基于被动声呐音频信号的水中目标识别综述

徐齐胜[1,2] 许可乐[1,2] 窦勇[1,2] 高彩丽[1,2] 乔鹏[1,2] 冯大为[1,2] 朱博青[1,2]

摘 要 基于被动声呐音频信号的水中目标识别是当前水下无人探测领域的重要技术难题, 在军事和民用领域都应用广泛. 本文从数据处理和识别方法两个层面系统阐述基于被动声呐信号进行水中目标识别的方法和流程. 在数据处理方面, 从基于被动声呐信号的水中目标识别基本流程、被动声呐音频信号分析的数理基础及其特征提取三个方面概述被动声呐信号处理的基本原理. 在识别方法层面, 全面分析基于机器学习算法的水中目标识别方法, 并聚焦以深度学习算法为核心的水中目标识别研究. 本文从有监督学习、无监督学习、自监督学习等多种学习范式对当前研究进展进行系统性的总结分析, 并从算法的标签数据需求、鲁棒性、可扩展性与适应性等多个维度分析这些方法的优缺点. 同时, 还总结该领域中较为广泛使用的公开数据集, 并分析公开数据集应具备的基本要素. 最后, 通过对水中目标识别过程的论述, 总结目前基于被动声呐音频信号的水中目标自动识别算法存在的困难与挑战, 并对该领域未来的发展方向进行展望.

关键词 被动声呐信号, 水中目标自动识别, 深度学习, 有监督学习, 自监督学习

引用格式 徐齐胜, 许可乐, 窦勇, 高彩丽, 乔鹏, 冯大为, 朱博青. 基于被动声呐音频信号的水中目标识别综述. 自动化学报, 2024, **50**(4): 649−673

DOI 10.16383/j.aas.c230153

A Review of Underwater Target Recognition Based on Passive Sonar Acoustic Signals

XU Qi-Sheng[1,2] XU Ke-Le[1,2] DOU Yong[1,2] GAO Cai-Li[1,2]
QIAO Peng[1,2] FENG Da-Wei[1,2] ZHU Bo-Qing[1,2]

Abstract Underwater target recognition based on passive sonar acoustic signals is a significant technical challenge in the field of underwater unmanned detection, with broad applications in both military and civilian domains. This paper provides a comprehensive exposition of the methodology and process involved in underwater target recognition using passive sonar acoustic signals, addressing data processing and recognition methods at two levels. Regarding data processing, this paper presents a thorough exploration of the fundamental principles of passive sonar signal processing from three key aspects: The underlying process of underwater target recognition based on passive sonar signals, the mathematical foundations of passive sonar acoustic signal analysis, and the extraction of relevant features. At the level of recognition methods, this paper offers a comprehensive analysis of underwater target recognition techniques based on machine learning algorithms, and focus on the research conducted with deep learning algorithms at its core. The paper systematically summarizes and analyzes the current research progress across various learning paradigms, including supervised learning, unsupervised learning and self-supervised learning, and analyzes their advantages and disadvantages in terms of the algorithm′s labeled data requirement, robustness, scalability and adaptability. Additionally, this paper provides an overview of widely-used public datasets in the field, and outlines the essential elements that such datasets should possess. Finally, by discussing the process of underwater target recognition, this paper summarizes the current difficulties and challenges in automatic underwater target recognition algorithms based on passive sonar acoustic signals, and offers insights into the future development direction of this field.

Key words Passive sonar signal, automatic underwater target recognition, deep learning, supervised learning, self-supervised learning

Citation Xu Qi-Sheng, Xu Ke-Le, Dou Yong, Gao Cai-Li, Qiao Peng, Feng Da-Wei, Zhu Bo-Qing. A review of underwater target recognition based on passive sonar acoustic signals. *Acta Automatica Sinica*, 2024, **50**(4): 649−673

随着人类对海洋资源开发利用的不断深入以及海上安全问题的日益突出, 水声目标识别 (Underwater acoustic target recognition, UATR) 作为海洋环境监测的一项基础性任务, 成为近年来水声信号处理领域中的研究热点之一. 目前该研究内容已广泛应用于海底目标定位与识别[1]、海岸线监视[2]、

收稿日期 2023-03-22 录用日期 2023-07-10
Manuscript received March 22, 2023; accepted July 10, 2023
本文责任编委 赫然
Recommended by Associate Editor HE Ran
1. 国防科技大学计算机学院 长沙 410073 2. 并行与分布处理国防科技重点实验室 长沙 410073
1. School of Computer Science, National University of Defense Technology, Changsha 410073 2. National Key Laboratory of Parallel and Distributed Processing, Changsha 410073

海洋生物行为的计数和分类[3]、船只识别[4] 以及潜艇、鱼雷的检测[5] 等领域. 相比于电磁信号, 基于声学信号进行分析是水中目标识别更加行之有效的方法, 主要有以下三个原因: 一是声波在水中的传播速度较快且衰减较慢, 相比之下电磁波在水中传播速度慢且衰减迅速; 二是水中目标通常使用声波信号进行通信; 三是声波在不同水域环境中具有更好的适应性, 无论是海洋、湖泊还是河流等不同水体环境, 声波传播的特性相对稳定, 使得被动声呐能够适应不同的水下环境进行目标识别. 而电磁信号在不同水体环境中的传播特性存在较大差异, 需要进行针对性的调整和适配. 一般而言, 声波信号的采集可以通过主动声呐和被动声呐获取. 特别地, 被动声呐具有干扰性小、效率高、可同时接收来自多个方向的声波等优点, 近年来被广泛部署, 是当前水中目标识别的主要数据来源. 然而, 由于海洋环境的复杂多变, 水中目标的声呐信号会受到许多干扰, 如海洋背景噪声、多路径效应、信号衰减等, 从而导致可用于研究和分析的被动声呐信号往往数量较少, 这大大增加了水中目标识别的挑战性. 为有效预处理和分析被动声呐信号, 设计高通用性和泛化性的特征提取方法、提升水中目标识别的准确率和时效性、降低模型的训练成本和复杂度、构建质量良好的公开可用水声数据集, 都是基于被动声呐音频信号的水中目标识别任务所面临的关键问题.

一般而言, 水中目标大致包含水面目标和水下目标两个大类. 其中水面目标主要是各种大型舰船、小型船以及浮标等; 水下目标则主要是各类海洋生物、潜艇、鱼雷等. 水中目标识别旨在通过非接触的方式实现目标类别的判断[6], 一般包括声学特征提取并据此进行信号的识别两个阶段. 该过程通常涉及信号处理、模式识别和机器学习等相关知识, 根据信号的特征 (如频率、振幅、持续时间和频谱特征) 进行目标的识别. 传统上, 该任务主要依赖于专业的声呐操作员进行人工听音判别来实现水中目标的识别. 然而, 该方式易受操作员的情绪、所处环境、健康状态以及外界天气等多种因素的影响, 从而导致错误的判断. 此外, 人工听音判别效率低, 难以适应复杂多变的海洋环境和无法满足日益增长的监测需求. 近年来机器学习特别是深度学习算法在许多领域中表现出强大的学习能力和优异的自动识别能力, 激励了学者探索该方法在水中目标识别任务中的应用, 目前基于机器学习的水中目标识别方法逐渐成为该领域的研究热点[7].

随着机器学习在基于被动声呐音频信号的水中目标识别任务的研究不断深入, 众多成果不断涌现, 近年来也出现了一些综述性工作. 例如, 文献 [8] 从水声特征提取的角度分析不同声学特征的提取方式和物理特性, 简单介绍部分目标识别方法. 文献 [9] 则从方法层面对水中目标识别研究进行综述. 相比之下, 本文同时从声学特征提取和方法层面对已有方法进行总结. 此外, 文献 [9] 只从宏观上分析不同深度学习方法的性能差异, 而本文从精度、鲁棒性、扩展性等多个维度对比不同方法的一般性能差异. 文献 [10] 从方法层面将基于机器学习的水中目标识别划分为基于统计学的方法、基于深度学习的方法和基于迁移学习的方法, 并进行相应的综述. 本文与文献 [10] 的区别主要体现在以下两个方面: 在内容层面, 本文总结当前主流的公开可用水声数据集, 在此基础上指出一个质量良好的水声数据集应该具备的特点; 在方法层面, 本文对已有方法进行更加细致和全面的总结. 例如本文加入了近期发展起来的两类重要方法, 即基于 Transformer 和基于自监督学习的水中目标识别方法, 这两类方法是当下广为关注并具有较大研究潜力的研究方向. 总之, 上述综述文章大多从特征或方法的某一维度出发对水中目标识别进行综述, 没有包含最新的研究进展. 此外, 当前的综述文章中缺少对不同方法的性能对比分析. 本文根据当前的研究进展, 系统阐述基于被动声呐音频信号进行水中目标识别的原理和方法, 对该领域的研究现状、存在的问题以及未来的发展趋势进行系统性的分析与讨论. 本文聚焦梳理基于被动声呐音频信号的水中目标识别的基本原理、方法以及最新成果, 突出将机器学习应用于水中目标自动识别的不同策略, 分析在此过程中存在的关键问题与挑战, 在此基础上对该领域未来的发展趋势进行总结与分析. 具体来说, 本文将基于被动声呐音频信号的 UATR 方法分为 7 大类: 基于传统机器学习的方法, 基于卷积神经网络的方法, 基于时延神经网络的方法, 基于循环神经网络的方法, 基于 Transformer 的方法, 基于迁移学习的方法, 基于无监督学习与自监督学习的方法, 如图 1 所示. 此外, 本文还对该领域中较为广泛使用的公开数据集进行总结与分析.

本文内容安排如下: 第 1 节从数据处理角度论述被动声呐信号处理的基本原理, 其中包括基于被动声呐信号进行水中目标识别的基本流程、被动声呐信号分析的数理基础以及特征提取方法; 第 2 节和第 3 节则从识别方法层面全面梳理基于被动声呐音频信号的水中目标识别方法的发展脉络和最新成果, 总结基于被动声呐音频信号的水中目标识别任务所面临的主要挑战, 指出探索“自学习–高效性–

跨模态融合”算法是解决技术瓶颈的有效手段; 第 4 节从现有公开可用的水声数据集角度论述, 指出一个质量良好的水声数据集应该具备的要素, 以便进一步促进该领域的发展; 第 5 节对全文内容进行总结, 从算法的精度、标签数据需求、可扩展性与实时性等多个维度, 论述水中目标自动识别需要重点研究的若干基础性问题和未来发展趋势.

1　被动声呐信号处理的基本原理概述

本节对基于被动声呐信号的水中目标识别基本流程、被动声呐信号分析的数理基础以及被动声呐信号的特征提取进行介绍, 这些是理解和分析当前基于被动声呐信号进行水中目标识别的背景知识.

1.1　基于被动声呐信号的水中目标识别基本流程

声呐 (sonar) 是利用声波在水中的传播和反射特性、通过电声转换和信号处理进行水中目标探测 (类型、位置、运动方向等) 和通讯的技术, 有主动式和被动式两种类型, 图 2 展示了它们的基本工作原理. 其中, 被动声呐是一种利用水听器 (hydrophone) 接收水下目标发出的声波信号, 从而实现水中目标探测与定位的技术. 其基本原理是: 当水中目标如潜艇、船舶、鱼类等运动时, 它们会在周围水域中产生声波信号, 这些信号会在水中传播并被水听器接收到. 主动声呐则是通过发射器主动发出声波脉冲并由接收器接收回波, 从而进行水中目标的探测与定位.

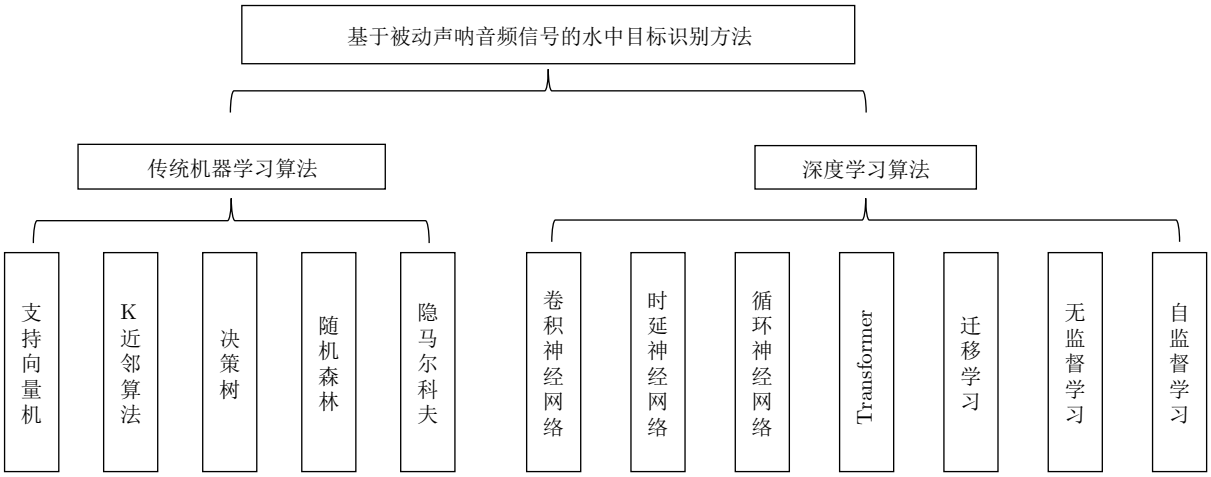
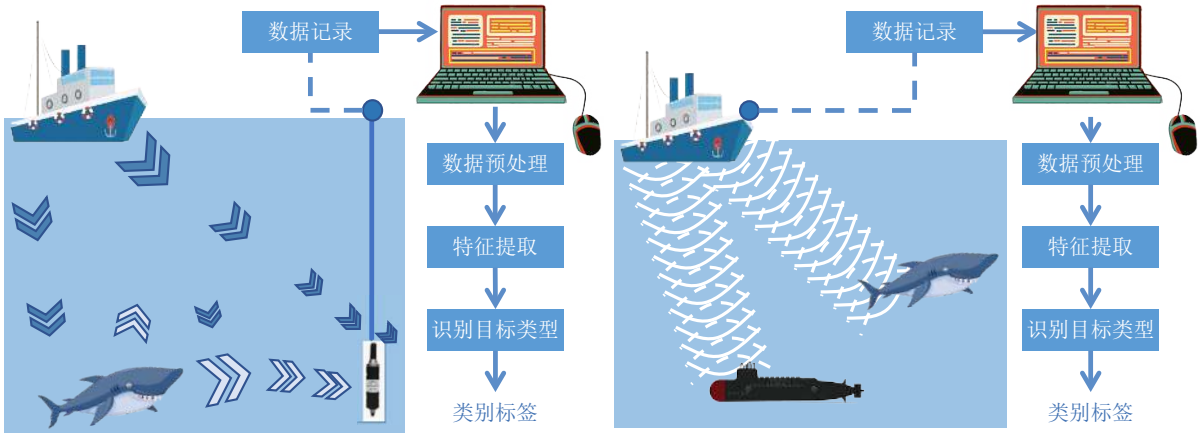


图 1　基于机器学习的水声目标识别方法

Fig. 1　Machine learning-based methods for UATR



(a) 基于被动声呐信号的水声目标识别基本原理
(a) Fundamental principles of UATR based on passive sonar signals

(b) 基于主动声呐信号的水声目标识别基本原理
(b) Fundamental principles of UATR based on active sonar signals

图 2　基于声呐信号的水声目标识别基本原理

Fig. 2　Fundamental principles of UATR based on sonar signals

一般而言, 对于接收到的声呐信号, 通过信号处理和相关算法可以识别水中目标的类型、位置和速度等信息, 从而实现基于声呐音频信号进行水中目标识别的任务. 相比于主动声呐, 被动声呐具有以下优势:

1) 隐蔽性高. 被动声呐只接收水下目标发出的声波信号, 不会主动发射任何声波, 因此不易暴露自己的位置. 而主动声呐需要发射声波信号, 可能被其他目标侦测到, 从而暴露自身位置.

2) 灵活性强. 被动声呐可以部署在船体上、水下电缆或浮标等位置, 安装和使用较为方便. 而主动声呐需要在水下目标附近进行发射, 需要有特定的发射设备和位置.

3) 可利用自然声源. 被动声呐可以利用自然声源 (如海豚、鲸鱼等) 或其他水下目标发出的声波信号进行侦测和定位, 由于其通常以静态的方式部署在不同的海洋环境中, 相对而言具有可探测范围广和受距离限制少的特点. 而主动声呐需要自身发射声波信号, 因此其侦测距离相对而言会受到更多限制.

因此, 当前主流水中目标识别研究所采用的数据集为被动声呐所采集的音频信号. 此类研究的基本原理是对被动声呐接收到的音频信号进行信号处理和特征提取, 得到与目标本质特性相关的可判别性特征, 并据此进行目标识别. 如图 3 所示, 该过程包括学习阶段和测试阶段. 其中学习阶段包括水中目标被动声呐信号的采集与预处理、特征选择与提取、样本选择以及分类器设计; 测试阶段包括信号采集与预处理、特征提取、分类决策以及输出识别结果. 信号采集是通过部署在船体上、水下电缆或浮标等位置的被动声呐来实现的. 信号预处理主要包括对信号进行放大、滤波等操作, 以去除背景噪声、提高数据的信噪比. 特征提取是从预处理的信号中提取出水中目标的特定识别特征, 常用的特征包括时域特征 (如振幅、相位、过零率等)、频域特征 (如频率谱、频率熵等) 和时频特征, 如梅尔倒谱系数 (Mel-scale frequency cepstral coefficients, MFCC)、伽马通滤波器倒谱系数 (Gammatone filter cepstral coefficient, GFCC)、LOFAR (Low frequency analysis recording) 谱和 DEMON (Detection of envelope modulation on noise) 谱等. 特征选择则是根据特征的判别性和相关性, 选取最优的特征进行目标识别, 常用方法有主成分分析 (Principal component analysis, PCA) 和线性判别分析 (Linear discriminant analysis, LDA) 等. 样本选择则是从原始被动声呐信号中选择最具代表性和差异性的样本, 同时保证不同类别的样本数量大致相当. 最后分类器设计与训练是指选用合适的分类算法对目标进行识别, 常用的分类算法包括支持向量机、决策树和神经网络等.

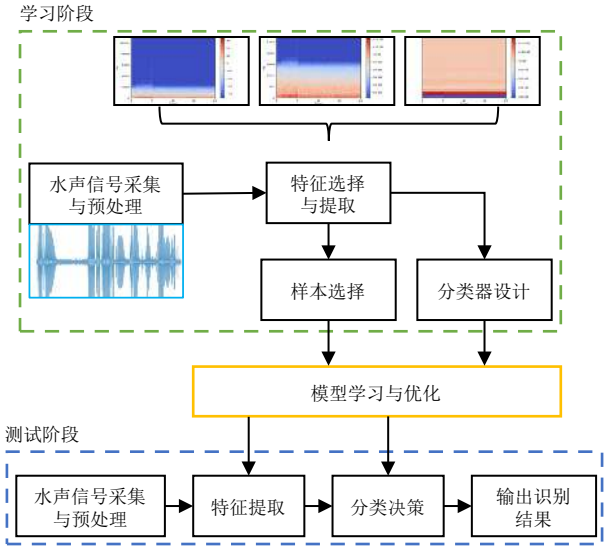


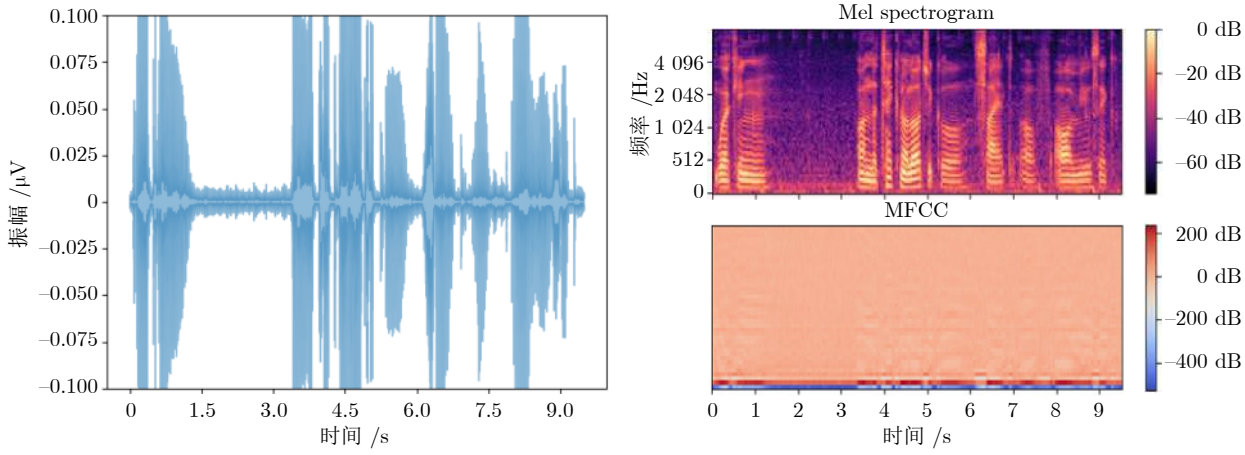
图 3 水声目标识别的基本流程

Fig. 3 Basic procedure of UATR

1.2 被动声呐信号分析的数理基础

水中声波信号由物体的运动产生, 如潜艇、船舶、鱼类等运动时会在周围水域中产生声波信号, 这些声波信号是基于被动声呐信号进行水中目标识别的基础. 声波信号可以理解为时域上的一维信号, 其在形式上可以被刻画为时域上的波形图, 如图 4(a) 所示. 波形图直观地表达了声音信号的基本时域特征, 如声波信号辐射强度 (振幅)、过零率[11] 等. 然而时域特征存在蕴含的声波信息有限、难以描述信号的周期性和谐波成分等问题. 因此, 在时域分析的基础上, 研究人员开始尝试分析信号的频率特性及其变化情况, 即对信号进行频域分析和时频分析. 特别地, 这些研究的数学基础为傅立叶变换 (Fourier transform, FT). 傅立叶变换假设任何连续信号都可以由不同频率的正弦函数和余弦函数叠加得到, 这些正弦函数和余弦函数统称为信号的分量. 通过对信号进行傅立叶变换, 可以将其从时域表示转换为频域表示, 以便更好地理解和处理信号的频率特性[12]. 例如对信号进行傅立叶变换和自相关函数运算, 可以得到信号的功率谱, 用以反映信号在某一特定频率值上的强度.

在实际应用中, 为满足傅立叶变换对信号平稳性的要求, 被动声呐信号通常会进行分帧与加窗的预处理操作, 以保证窗口内的信号具有短时平稳性.



(a) 被动声呐信号时域波形图
(a) Time-domain waveform of passive sonar signal

(b) 被动声呐信号的 Mel 频谱和 MFCC
(b) Mel spectrogram and MFCC of passive sonar signal

图 4　被动声呐信号的特征图示例

Fig. 4　The illustrative feature examples of passive sonar signals

对原始信号进行分帧、加窗, 然后逐个窗口进行傅立叶变换的过程称为短时傅立叶变换 (Short-time Fourier transform, STFT). 短时傅立叶变换是频域特征和时频特征提取的基础, 例如对信号做短时傅立叶变换, 并做进一步的取模等运算, 可以得到被动声呐信号的梅尔 (Mel) 频谱和 MFCC, 图 4(b) 展示了其 Mel 频谱和 MFCC. 谱图通常以时间为横坐标、频率为纵坐标、振幅值为 Z 坐标绘制, 同时反映信号的多维信息. 此外, 谐波分析也是被动声呐信号的一个重要分析方法[13−15], 其中谐波信息可以通过提取信号的倒谱表示来获得, 倒谱则是通过对信号的 STFT 谱取对数、做逆傅立叶变换得到的, 用于获得频谱中的周期结构[16].

由于人耳听觉系统具有出色的信号辨识能力, 因此人们设计一组模拟人耳听觉系统的被动声呐信号分析方法, 即梅尔尺度与梅尔频谱. 梅尔尺度是一种对数尺度[17], 用于表示声音频率的非线性度量, 它是基于人耳听觉系统对声音频率的感知方式提出的. 事实上, 人耳听觉系统对频率的感知大致遵循对数分布, 表现为对低频的变化敏感, 对高频的变化迟钝[18], 因此基于对数运算的梅尔尺度能够模拟人耳的听觉特性. 梅尔频率与原始频率之间的转换关系如式 (1) 所示, 其中 m 表示梅尔频率, f 表示原始频率. 梅尔频谱是基于梅尔尺度对声音信号进行频谱分析的结果, 具体而言, 它可以通过计算快速傅立叶变换并将其结果与一个三角滤波器组卷积得到. 然而, 梅尔滤波器具有固定的带宽, 无法精细地模拟人耳听觉特性. 相比之下, Gammatone 滤波器基于人耳耳蜗对不同频率音频信号的敏感度作出响应, 具有高通性和带宽变化, 能更好地模拟人耳基底膜的滤波特性.

$$m = 2\,595\lg\left(1 + \frac{f}{700}\right) \tag{1}$$

小波变换 (Wavelet transform, WT) 是一种在时域和频域上都有良好性能的信号分析方法, 也被广泛应用于被动声呐信号的处理. 小波变换通过对信号进行一系列小波基函数的线性组合来表示信号, 小波基函数是一组具有一定局部性质和频率性质的基函数, 可以将信号分解为具有不同频率和时间分辨率的小波子带, 从而实现时频分析. 由于小波基函数是以有限长度的信号为基础的[19], 因此小波变换不受傅立叶变换的局限, 可以很好地处理非平稳信号. 此外, 小波变换还可以提供更好的时频分析精度, 从而实现对信号的多分辨率表示. 但需要注意的是, 小波变换也存在如计算量较大以及容易产生边缘效应等问题.

另一种常用的被动声呐信号分析与处理工具为 Gabor 滤波器, 它的基本原理是将一个带有高斯包络的正弦波作为滤波器的模板. 该模板可以在时域和频域上进行调整, 以适应不同的信号分析需求. 具体而言, Gabor 滤波器可以通过调整其中心频率和带宽来选择不同的频率范围、调整其中心位置和时间分辨率来选择不同的时间范围. 这种灵活性使得 Gabor 滤波器非常适合用于分析具有时变特性的水下声音信号. 近年来基于 Gabor 滤波器进行声学分析的研究日益丰富, 例如环境音的识别[20]、音乐流派识别[21] 和语音分析[22−23] 等.

上述方法构成了被动声呐信号分析与处理的数理基础, 根据被动声呐信号的特点和具体应用场景, 合理选择不同的处理方法对于被动声呐信号识

别的性能而言至关重要. 下面进一步讨论基于这些数理基础所发展出来的一系列被动声呐信号特征提取方法.

1.3 被动声呐信号的特征提取

被动声呐信号的特征提取是指借助相关的数理分析方法, 从水中目标发出的声波信号中提取出有用的特征信息, 以实现目标类型识别任务. 特征提取旨在通过提取出有效的特征信息帮助我们了解目标的物理属性和运动状态, 以便更好地进行目标的分析和判别. 本文依据被动声呐信号特征提取的主要发展脉络对其进行分类, 并依次展开介绍每一种方法.

1.3.1 基于水中目标固有物理机理的被动声呐信号特征提取

基于水中目标固有物理机理的音频特征提取, 是一种根据水中目标所发出的声波和产生的水动力学效应的物理特性来提取目标特征的方法. 具体来说, 水中目标在运动时会产生一些特有的水动力学效应, 如水流的涡旋结构、气泡的形成、漩涡等. 这些效应会改变水中声波的传播特性, 进而影响声波信号在水中的传播和接收. 因此, 通过对这些声波信号和水动力学效应的分析, 可以提取出与目标本质特性相关的特征, 如振幅值、目标声呐参数、接收信号的线谱结构以及各类目标的机动特点等, 从而实现对水中目标的高精度识别.

文献 [24–26] 直接从水中目标辐射噪声的波形图中提取过零率、振幅包络线等特征用以进行目标的识别. Rajagopal 等[27] 根据对船舶噪声的充分了解, 提出检测线谱的方法, 首先选取极具物理意义和现实意义的特征量进行目标识别, 包括螺旋桨叶片数、螺旋桨转速、推进器类型、目标壳体辐射低频噪声、活塞松动产生的谐音基频、注水器噪声、最大速度、槽极噪声和传动装置类型等 9 个特征. 这些物理特征可以清楚地表现出船舶框架结构, 从而实现 4 类目标的识别. Lourens[28] 则是将研究重点放在螺旋桨转速上, 同时描述齿轮机箱的噪声性质, 在此基础上提出倒频谱特征作为检测噪音的手段. Liu 等[29] 则提出一种基于薄壳振动及模态分解理论的壳体振动模型进行水下目标辐射噪声的本征模态特征提取方法.

1.3.2 基于时域、频域和时频分析的被动声呐信号特征提取

由于海洋环境的日益复杂和各种声隐技术的使用, 极大影响了水中目标所产生的辐射噪声的物理机理, 因此依靠基于目标固有物理机理所提取的特征来识别水中目标的正确率已经不能满足现实需要. 随着信号处理技术的发展, 研究者们逐渐将目光投向能同时表达更多信息的时域特征提取、频域特征提取和时频特征提取.

时域特征提取是从时域声呐音频信号中提取特征的基础步骤, 它包括对原始声呐音频信号进行预处理和特征提取两个阶段. 常用的时域信号预处理技术包括滑动窗口、加窗和滤波等方法, 旨在提高信号质量和增强目标信息. 常用的时域特征主要有振幅、能量、时长、过零率等. 这些特征从不同角度描述声音的时长、强度、节奏等方面的信息, 从而反映声呐音频信号的时域波形和时序特征. 通过提取这些特征, 可以获得关键的时域信息, 为后续的目标识别提供有力支持. 时域特征提取以其简单直观、计算效率高和对目标的时序特征敏感等优势而被广泛使用. 然而, 它也存在一些限制, 包括对噪声和干扰的敏感以及难以提取复杂目标的细节特征等.

频域特征提取是将时域声呐音频信号转换为频域表示的过程, 通常通过 FT 或滤波器设计来实现. 通过 FT, 可以获得声呐信号在频域上的能量分布和频谱特性, 从而更好地描述声呐信号的频率成分和频率响应. 常用的频域特征包括频谱形状、频带能量分布、频率峰值等, 这些特征可以用于描述声音的频率成分和谱线密度等信息. 另外, 滤波器设计也是频域声呐音频信号特征提取中的重要内容. 通过设计不同类型的滤波器, 可以在频域上选择感兴趣的频带, 并去除干扰信号. 常见的滤波器设计方法包括低通滤波器、高通滤波器 (如 Gammatone 滤波器) 和带通滤波器 (如梅尔滤波器) 等, 它们能够帮助提取感兴趣频率范围内的目标信号, 并减弱或排除其他频率的干扰. 频域特征提取从频域的角度对被动声呐音频信号进行特征提取, 相较于时域特征提取, 具有以下优势: 首先, 频域特征提取能够提供声呐信号在不同频率下的能量分布和频谱特性, 从而更全面地描述声呐信号的频率信息; 其次, 频域特征具有较好的抗噪声能力, 能够减少噪声对目标识别的影响; 再次, 频域特征提取还能帮助识别目标的频率特征, 在不同类型的水中目标识别中具有重要意义.

时频声呐音频信号特征提取是一种综合利用时域和频域信息的方法, 能够提供更全面、准确的声呐信号描述. 时频特征提取的原理是基于声呐信号在时域和频域上的特征进行联合分析, 通过将这两个领域的特征进行组合, 能够更全面地描述声呐信号的时序和频率特性. 常用的时频分析方法有 STFT、LOFAR 谱分析、DEMON 谱分析、高阶谱分析、小波变换、Hilbert-Huang 变换 (Hilbert-

Huang transform, HHT)、倒谱分析以及 Gabor 滤波等. Das 等[30] 采用 STFT 对被动声呐信号进行处理, 提取其光谱特征和倒谱系数. 文献 [31–33] 则基于小波变换进行舰船辐射噪声的时频特征提取, 实验表明小波变换使信号的谱类别特征和波形结构特征有了明显的增强, 更具判别性. Wei 等[34] 结合小波特征和 PCA 以实现对特征降维的目的. Xu 等[35] 设计一种基于不确定性估计的可信多表征学习方法, 用以提升时频特征的判别性. 相比于时域和频域特征提取, 时频特征提取具有以下优势: 首先, 时频特征提取能够捕捉到声呐信号在不同时间和频率上的变化情况, 提供了更加丰富的信息; 其次, 时频特征具有较好的抗噪声能力, 能够减少噪声对目标识别的影响; 再次, 时频特征还能够提取目标的时序和频率特征, 对于不同类型的水中目标识别具有重要意义.

1.3.3 基于声音生成感知模型的被动声呐信号特征提取

基于声音生成感知模型的水中音频特征提取是一种利用人耳听觉感知机制的特征提取方法[36], 它的基本原理是人类听觉系统能够感知不同频率范围内的声音, 并对其进行不同的处理. 具体来说, 人耳会将声音分解成多个频带, 每个频带内的声音信号会被独立地处理. 因此, 基于声音生成感知模型的特征提取方法也采用了这种分频带的思想. 在具体实现上该方法使用一组带通滤波器将声音信号分解成多个频带, 然后对每个频带内的信号进行能量特征的提取, 以捕捉声音的关键信息. 该方法从听觉的生理机制、耳蜗的频率分解特性、掩蔽效应、临界带宽等听觉特性出发, 构建基于响度、音调和音色的相应特征, 以期获得接近人耳听觉系统对声音的良好辨识能力. 梅尔尺度和梅尔频谱正是基于这一思路进行设计的. 此外, 由于能量特征计算速度很快, 因此该方式适用于实时处理. 基于上述特性, 该方法一直是音频表征提取的研究热点.

早期, Békésy[37] 通过频闪观测仪发现了耳蜗基底膜上的行波及基底膜的频率分解作用, 据此建立了最早的耳蜗一维传输模型. Johnstone 等[38] 采用 Mossbauer 技术对耳蜗中的基底膜振动进行测量研究, 得到比文献 [37] 相对更为精确的实验结果. Zwislocki[39] 则建立一维传输线模型来解释文献 [37] 的实验结果. 费鸿博等[40] 则基于梅尔频谱提出一种可分离方法, 进行更精细的声音特征提取. 随着神经学对人耳听觉系统认识的不断深入, 文献 [41–42] 进一步设计更加精细的方法用于模拟人耳听觉系统的功能, 具体来说, 借助一组基于卷积的滤波器模拟人的听觉皮层、听觉中枢等区域的功能, 将原始时域音频信号分解为一系列不同频率的音频分量信号, 同时卷积核的大小可变, 用以模拟听觉系统受到声音刺激后对不同波长分量的感兴趣程度.

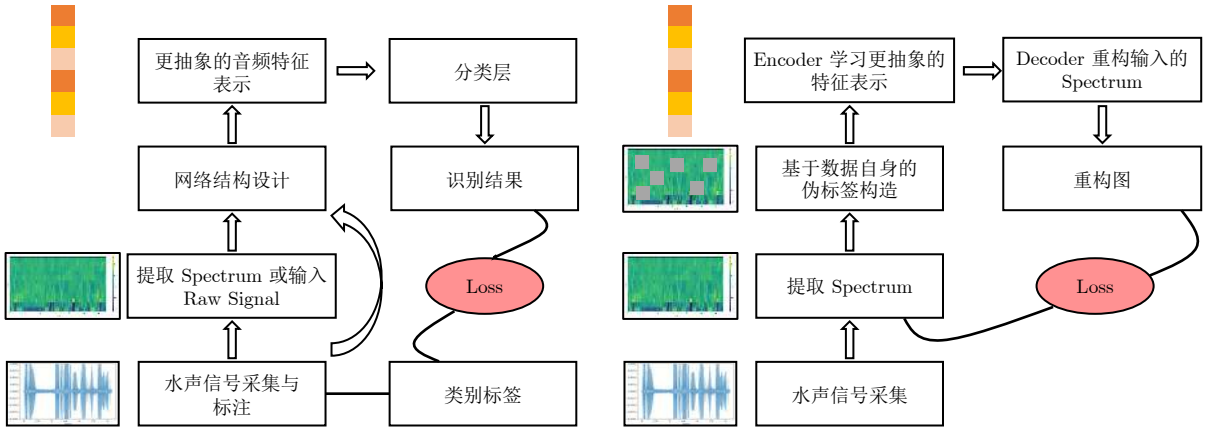
1.3.4 基于有监督深度学习的被动声呐信号特征提取

传统音频特征提取方法往往需要专业的领域知识和专家经验来设计合适的手工特征提取器, 然而由于海洋环境的复杂多变, 从中获取足够的水中目标先验知识是非常困难的. 近年来, 在水中目标被动声呐信号特征提取领域, 深度学习方法作为一种直接从原始数据构建分层表征的方法得到了广泛的研究, 主要包括基于有监督学习的水中音频特征提取和基于自监督学习的水中音频特征提取.

基于有监督深度学习的被动声呐信号特征提取是指通过使用带标签的音频数据集来训练深度神经网络, 利用音频的标签信息驱动模型学习最优的音频特征, 图 5(a) 展示了该方法的一般范式. 由于卷积神经网络 (Convolution neural network, CNN) 具有空间局部性和平移不变性等优点, 文献 [43–45] 利用 CNN 从声音的原始音频或其频谱中进一步学习更高层次的特征以增强特征的判别性. 文献 [46] 在 CNN 中引入注意力机制, 以更好地捕获频谱中更大范围的上下文信息. 然而, 并没有直接证据表明注意力机制对声音特征的判别性增强必须依赖于 CNN. 基于这一认知, Gong 等[47] 在计算机视觉领域的视觉转换器 (Vision transformer, ViT) 的启发下, 首次提出完全基于注意力机制的声音频谱转换器 (Audio spectrogram transformer, AST). Yang 等[48] 则从水中目标的多维属性角度进行考虑, 设计一种基于多属性相关度感知的深度学习方法用以捕捉水中音频信号的特征. 相比于传统的特征提取方法, 基于有监督的深度学习方法可以更好地捕捉水中音频信号的复杂特征, 从而提高识别性能. 然而该方法需要大量的标注数据集来进行训练, 并且在实际应用中需要考虑方法的实时性和鲁棒性等问题.

1.3.5 基于自监督学习的被动声呐信号特征提取

基于自监督学习的被动声呐信号特征提取是一种使用无标签数据进行训练的深度学习方法, 旨在从音频信号中学习判别性特征. 相比基于有监督深度学习的音频特征提取方法, 该方法不需要标注数据, 从而数据的获取和准备更加便捷. 此外, 由于该方法从数据自身出发构建监督信号用以指导网络的学习, 能够更好地利用数据的上下文信息实现更具判别性的音频特征提取. 同时, 基于自监督学习的



(a) 基于有监督学习的水中音频信号特征提取范式
(a) Supervised learning-based paradigm for underwater acoustic signals feature extraction

(b) 基于自监督学习的水中音频信号特征提取范式
(b) Self-supervised learning-based paradigm for underwater acoustic signals feature extraction

图 5 基于深度学习的水中音频信号特征提取范式

Fig. 5 Deep learning-based paradigm for underwater acoustic signals feature extraction

方法还具有更好的泛化能力, 可以应用于不同水下环境的目标识别任务.

该方法的基本思想是利用自编码器 (autoencoder) 的网络结构从数据自身的内在联系出发以自适应提取最优的音频特征, 训练过程中模型的输入和输出都是相同的音频信号, 但在网络的中间层提取出的特征可以用于后续的目标识别等任务, 图 5(b) 展示了该方法的基本流程. 如 Gong 等[49] 设计一种联合判别与掩码重构的策略, 从无标签数据中学习音频的抽象特征. 具体来说, 该方法首先将声音的 Log-Mel 谱图切分成相同大小的图像块 (patch), 然后随机掩码部分 patches, 并将未被掩码的 patches 输入基于自编码器架构的 ViT 模型中学习、重构被掩码的 patches, 最后利用重构的 patches 与原始 patches 进行损失函数计算, 以指导模型学习更优异的特征表示. 受计算机视觉领域 MAE[50] 的启发, Baade 等[51] 在 Gong 等[49] 工作的基础上设计一种高比率掩码的策略, 在实现自监督音频特征提取的同时, 大大加速了模型的训练. Ghosh 等[52] 则提出一种基于对比学习的多尺度声音频谱转换器, 该方法设计了教师网络和学生网络两个子网络, 通过计算教师网络与学生网络输出之间的对比损失来指导模型学习更好的音频表征. 每个子网络在实现上逐层以 2 倍的比例扩大 patches 的大小, 从而更好地学习信号的全局结构与局部特性.

2 基于传统机器学习算法的水中目标识别

基于传统机器学习的水中目标识别主要分为 3 个模块: 音频信号预处理、特征提取、目标识别. 其中, 信号预处理是基础, 特征提取是关键, 目标识别是最终目的. 音频信号预处理目的是为了消除噪声、提高音频信号的质量和可分析性, 主要方法有降噪、滤波、压缩、分解等. 特征提取是整个水中目标自动识别系统中最为核心的部分, 如何提取出具有足够判别性的音频特征对提高系统识别性能具有关键作用, 当前主流的特征提取方法见第 1.3 节. 影响水中目标识别性能的另一个关键问题在于如何选择合适的分类器. 目前, 基于传统机器学习的水中目标识别研究中, 主要的特征识别算法有 K 近邻算法 (K-nearest neighbor, KNN)、支持向量机 (Support vector machine, SVM)、决策树 (Decision tree, DT)、随机森林 (Random forest, RF) 和基于隐马尔科夫模型 (Hidden Markov model, HMM) 的方法等. 在本节中, 将重点讨论用于特征识别的机器学习算法及其内在联系与发展趋势.

2.1 KNN 算法

KNN 是一种基于实例的学习算法, 在水中目标识别中常用于对被动声呐接收器接收到的声波信号进行分类, 以实现水中目标的识别. 该算法的工作原理是根据给定的训练集, 在训练集中寻找与新输入实例最邻近的 K 个实例, 然后将新实例划分给这 K 个实例中最具有代表性的类别. 在 KNN 算法中, K 是一个可学习的参数, 选择合适的 K 值对算法性能至关重要. KNN 算法的思路简单且易于实现, 但其收敛速度相对较慢.

2.2 SVM 算法

SVM 是一种基于核函数的有监督学习模型, 在

水中目标识别中, SVM 可以用于将水中目标的声学特征与预定义的类别进行识别. 该算法依据 Vapnik-Chervonenkis (VC) 理论和结构风险最小化原理, 旨在构建一个最优超平面以实现将数据集分割成两个部分, 使得分割超平面两侧的样本尽可能的远. 其本质上是从有限的样本数据中搜索一种最优的折中方案, 以实现获取最佳泛化性的目的. 此外, 通过使用核函数, SVM 不仅可以将在低维空间线性不可分的数据映射到更高维的空间, 转化为线性可分的, 还能在一定程度上缓解高维数据带来的维数灾难问题.

2.3　决策树与随机森林算法

DT 是一种基于树结构的算法, 采用非常直观的方式对事物进行分类或标注. 在水中目标识别中, 决策树可以用于根据提取的特征来对目标类型进行识别. 该算法基于训练数据的特征进行树结构的构建, 其中每个节点表示一个特征, 每个分支表示一个可能值, 最终的叶节点表示一个类别. 该算法直观清晰, 但随着深度的增加, 容易陷入过拟合. 随机森林是一种基于决策树的集成学习算法, 它从训练数据集中随机抽取一部分数据进行决策树的构建, 然后重复这个过程, 构建多棵决策树, 最后根据所有决策树的结果进行投票以决定最终的识别结果, 这在一定程度上缓解了过拟合的风险.

2.4　隐马尔科夫模型

HMM 是一种基于概率的时序统计模型, 它用来描述一个含有隐含参数的马尔科夫过程, 从可观察的参数中确定该过程的隐含参数, 从而预测一个序列的概率. 在水中目标识别中, 可以将不同目标的声学特征作为观察数据序列输入到 HMM 中进行建模. 具体来说, 可以将不同水中目标的声学特征提取为一个向量序列, 然后将这个向量序列作为观察数据序列输入到模型中, 模型输出一个对应于每个目标的概率分布, 表示该目标所生成的声学特征序列的概率. 在实际应用中, 可以使用基于贝叶斯准则的后验概率最大化来进行目标识别.

2.5　算法的内在联系与发展趋势

KNN、SVM、决策树和随机森林都是有监督学习算法, 它们都是基于训练数据集进行模型训练, 然后用于预测新的输入实例的类别或值. 而隐马尔科夫模型是一种时序模型, 它用于模拟一个隐藏的马尔科夫链, 从而预测一个序列的概率. 总体来说, 这些算法具有基本一致的工作机制[53], 都是基于训练数据学习出一种分类边界. 表 1 列出了部分基于传统机器学习算法进行水声目标识别的研究.

表 1　典型传统机器学习的水声目标识别算法
Table 1　Typical traditional machine learning algorithms for UATR

年份	机器学习算法	音频特征	数据集
1992	Naive Bayes[54]	目标固有物理机理特征	自建数据集
2016	DT[55]	时域、频域特征	仿真数据集
2014	SVM[56–62]	基于小波变换的时频特征	自建数据集
2016		MFCC	真实数据集
2017		GFCC	历史数据集
2017		改进的 GFCC	舰船数据集
2018		过零率	鱼类数据集
2019		融合表征	自建数据集
2022		LOFAR 谱	ShipsEar
2018	KNN[60, 62]	MFCC	鱼类数据集
2022		LOFAR 谱	ShipsEar
2011	SVDD[63]	—	舰船数据集
2014	HMM[56, 64]	MFCC	—
2018			机器音频数据集

然而, 需要注意的是基于传统机器学习的水中目标识别模型本质上是一种浅层结构, 模型的信息容量和学习能力有限. 随着海洋环境的日益复杂和各种技术的干扰, 基于此类方法的识别分类精度难以满足使用需求, 因此目前主流的研究方向为基于深度学习算法的水中目标识别.

3　基于深度学习算法的水中目标识别

近年来深度学习算法在许多领域表现出强大的自动特征提取和优化能力, 为基于被动声呐信号的水中目标识别研究开辟了一个新的发展方向, 并逐渐成为该领域的研究热点. 相比于传统机器学习算法, 基于深度学习算法的水中目标识别具有以下优势:

1) 深度学习算法可以从原始数据中自动学习音频特征, 避免人工选择特征的主观性;

2) 深度学习算法具有强大的信息表达能力, 可以处理高维数据和非线性关系, 对于复杂的海洋环境和水中目标被动声呐信号具有更好的适应性;

3) 深度学习算法可以处理大规模数据, 能够有效利用数据资源, 从而提高识别的准确率和效率.

因此, 目前利用被动声呐信号的水中目标识别研究主流方案大多都是基于深度学习算法展开的, 并取得了良好的研究成果. 从最初基于卷积神经网络的水中目标识别方法, 到后来更有利于捕获全局依赖的循环神经网络 (Recurrent neural network, RNN)、时延神经网络 (Time delay neural net-

works, TDNN)、基于预训练模型的迁移学习 (Transfer learning, TL) 等方法, 再到近年来很有前景的基于 Transformer 的方法以及自监督学习 (Self-supervised learning, SSL) 方法, 图 6 列出了部分有代表性的基于深度学习的水声目标识别算法的发展历程. 按照学习范式的不同, 可以将基于深度学习的水声目标识别方法分为有监督范式 (图中轴线上方) 和无监督范式 (图中轴线下方); 根据网络结构的不同, 将深度学习方法划分为基于卷积神经网络的方法 (图中轴线上方黑色不加粗)、基于时延神经网络的方法 (图中轴线上方蓝色不加粗)、基于循环神经网络的方法 (图中轴线上方橙色不加粗)、基于 Transformer 的方法 (图中轴线上方橙色加粗)、基于迁移学习的方法 (图中轴线上方黑色加粗)、基于无监督学习的方法 (图中轴线下方黑色不加粗) 和基于自监督学习的方法 (图中轴线下方蓝色加粗).

3.1 基于卷积神经网络的水中目标识别

CNN 是一类高度非线性的深度学习模型, 逐层扩大感受野、权值共享等特性使其能多尺度、细粒度地提取数据的特征, 在图像识别[65] 等领域取得了非常先进的成果. 它通过将层次化的特征提取和目标识别结合在一起, 从而同时具备自动特征提取器和分类器的功能. 由于 CNN 在计算机视觉与自然语言处理应用中表现出巨大的性能, 激发了该方法在水中目标识别领域的应用. 基于卷积神经网络的水中目标识别方法主要采用卷积神经网络的架构, 以最大化识别准确率为目标, 旨在通过网络结构的设计与训练从水中目标的被动声呐信号中提取更优异的音频特征, 从而实现对目标类型的识别. 图 7 展示了该方法的基本架构, 其中网络的核心由一系列卷积层和池化层构成, 分类器由多个全连接层构成. 基于 CNN 的水中目标识别研究主要可以划分为优化网络的输入[41−44, 66−69]、设计不同的网络结构[70−73] 以及在网络中加入新的机制以学习更优异的高层次目标特征[46] 这三种. 但需要注意的是该方法存在对标签数据的需求较大、提取的特征相对抽象、结果的可解释性较差的局限.

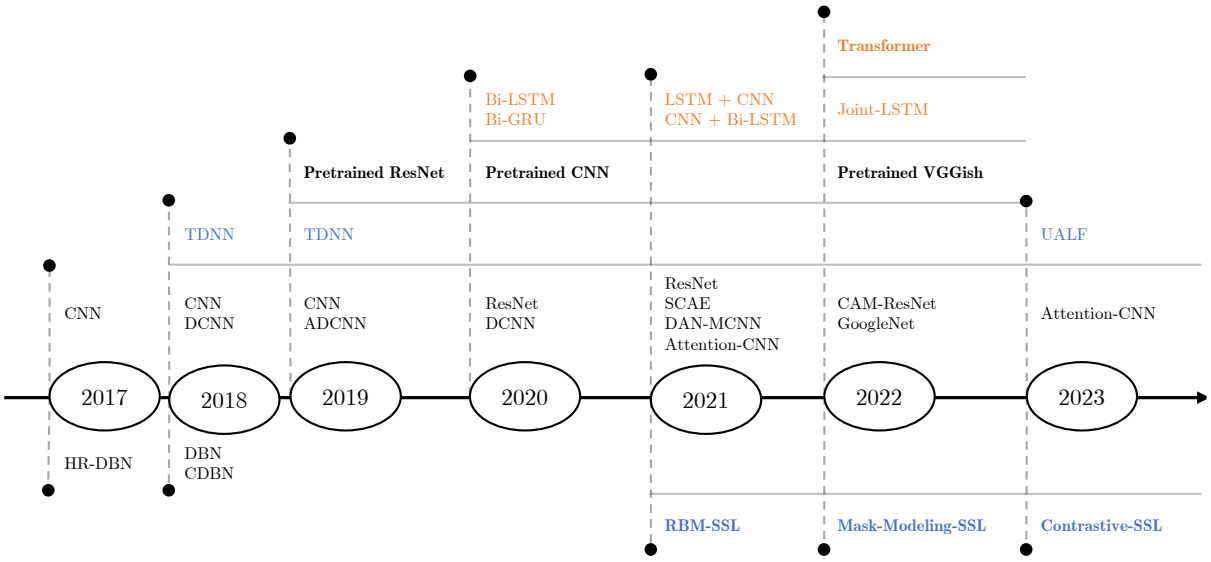


图 6 基于深度学习的水声目标识别主流算法模型发展时间轴线

Fig. 6 Timeline: Evolution of mainstream deep learning algorithms for UATR

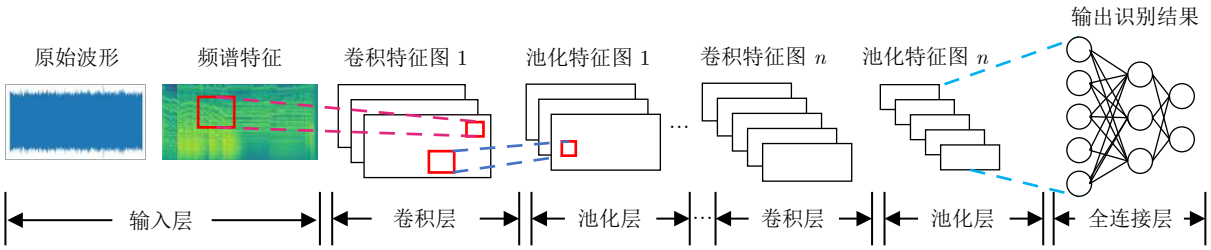


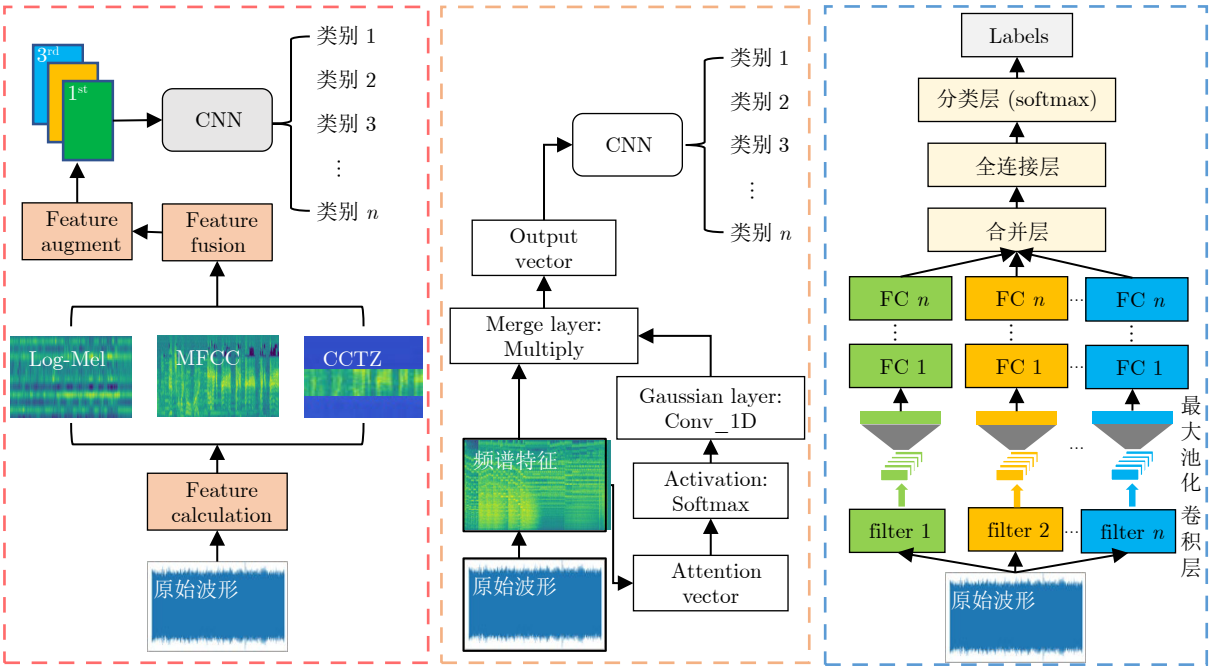
图 7 基于 CNN 的水声目标识别方法基本架构

Fig. 7 Basic framework of CNN-based methods for UATR

文献 [66−69] 将水中目标的被动声呐信号转换为频谱特征, 然后输入到 CNN 中学习更抽象的音频特征并据此识别不同目标, 其中用到的频谱特征有幅度谱、MFCC 谱、LOFAR 谱等, 用到的 CNN 网络结构有自建网络、ResNet、VGG 等. 注意到上述方法使用的池化策略均为平均池化, 这在一定程度上减弱了不同信号分量的差异. Cao 等[70] 基于二阶池化策略设计一种端到端的 CNN 网络, 利用常数 Q 变换 (Constant-Q transform, CQT) 从水中目标辐射噪声中提取时间相关性, 并据此进行目标识别. 由于二阶池化策略可以捕捉一个频点上所有 CNN 滤波器的时间相关性并保留它们的差异性, 从而实现模型性能的提升. Hu 等[71] 使用极限学习机 (Extreme learning machines, ELM) 替换 CNN 的全连接层, 在民用客船数据集上识别精度可达 93.04%. Wang 等[74] 提出一种基于注意力机制的多分支 CNN, 其中注意力机制用以捕捉音频特征图中重要的信息, 多路分支用以加速网络的训练过程, 该方法在 ShipsEar 数据集上实现了 2.4% 的性能提升.

理论上, 可以通过增加网络层数或神经元的个数来提升深度学习算法的性能, 但实际应用中会出现两个问题: 一是参数过多, 会导致计算复杂度增加, 并且当数据集有限时, 容易出现过拟合问题; 二是随着网络深度的增加, 反向传播算法在更新参数时, 可能出现“梯度消失”问题. 针对上述问题, Zheng 等[72] 使用 GoogleNet (一种基于稀疏结构设计的网络) 作为主网络从水中目标声波的时频谱中提取更抽象的特征, 旨在增大背景噪声与目标信号的辨识度. 结果表明, 在信噪比为 −10 dB 时, 所提方法的识别能力较高. 然而该方法只使用了一种音频特征作为输入, 而未对比在其他音频特征上的性能. Irfan 等[73] 提出一种基于可分离卷积自编码器 (Separable convolution autoencoder, SCAE) 的网络, 并使用 6 种不同的音频特征 (包括 Cepstrum 谱、Mel 谱、MFCC 谱、CQT 谱、GFCC 谱和 Wavelet packets) 对所提方法进行性能分析, 证明了该方法的性能优于大部分对比实验的方法. 此外, 可分离卷积降低了模型的参数量和计算复杂度, 提升了模型的训练效率.

通常情况下, 基于不同方法所提取的目标音频特征在信息表达上具有不同的侧重点, 将多种音频特征进行融合可以综合它们的优点, 从而获得更好的识别效果[75]. Hong 等[76] 基于 ResNet18 设计一种具有三通道输入的残差网络用于水中目标识别. 如图 8(a) 所示, 其中 Log-Mel 谱作为第一通道,



(a) 基于融合特征的优化方法
(a) Optimization method based on fusion feature

(b) 基于注意力机制的优化方法
(b) Optimization method based on attention mechanism

(c) 模拟人耳听觉感知的优化方法
(c) Optimization method based on simulating human auditory perception

图 8　基于 CNN 的水声目标识别主流优化方法

Fig. 8　Mainstream optimization methods for CNN-based UATR

MFCC 谱作为第二通道以及由色度 (chroma)、对比度 (contrast)、音网图 (tonnetz) 和过零率 (zero-cross ratio) 等组成 CCTZ 特征作为第三通道, 在网络中进一步加入特征融合层和频谱增强层, 得到三通道的声学特征, 然后输入到 CNN 中进一步学习更优异的音频特征, 该方法在 ShipsEar 数据集上准确率提升至 94.3%. 但是残差网络对数据较为敏感, 容易受干扰信息的影响, 并且网络结构需要随着任务场景的改变进行重新调整.

近年来, 在基于残差网络的图像分类和目标检测等任务中, 通过引入注意力机制极大缓解了残差网络受干扰信息影响的问题, 并取得了可喜的进展[77−80], 这促进了其在水中目标识别领域的应用. Xiao 等[46] 提出在输入层后面嵌入注意力层的水中目标识别方法. 如图 8(b) 所示, 通过引入注意力机制以抑制环境噪声和海上舰船干扰, 更好地保留与目标特性相关的特征, 从而实现了较高的目标识别精度. Liu 等[81] 设计一种基于双注意力网络 (Dual attention networks, DAN) 和多分辨率卷积神经网络 (Multiresolution convolutional neural network, MCNN) 的架构, 其中 DAN 用以更好地捕捉音频的局部特性和全局依赖性, 并采用动态加权的策略以强调感兴趣区域; MCNN 用以模拟人耳的听觉感知机制. 该方法在实现上采用 Inception 模型的多分辨率池化卷积方案, 构建 MCNN 架构以实现更好地适应三维聚合特征的时频结构, 同时采用位置注意力模块和空间注意力模块并行学习, 使网络兼顾音频的局部特性和全局依赖. 实验表明, 该方法在 ShipsEar 数据集上的平均识别准确率可达 95.6%. Xue 等[82] 设计一种基于通道注意力机制 (Channel attention mechanism, CAM) 的残差神经网络 (Residual network, ResNet), 具体来说使用一组一维卷积滤波器将水中目标的声波信号分解为不同频率的分量, 然后使用两层残差块堆叠的结构来提取更抽象的音频特征, 最后通过在残差块后面加入通道注意力机制, 大大减少了海洋背景噪声和多目标噪声的干扰, 从而获得更好的识别效果. Li 等[83] 基于类似的网络结构, 将多种音频表征融合并使用频谱增强 (SpecAugment) 技术[84] 对融合后的音频特征进行增强, 然后堆叠 3 层基于通道注意力的残差块以优化特征. 此外, 该方法采用交叉熵损失函数和中心损失函数作为联合损失函数指导网络的学习. 其中, 中心损失函数通过为每个类提供一个中心, 使得同一类的样本特征分布在类中心附近, 同时抑制了交叉熵损失函数类内变化明显的问题.

上述方法的模型输入均为基于手工设计的特征提取器所提取的音频特征, 对原始音频信号具有一定的压缩与损失. Doan 等[43] 提出基于稠密卷积神经网络 (Dense convolutional neural network, DCNN) 的水中目标识别方法, 其中 DCNN 被用来自动提取音频的特征, 无需专业的领域知识和专家经验的干预. 同时, 使用跳跃连接技术的架构允许不同网络层之间复用在不同尺度下提取的特征图, 从而避免了在一个卷积神经网络中顺序堆叠多个卷积层和激活层所导致的梯度消失问题. 由于水中目标本身的物理机理不同以及所处的水下环境复杂多变, 不同目标的声波具有不同的频率和波段, 现有的特征提取方式往往面临分辨率固定而无法很好地将目标的声波信号正确区分开的问题. Miao 等[44] 采用各项异性的线调频 Chirplet 变换以获得能清晰准确地刻画音频信号频率随时间变化的谱图, 然后利用 5 个膨胀率不同的卷积层并行地提取多尺度音频特征, 最后将得到的特征进行融合并输入分类器进行目标类型的识别. 此外, 该方法还设计一种前向特征融合的高效特征金字塔以降低特征融合过程的模型复杂度, 在提高识别性能的同时减少了计算时间. Luo 等[85] 提出基于多分辨率时频特征分析的水中目标识别方法, 并设计一种基于条件卷积生成对抗网络的数据增强策略, 用以增大训练样本规模. 该方法使用的骨干网络为 ResNet, 在 ShipsEar 数据集上实现了 96.32% 的识别精度.

文献 [41−42] 等受人耳对声音频率感知神经机制的启发, 提出模拟人耳听觉系统的卷积神经网络方法, 用于水声目标的类型识别. 如图 8(c) 所示, 该方法借助一组基于卷积运算的滤波器模拟人的听觉皮层、听觉中枢等区域的功能, 将原始时域音频信号分解为一系列不同频率的音频分量信号, 同时卷积核的大小可变, 用以模拟听觉系统受到声音刺激后对不同波长分量的感兴趣程度. 然后在网络的末端堆叠最大池化层和全连接层以提取分解信号的幅值, 并使用一个融合层来合并每个分解信号的特征, 最后将学习到的特征输入到 softmax 层输出类型识别结果. 考虑到传统卷积神经网络存在卷积层和全连接层参数众多导致计算复杂度高、训练效率低的问题, 文献 [42] 设计一种具有初始结构和残差连接的深度架构[86] 作为方法的实现, 既保证了识别精度又提高了训练效率. 表 2 列举了主要的基于 CNN 的水声目标识别方法.

3.2 基于时延神经网络的水中目标识别

TDNN 本质上可以理解为一个一维的 CNN, 通常被用于时序数据的建模和处理. TDNN 的基本思想是通过构建多个时延单元, 对输入信号进行时间平移后的叠加, 并对结果进行线性变换, 最终输

表 2　基于卷积神经网络的水声目标识别方法

Table 2　Convolutional neural network-based methods for UATR

年份	技术特点	模型优劣分析	数据集来源	样本大小
2017	卷积神经网络[66]	自动提取音频表征, 提高了模型的精度	Historic Naval Sound and Video database	16 类
2018	卷积神经网络[71]	使用极限学习机代替全连接层, 提高了模型的识别精度	私有数据集	3 类
2019	卷积神经网络[70]	使用二阶池化策略, 更好地保留了信号分量的差异性	中国南海	5 类
	一种基于声音生成感知机制的卷积神经网络[41]	模拟听觉系统实现多尺度音频表征学习, 使得表征更具判别性	Ocean Networks Canada	4 类
2020	基于 ResNet 的声音生成感知模型[42]	使用全局平均池化代替全连接层, 极大地减少了参数, 提高了模型的训练效率	Ocean Networks Canada	4 类
	一种稠密卷积神经网络 DCNN[43]	使用 DCNN 自动提取音频特征, 降低了人工干预对性能的影响	私有数据集	12 类
2021	一种具有稀疏结构的 GoogleNet[72]	稀疏结构的网络设计减少了参数量, 提升模型的训练效率	仿真数据集	3 类
	一种基于可分离卷积自编码器的 SCAE 模型[73]	使用音频的融合表征进行分析, 证明了方法的鲁棒性	DeepShip	5 类
	残差神经网络[76]	融合表征使得学习到的音频表征更具判别性, 提升了模型的性能	ShipsEar	5 类
	基于注意力机制的深度神经网络[46]	使用注意力机制抑制了海洋环境噪声和其他舰船信号的干扰, 提升模型的识别能力	中国南海	4 类
	基于双注意力机制和多分辨率卷积神经网络架构[81]	多分辨率卷积网络使得音频表征更具判别性, 双注意力机制有利于同时关注局部信息与全局信息	ShipsEar	5 类
	基于多分辨率的时频特征提取与数据增强的水中目标识别方法[85]	多分辨率卷积网络使得音频表征更具判别性, 数据增强增大了模型的训练样本规模, 从而提升了模型的识别性能	ShipsEar	5 类
2022	基于通道注意力机制的残差网络[82]	通道注意力机制的使用使得学习到的音频表征更具判别性和鲁棒性, 提升了模型的性能	私有数据集	4 类
	一种基于融合表征与通道注意力机制的残差网络[83]	融合表征与通道注意力机制的使用使得学习到的音频表征更具判别性和鲁棒性, 提升了模型的性能	DeepShip ShipsEar	5 类
2023	基于注意力机制的多分支 CNN[74]	注意力机制用以捕捉特征图中的重要信息, 多分支策略的使用提升了模型的训练效率	ShipsEar	5 类

出一个特征向量. 这个特征向量可以用于进行水中目标的识别任务. 相比于传统的识别方法, 基于 TDNN 的方法能够兼顾时域信息和频域信息对时序声音信号进行建模, 有效地利用时序信息来捕捉不同目标的动态特征变化, 并且在处理长时间序列时具有更好的性能. 同时, TDNN 还可以通过设置不同的神经元数量和层数来适应不同的任务需求. 考虑到 TDNN 的上述优势, 因此基于 TDNN 的水中目标识别引起了学者的关注.

Ren 等[87] 采用 TDNN 对水中目标进行识别, 该方法使用一种更能反映目标辐射信号频谱分布的小波包分量谱 (Wavelet packet component spectrum, WPCS) 特征作为输入, 实验结果表明相比于其他音频特征, WPCS 特征的性能更好. 文献 [88] 设计一种基于可学习前端 (Underwater acoustic learnable front, UALF) 的水中目标识别方法. UALF 设计一组可学习的一维卷积滤波器用以提取信号中不同频率的信号分量, 然后进一步执行池化操作并输出信号的时频特征, 用以支持后续网络的学习. 由于 UALF 从原始音频信号中自适应地学习合适的特征提取参数, 从而实现更具判别性的音频特征提取. 在 QLED、ShipsEar、DeepShip 数据集上进行实验, 结果表明 UALF 学习到的特征比手工特征器所提取的 STFT 谱、FBank 谱等表现出更好的识别性能.

3.3　基于循环神经网络的水中目标识别

基于 RNN 的水中目标识别是另一种基于深度学习的水中目标识别方法. RNN 是一种能够对序列数据进行建模的神经网络, 其内部包含一个循环结构, 可以将当前时间步的输入与上一个时间步的输出结合起来进行计算. 在水中目标识别中, RNN 可以用于建立从音频序列到目标类别的映射. 长短期记忆模型 (Long short-term memory, LSTM) 和双向长短期记忆模型 (Bi-directional LSTM, Bi-LSTM) 是两种主流的 RNN 架构, 由于其细胞状态能够决定哪些时间状态应该被留下哪些应该被遗忘, 所以在处理水下声音信号这种时序数据时具有更大的优势. 此外, 水中目标所产生的被动声呐信号的分析在很大程度上依赖于局部时频信息和时间序列相关信息, 与 RNN 的特性十分契合. 因此, 有学者将 RNN 应用于水中目标识别.

Li 等[89] 首次提出基于向量传感器原始音频数据的 Bi-LSTM 方法用于水中舰船目标的识别. 该方法直接将向量传感器数据输入到模型中自动学习音频特征, 在一定程度上避免了人工特征提取所带来的信息损失. 此外, Bi-LSTM 使得音频特征同时具有过去和未来的信息作为补充, 更具判别性. Wang 等[90] 提出一种混合时序网络用于水中目标识别, 该网络由双向门控单元 (Bi-direction gated recurrent unit, Bi-GRU) 和多层门控单元 (Gated recurrent unit, GRU) 组合而成, 并通过级联顺序对网络参数进行优化以学习更高层次的目标特征. 实验结果表明, 该方法在具有 4 层 Bi-GRU 和 4 层 GRU 的网络结构上具有良好的抗环境干扰能力和识别性能. Qi 等[91] 则采用 LSTM 模型用以学习音频的相位和频谱特征, 并将学习到的特征进行融合以提升模型的识别性能. 受卷积运算可以很好地学习局部特征, 而 RNN 可以利用数据的时序信息来学习上下文依赖的启发, Kamal 等[92] 提出一种基于 CNN 与 Bi-LSTM 融合的水中目标识别方法. 如图 9 所示, 该方法使用一组可学习的滤波器用以提取被动声呐音频信号的时频特征, 然后将时频特征输入到卷积层执行卷积运算, 接着使用 Bi-LSTM 从当前时刻的之前、之后两个方向捕捉序列的时域特征, 最后使用选择注意力层选取最有效的特征用以执行目标识别任务. Han 等[93] 则设计一种基于一维卷积和 LSTM 相结合的联合网络进行水中目标识别, 其中一维卷积用于减少模型的参数量, LSTM 能同时关注历史信息和当前信息, 有利于更具判别性的时域特征提取.

基于 RNN 的水中目标识别和基于 TDNN 的水中目标识别都是从水中目标所产生的被动声呐信号中提取其时序特征并据此进行目标识别的方法. 然而 TDNN 是一种前向结构的神经网络, 通过卷积和非线性变换来提取输入序列中的局部特性, 主要用于对固定长度的被动声呐信号进行建模. 而 RNN 的主要结构是循环单元, 通过反馈连接将过去的信息进行记忆和传递, 更擅长捕捉信号中的长期依赖关系.

3.4 基于 Transformer 的水中目标识别

Transformer 是一种完全基于自注意力机制的网络架构. 与传统的 RNN 相比, 它可以同时捕捉输入序列中不同位置间的关系, 避免了传统模型中的顺序依赖性问题. 近年来, 在自然语言处理、计算机视觉等领域, Transformer 取得了出色的性能. 此外, 自注意力机制的使用使得 Transformer 可以并行计算, 从而加快训练速度, 并且能够更好地捕捉长距离的依赖关系. 这些优势促使学者将其应用于水中目标识别领域. 在基于 Transformer 的水中目标识别中, 声学信号通常被转换为声谱图或梅尔频谱等表示形式, 然后输入到 Transformer 网络中进行学习. Transformer 的自注意力机制能够捕捉输入信号中不同位置间的依赖关系, 并学习到目标的高级特征表示, 图 10 给出了该方法的基本架构.

Li 等[94] 首次探索将 Transformer 引入水中目标识别领域, 提出频谱转换模型 (Spectrogram transformer model, STM) 用于水中目标识别. 该方法首先提取水中目标所产生的被动声呐信号的时频谱 (包括 STFT 谱、Fbank 谱、MFCC), 并从时域和频域维度将其划分为重叠度为 6 的 16 × 16 大小的图像块, 然后使用一个线性编码层将每个图像块编码为一维的向量序列, 输入到 Transformer 模型中学习更抽象的音频表征. 由于 Transformer 架构可以更好地捕捉长距离的时序信息和全局依赖关系, 与最先进的基线 CNN、CRNN 以及 ResNet18 进行对比, 该方法在 ShipsEar 数据集上的精度分别提升了 13.7%、3.1%、1.8%. Feng 等[95]则在 Transformer 模型的基础上设计一种新的逐层聚合的 Token 机制 (Progressive Token embedding strategy, PTES), 通过多头自注意力机制捕捉全局信息, 通过逐层聚合的 Token 机制分层聚合局部特性, 学习更精细的声音特征表示, 从而提升模型的识别精度.

需要注意的是, 基于 Transformer 的水中目标

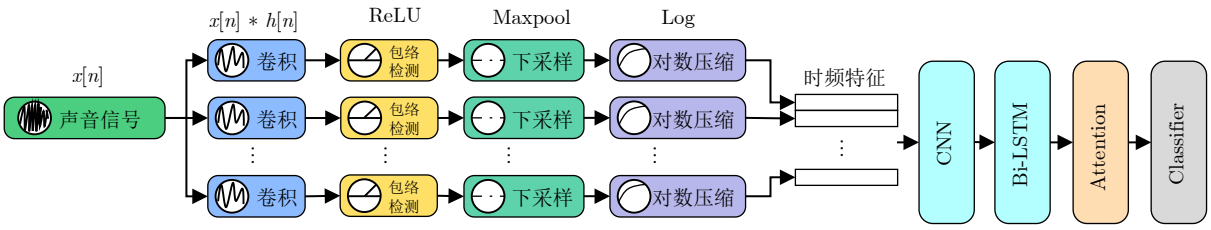


图 9 基于 CNN 与 Bi-LSTM 融合的水声目标识别方法网络架构

Fig. 9 Network framework of UATR methods based on the fusion of CNN and Bi-LSTM

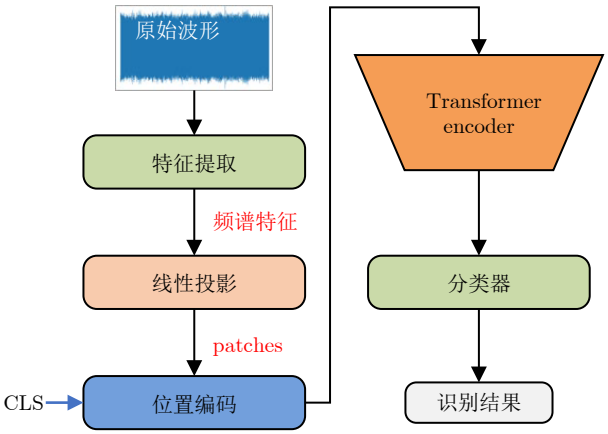


图 10　基于 Transformer 的水声目标识别方法基本架构

Fig. 10　Basic framework of Transformer-based methods for UATR

识别往往通过大规模的数据训练和优化网络参数以达到良好的性能, 这极大限制了它的应用场景. 尽管存在上述问题, 但它具有巨大的潜力. 随着技术的进一步发展和更多数据的积累, 基于 Transformer 的水中目标识别方法有望成为未来水中目标识别的重要研究方向, 为水下环境中的目标监测、海洋资源调查和海洋工程等领域提供更高效和更精确的解决方案. 表 3 列举了主要的基于 TDNN、RNN 和 Transformer 的水声目标识别方法.

3.5　基于迁移学习的水中目标识别

虽然深度学习方法在水中目标识别任务上已经取得了良好的性能, 但它的成功往往需要大量质量良好的标注数据来支撑模型的训练. 由于海洋环境的复杂性和声音隐身技术的应用, 水中目标所产生的被动声呐音频信号往往需要专业的领域知识和丰富的专家经验才能得到质量较高的音频类别标注, 这使得音频标注数据集的规模一般比较小, 难以支撑大规模的深度神经网络模型的学习. 现有的一些研究表明, 迁移学习对于解决模型训练数据不足的问题十分有效[96–97]. TL 通常在一个大规模的相关数据集 (源域) 上训练一个模型, 然后使用小规模目标域的数据集来微调源域上训练好的模型参数, 从而使模型收敛到目标域. TL 可以利用在源域上学习到的知识来加速水中目标识别的训练过程, 从而缓解水中环境数据稀缺或不平衡的问题. 因此, 许多学者开始探索将迁移学习引入水中目标识别领域.

文献 [98–99] 利用在 ImageNet[65] 数据集上训练所得到的性能优异的网络作为预训练模型 (Pre-trained model), 然后利用小规模的音频标注数据集对模型进行微调, 使模型收敛到相应的音频识别任务. 其中, 文献 [96] 利用在计算机视觉领域性能表现优异的 ResNext101[100] 和 Xception[101] 作为预训练模型, 然后采用小样本标注数据进行模型微调, 结果表明该方法的精度提高了 20%. 文献 [99] 则在分别采用 VGG16、ResNet 以及 DenseNet 作为预训练模型并进行微调的基础上, 设计一种模型集成机制以进一步提升识别性能, 精度可达 96.56%.

然而当预训练模型与下游任务属于不同的领域时, 例如使用在图像数据集 ImageNet 训练的模型初始化模型参数, 由于图像数据与音频数据本身的固有偏差会降低其在下游任务上的性能表现. 因此文献 [102] 和文献 [103] 先后探索基于音频大型数据集 AudioSet 的预训练模型, 并将其迁移到其他音频相关的下游任务上. 文献 [102] 和文献 [103] 都采用基于神经网络和音频信号的时频分析方法进行网络结构的设计, 以提取更优秀的音频特征. 其中前者基于自建卷积神经网络的架构, 后者则以 VGGish 作为骨干网络. 此外, 为增强结果的可解释性, 后者参考有限冲击响应滤波器的计算模式, 设计了基于一维卷积运算的网络滤波器并馈入注意力机制, 通过网络优化自动挖掘出适合当前目标识别任务的音频特征. 在 ShipsEar 数据集上的结果表明, 该模型能够自适应感知水中目标的频域特征, 在各种目标识别任务中表现出有竞争力的性能, 特别是那些对泛化性要求高的任务. 表 4 列举了主要的基于 TL 的水声目标识别方法.

3.6　基于无监督学习和自监督学习的水中目标识别方法

基于无监督学习和自监督学习的水中目标识别方法主要是通过数据自身的统计规律和特征分布来进行特征提取与模型训练, 从而避免了需要大量标签数据的瓶颈问题. 与监督学习方法相比, 无监督学习和自监督学习方法的一个共同特点是不需要大量标注数据, 因此具有更好的可扩展性和适用性.

3.6.1　基于无监督学习的水中目标识别

基于无监督学习的水中目标识别方法主要基于自编码器、聚类分析等策略从不含标签信息的数据中进行学习. 其中, 自编码器可以用于特征的无监督学习, 通过重构输入数据来学习水中目标所产生的被动声呐音频信号的特征表示. 聚类分析则是将未标注的数据分成不同的类别, 从而获得数据的特征分布信息.

深度置信网络 (Deep belief network, DBN) 是一种类似于自编码器的网络架构, 被广泛应用于水中目标识别. DBN 是一种由多个受限玻尔兹曼机

表 3 基于时延神经网络、循环神经网络和 Transformer 的水声目标识别方法

Table 3 Time delay neural networks-based, recurrent neural network-based and Transformer-based methods for UATR

年份	技术特点	模型优劣分析	数据集来源	样本大小
2019	基于时延神经网络的 UATR[87]	时延神经网络能够学习音频时序信息, 从而提高模型的识别能力	私有数据集	2 类
2022	一种可学习前端[88]	可学习的一维卷积滤波器可以实现更具判别性的音频特征提取, 表现出比传统手工提取的特征更好的性能	QLED, ShipsEar, DeepShip	QLED 2 类 ShipsEar 5 类 DeepShip 4 类
2020	采用 Bi-LSTM 同时考虑过去与未来信息的 UATR[89]	使用双向注意力机制能够同时学习到历史和后续的时序信息, 从而使得音频表征蕴含信息更丰富以及判别性更高, 然而该方法复杂度较高	Sea Trial	2 类
	基于 Bi-GRU 的混合时序网络[90]	混合时序网络从多个维度关注时序信息, 从而学习到更具判别性的音频特征, 提高模型的识别能力	私有数据集	3 类
2021	采用 LSTM 融合音频表征[91]	该方法能够同时学习音频的相位和频谱特征, 并加以融合, 从而提升模型的识别性能	私有数据集	2 类
	CNN 与 Bi-LSTM 组合的 UATR[92]	CNN 与 Bi-LSTM 组合可以提取出同时关注局部特性和时序上下文依赖的音频特征, 提高了模型的识别能力	私有数据集	3 类
2022	一维卷积与 LSTM 组合的 UATR[93]	首次采用一维卷积和 LSTM 的组合网络提取音频表征, 能够在提高音频识别率的同时降低模型的参数量, 然而该方法稳定性有待提高	ShipsEar	5 类
2022	Transformer[94–95]	增强了模型的泛化性和学习能力, 提高了模型的识别准确率	ShipsEar	5 类
		加入逐层聚合的 Token 机制, 同时兼顾全局信息和局部特性, 提高了模型的识别准确率	ShipsEar, DeepShip	5 类

表 4 基于迁移学习的水声目标识别方法

Table 4 Transfer learning-based methods for UATR

年份	技术特点	模型优劣分析	数据集来源	样本大小
2019	基于 ResNet 的迁移学习[98]	在保证较高性能的同时减少对标签样本的需求, 但不同领域任务的数据特征分布存在固有偏差	Whale FM website	16 类
2020	基于 ResNet 的迁移学习[99]	在预训练模型的基础上设计模型集成机制, 提升识别性能的同时减少了对标签样本的需求, 但不同领域任务的数据特征分布存在固有偏差	私有数据集	2 类
	基于 CNN 的迁移学习[102]	使用 AudioSet 音频数据集进行预训练, 减轻了不同领域任务的数据特征分布所存在的固有偏差	—	—
2022	基于 VGGish 的迁移学习[103]	除了使用 AudioSet 数据集进行预训练, 还设计基于时频分析与注意力机制结合的特征提取模块, 提高了模型的泛化能力	ShipsEar	5 类

(Restricted Boltzmann machine, RBM) 组成的无监督学习深度神经网络, 这些 RBM 依次训练, 以便逐层生成高层次的特征表示来增强特征的判别性, 最后使用反向传播算法进行微调, 以进一步提高识别精度.

在水中目标识别中, 深度置信网络可以从被动声呐采集到的声波信号中自动学习特征表示, 从而实现目标类型的识别. 该方法的优点是可以自动提取特征, 避免了手动设计特征的困难和复杂性, 并且在处理大规模数据时可以获得较高的准确性和泛化能力. 由于 DBN 具有上述优势, 近年来基于 DBN 的水中目标识别研究十分广泛. 文献 [66, 104–105] 利用 DBN 在无标签的舰船辐射噪声信号上进行预训练, 然后在预训练好的 DBN 模型后面加入分类层进行模型的微调, 其中文献 [104] 在包含 40 个类别的 1 000 个样本中精度达到了 90.23%. 然而, 该方法在小数据集上微调的迭代次数较多, 可能存在过拟合的风险. 杨宏晖等[106] 提出一种混合正则化深度置信网络 (Hybrid regularization deep belief network, HR-DBN) 用于水中目标识别, 其中最大互信息组正则化策略旨在提高隐含层的稀疏度, 增强所学到的声音特征的判别性; 数据驱动正则化策略则是利用大量的无标签样本进行预训练, 从中学习水中目标的先验知识与通用表征, 引导网络更好地学习. 在该工作的基础上, Yang 等[107] 进一步提出一种基于 DBN 与竞争学习机制结合的水中目标识别方法 —— 结合竞争机制的深度置信网络 (Competitive deep belief network, CDBN). 具体来说, 该方法首先利用大量无标签音频数据以无监督学习的方式预训练 RBM. 其次, 对于隐藏层, 该

架构根据不同类别对应的得分对隐藏层单元进行分组. 然后, 通过在分组的隐藏层单元之间添加横向连接, 构建了具有组内增强和组间抑制机制的竞争层, 组成竞争性受限玻尔兹曼机 (Competitive restricted Boltzmann machine, CRBM). 最后, 将 CRBM 堆叠构建 CDBN, 并对整个模型进行微调, 以最大化其预测水中目标的概率. 该方法通过增加竞争层, 可以迫使网络学习到更具有判别性的音频特征. 然而当隐藏层神经元过多时, 计算任意两个特征之间的互信息是低效的. 基于此, Shen 等[108] 提出一种压缩的 CDBN 用于船舶辐射噪声的特征学习, 使用竞争学习的机制使得同类别样本的特征更加聚集, 并采用基于互信息的剪枝策略去除网络的冗余参数. 结果表明该方法的识别精度比 CDBN 提高了 5.3%.

受自编码器思想的启发, Cao 等[109] 使用堆叠自编码器架构进行音频信号表征学习, 并在网络末端使用 softmax 层进行信号识别. 其中堆叠自编码器的基本结构为稀疏自编码器, 并以无监督贪婪范式进行逐层训练, 在包含 3 类的海洋测试数据集上精度达到了 94.12%. Luo 和 Feng[110] 设计一种基于 RBM 进行预训练、级联 BP 神经网络进行水中目标识别的方法. 该方法将信号的 MFCC 和 GFCC 归一化频谱作为输入, 使用 4 层 RBM 进行更抽象的音频特征学习, 并将得到的音频特征输入到 BP 神经网络分类器中进行信号识别. 在两个真实舰船辐射噪声数据集上对该方法进行测试, 结果表明该方法比手工设计的特征提取方法具有更好的识别精度和鲁棒性.

3.6.2　基于自监督学习的水中目标识别

基于 SSL 的水中目标识别是利用数据本身的内在结构和特性进行学习, 从而实现对目标类型识别的方法. 该方法通过设计代理任务来进行模型训练, 从而消除对人工标注的数据标签的需求. 常用的代理任务有对比学习任务和预测任务. 其中, 对比学习旨在将来自同一样本的不同视图进行比较, 以学习样本之间的相似性和差异性. 在水中目标识别中, 可以设计对比学习任务, 如同一声音信号在时间或频域上的不同切片进行对比. 通过对比学习的训练, 网络可以学习到区分目标和背景的特征表示. 预测任务则是通过模型对未来或缺失的部分进行预测, 以学习数据的内在结构. 在水中目标识别中, 可以设计预测任务, 如预测声音信号的下一个时间步或缺失的频谱区域. 通过预测任务的训练, 网络可以学习到对目标关键特征的建模能力.

Luo 等[111] 提出一种基于 RBM 的自编码器与重构输入的水中目标识别方法. 如图 11 所示, 自编码器是一个由多层 RBM 堆叠而成的结构, 用以逐层提取更抽象的音频特征. 自解码器与自编码器在结构上对称, 用以逐层重构原始输入. 最后将重构

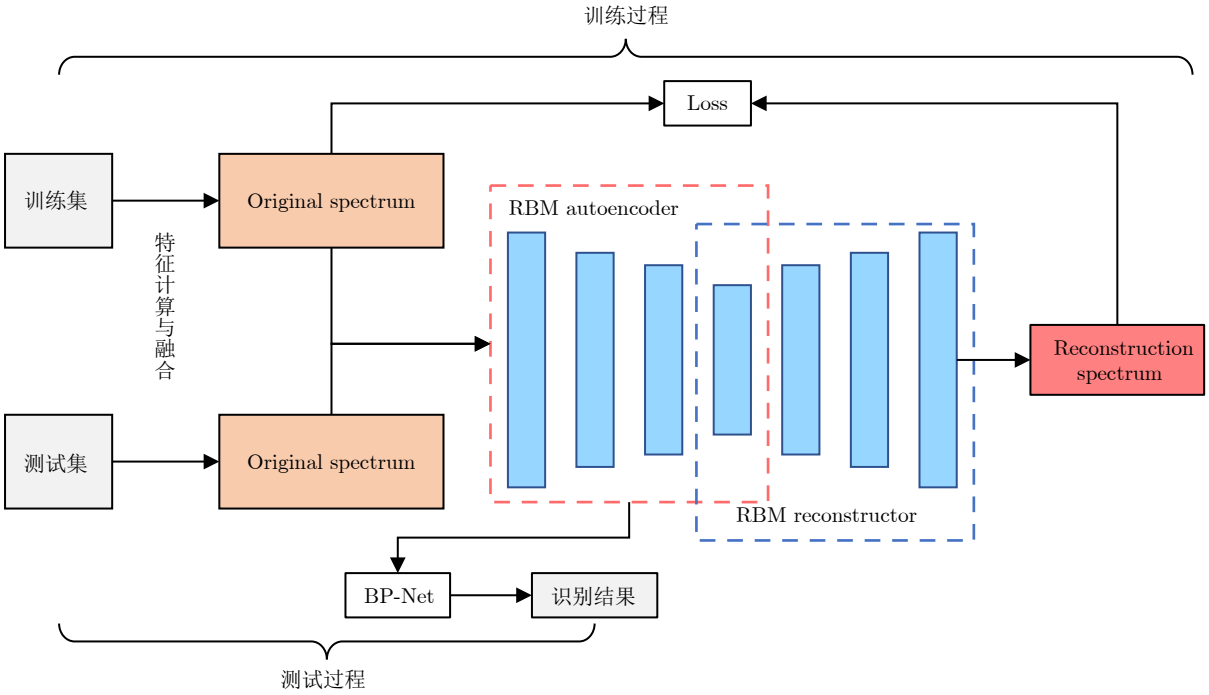


图 11　基于 RBM 自编码器重构的水声目标识别方法架构

Fig. 11　The framework of RBM autoencoder-based reconstruction methods for UATR

的声音特征与原始输入进行对比构成一组自监督信号, 以指导网络的学习. 该方法融合功率谱和 DEMON 谱作为模型的输入, 在 ShipsEar 数据集上取得了 92.6% 的识别性能. Sun 和 Luo[112] 将自监督对比学习的思想引入水中目标识别领域, 提出对比编码学习 (Contrastive coding for UATR, CCU) 的方法.

近年来, 另一种基于掩码建模的 SSL 方法在音频分类任务上表现出良好的性能. Gong 等[49] 提出一种联合判别与掩码重构的自监督学习方法 (Self-supervised audio spectrogram Transformer, SSAST) 用于音频与语音分类. 图 12 展示了该方法的网络架构, 首先将音频信号转换为频谱特征并将频谱图切分成大小相等且互不重叠的 patch, 对 patch 执行随机掩码操作 (图中灰色的 patch), 经过线性投影层将 patch 编码为一维向量, 在编码向量中加入每个 patch 对应的位置编码作为最终的模型输入, 送入 Transformer encoder 中学习更高层次的目标特征. Transformer encoder 的输出: 1) 在预训练阶段, 输入到 Reconstruct head 和 Classification head 中分别重构掩码的 patch 并对恢复的 patch 进行分类, 通过评估重构效果和分类精度来指导网络的反向传播; 2) 在微调阶段, 用于音频的分类. 然而, 该方法在水中目标识别领域的应用仍处于探索中. 基于这样的观察, 文献 [113−114] 率先将掩码建模的思想引入水中目标识别任务中, 提出掩码建模与多表征重构的方法用于水中目标识别, 其基本处理流程与 SSAST 类似, 都包含频谱转换、patch 的切分、随机掩码与重构 (预训练阶段)、微调等过程. 其中输入的频谱特征为 Log-Mel, 使用两个 decoder 分别用于重构被掩码的 Log-Mel 特征以及预测 Grammatone 频谱, 通过评估重构效果和预测效果来指导网络的训练, 该方法在 DeepShip 数据集上实现了 78.03% 的识别精度.

总的来说, 基于自监督学习的方法可以充分利用未标记数据进行训练, 从而提高水中目标识别的性能和泛化能力. 这种方法在数据量有限或难以获得标记数据的情况下尤为有用, 并且能够有效应对水中环境的复杂性和变化性. 而基于无监督学习的方法通常使用聚类或降维等无监督学习方法, 并且没有监督信号参与训练. 此外, 基于自监督学习的方法通常能够提取更丰富的特征, 但需要更多的计

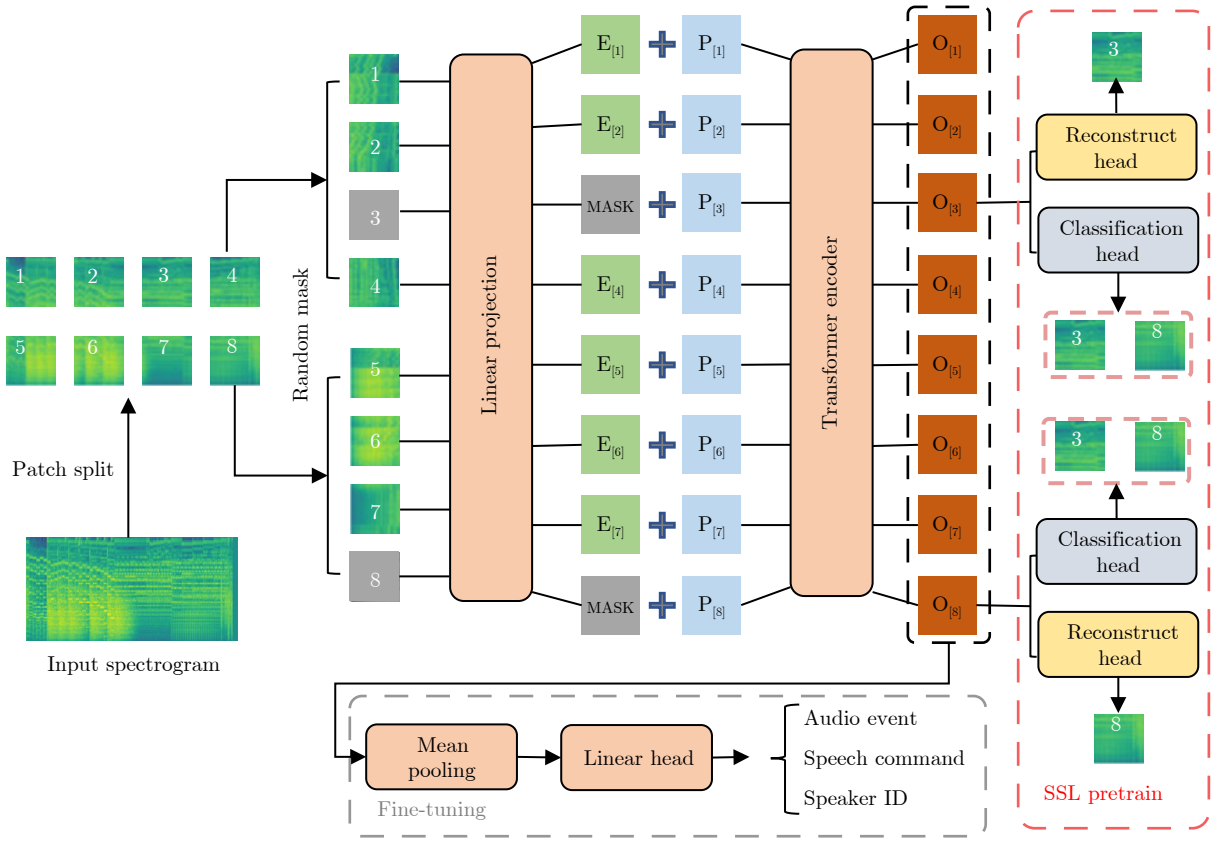


图 12 SSAST 的网络结构

Fig. 12 The network architecture of SSAST

算资源和时间, 而基于无监督学习的方法则更加简单快速, 但通常提取的特征较为简单. 表 5 列举了主要的基于无监督学习和自监督学习的水声目标识别方法.

3.7　深度学习算法与传统机器学习算法的比较

基于机器学习算法进行水中目标识别的研究已成为当前采用被动声呐音频信号实现水中目标识别任务的主流方法. 在第 2 节和第 3 节中, 全面梳理了基于传统机器学习算法和基于深度学习算法的水中目标识别技术. 传统机器学习算法通过手工设计和选择特征, 能够根据领域知识和专家经验进行目标识别, 并且具有较高的计算效率. 然而, 传统算法通常需要依赖专业知识, 手动提取特征, 因此在处理复杂的数据模式时存在一定的限制. 此外, 传统机器学习算法的网络结构往往比较浅, 模型性能上界有限.

相比之下, 深度学习算法具有自动特征学习的能力, 能够从原始数据中学习到更有效的特征表示, 适应复杂的数据模式和非线性关系. 深度学习算法具有深层的网络结构, 具有较大的网络容量, 可以更好地捕捉数据中的复杂关系. 它具有较强的表达能力和学习能力, 可以通过大规模数据和增加模型复杂度来提高性能. 然而, 深度学习算法对于标记数据的需求较高且计算复杂度较高. 此外, 由于某些深度学习模型的复杂性, 它们可能是黑盒模型, 难以解释和理解模型的决策过程.

总的来说, 在具备领域知识和对特定问题有深入理解的情况下, 传统机器学习算法可以提供较好的性能和可解释性, 尤其适用于处理相对简单的数据模式. 而对于复杂的数据模式和大规模数据集, 深度学习算法能够更好地发挥其自动特征学习能力. 然而, 研究人员需要权衡所需的标记数据量和计算资源, 并在应用中注意深度学习模型的可解释性问题.

4　水中目标识别相关数据集

近年来, 机器学习特别是深度学习的快速发展使数据驱动学习在水中目标识别领域取得了优异的效果, 并逐渐成为该领域研究的主流方法. 然而, 深度学习对训练数据的需求是巨大的, 因为它们需要大量带标签的数据来指导模型学习到正确的知识. 由于水声标注数据的获取涉及复杂的技术、高昂的成本以及潜在的国防安全敏感信息, 大部分数据集并不公开. 因此, 许多水中目标识别的研究是基于私人搜集的未公开数据集[104−107]、仿真数据集[72] 或者基于有限的真实数据进行数据增强[115−117] 的, 但由于缺乏合适大小的真实数据集, 它们所达到的精度仍然不能令人满意. 此外, 由于所使用的数据集不同, 不同方法之间的性能对比也难以令人信服. 为了发展更准确的水声目标识别技术以及便于不同方法之间的性能对比, 陆续推出了一些公开可用的水声数据集. 其中较为有代表性的为 Santos-Domínguez 等[118] 提出的 ShipsEar 数据集和 Irfan 等[73] 提出的 DeepShip 数据集. 表 6 总结了常用的公开水声数据集.

相比于计算机视觉领域, 目前水声标注数据集的数量和规模仍有待发展. 对于一个良好的数据集而言, 它需要具备良好的可读性、足够的完整性、可靠性和结果的可复现性. 良好的可读性便于使用者轻松理解数据的含义; 足够的完整性确保数据蕴含完整的信息, 便于指导网络正确的学习; 可靠性要求数据具有较高的质量; 可复现性确保多次基于数据的分析结果基本一致.

表 5　基于无监督学习和自监督学习的水声目标识别方法

Table 5　Unsupervised and self-supervised learning-based methods for UATR

年份	技术特点	模型优劣分析	数据集来源	样本大小
2013	深度置信网络[104−108]	对标注数据集的需求小, 但由于训练数据少, 容易出现过拟合的风险	私有数据集	40 类
2017		加入混合正则化策略, 增强了所学到音频表征的判别性, 提高了模型的识别准确率		3 类
2018		加入竞争机制, 增强了所学到音频表征的判别性, 提高了模型的识别准确率		2 类
2018		加入压缩机制, 减少了模型的冗余参数, 提升了模型的识别准确率	中国南海	2 类
2021	基于 RBM 自编码器与重构的 SSL[111]	降低了模型对标签数据的需求, 增强了模型的泛化性和可扩展性	ShipsEar	5 类
2022	基于掩码建模与重构的 SSL[113]	使用掩码建模与多表征重构策略, 提升了模型对特征的学习能力, 从而提升了识别性能	DeepShip	5 类
2023	基于自监督对比学习的 SSL[112]	降低了模型对标签数据的需求, 增强了学习到的音频特征的泛化性和对数据的适应能力	ShipsEar, DeepShip	5 类

表 6 常用的公开水声数据集总结

Table 6 Summary of commonly used public underwater acoustic signal datasets

数据集名称	数据结构				数据类型	采样频率 (kHz)	获取地址
	类别	样本数 (个)	持续时间 (样本)	总时间			
DeepShip	Cargo Ship	110	180 ~ 610 s	10 h 40 min	音频	32	DeepShip
	Tug	70	180 ~ 1140 s	11 h 17 min			
	Passenger Ship	70	6 ~ 1530 s	12 h 22 min			
	Tanker	70	6 ~ 700 s	12 h 45 min			
ShipsEar	A	16	—	1729 s	音频	32	ShipsEar
	B	17		1435 s			
	C	27		4054 s			
	D	9		2041 s			
	E	12		923 s			
Ocean Network Canada (ONC)	Background noise	17000	3 s	8 h 30 min	音频	—	ONC
	Cargo	17000		8 h 30 min			
	Passenger Ship	17000		8 h 30 min			
	Pleasure craft	17000		8 h 30 min			
	Tug	17000		8 h 30 min			
Five-element acoustic dataset	9 个类别	360	0.5 s	180 s	音频	—	Five-element···
Historic Naval Sound and Video	16 个类别	—	2.5 s	—	音视频	10.24	Historic Naval···
DIDSON	8 个类别	524	—	—	—	—	DIDSON
Whale FM	Pilot whale	10858	1 ~ 8 s	5 h 35 min	音频	—	Whale FM
	Killer whale	4673					

注: 1. 官方给出的 DeepShip 数据集只包含 Cargo Ship、Tug、Passenger Ship 和 Tanker 这 4 个类别. 在实际研究中, 学者通常会自定义一个新的类别 "Background noise" 作为第 5 类.

2. 获取地址的访问时间为 2023-07-20.

5 结论与展望

本文对基于被动声呐音频信号的水中目标识别的相关研究进行综述. 首先从数据的角度阐述了当前水中目标识别主要使用的数据类型为被动声呐音频信号, 并对音频信号处理中所涉及的关键技术进行了概述, 包括采用被动声呐音频信号进行水中目标识别的基本原理、被动声呐音频信号分析的数理基础以及系统介绍了相关研究中所使用的音频特征提取方法, 为后续介绍机器学习方法在水中目标识别任务中的应用提供了必要的背景知识. 然后分别从传统机器学习和深度学习的角度全面分析了水中目标识别任务的相关进展, 发现由于海洋环境的复杂性和各种声音隐身技术的应用, 基于深度学习的水中目标识别方法逐渐成为主流研究方法. 按照深度学习的模型结构将这些方法分为: 1) 基于卷积神经网络的方法; 2) 基于时延神经网络的方法; 3) 基于循环神经网络的方法; 4) 基于 Transformer 的方法; 5) 基于迁移学习的方法; 6) 基于无监督和自监督学习的方法. 对相关方法进行上述分类, 可以确保在涵盖所有主流方法的同时又能实现每个类别之间不会存在交集的目的, 分类脉络更清晰. 图 13 展示了这些方法在水中目标识别任务上的性能对比, 从图中可以发现, 基于自监督学习的方法在性能上足以媲美有监督学习的方法, 并且由于该方法对标签数据需求小、泛化性和可扩展性高等优势, 近年来自监督学习方法逐渐成为基于被动声呐音频信号的水中目标识别任务的研究热点.

然而需要注意的是, 虽然近年来深度学习方法在很大程度上提高了水中目标识别的精度和速度, 但距离真正实时、鲁棒、精准和可持续学习的识别系统, 仍存在较大的提升空间. 主要表现在:

1) 公开可获得的被动声呐数据集及其标注的显著稀缺性. 由于海洋环境的复杂性、处理与标注数据集的高昂成本以及潜在的国防敏感信息等因素[119], 使得该类数据集通常不可公开获取. 这使得这类研究在很大程度上失去了对比意义, 因为如果没有一个共同的数据集, 对解决方案进行比较和基准测试难以进行.

2) 噪声标签的普遍性. 由于被动声呐数据的标注成本高昂, 使用廉价的数据收集方式 (比如在线查询和众包等) 成为可行的替代方案. 然而这些

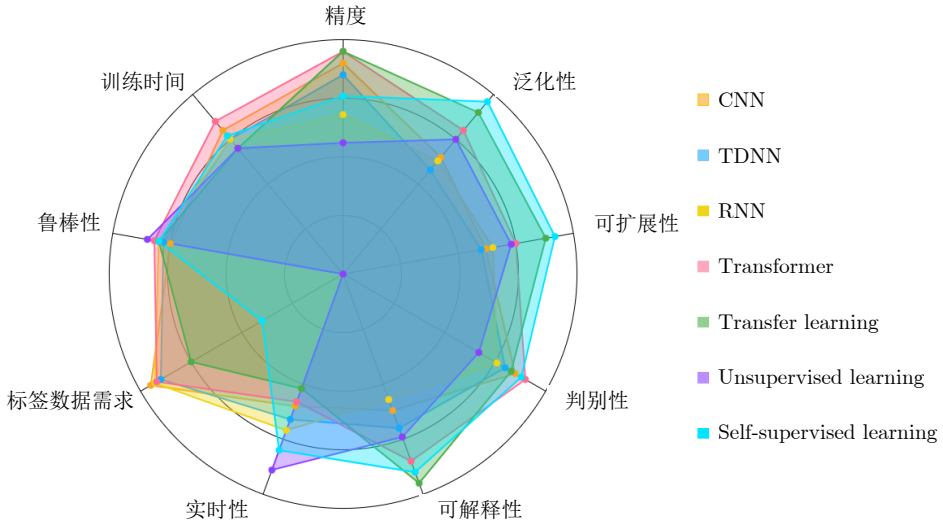


图 13　不同深度学习方法在水声目标识别领域的性能对比

Fig. 13　Performance comparison of various deep learning methods for UATR

方式会引入大量的噪声标签, 甚至是专家标注的数据集中也可能出现噪声标签, 深度学习由于其强大的拟合能力, 很容易受到这些噪声标签的干扰. 因此, 在将数据用于模型训练之前, 进行噪声的清洗是一项十分重要的工作[120−121].

3) 具有判别性和泛化性的水中目标通用音频特征提取方法仍处于探索中. 目前许多研究所采用的水中目标音频特征往往是基于手工制作、特征提取器进行提取的, 然而这类参数固定的特征提取器难以自动适应数据的特点. 其次, 被动声呐数据受采集时间、季节、天气、地理区域、传感器类型、海洋深度等影响, 往往需要专业的领域知识和专家经验来选取合适的音频特征, 以适应相应的任务场景. 此外, 虽然有些研究开始采用深度学习方法自动提取音频特征, 但所设计的提取策略也仅在私人数据集或单一数据集上取得相对不错的效果, 在其他数据集上的性能仍有待验证. 因此, 探索具有判别性和泛化性的通用音频特征提取方法是一项十分有意义的工作.

4) 模型持续学习能力的探索. 现有研究主要聚焦于设计合适的深度学习策略以提升模型的识别能力, 然而这些方法在模型训练结束后, 对知识的学习过程也随之结束. 此外, 水中目标所处的海洋环境是动态变化的, 这种参数固定的模型难以适应这样的任务场景. 因此, 探讨模型的持续学习问题是一个非常具有现实意义的问题.

此外, 在第 4 节, 总结了文献中常用的一些被动声呐音频公开数据集, 并指出一个良好的数据集应该具备的特点, 为后续搭建被动声呐水声数据集提供了指导性意见. 同时, 本文认为未来的工作应该明确所使用数据集的获取条件和限制, 同时最好能在公开数据集上进一步测试模型的性能, 以便更好地进行性能对比.

总的来说, 高精准、可扩展性、鲁棒性、实时性和可持续学习性仍然是未来基于被动声呐音频信号的水中目标识别任务的重要挑战. 同时, 如何将已有的成果应用于生活实际、实现模型压缩和跨平台部署等也是亟需解决的问题.

References

1 Cho H, Gu J, Yu S C. Robust sonar-based underwater object recognition against angle-of-view variation. *IEEE Sensors Journal*, 2016, **16**(4): 1013−1025

2 Dästner K, Roseneckh-Köhler B V H Z, Opitz F, Rottmaier M, Schmid E. Machine learning techniques for enhancing maritime surveillance based on GMTI radar and AIS. In: Proceedings of the 19th International Radar Symposium (IRS). Bonn, Germany: IEEE, 2018. 1−10

3 Terayama K, Shin K, Mizuno K, Tsuda K. Integration of sonar and optical camera images using deep neural network for fish monitoring. *Aquacultural Engineering*, 2019, **86**: Article No. 102000

4 Choi J, Choo Y, Lee K. Acoustic classification of surface and underwater vessels in the ocean using supervised machine learning. *Sensors*, 2019, **19**(16): Article No. 3492

5 Marszal J, Salamon R. Detection range of intercept sonar for CWFM signals. *Archives of Acoustics*, 2014, **39**: 215−230

6 Meng Qing-Xin. Research on Passive Recognition Methods of Marine Targets [Ph.D. dissertation], Harbin Engineering University, China, 2016.
(孟庆昕. 海上目标被动识别方法研究 [博士学位论文], 哈尔滨工程大学, 中国, 2016.)

7 Yang H, Lee K, Choo Y, Kim K. Underwater acoustic research trends with machine learning: General background. *Journal of Ocean Engineering and Technology*, 2020, **34**(2): 147−154

8 Fang Shi-Liang, Du Shuan-Ping, Luo Xi-Wei, Han Ning, Xu Xiao-Nan. Development of underwater acoustic target feature analysis and recognition technology. *Bulletin of Chinese Academy of Sciences*, 2019, **34**(3): 297−305

(方世良, 杜栓平, 罗昕炜, 韩宁, 徐晓男. 水声目标特征分析与识别技术. 中国科学院院刊, 2019, **34**(3): 297−305)

9 Domingos L C, Santos P E, Skelton P S, Brinkworth R S, Sammut K. A survey of underwater acoustic data classification methods using deep learning for shoreline surveillance. *Sensors*, 2022, **22**(6): Article No. 2181

10 Luo X W, Chen L, Zhou H L, Cao H L. A survey of underwater acoustic target recognition methods based on machine learning. *Journal of Marine Science and Engineering*, 2023, **11**(2): Article No. 384

11 Kumar S, Phadikar S, Majumder K. Modified segmentation algorithm based on short term energy & zero crossing rate for Maithili speech signal. In: Proceedings of the International Conference on Accessibility to Digital World (ICADW). Guwahati, India: IEEE, 2016. 169−172

12 Boashash B. *Time-frequency Signal Analysis and Processing: A Comprehensive Reference*. Pittsburgh: Academic Press, 2015.

13 Gopalan K. Robust watermarking of music signals by cepstrum modification. In: Proceedings of the IEEE International Symposium on Circuits and Systems. Kobe, Japan: IEEE, 2005. 4413−4416

14 Lee C H, Shih J L, Yu K M, Lin H S. Automatic music genre classification based on modulation spectral analysis of spectral and cepstral features. *IEEE Transactions on Multimedia*, 2009, **11**(4): 670−682

15 Tsai W H, Lin H P. Background music removal based on cepstrum transformation for popular singer identification. *IEEE Transactions on Audio, Speech, and Language Processing*, 2011, **19**(5): 1196−1205

16 Oppenheim A V, Schafer R W. From frequency to quefrency: A history of the cepstrum. *IEEE Signal Processing Magazine*, 2004, **21**(5): 95−106

17 Stevens S S, Volkmann J, Newman E B. A scale for the measurement of the psychological magnitude pitch. *The Journal of the Acoustical Society of America*, 1937, **8**(3): 185−190

18 Oxenham A J. How we hear: The perception and neural coding of sound. *Annual Review of Psychology*, 2018, **69**: 27−50

19 Donald D A, Everingham Y L, McKinna L W, Coomans D. Feature selection in the wavelet domain: Adaptive wavelets. *Comprehensive Chemometrics*, 2009: 647−679

20 Souli S, Lachiri Z. Environmental sound classification using log-Gabor filter. In: Proceedings of the IEEE 11th International Conference on Signal Processing. Beijing, China: IEEE, 2012. 144−147

21 Costa Y, Oliveira L, Koerich A, Gouyon F. Music genre recognition using Gabor filters and LPQ texture descriptors. In: Proceedings of the Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. Havana, Cuba: Springer, 2013. 67−74

22 Ezzat T, Bouvrie J V, Poggio T A. Spectro-temporal analysis of speech using 2-D Gabor filters. In: Proceedings of 8th Annual Conference of the International Speech Communication Association. Antwerp, Belgium: ISCA, 2007. 506−509

23 He L, Lech M, Maddage N, Allen N. Stress and emotion recognition using log-Gabor filter analysis of speech spectrograms. In: Proceedings of the 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops. Amsterdam, Netherlands: IEEE, 2009. 1−6

24 Cai Yue-Bin, Zhang Ming-Zhi, Shi Xi-Zhi, Lin Liang-Ji. The feature extraction and classification of ocean acoustic signals based on wave structure. *Acta Electronica Sinica*, 1999, **27**(6): 129−130
(蔡悦斌, 张明之, 史习智, 林良骥. 舰船噪声波形结构特征提取及分类研究. 电子学报, 1999, **27**(6): 129−130)

25 Meng Q X, Yang S. A wave structure based method for recognition of marine acoustic target signals. *The Journal of the Acoustical Society of America*, 2015, **137**(4): 2242

26 Meng Q X, Yang S, Piao S C. The classification of underwater acoustic target signals based on wave structure and support vector machine. *The Journal of the Acoustical Society of America*, 2014, **136**(4_Supplement): 2265

27 Rajagopal R, Sankaranarayanan B, Rao P R. Target classification in a passive sonar——An expert system approach. In: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing. Albuquerque, USA: IEEE, 1990. 2911−2914

28 Lourens J G. Classification of ships using underwater radiated noise. In: Proceedings of the Southern African Conference on Communications and Signal Processing (COMSIG). Pretoria, South Africa: IEEE, 1988. 130−134

29 Liu Q Y, Fang S L, Cheng Q, Cao J, An L, Luo X W. Intrinsic mode characteristic analysis and extraction in underwater cylindrical shell acoustic radiation. *Science China Physics, Mechanics and Astronomy*, 2013, **56**: 1339−1345

30 Das A, Kumar A, Bahl R. Marine vessel classification based on passive sonar data: The cepstrum-based approach. *IET Radar, Sonar & Navigation*, 2013, **7**(1): 87−93

31 Wang S G, Zeng X Y. Robust underwater noise targets classification using auditory inspired time-frequency analysis. *Applied Acoustics*, 2014, **78**: 68−76

32 Kim K I, Pak M I, Chon B P, Ri C H. A method for underwater acoustic signal classification using convolutional neural network combined with discrete wavelet transform. *International Journal of Wavelets, Multiresolution and Information Processing*, 2021, **19**(4): Article No. 2050092

33 Qiao W B, Khishe M, Ravakhah S. Underwater targets classification using local wavelet acoustic pattern and multi-layer perceptron neural network optimized by modified whale optimization algorithm. *Ocean Engineering*, 2021, **219**: Article No. 108415

34 Wei X, Li G H, Wang Z Q. Underwater target recognition based on wavelet packet and principal component analysis. *Computer Simulation*, 2011, **28**(8): 8−290

35 Xu K L, You K, Feng M, Zhu B Q. Trust-worth multi-representation learning for audio classification with uncertainty estimation. *The Journal of the Acoustical Society of America*, 2023, **153**(3_supplement): Article No. 125

36 Xu Xin-Zhou, Luo Xi-Wei, Fang Shi-Liang, Zhao Li. Research process of underwater target recognition based on auditory perceiption mechanism. *Technical Acoustics*, 2013, **32**(2): 151−158
(徐新洲, 罗昕炜, 方世良, 赵力. 基于听觉感知机理的水下目标识别研究进展. 声学技术, 2013, **32**(2): 151−158)

37 Békésy G V. On the elasticity of the cochlear partition. *The Journal of the Acoustical Society of America*, 1948, **20**(3): 227−241

38 Johnstone B M, Yates G K. Basilar membrane tuning curves in the guinea pig. *The Journal of the Acoustical Society of America*, 1974, **55**(3): 584−587

39 Zwislocki J J. Five decades of research on cochlear mechanics. *The Journal of the Acoustical Society of America*, 1980, **67**(5): 1679−1685

40 Fei Hong-Bo, Wu Wei-Guan, Li Ping, Cao Yi. Acoustic scene classification method based on Mel-spectrogram separation and LSCNet. *Journal of Harbin Institute of Technology*, 2022, **54**(5): 124−130
(费鸿博, 吴伟官, 李平, 曹毅. 基于梅尔频谱分离和 LSCNet 的声学场景分类方法. 哈尔滨工业大学学报, 2022, **54**(5): 124−130)

41 Yang H H, Li J H, Shen S, Xu G H. A deep convolutional neural network inspired by auditory perception for underwater acoustic target recognition. *Sensors*, 2019, **19**(5): Article No. 1104

42 Shen S, Yang H H, Yao X H, Li J H, Xu G H, Sheng M P. Ship type classification by convolutional neural networks with auditory-like mechanisms. *Sensors*, 2020, **20**(1): Article No. 253

43 Doan V S, Huynh-The T, Kim D S. Underwater acoustic target classification based on dense convolutional neural network. *IEEE Geoscience and Remote Sensing Letters*, 2020, **19**: Article No. 1500905

44 Miao Y C, Zakharov Y V, Sun H X, Li J H, Wang J F. Underwater acoustic signal classification based on sparse time-fre-

quency representation and deep learning. *IEEE Journal of Oceanic Engineering*, 2021, **46**(3): 952−962

45 Hu G, Wang K J, Liu L L. Underwater acoustic target recognition based on depthwise separable convolution neural networks. *Sensors*, 2021, **21**(4): Article No. 1429

46 Xiao X, Wang W B, Ren Q Y, Gerstoft P, Ma L. Underwater acoustic target recognition using attention-based deep neural network. *JASA Express Letters*, 2021, **1**(10): Article No. 106001

47 Gong Y, Chung Y A, Glass J. AST: Audio spectrogram Transformer. arXiv preprint arXiv: 2104.01778, 2021.

48 Yang H H, Li J H, Sheng M P. Underwater acoustic target multi-attribute correlation perception method based on deep learning. *Applied Acoustics*, 2022, **190**: Article No. 108644

49 Gong Y, Lai C I J, Chung Y A, Glass J. SSAST: Self-supervised audio spectrogram transformer. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI, 2022. 10699−10709

50 He K M, Chen X L, Xie S N, Li Y H, Dollár P, Girshick R. Masked autoencoders are scalable vision learners. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE, 2022. 15979−15988

51 Baade A, Peng P Y, Harwath D. MAE-AST: Masked autoencoding audio spectrogram Transformer. arXiv preprint arXiv: 2203.16691, 2022.

52 Ghosh S, Seth A, Umesh S, Manocha D. MAST: Multiscale audio spectrogram Transformers. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Rhodes Island, Greece: IEEE, 2023. 1−5

53 Domingos P. A few useful things to know about machine learning. *Communications of the ACM*, 2012, **55**(10): 78−87

54 Kelly J G, Carpenter R N, Tague J A. Object classification and acoustic imaging with active sonar. *The Journal of the Acoustical Society of America*, 1992, **91**(4): 2073−2081

55 Shi H, Xiong J Y, Zhou C Y, Yang S. A new recognition and classification algorithm of underwater acoustic signals based on multi-domain features combination. In: Proceedings of the IEEE/OES China Ocean Acoustics (COA). Harbin, China: IEEE, 2016. 1−7

56 Li H T, Cheng Y S, Dai W G, Li Z Z. A method based on wavelet packets-fractal and SVM for underwater acoustic signals recognition. In: Proceedings of the 12th International Conference on Signal Processing (ICSP). Hangzhou, China: IEEE, 2014. 2169−2173

57 Sherin B M, Supriya M H. SOS based selection and parameter optimization for underwater target classification. In: Proceedings of the OCEANS 2016 MTS/IEEE Monterey. Monterey, USA: IEEE, 2016. 1−4

58 Lian Z X, Xu K, Wan J W, Li G. Underwater acoustic target classification based on modified GFCC features. In: Proceedings of the IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). Chongqing, China: IEEE, 2017. 258−262

59 Lian Z X, Xu K, Wan J W, Li G, Chen Y. Underwater acoustic target recognition based on Gammatone filterbank and instantaneous frequency. In: Proceedings of the IEEE 9th International Conference on Communication Software and Networks (ICCSN). Guangzhou, China: IEEE, 2017. 1207−1211

60 Feroze K, Sultan S, Shahid S, Mahmood F. Classification of underwater acoustic signals using multi-classifiers. In: Proceedings of the 15th International Bhurban Conference on Applied Sciences and Technology (IBCAST). Islamabad, Pakistan: IEEE, 2018. 723−728

61 Li Y X, Chen X, Yu J, Yang X H. A fusion frequency feature extraction method for underwater acoustic signal based on variational mode decomposition, duffing chaotic oscillator and a kind of permutation entropy. *Electronics*, 2019, **8**(1): Article No. 61

62 Aksüren İ G, Hocaoğlu A K. Automatic target classification using underwater acoustic signals. In: Proceedings of the 30th Signal Processing and Communications Applications Conference (SIU). Safranbolu, Turkey: IEEE, 2022. 1−4

63 Kim T, Bae K. HMM-based underwater target classification with synthesized active sonar signals. In: Proceedings of the 19th European Signal Processing Conference. Barcelona, Spain: IEEE, 2011. 1805−1808

64 Li H F, Pan Y, Li J Q. Classification of underwater acoustic target using auditory spectrum feature and SVDD ensemble. In: Proceedings of the OCEANS-MTS/IEEE Kobe Techno-Oceans (OTO). Kobe, Japan: IEEE, 2018. 1−4

65 Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks. *Communication of the ACM*, 2017, **60**(6): 84−90

66 Yue H, Zhang L L, Wang D Z, Wang Y X, Lu Z Q. The classification of underwater acoustic targets based on deep learning methods. In: Proceedings of the 2nd International Conference on Control, Automation and Artificial Intelligence (CAAI 2017). Hainan, China: Atlantis Press, 2017. 526−529

67 Kirsebom O S, Frazao F, Simard Y, Roy N, Matwin S, Giard S. Performance of a deep neural network at detecting North Atlantic right whale upcalls. *The Journal of the Acoustical Society of America*, 2020, **147**(4): 2636−2646

68 Yin X H, Sun X D, Liu P S, Wang L, Tang R C. Underwater acoustic target classification based on LOFAR spectrum and convolutional neural network. In: Proceedings of the 2nd International Conference on Artificial Intelligence and Advanced Manufacture (AIAM). Manchester, United Kingdom: ACM, 2020. 59−63

69 Jiang J J, Shi T, Huang M, Xiao Z Z. Multi-scale spectral feature extraction for underwater acoustic target recognition. *Measurement*, 2020, **166**: Article No. 108227

70 Cao X, Togneri R, Zhang X M, Yu Y. Convolutional neural network with second-order pooling for underwater target classification. *IEEE Sensors Journal*, 2019, **19**(8): 3058−3066

71 Hu G, Wang K J, Peng Y, Qiu M R, Shi J F, Liu L L. Deep learning methods for underwater target feature extraction and recognition. *Computational Intelligence and Neuroscience*, 2018, **2018**: Article No. 1214301

72 Zheng Y L, Gong Q Y, Zhang S F. Time-frequency feature-based underwater target detection with deep neural network in shallow sea. *Journal of Physics: Conference Series*, 2021, **1756**: Article No. 012006

73 Irfan M, Zheng J B, Ali S, Iqbal M, Masood Z, Hamid U. DeepShip: An underwater acoustic benchmark dataset and a separable convolution based autoencoder for classification. *Expert Systems With Applications*, 2021, **183**: Article No. 115270

74 Wang B, Zhang W, Zhu Y N, Wu C X, Zhang S Z. An underwater acoustic target recognition method based on AMNet. *IEEE Geoscience and Remote Sensing Letters*, 2023, **20**: Article No. 5501105

75 Li Y X, Gao P Y, Tang B Z, Yi Y M, Zhang J J. Double feature extraction method of ship-radiated noise signal based on slope entropy and permutation entropy. *Entropy*, 2022, **24**(1): Article No. 22

76 Hong F, Liu C W, Guo L J, Chen F, Feng H H. Underwater acoustic target recognition with a residual network and the optimized feature extraction method. *Applied Sciences*, 2021, **11**(4): Article No. 1442

77 Chen S H, Tan X L, Wang B, Lu H C, Hu X L, Fu Y. Reverse attention-based residual network for salient object detection. *IEEE Transactions on Image Processing*, 2020, **29**: 3763−3776

78 Lu Z Y, Xu B, Sun L, Zhan T M, Tang S Z. 3-D channel and spatial attention based multiscale spatial-spectral residual network for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2020, **13**: 4311−4324

79 Fan R Y, Wang L Z, Feng R Y, Zhu Y Q. Attention based residual network for high-resolution remote sensing imagery scene classification. In: Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Yokohama,

Japan: IEEE, 2019. 1346−1349

80 Tripathi A M, Mishra A. Environment sound classification using an attention-based residual neural network. *Neurocomputing*, 2021, **460**: 409−423

81 Liu C W, Hong F, Feng H H, Hu M L. Underwater acoustic target recognition based on dual attention networks and multiresolution convolutional neural networks. In: Proceedings of the OCEANS 2021: San Diego-Porto. San Diego, USA: IEEE, 2021. 1−5

82 Xue L Z, Zeng X Y, Jin A Q. A novel deep-learning method with channel attention mechanism for underwater target recognition. *Sensors*, 2022, **22**(15): Article No. 5492

83 Li J, Wang B X, Cui X R, Li S B, Liu J H. Underwater acoustic target recognition based on attention residual network. *Entropy*, 2022, **24**(11): Article No. 1657

84 Park D S, Chan W, Zhang Y, Chiu C C, Zoph B, Cubuk E D, et al. SpecAugment: A simple data augmentation method for automatic speech recognition. arXiv preprint arXiv: 1904.08779, 2019.

85 Luo X W, Zhang M H, Liu T, Huang M, Xu X G. An underwater acoustic target recognition method based on spectrograms with different resolutions. *Journal of Marine Science and Engineering*, 2021, **9**(11): Article No. 1246

86 Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, inception-ResNet and the impact of residual connections on learning. arXiv preprint arXiv: 1602.07261, 2016.

87 Ren J W, Huang Z Q, Li C, Guo X Y, Xu J. Feature analysis of passive underwater targets recognition based on deep neural network. In: Proceedings of the OCEANS 2019-Marseille. Marseille, France: IEEE, 2019. 1−5

88 Ren J W, Xie Y, Zhang X W, Xu J. UALF: A learnable front-end for intelligent underwater acoustic classification system. *Ocean Engineering*, 2022, **264**: Article No. 112394

89 Li S C, Yang S Y, Liang J H. Recognition of ships based on vector sensor and bidirectional long short-term memory networks. *Applied Acoustics*, 2020, **164**: Article No. 107248

90 Wang Y, Zhang H, Xu L W, Cao C H, Gulliver T A. Adoption of hybrid time series neural network in the underwater acoustic signal modulation identification. *Journal of the Franklin Institute*, 2020, **357**(18): 13906−13922

91 Qi P Y, Sun J G, Long Y F, Zhang L G, Tian Y. Underwater acoustic target recognition with fusion feature. In: Proceedings of the 28th International Conference on Neural Information Processing. Sanur, Indonesia: Springer, 2021. 609−620

92 Kamal S, Chandran C S, Supriya M H. Passive sonar automated target classifier for shallow waters using end-to-end learnable deep convolutional LSTMs. *Engineering Science and Technology, an International Journal*, 2021, **24**(4): 860−871

93 Han X C, Ren C X, Wang L M, Bai Y J. Underwater acoustic target recognition method based on a joint neural network. *PLoS ONE*, 2022, **17**(4): Article No. e0266425

94 Li P, Wu J, Wang Y X, Lan Q, Xiao W B. STM: Spectrogram Transformer model for underwater acoustic target recognition. *Journal of Marine Science and Engineering*, 2022, **10**(10): Article No. 1428

95 Feng S, Zhu X Q. A Transformer-based deep learning network for underwater acoustic target recognition. *IEEE Geoscience and Remote Sensing Letters*, 2022, **19**: Article No. 1505805

96 Ying J J C, Lin B H, Tseng V S, Hsieh S Y. Transfer learning on high variety domains for activity recognition. In: Proceedings of the ASE BigData & SocialInformatics. Taiwan, China: ACM, 2015. 1−6

97 Zhang Y K, Guo X S, Leung H, Li L. Cross-task and cross-domain SAR target recognition: A meta-transfer learning approach. *Pattern Recognition*, 2023, **138**: Article No. 109402

98 Zhang L L, Wang D Z, Bao C C, Wang Y X, Xu K L. Large-scale whale-call classification by transfer learning on multi-scale waveforms and time-frequency features. *Applied Sciences*, 2019, **9**(5): Article No. 1020

99 Zhong M, Castellote M, Dodhia R, Ferres J L, Keogh M, Brewer A. Beluga whale acoustic signal classification using deep learning neural network models. *The Journal of the Acoustical Society of America*, 2020, **147**(3): 1834−1841

100 Hitawala S. Evaluating ResNeXt model architecture for image classification. arXiv preprint arXiv: 1805.08700, 2018.

101 Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016. 2818−2826

102 Kong Q Q, Cao Y, Iqbal T, Wang Y X, Wang W W, Plumbley M D. PANNs: Large-scale pretrained audio neural networks for audio pattern recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2020, **28**: 2880−2894

103 Li D H, Liu F, Shen T S, Chen L, Yang X D, Zhao D X. Generalizable underwater acoustic target recognition using feature extraction module of neural network. *Applied Sciences*, 2022, **12**(21): Article No. 10804

104 Kamal S, Mohammed S K, Pillai P R S, Supriya M H. Deep learning architectures for underwater target recognition. In: Proceedings of the Ocean Electronics (SYMPOL). Kochi, India: IEEE, 2013. 48−54

105 Seok J. Active sonar target classification using multi-aspect sensing and deep belief networks. *International Journal of Engineering Research and Technology*, 2018, **11**: 1999−2008

106 Yang Hong-Hui, Shen Sheng, Yao Xiao-Hui, Han Zhen. Hybrid regularized deep belief network for underwater acoustic target feature learning and recognition. *Journal of Northwestern Polytechnical University*, 2017, **35**(2): 220−225
(杨宏晖, 申昇, 姚晓辉, 韩振. 用于水声目标特征学习与识别的混合正则化深度置信网络. 西北工业大学学报, 2017, **35**(2): 220−225)

107 Yang H H, Shen S, Yao X H, Sheng M P, Wang C. Competitive deep-belief networks for underwater acoustic target recognition. *Sensors*, 2018, **18**(4): Article No. 952

108 Shen S, Yang H H, Sheng M P. Compression of a deep competitive network based on mutual information for underwater acoustic targets recognition. *Entropy*, 2018, **20**(4): Article No. 243

109 Cao X, Zhang X M, Yu Y, Niu L T. Deep learning-based recognition of underwater target. In: Proceedings of the IEEE International Conference on Digital Signal Processing (DSP). Beijing, China: IEEE, 2016. 89−93

110 Luo X W, Feng Y L. An underwater acoustic target recognition method based on restricted Boltzmann machine. *Sensors*, 2020, **20**(18): Article No. 5399

111 Luo X W, Feng Y L, Zhang M H. An underwater acoustic target recognition method based on combined feature with automatic coding and reconstruction. *IEEE Access*, 2021, **9**: 63841−63854

112 Sun B G, Luo X W. Underwater acoustic target recognition based on automatic feature and contrastive coding. *IET Radar, Sonar & Navigation*, 2023, **17**(8): 1277−1285

113 You K, Xu K L, Feng M, Zhu B Q. Underwater acoustic classification using masked modeling-based swin Transformer. *The Journal of the Acoustical Society of America*, 2022, **152**(4_Supplement): Article No. 296

114 Xu K L, Xu Q S, You K, Zhu B Q, Feng M, Feng D W, et al. Self-supervised learning-based underwater acoustical signal classification via mask modeling. *The Journal of the Acoustical Society of America*, 2023, **154**(1): 5−15

115 Le H T, Phung S L, Chapple P B, Bouzerdoum A, Ritz C H, Tran L C. Deep Gabor neural network for automatic detection of mine-like objects in sonar imagery. *IEEE Access*, 2020, **8**: 94126−94139

116 Huo G Y, Wu Z Y, Li J B. Underwater object classification in sidescan sonar images using deep transfer learning and semisynthetic training data. *IEEE Access*, 2020, **8**: 47407−47418

117 Berg H, Hjelmervik K T. Classification of anti-submarine warfare sonar targets using a deep neural network. In: Proceedings

of the OCEANS 2018 MTS/IEEE Charleston. Charleston, USA: IEEE, 2018. 1–5

118 Santos-Domínguez D, Torres-Guijarro S, Cardenal-López A, Pena-Gimenez A. ShipsEar: An underwater vessel noise database. *Applied Acoustics*, 2016, **113**: 64–69

119 Liu Mei-Qin, Han Xue-Yan, Zhang Sen-Lin, Zheng Rong-Hao, Lan Jian. Research status and prospect of target tracking technologies via underwater sensor networks. *Acta Automatic Sinica*, 2021, **47**(2): 235–251
(刘妹琴, 韩学艳, 张森林, 郑荣濠, 兰剑. 基于水下传感器网络的目标跟踪技术研究现状与展望. 自动化学报, 2021, **47**(2): 235–251)

120 Xu K L, Zhu B Q, Kong Q Q, Mi H B, Ding B, Wang D Z, et al. General audio tagging with ensembling convolutional neural networks and statistical features. *The Journal of the Acoustical Society of America*, 2019, **145**(6): EL521–EL527

121 Zhu B Q, Xu K L, Kong Q Q, Wang H M, Peng Y X. Audio tagging by cross filtering noisy labels. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2020, **28**: 2073–2083

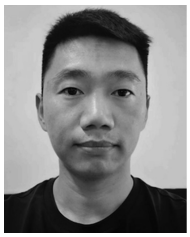
徐齐胜 国防科技大学计算机学院硕士研究生. 2021 年获得武汉大学学士学位. 主要研究方向为音频信号处理, 并行计算.
E-mail: qishengxu@nudt.edu.cn
(**XU Qi-Sheng** Master student at the School of Computer Science, National University of Defense Technology. He received his bachelor degree from Wuhan University in 2021. His research interest covers audio signal processing and parallel computing.)

许可乐 国防科技大学计算机学院副教授. 2017 年获得法国巴黎六大博士学位. 主要研究方向为音频信号处理, 机器学习和智能软件系统. 本文通信作者.
E-mail: xukelele@163.com
(**XU Ke-Le** Associate professor at the School of Computer Science, National University of Defense Technology. He received his Ph.D. degree from Paris VI University in 2017. His research interest covers audio signal processing, machine learning, and intelligent software systems. Corresponding author of this paper.)

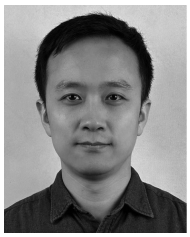
窦 勇 国防科技大学并行与分布处理国防科技重点实验室教授. 1995 年获得国防科技大学博士学位. 主要研究方向为高性能计算, 智能计算, 机器学习和深度学习.
E-mail: yongdou@nudt.edu.cn
(**DOU Yong** Professor at the National Key Laboratory of Parallel and Distributed Processing, National University of Defense Technology. He received his Ph.D. degree from National University of Defense Technology in 1995. His research interest covers high performance computing, intelligence computing, machine learning, and deep learning.)

高彩丽 国防科技大学计算机学院硕士研究生. 2021 年获得南昌大学学士学位. 主要研究方向为人脸伪造检测, 并行优化.
E-mail: gaocl@nudt.edu.cn
(**GAO Cai-Li** Master student at the School of Computer Science, National University of Defense Technology. He received his bachelor degree from Nanchang University in 2021. His research interest covers face forgery detection and parallel optimization.)

乔 鹏 国防科技大学并行与分布处理国防科技重点实验室助理研究员. 2018 年获得国防科技大学博士学位. 主要研究方向为高性能计算, 图像恢复和深度强化学习.
E-mail: pengqiao@nudt.edu.cn
(**QIAO Peng** Assistant researcher at the National Key Laboratory of Parallel and Distributed Processing, National University of Defense Technology. He received his Ph.D. degree from National University of Defense Technology in 2018. His research interest covers high performance computing, image restoration, and deep reinforcement learning.)



冯大为 国防科技大学计算机学院副教授. 2014 年获得法国巴黎第十一大学博士学位. 主要研究方向为分布计算与智能软件系统.
E-mail: dafeng@nudt.edu.cn
(**FENG Da-Wei** Associate professor at the School of Computer Science, National University of Defense Technology. He received his Ph.D. degree from Paris-Sud University in 2014. His research interest covers distributed computing and intelligent software systems.)



朱博青 国防科技大学博士研究生. 2019 年获得国防科技大学硕士学位. 主要研究方向为多模态机器学习, 持续学习和计算声学.
E-mail: zhuboq@gmail.com
(**ZHU Bo-Qing** Ph.D. candidate at the School of Computer Science, National University of Defense Technology. He received his master degree from National University of Defense Technology in 2019. His research interest covers multi-modal machine learning, continual learning, and computational acoustics.)