有均等的机会被选取。在类平衡采样的训练下,尾部类别样本被选取的概率将增加,而头部类别样本被选取的概率则会降低,从而促使类平衡分支的专家分类器更关注尾部类别样本的特征并提高稀有样本的识别率。如图3(a)所示,给定声呐图像识别数据集中的 N 类类别 $D = \{C_1, C_2, \dots, C_N\}$ 和第 i 类类别 C_i 的样本数量为 M_i ,类平衡分支的样本采样步骤为:

① 根据类别数计算类别 i 被采样的概率 $P_i = \frac{1}{N}$,并根据概率随机选取一个类别;

② 根据选取的类别 i 的样本数 M_i 计算该类别下样本 j 被均匀采样的概率 $P_{ij} = \frac{1}{M_i}$,并根据概率随机选取一个样本。通过重复上述采样步骤可以获得一个训练批次的样本进行训练。

根据上述步骤,在类平衡分支中,任一样本 j 被选取的概率可以表示为: $P_j = \frac{1}{NM_i}$ 。

(3)类弱平衡分支。为了增加不同专家之间的多样性,类弱平衡分支被设计用来缓解类平衡分支采样中容易出现欠采样头部类别和过采样尾部类别的现象。类弱平衡分支的数据输入来源于类弱平衡采样的方式,通过基本保持头部类别样本的分布和轻微提高尾部类别样本采样概率相结合的方式降低原本数据集中的不平衡因子,使得专家分类器同时兼顾头部类别和尾部类别的分类,从而能够维持头部类别的样本识别率的情况下提高尾部类别的准确率。具体地,如图3(b)所示,类弱平衡采样的步骤为:

① 给定样本总数 M_{all} ,计算类别样本平均数: $M_{ave} = \frac{M_{all}}{N}$;

② 根据每一类别的样本数量计算类别 i 被采样的概率 $P_i = \frac{M_i}{M_{all}}$,并根据概率随机选取一个类别;

③ 若选取的类别 i 的样本数 $M_i \geq M_{ave}$,则

计算该类别下样本 j 被均匀采样的概率 $P_{ij} = \frac{1}{M_i}$, 并根据概率随机选取一个样本; 若选取的类别 i 的样本数 $M_i < M_{ave}$, 则计算该类别下样本 j 被均匀采样的概率 $P_{ij} = \frac{1}{M_i}$, 并根据概率随机选取两个样本。

通过重复上述采样步骤可以获得一个训练批次的样本进行训练。通过上述采样步骤, 头部类别样本被采样的概率基本维持不变, 尾部类别样本被采样的概率提高一倍, 相当于数据

分布的不平衡因子变为原始数据集的 $\frac{1}{2}$, 减弱了原始数据分布的不平衡并增加了分类器专家的多样性。

(4) 共享权重。在多专家网络中, 所有分支都共享模型的骨干网络并独立配备全连接专家分类器。通过共享骨干网络权重, 再平衡分支可以利用传统分支训练的骨干网络更好地提取图像的特征表达, 也能够极大地降低模型所需要训练的参数和推理时的计算复杂度。

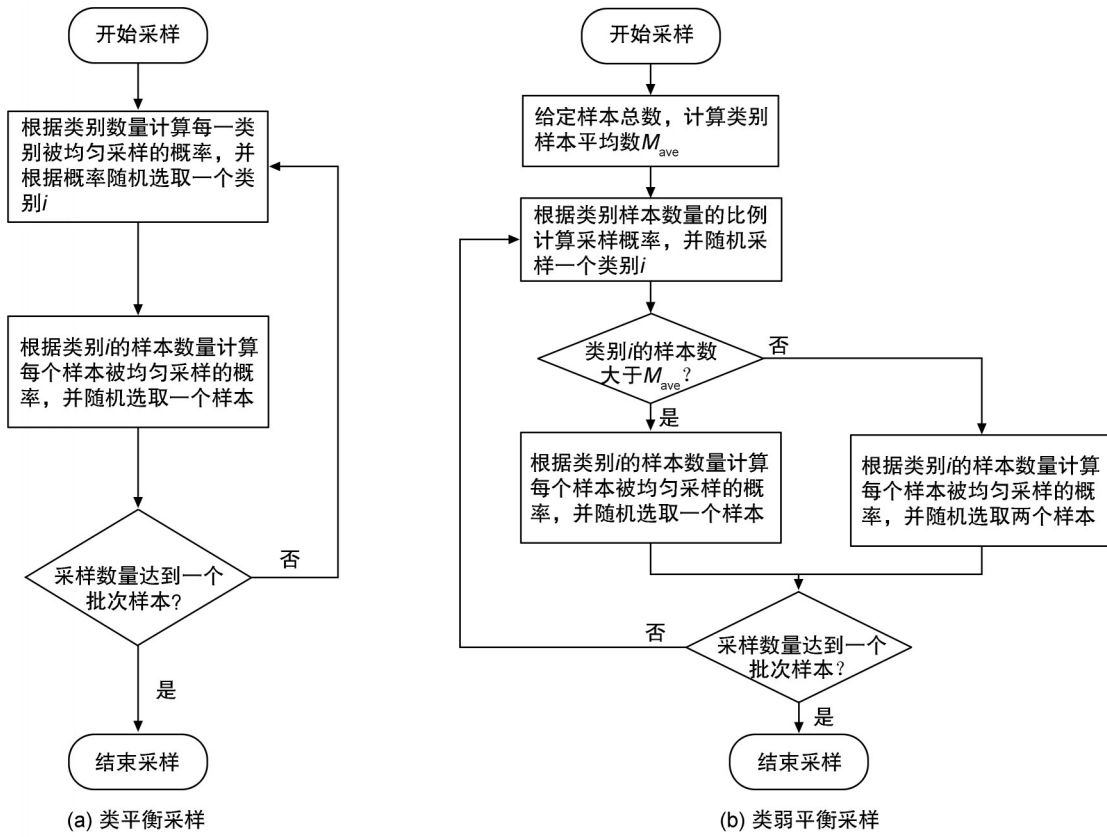


图3 类平衡采样和类弱平衡采样的流程图

Fig. 3 The flow charts of class-balance sampling and weakly class-balance sampling

2.4 动态学习策略

动态学习策略的目的是在训练过程中自动转移模型在传统分支和再平衡分支之间的学习注意力以提高模型的特征表达能力和分类器学习能力。在本文中, 按照“特征学习—分类器学习—特征表达能力恢复”的思想, 我们设计了基于正弦函数的动态权重调整策略来平衡三个分支之间的特征学习和分类器学习。给定多专家网络中传统分支所对应分类器的输出 logits 向量 z_u , 其预测概率 P_u 和损失值 L_u 可以表示为:

$$P_u^k = \frac{e^{z_u^k}}{\sum_{m=1}^N e^{z_u^m}}, \quad (1)$$

$$L_u = -\sum_{k=1}^N y_u^k \log(P_u^k), \quad (2)$$

其中 y_u 为真实标签的 one-hot 向量。类似地, 给定类平衡分支分类器和类弱平衡分支分类器的输出 logits 向量 z_{cb} 和 z_{wcb} , 通过交叉熵损失函数可以计算得到每个分支的损失函数值 L_{cb} 和 L_{wcb} 。那么, 基于动态学习策略的多专家网络的损失函数可以表示为:

$$L_{\text{all}} = \alpha \left(\frac{L_{\text{cb}} + L_{\text{wcb}}}{2} \right) + (1 - \alpha) L_u, \quad (3)$$

其中 α 为动态权重。通过上述公式可以发现, 动态权重同时作用到所有的三个分支, 其中类平衡分支和类弱平衡分支组合在一起形成再平衡损失值 $\left(\frac{L_{\text{cb}} + L_{\text{wcb}}}{2} \right)$ 与传统分支的损失值 L_u

在动态权重的调节下形成模型的总损失 L_{all} 。特别地, 为了减少超参设置, 类平衡分支和类弱平衡分支在总损失中具有相同的权重, 且取二者的平均值以防止尾部类别权重过大所导致的模型优化困难和过拟合尾部类别。具体地, 给定模型训练迭代的总次数 T_{max} , 在模型训练的第 t 次迭代时, 动态权重 α 的权值为:

$$\alpha = \frac{1}{2} \cos \left(\frac{2\pi t}{T_{\text{max}}} - \pi \right) + \frac{1}{2}. \quad (4)$$

动态权重 α 随训练迭代次数 t 的变化趋势如图 4 所示。从图中可以发现, 当 $0 \leq t < \frac{T_{\text{max}}}{4}$ 时, α 从 0 逐渐增到 $\frac{1}{2}$ 。此时, 在模型训练中, 传统分支将占据主导地位, 模型在训练的初始阶段将主要训练传统分支的分类器和提高模型的特征表达能力。当 $\frac{T_{\text{max}}}{4} \leq t < \frac{3T_{\text{max}}}{4}$ 时, $\frac{1}{2} \leq \alpha \leq 1$ 。此时, 模型的训练注意力从传统分支逐步转移到再平衡分支, 再平衡分支在模型训练中占据主导地位, 促使模型提升再平衡分类器对尾部类别的识别能力。当 $\frac{3T_{\text{max}}}{4} \leq t \leq T_{\text{max}}$ 时, α 从 $\frac{1}{2}$ 逐渐降低为 0。此时, 模型的训练注意力从再平衡分支重新转移回传统分支, 模型将主要恢复被再平衡训练所损害的特征表达能力并进一步充分训练传统分支的分类器。在此阶段, 再平衡分类器的权重参数将被冻结。

在推理阶段, 模型将对三个分类器专家的输出预测进行集成。与公式(3)类似, 集成后的 logits 向量 z_{all} 可以计算为:

$$z_{\text{all}} = \alpha \left(\frac{z_{\text{cb}} + z_{\text{wcb}}}{2} \right) + (1 - \alpha) z_u, \alpha = \frac{1}{2}. \quad (5)$$

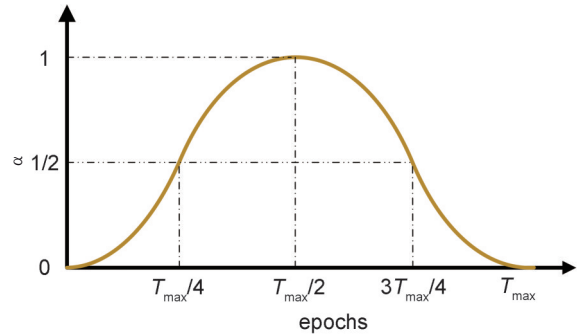


图4 动态权重 α 随训练迭代次数 t 的变化趋势

Fig. 4 The changes of α value with the increase of epoch number t

3 实验

3.1 实验设置

数据集。为与其他模型进行公平的比较, 本文模型将在 3 个公开声呐图像识别数据集上进行测试: KLSG^[36]、FLSMDD^[37] 和 NKSID^[14]。KLSG 数据集是一种真实环境采集的侧扫声呐图像数据集, 该数据集中包含两类图像: 385 张沉船图像和 62 张飞机图像, 其不平衡因子为 6.21。FLSMDD 数据集一共 2 364 张声呐图像, 包含常见海洋垃圾等 10 个类别, 例如: 瓶子、链条、钩子、螺旋桨、轮胎等, 其不平衡因子为 6.91。NKSID 数据集为前视声呐图像数据集, 包含 2 617 张声呐图像和 8 类常见水下物体, 例如螺旋桨、渔网、管道、轮胎等, 该数据集不平衡程度更为严重, 不平衡因子为 47.55。

实验测度。结合长尾分布识别和声呐图像识别的特点, 与 BETL^[13] 等模型类似, 本文选取了 Top-1 Acc (准确率)、G-mean 和 Macro-F1 作为模型的评估测度。Top-1 Acc 定义为正确预测的正例样本数在总预测的正例样本数中的占比, 主要衡量模型对所有样本的识别准确率。但在不平衡分布中, Top-1 Acc 容易受到头部类别性能的影响。G-mean 通过平等考虑每一类别的分类精度来更好地反映模型对尾部类别的分类精度。Macro-F1 测度以精确率 P 和召回率 R 为基础计算 F1 值, 能够更为全面的反映模型对尾部类别的识别性能。Macro-F1 是多分类任务中的一种拓展。它先对每个类别独立计算 F1 值, 然后求所有 F1 值的算术平均值。Macro-F1 的计算方法为:

$$F1_i = \frac{2 \times P_i \times R_i}{P_i + R_i}, \quad (6)$$

$$\text{Macro-F1} = \frac{1}{K} \sum_{i=1}^K F1_i, \quad (7)$$

其中 K 为 $F1$ 值的总个数。此外,在结果可视化中,本文还采用了混淆矩阵的方式对模型的性能进行评测。

实现细节。与 BETL 模型^[13]保持一致,我们以 ResNet18 为模型骨干网络继续特征提取。所有模型训练 160 个纪元,模型通过 SGD 优化算法进行训练,其中动量为 0.9,权重衰减参数设置为 0.001。初始学习率为 0.01,当模型训练至 100 和 140 纪元时,学习率将变为原来的 0.1 倍。所有实验均在 Ubuntu 20.04 操作系统下进行,模型基于 Python 3.7.0 和 Pytorch 1.12.0 在 Nvidia GeForce RTX 3090Ti 上进行训练。

3.2 与最新方法的对比

为了比较的公平性,我们将本文 MEDC-SI 模型与最新的声呐图像分类方法在相同骨干网络框架下进行了对比,包括 CNN-SVM^[17]、DTL^[38]、DTL-RW^[39]、CE-DRS^[40]、LTR-WB^[41]、BETL^[13]、PLUD^[14]等。与最新模型在 KLSG 和 FLSMDD 数据集上的性能对比如表 1 所示,其中*表示本文重现结果。

表 1 中可以看出,本文 MEDC-SI 模型在 KLSG 和 FLSMDD 数据集上都取得了更优的性能。在 TOP-1 Acc 测度上,本文模型在 KLSG 数据集上取得了 91.51% 的准确率,比最新的 BETL 和 PLUD 模型分别提高了 3.37% 和 1.63%。通过对比 DTL 和 DTL-RS 模型的性能可以发现,对于长尾分布的数据,再平衡的训练方式可以一定程度提高模型对尾部样本的识别能力。例如,在 FLSMDD 数据集上,DTL-RS 的 G -mean 测度性能比 DTL 提高了 3.62%。类似地,通过多平衡采样和剪枝, BETL 模型比 LTR-WB 模型在 G -mean 测度和 $F1$ 测度上的性能分别提高了 0.35% 和 0.84%。本文采用多个不平衡专家相互协同的方式,提高了模型对尾部类别的识别能力。在 G -mean 测度和 $F1$ 测度上,本文 MEDC-SI 模型比 BETL 模型的性能分别提高了 2.29% 和 2.32%。此外,模型在具有更大不平衡因子的 NKSID 数据集上进行了测试。表 2 中展示了不同模型在 NKSID 数据集上的性能。

表 1 模型在 KLSG 和 FLSMDD 数据集上的性能对比结果(%)

Table 1 The comparison results of our model and other SOTA models on KLSG and FLSMDD datasets (%)

模型	KLSG			FLSMDD		
	Acc	G -mean	$F1$	Acc	G -mean	$F1$
CNN-SVM	—	74.47	76.62	—	67.56	72.42
DTL	—	77.80	76.32	—	92.36	92.02
DTL-RS	—	73.04	67.62	—	95.98	95.72
CE-DRS	—	53.94	47.41	—	75.91	75.65
LTR-WB	—	54.95	42.54	—	97.04	96.56
BETL*	88.14	86.07	79.15	96.79	97.39	97.40
PLUD*	89.88	89.46	82.12	99.22	99.14	98.86
MEDC-SI	91.51	91.96	84.76	99.74	99.68	99.72

注:*表示本文重现结果,—表示该方法没有给出 Acc 结果。

表 2 模型在 NKSID 数据集上的性能对比结果(%)

Table 2 The comparison results of our model and other SOTA models on NKSID dataset (%)

模型	NKSID		
	Acc	G -mean	$F1$
BETL*	92.12	84.46	85.22
PLUD*	95.47	87.82	90.34
MEDC-SI	96.19	88.83	91.23

注:*表示本文重现结果。

从表 2 中可以看出,本文 MEDC-SI 模型在 NKSID 数据集上也取得了最优的性能。当数据集出现严重不平衡时,模型对于稀有类别的学习将更加困难。本文 MEDC-SI 模型通过多专家网络和动态学习策略能够在保持模型对头部类别识别性能的同时有效提升尾部类别的识别性能。在 Top-1 Acc 准确率测度上,本文 MEDC-SI 模型比 BETL 和 PLUD 模型分别提升了 4.07% 和 0.72%。在更能反映模型对尾部类别识别性能的 G mean 和 $F1$ 测度上,本文模型比最新的 PLUD 模型分别提升了 1.01% 和 0.89%,这充分证明了本文模型的有效性。

3.3 消融性实验

为了充分探索本文模型中所提出模块的影响,本文设计了两个消融性实验来验证多专家网络和动态学习策略对模型性能的影响。表 3 展示了多专家网络不同变种在 FLSMDD 数据集上对模型性能的影响。

如表 3 所示,我们分别给多专家网络配备了 4 种不同的分支。当在多专家网络中仅配置传统分支和逆采样分支时,由于逆采样分支会

过度伤害头部类别的性能,其在 G -mean 和 $F1$ 测度上的性能分别为 97.74% 和 97.61%。当将逆采样分支分别替换为类弱平衡分支和类平衡分支时,其 G -mean 测度上的性能分别提升了 0.9% 和 1.44%。特别地,当将传统分支、类平衡分支和类弱平衡分支三者组合时,模型不仅能够精准地识别头部类别,更能提升尾部类别的识别性能。与逆采样分支相比,其在 G -mean 和 $F1$ 测度上的性能分别提升了 1.94% 和 2.11%。

表 3 多专家网络不同变种在 FLSMDD 数据集上对模型性能的影响(%)

Table 3 The impact of different variants of multi-expert networks on FLSMDD dataset (%)

传统分支	类平衡分支	类弱平衡分支	其他采样	G -mean	$F1$
✓			逆采样	97.74	97.61
✓		✓		98.64	98.64
✓	✓			99.18	98.92
✓	✓	✓		99.68	99.72

注:✓表示采用该模块,空白表示没有采用相应模块。

为了验证动态学习策略的有效性,我们设计了多种不同的学习策略来对模型进行训练,包括相同权重策略、线性衰减策略、分段线性策略^[25]等。表 4 展示了不同学习策略对模型性能的影响。

表 4 不同学习策略对模型性能的影响(%)

Table 4 The impact of different learning strategies on model performance (%)

策略	G -mean	$F1$
相同权重策略	95.85	94.25
线性衰减策略	97.05	96.22
分段线性策略	99.33	99.44
MEDC-SI	99.68	99.72

从表 4 中结果可以发现,线性衰减策略的 G mean 测度性能要低于分段线性策略和本文模型,这说明单调的权重衰减策略并不是模型训练优化的最优选择,因为当模型侧重于再平衡分支的训练时会不可避免地损害骨干网络的特征表达能力。当采用分段线性策略时,模型的 G -mean 测度性能比线性衰减策略和相同权重策略分别提升了 3.48% 和 5.19%。而本文 MEDC-SI 模型比分段线性策略的 G -mean 和 $F1$ 测度性能提升了 0.35% 和 0.28%,且在这四种策略中取得了最优的性能。

3.4 可视化结果分析

为了充分地证明本文 MEDC-SI 模型的优点,我们进一步将模型在 FLSMDD 和 NKSID 数据集上的混淆矩阵和预测结果进行了可视化。图 5 展示了模型分别在 FLSMDD 和 NKSID 数据集上的混淆矩阵。

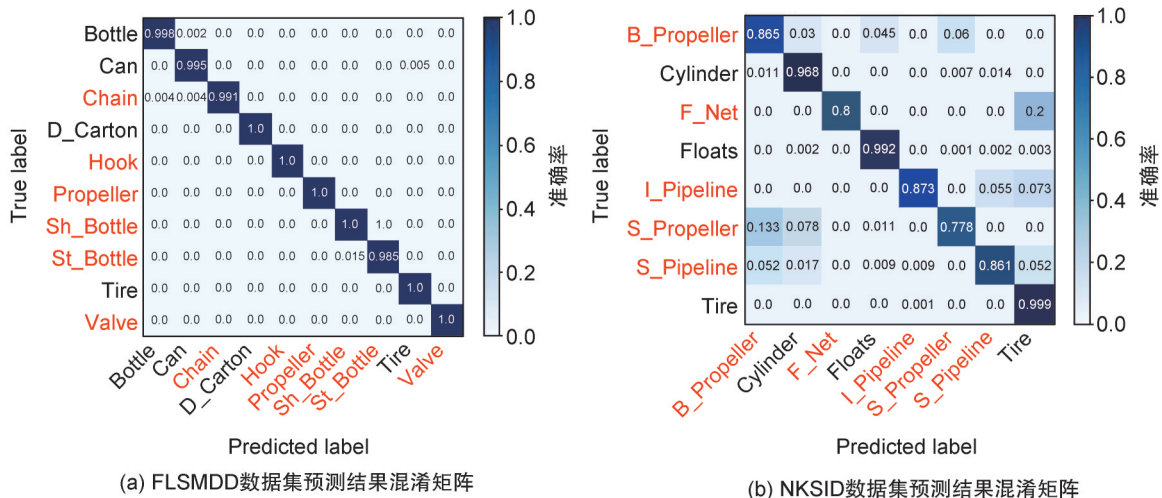


图 5 混淆矩阵结果图

Fig. 5 Results of confusion matrix on FLSMDD and NKSID datasets

在混淆矩阵图中,我们分别展示了每一类别的具体识别准确率的情况。具体地,在图 5 中,

水平轴和垂直轴分别为各个类别的预测标签和真实标签,混淆矩阵中的对角线的数值展示了每

一类别的识别准确率。为了更好地展示模型对尾部类别的识别效果,我们将尾部类别标签用红色进行标注。从图5中可以发现,在FLSMDD数据集上,本文MEDC-SI模型在头部类别和尾部类别的识别上都达到了较好的效果。在NKSID数据集上,受到极度不平衡数据的影响,模型会

将部分尾部类别错误地辨别为头部类别,但整体尾部类别识别准确率相比其他模型依然得到一定提升。图6展示了本文MEDC-SI模型和BETL模型对于部分测试样本的识别结果。从图中可以发现本文MEDC-SI模型能够同时更好地辨别头部类别样本和尾部类别样本。

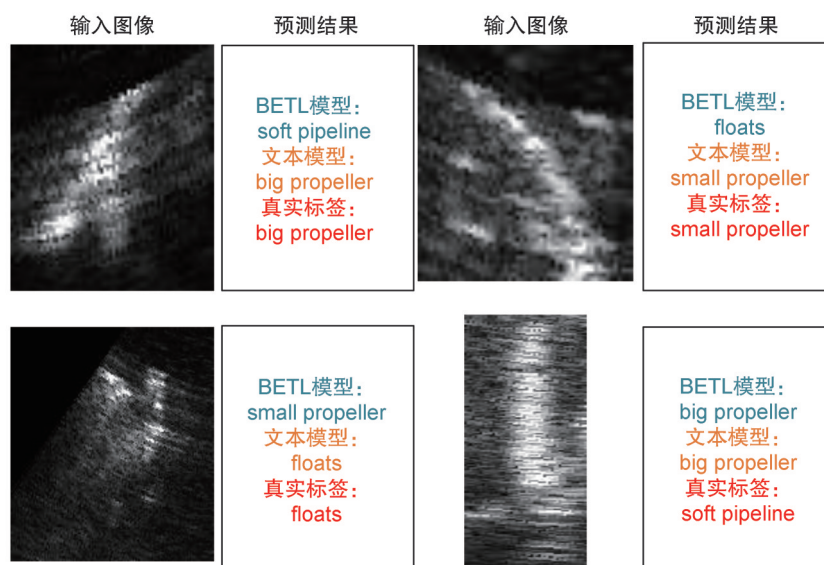


图6 BETL模型和本文MEDC-SI模型在NKSID数据集上的测试结果样例

Fig. 6 The samples of predicted results of BETL and our model on NKSID dataset

4 结论

本文提出了一个多专家动态协作模型来实现不平衡声呐图像识别。为了提高模型对稀有类别样本的识别率,本文设计了一个多专家网络,通过采取多种再平衡采样方式来训练多个具有特定专长的专家,促使专家间合作实现长尾声呐图像识别。为了提高每个专家的辨别能力,本文还设计了一种动态学习策略,通过在训练中转移模型对不同专家的注意力来提高模型的特征学习能力和分类器识别能力。最后,模型在三个声呐图像数据集上进行了充足的实验和消融性研究以证明模型的有效性。在未来的研究中,不平衡声呐图像的开放集研究是一个值得进一步探索的任务。

参考文献:

[1] 黄海宁,李宝奇,刘纪元,等.声呐图像水下目标识别综述与展望[J].电子与信息学报,2024,46(5):1742-1760. DOI: 10.11999/JEIT231207.
HUANG H N, LI B Q, LIU J Y, et al. Sonar Image Un-

derwater Target Recognition: A Comprehensive Overview and Prospects[J]. *J Electron Inf Technol*, 2024, **46** (5): 1742-1760. DOI: 10.11999/JEIT231207.

- [2] MA Q X, JIANG L Y, YU W X. Lambertian-based Adversarial Attacks on Deep-learning-based Underwater Side-scan Sonar Image Classification[J]. *Pattern Recognit*, 2023, **138**: 109363. DOI: 10.1016/j.patcog.2023.109363.
- [3] VALDENEGRO-TORO M, PRECIADO-GRIJALVA A, WEHBE B. Pre-trained Models for Sonar Images[C]// OCEANS 2021. San Diego-Porto: IEEE, 2021: 1-8. DOI: 10.23919/OCEANS44145.2021.9705825.
- [4] CHO H, GU J, YU S C. Robust Sonar-based Underwater Object Recognition Against Angle-of-view Variation[J]. *IEEE Sens J*, 2016, **16**(4): 1013-1025. DOI: 10.1109/JSEN.2015.2496945.
- [5] MARSZAL J, SALAMON R. Detection Range of Intercept Sonar for CWFM Signals[J]. *Arch Acoust*, 2015, **39** (2): 215-230. DOI: 10.2478/aoa-2014-0026.
- [6] FEI T, KRAUS D, ZOUBIR A M. Contributions to Automatic Target Recognition Systems for Underwater Mine Classification[J]. *IEEE Trans Geosci Remote Sens*, 2015, **53**(1): 505-518. DOI: 10.1109/TGRS.2014.2324971.

- [7] MYERS V, FAWCETT J. A Template Matching Procedure for Automatic Target Recognition in Synthetic Aperture Sonar Imagery[J]. *IEEE Signal Process Lett*, 2010, **17**(7): 683–686. DOI: 10.1109/LSP.2010.2051574.
- [8] BARNGROVER C M. Automated Detection of Mine-like Objects in Side Scan Sonar Imagery[M]. San Diego: University of California, 2014.
- [9] CHHABRA P S, PAILHAS Y, WALLACE A M, *et al.* Target Classification in SAS Imagery Using Orthogonal Basis Selection[C]//OCEANS 2017–Aberdeen. New York: IEEE, 2017: 1–10. DOI: 10.1109/OCEANSE.2017.8084647.
- [10] WILLIAMS D P. Underwater Target Classification in Synthetic Aperture Sonar Imagery Using Deep Convolutional Neural Networks[C]//2016 23rd International Conference on Pattern Recognition (ICPR). New York: IEEE, 2016: 2497–2502. DOI: 10.1109/ICPR.2016.7900011.
- [11] CHENG Z, HUO G Y, LI H S. A Multi-domain Collaborative Transfer Learning Method with Multi-scale Repeated Attention Mechanism for Underwater Side-scan Sonar Image Classification[J]. *Remote Sens*, 2022, **14**(2): 355. DOI: 10.3390/rs14020355.
- [12] XU H P, YANG L Z, ZHANG M X. Unsupervised Classification Based on Deep Adaptation Network for Sonar Images[J]. *J Electron Imag*, 2023, **32**: 013029. DOI: 10.1117/1.JEL.32.1.013029.
- [13] JIAO W P, ZHANG J L. Sonar Images Classification while Facing Long-tail and Few-shot[J]. *IEEE Trans Geosci Remote Sens*, 2022, **60**: 4210420. DOI: 10.1109/TGRS.2022.3211847.
- [14] JIAO W P, ZHANG J L, ZHANG C Y. Open-set Recognition with Long-tail Sonar Images[J]. *Expert Syst Appl*, 2024, **249**: 123495. DOI: 10.1016/j.eswa.2024.123495.
- [15] YANG Y X, WEI Y, GUO J, *et al.* Recognition of Side-scan Sonar Images under Long-tail Distribution[C]//International Conference on Computer Vision and Image Processing (CVIP 2024). Bellingham: SPIE, 2025: 33. DOI: 10.1117/12.3058318.
- [16] LI S B, MA J F, WU Y L, *et al.* SSS Small Target Detection via Combining Weighted Sparse Model with Shadow Characteristics[J]. *IEEE Trans Geosci Remote Sens*, 2023, **61**: 5911711. DOI: 10.1109/TGRS.2023.3285436.
- [17] ZHU P 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). New York: IEEE, 2017: 2724–2731. DOI: 10.1109/CDC.2017.8264055.
- [18] HE K M, ZHANG X Y, REN S Q, *et al.* Deep Residual Learning for Image Recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2016: 770–778. DOI: 10.1109/CVPR.2016.90.
- [19] HOWARD A G, ZHU M, CHEN B, *et al.* Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications[EB/OL]. (2017–04–17)[2025–03–15]. <https://arXiv.org/abs/1704.04861>.
- [20] HUANG G, LIU Z, VAN DER MAATEN L, *et al.* Densely Connected Convolutional Networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2017: 2261–2269. DOI: 10.1109/CVPR.2017.243.
- [21] ARRIAGA O, VALDENEGRO-TORO M, PLÖGER P. Real-time Convolutional Neural Networks for Emotion and Gender Classification[J]. (2017–10–20)[2025–03–15]. <https://arXiv.org/abs/1710.07557>.
- [22] YANG Z W, ZHAO J H, YU Y C, *et al.* A Sample Augmentation Method for Side-scan Sonar Full-class Images that Can be Used for Detection and Segmentation [J]. *IEEE Trans Geosci Remote Sens*, 2024, **62**: 5908111. DOI: 10.1109/TGRS.2024.3371051.
- [23] XU H L, BAI Z Y, ZHANG X Y, *et al.* MFSANet: Zero-shot Side-scan Sonar Image Recognition Based on Style Transfer[J]. *IEEE Geosci Remote Sens Lett*, 2023, **20**: 1503105. DOI: 10.1109/LGRS.2023.3318051.
- [24] 陈禹乐, 李博, 梁红, 等. 小样本下基于深度学习的声呐图像分类研究[J]. *西北工业大学学报*, 2022, **40**(4): 739–745. DOI: 10.1051/jnwpu/20224040739.
- [25] CHEN Y L, LI B, LIANG H, *et al.* Research on Sonar Image Few-shot Classification Based on Deep Learning [J]. *J Northwest Polytech Univ*, 2022, **40**(4): 739–745. DOI: [org/10.1051/jnwpu/20224040739](https://doi.org/10.1051/jnwpu/20224040739).
- [26] ZHOU H, LUO T J, HE Y M. Dynamic Collaborative Learning with Heterogeneous Knowledge Transfer for Long-tailed Visual Recognition[J]. *Inf Fusion*, 2025, **115**: 102734. DOI: 10.1016/j.inffus.2024.102734.
- [27] DU C Q, WANG Y L, SONG S J, *et al.* Probabilistic Contrastive Learning for Long-tailed Visual Recognition[J]. *IEEE Trans Pattern Anal Mach Intell*, 2024, **46**(9): 5890–5904. DOI: 10.1109/TPAMI.2024.3369102.
- [28] ZHOU B Y, CUI Q, WEI X S, *et al.* BBN: Bilateral-branch Network with Cumulative Learning for Long-tailed Visual Recognition[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2020: 9716–9725. DOI: 10.1109/cvpr42600.2020.00974.
- [29] REN J, YU C, MA X, *et al.* Balanced Meta-softmax for

- Long-tailed Visual Recognition[C]//Advances in Neural Information Processing Systems. Red Hook: Curran Associates. 2020, **33**: 4175–4186.
- [29] GAO J, ZHAO H, LI Z, *et al.* Enhancing Minority Classes by Mixing: An Adaptive Optimal Transport Approach for Long-tailed Classification[C]//Advances in Neural Information Processing Systems. Red Hook: Curran Associates. 2023, **36**: 60329–60348.
- [30] LI M K, HU Z K, LU Y, *et al.* Feature Fusion from Head to Tail for Long-tailed Visual Recognition[J]. *Proc AAAI Conf Artif Intell*, 2024, **38**(12): 13581–13589. DOI: 10.1609/aaai.v38i12.29262.
- [31] ZHANG H, CISSE M, DAUPHIN Y N, *et al.* Mixup: Beyond Empirical Risk Minimization[C/OL]//International Conference on Learning Representations, 2018. (2017–11–06) [2025–03–15]. <https://openreview.net/forum?id=r1Ddp1-Rb>.
- [32] CHOU H P, CHANG S C, PAN J Y, *et al.* Remix: Rebalanced Mixup[M]//Computer Vision-ECCV 2020 Workshops. Cham: Springer International Publishing, 2020: 95–110. DOI: 10.1007/978-3-030-65414-6_9.
- [33] TAN J R, WANG C B, LI B Y, *et al.* Equalization Loss for Long-tailed Object Recognition[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2020: 11659–11668. DOI: 10.1109/cvpr42600.2020.01168.
- [34] JAMAL M A, BROWN M, YANG M H, *et al.* Rethinking Class-balanced Methods for Long-tailed Visual Recognition from a Domain Adaptation Perspective[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2020: 7607–7616. DOI: 10.1109/CVPR42600.2020.00763.
- [35] HUANG C, ZHAO J H, ZHANG H M, *et al.* Seg2Sonar: a Full-class Sample Synthesis Method Applied to Underwater Sonar Image Target Detection, Recognition, and Segmentation Tasks[J]. *IEEE Trans Geosci Remote Sens*, 2024, **62**: 5909319. DOI: 10.1109/TGRS.2024.3363875.
- [36] HUO G Y, WU Z Y, LI J B. 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.
- [37] VALDENEGRO-TORO M. Deep Neural Networks for Marine Debris Detection in Sonar Images[EB/OL]. (2019-05-13) [2025-03-15]. <https://arXiv.org/abs/1905.05241>.
- [38] YE X F, LI C L, ZHANG S Y, *et al.* Research on Sidescan Sonar Image Target Classification Method Based on Transfer Learning[C]//OCEANS 2018 MTS/IEEE Charleston. New York: IEEE, 2018: 1–6. DOI: 10.1109/OCEANS.2018.8604691.
- [39] WANG Y X, RAMANAN D, HEBERT M. Learning to Model the Tail[C]//Advances in Neural Information Processing Systems. Red Hook: Curran Associates. 2017: 30.
- [40] CAO K, WEI C, GAIDON A, *et al.* Learning Imbalanced Datasets with Label-distribution-aware Margin Loss[C]//Advances in Neural Information Processing Systems. Red Hook: Curran Associates. 2019: 32.
- [41] ALSHAMMARI S, WANG Y X, RAMANAN D, *et al.* Long-Tailed Recognition via Weight Balancing[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2022: 6887–6897. DOI: 10.1109/CVPR52688.2022.00677.