imts Intelligent Marine
Technology and Systems

Check for updates

Advances and applications of machine learning in underwater acoustics

Haiqiang Niu[1][*] , Xiaolei Li[2], Yonglin Zhang[1] and Ji Xu[1]

Abstract

Recent advancements in machine learning (ML) techniques applied to underwater acoustics have significantly impacted various aspects of this field, such as source localization, target recognition, communication, and geoacoustic inversion. This review provides a comprehensive summary and evaluation of these developments. As a data-driven approach, ML played a pivotal role in discerning intricate relationships between input features and desired labels based on the provided training dataset. They are achieving success in ocean acoustic applications through ML hinges on several critical factors, including well-designed input feature preprocessing, appropriate labels, choice of ML models, effective training strategy, and availability of ample training and validation datasets. This review highlights noteworthy results from published studies to illustrate the effectiveness of ML methods in diverse application scenarios. In addition, it delves into the essential techniques employed within these applications. To understand the utility of ML in underwater acoustics, one must analyze its advantages and limitations. This assessment will aid in identifying scenarios where ML excels and those where it may face challenges. In addition, it provides insights into promising avenues for future research, shedding light on potential research directions that warrant exploration.

Keywords Machine learning, Underwater acoustics, Source localization, Target recognition, Communication, Geoacoustic inversion

1 Introduction

In recent years, machine learning (ML), particularly deep learning (DL), has achieved remarkable breakthroughs in various fields, including image processing and speech recognition. ML has garnered widespread attention and witnessed significant development across different natural science domains. Bianco et al. (2019) provided a comprehensive overview of ML applications in multiple acoustic environments. In contrast, Niu et al. (2019a, b) examined ML techniques explicitly applied to underwater source localization. In contrast to prior research

efforts, this study takes a distinct approach by delivering an extensive review and summarization of the advancements and distinctive features of ML methodologies in various noteworthy underwater acoustic applications from recent years. Within underwater acoustics, ML has subdomain applications such as source localization, target recognition, communication, geoacoustic inversion, direction-of-arrival estimation, and line spectrum enhancement. This study primarily concentrates on the first four underwater acoustic challenges, offering a comprehensive summary of the research landscape, data preprocessing methods, ML models, learning strategies, dataset characteristics, and other pertinent aspects based on existing literature. In Section 2 through 5, we review and analyze source localization, target recognition, communication, and geoacoustic inversion, respectively.

Following this analysis, we will expound on the potential benefits of employing ML in underwater acoustics. In addition, we outline the primary constraints and

*Correspondence:
Haiqiang Niu
nhq@mail.ioa.ac.cn
[1] Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China
[2] Department of Marine Technology, Ocean University of China, Qingdao 266100, China

obstacles that this integration encounters. Considering the evolution of ML techniques and the distinctive attributes of underwater acoustic scenarios, Section 6 provides a set of prospective research avenues for advancing the field of underwater acoustics through ML applications.

2 Source localization

In underwater environments, when the sound source is located in different locations, the resulting sound field received varies, enabling the utilization of the received sound field for passive sound source localization. This process involves establishing a mapping relationship between the received sound field and the sound source's location. Unquestionably, this mapping can be approximated by ML models when a significant amount of labeled training data is accessible.

ML has been applied to passive sound source localization since the early 1990s. An example of such early work can be found in Steinberg et al. (1991). These authors employed neural network techniques to localize an acoustic point source within a homogeneous medium. This pioneering study demonstrates the early adoption of ML methods for passive sound source localization tasks. In the same year, Ozard et al. (1991) applied associative feedforward neural networks with no hidden layers to localize a source in range and depth using the acoustic signal arriving at a vertical array of sensors. Although the number of hidden layers of neural networks used was no more than one, Steinberg et al. (1991) found a general characteristic of supervised learning methods: good interpolation ability and poor extrapolation ability. However, limited by the hardware capabilities and algorithms available then, ML methods faced challenges in dealing with source localization problems in realistic ocean environments. Moreover, matched-field processing (MFP) stood as the prevailing passive localization algorithm and was undergoing rapid development during that period. Consequently, ML methods received little attention in underwater acoustics for an extended duration after that. Although MFP-related methods have made significant progress after decades of development and have been widely used in relevant engineering practices, they still encounter numerous difficulties and challenges in real-world applications, such as environmental mismatch problems. It is worth noting that ocean waveguides exhibit intricate time-varying and space-varying characteristics, and precise measurement-based determination of the ocean environment parameters is challenging. Consequently, achieving accurate modeling of the ocean environment is a formidable task. While the environmental mismatch issue in MFP can be mitigated by incorporating the uncertainty of environmental parameters, such as environmental focusing (Collins and Kuperman
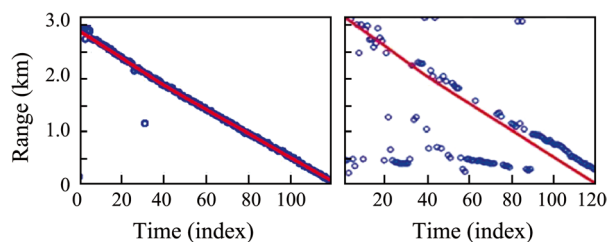
Fig. 1 The figure on the left illustrates range predictions on Test-Data-1 by RF method for the frequency range of 300–950 Hz with a 10 Hz increment. Meanwhile, the figure on the right depicts the range predictions generated by Bartlett matched-field processing using the same dataset. The red lines in both figures correspond to the GPS-derived results. (Adapted from Niu et al. 2017b)

1991; Gerstoft 1994; Gingras and Gerstoft 1995), Bayesian tracking (Dosso and Wilmut 2008, 2009), and stochastic matched-field localization (Finette and Mignerey 2018), these approaches often involve high computational costs, impeding real-time processing capabilities.

Owing to the rapid advancement of computer hardware and ML theory, ML-related methods in underwater source localization have experienced a resurgence. This resurgence has also opened up a new avenue for addressing the environmental mismatch problem in MFP. Lefort et al. (2017) studied the localization performance of a nonlinear regression algorithm in fluctuating ocean environments using data from water tank experiments. The results demonstrate the advantages and potential of ML algorithms in the context of underwater source localization. Simultaneously, Niu et al. (2017a, b) introduced a practical class of ML-based underwater source localization methods. They systematically analyzed the sound source localization performance of three ML models: feedforward neural network (FNN), support vector machines (SVM), and random forest (RF), based on the Noisy09 experiment dataset. This research marked a significant milestone as it systematically verified the feasibility of employing ML for underwater source ranging using sea trial data. Niu et al. (2017a, b) studies demonstrate that an underwater source localization model trained directly using measured sound field data in test waters can effectively alleviate the environmental mismatch problem. Please refer to Fig. 1 for a visual representation of the findings.

Wang and Peng (2018) trained a generalized regression neural network (GRNN) and an FNN for sound source ranging, utilizing a portion of the data from the SWellEx-96 experiment as the training dataset. The outcomes reveal that both the GRNN and FNN exhibit a commendable localization performance surpassing that of MFP, as depicted in Fig. 2. This outcome provides further evidence that the environmental mismatch problem

Niu *et al. Intelligent Marine Technology and Systems* (2023) 1:8

Page 3 of 23

can be substantially mitigated by incorporating measured data from the test waters as training samples.

While ML models can be trained using experimental data, there is a limitation due to the scarcity of ocean acoustic experimental data that includes source location labels. This scarcity makes training ML models for ocean sound source localization cumbersome. Huang et al. (2018) employed numerical simulations to generate synthetic training data to address this issue. They noted that the simulation data can be effectively incorporated to enhance performance when experimental data are insufficient, as long as the test environment aligns with the simulation data. The data processing outcomes from the Yellow Sea experiment support this assertion. In this experiment, only simulation data were employed to train a deep neural network (DNN) for source localization,

methods. Liu et al. (2020a, b) introduced a novel multi-task learning (MTL) approach, incorporating adaptively weighted losses within a convolutional neural network (CNN) for source localization in deep-ocean environments. Simulation results and tests conducted using real data from the South China Sea experiment demonstrate that, compared to conventional MFP, CNN with MTL exhibits superior performance and increased robustness, particularly in scenarios involving array tilt within the deep-ocean environment (as depicted in Fig. 4). Importantly, because of the offline nature of the training process, ML models offer improved real-time performance compared with environmental focusing and Bayesian tracking.

In ocean areas with limited environmental data, there is a need for both measured acoustic data and suitable
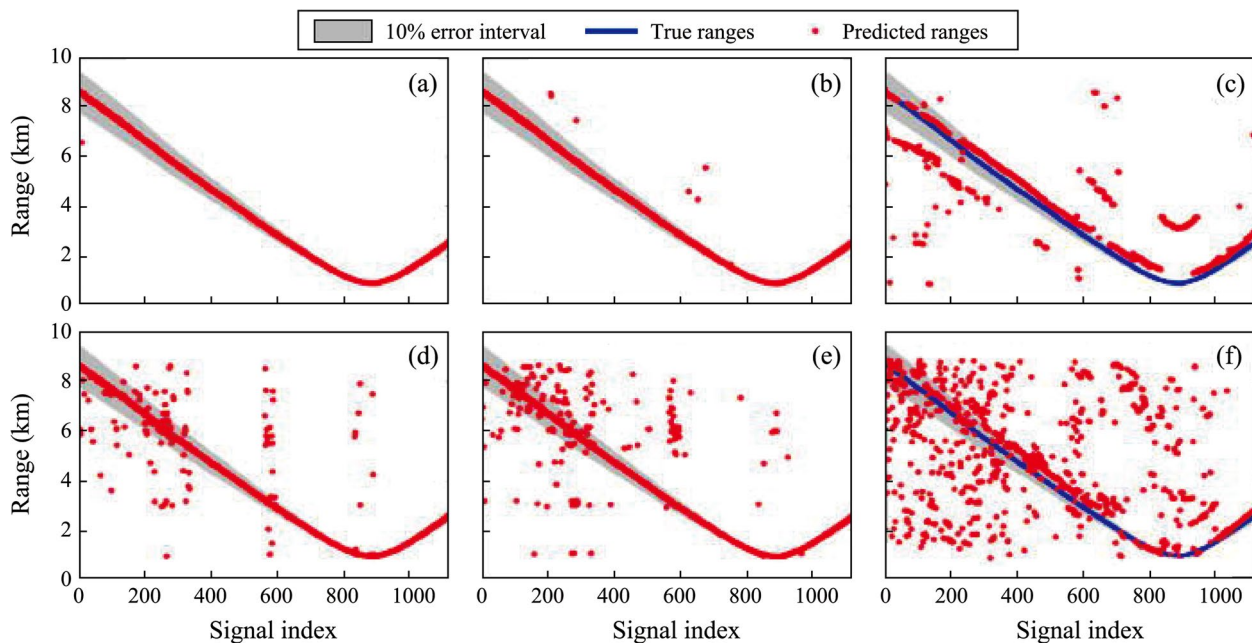

Fig. 2 Localization results were obtained from the complete 75 min dataset using GRNN, FNN, and MFP. The top section illustrates the narrowband scenario at 232 Hz for the shallow source, while the bottom section depicts the narrowband scenario at 338 Hz for the deep source. In both scenarios, **a** and **d** represent the results obtained with GRNN; **b** and **e** correspond to the outcomes achieved with FNN; and **c** and **f** indicate the results obtained through MFP. (Adapted from Wang and Peng 2018)

and the results, as shown in Fig. 3, reveal that the source-ranging performance of the DNN surpasses that of MFP.

Similar to environmental focusing (Collins and Kuperman 1991; Gerstoft 1994; Gingras and Gerstoft 1995; Collins and Kuperman 1991) and Bayesian tracking (Dosso and Wilmut 2008, 2009), the robustness of ML ranging and localization models can be significantly enhanced by considering the distribution of environmental parameters (Niu et al. 2019a, b; Liu et al. 2020a, b) during the preparation of training data using numerical

environmental models to generate an extensive set of accurately labeled training data for ML models. Wang et al. (2019a) employed deep transfer learning (DTL) for sources ranging in uncharted deep-sea regions to address this challenge. DTL facilitates the transfer of predictive capabilities from a trained DNN to a new, similar environment by sharing some DNN parameters while relearning others. Within the framework of DTL, Wang et al. (2019a) initially trained a pretrained CNN using replicated sound field data generated from historical

Niu *et al. Intelligent Marine Technology and Systems*   (2023) 1:8

Page 4 of 23
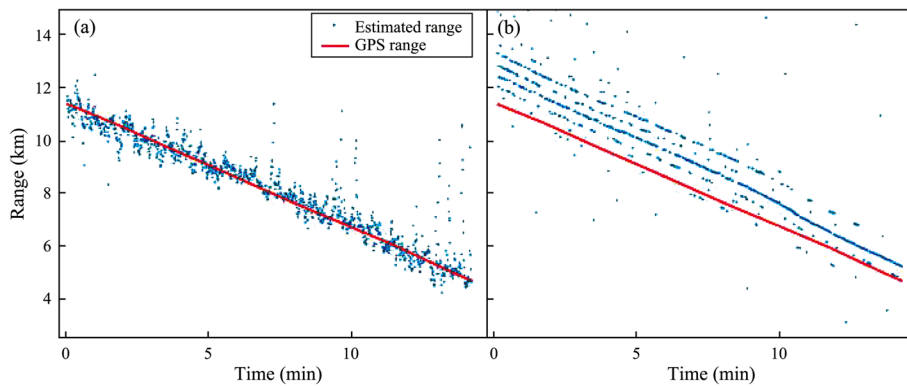
Fig. 3 Source ranging based on experimental data. **a** The results were derived from a DNN trained using simulated data at a water depth of 35.5 and 36 m and **b** the results generated by the MFP technique, especially applied to a water depth of 36 m. (Adapted from Huang et al. 2018)

environment information. Subsequently, they fine-tuned specific parameters of the CNN using a limited dataset collected at sea for source-ranging purposes. Although DTL has exhibited promise in improving ranging performance in data-poor regions, it encounters challenges when labeled acoustic field data are unavailable for such areas. An alternative approach to enhance the ranging performance of ML models in unfamiliar environments is to bolster their generalization capabilities. Taking the

FNN as an example, it is well established that the generalization ability of FNNs can be improved by applying the early stopping technique. A fundamental concern revolves around determining the optimal stopping point during FNN training to ensure optimal ranging performance in the testing environment. Chi et al. (2019) introduced a fitting-based early stopping (FEAST) method to evaluate the FNN's ranging error on test data where the source-to-receiver distance is unknown. The core
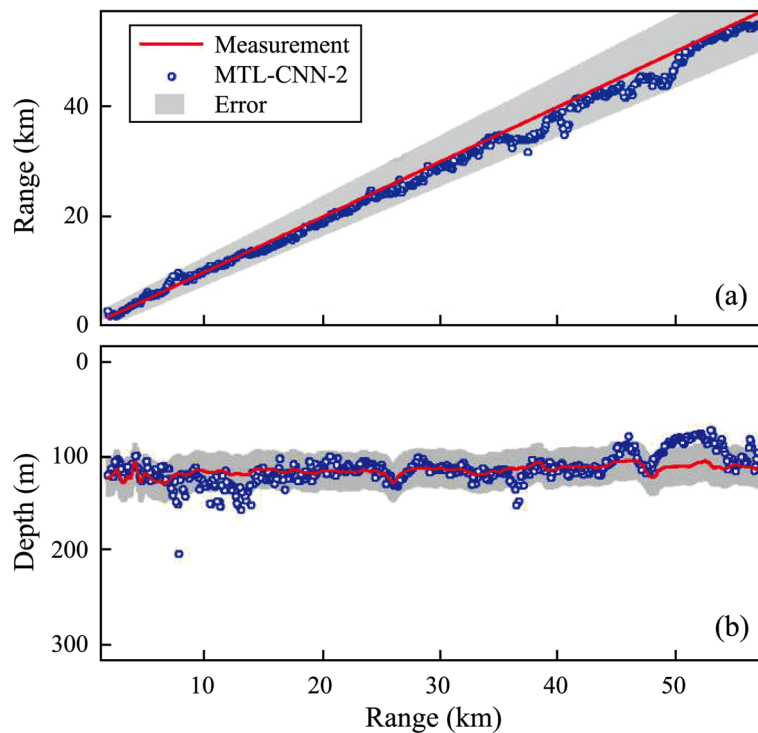
Fig. 4 The MTL–CNN-2 model's predictions for both ranges **a** and depths **b** are based on real data obtained from the South China Sea experiment. (Adapted from Liu et al. 2020b)

concept of FEAST is as follows: In the testing environment, testing data samples are sequentially fed into the FNN based on their chronological order. The FNN output results are then fitted with a simple curve on the time-distance plane. Assuming that the source trajectory adheres to the constraints of a simple curve, the deviation between the FNN's output results and the fitted curve indicates the FNN's ranging error. Using FEAST, training is halted when the evaluated ranging error reaches its minimum on the test data. The effectiveness of FEAST is demonstrated using data from the SWellEx-96 experiment.

Previous research on ML-based ranging has focused on range-independent ocean waveguides, with limited exploration into range-dependent scenarios. In contrast to range-independent waveguides, generating training data for range-dependent waveguides using numerical methods poses more significant challenges. On one hand, computing the sound field in a range-dependent ocean waveguide is time-consuming. However, describing diverse range-dependent waveguides with finite parameters is a complex task. Li et al. (2020b) introduced a novel random mode-coupling matrix model to address this challenge. This model was designed to facilitate training data generation for range-dependent waveguides. The proposed model was applied to recover Acoustic Interference Striations (AISs) within a nonlinear internal wave environment using a U-Net, as illustrated in Fig. 5. The random mode-coupling matrix model uses random sampling to construct the mode-coupling matrix, combining the mathematical framework of the mode-coupling matrix with statistical principles. Consequently, the preparation speed of training data for the random mode-coupling matrix model significantly outperforms traditional simulation methods.

Continued advancements in ML application to source localization have led to significant progress in recent years. Researchers have successfully used ML methods for single hydrophone-based source localization (Niu et al. 2019a, b; Liu et al. 2021c; Goldwater et al. 2023) as well as for the localization of multiple sources (Liu et al. 2021d). Liu et al. (2021d) introduced the application of a gated feedback recurrent unit network (GFGRU) for multiple source localization within the direct arrival zone of the deep ocean. The results indicate that GFGRU exhibits behavior similar to that of SBL and offers modest improvements in localization performance compared with Bartlett MFP and FNN, particularly in scenarios involving array tilt mismatch. In a real experimental dataset collected in the South China Sea, GFGRU, unlike Bartlett MFP, demonstrates reduced ambiguity

in multisource localization and effectively distinguishes between two closely spaced sources, as illustrated in Fig. 6.

Some studies (Van Komen et al. 2020; Neilsen et al. 2021) have also explored using time series or long-time time-frequency spectrograms as input features to estimate source locations and seedbed types concurrently. Furthermore, researchers have applied ML methods to estimate modal wavenumbers (Niu et al. 2020; Li et al. 2023b), which can be employed for source localization. A summary of the ML-based source localization methods is provided in Table 1.

3 Target recognition
3.1 Background
Underwater acoustic target recognition is a vital element in underwater acoustics. Its primary objective is to identify underwater targets by analyzing their emitting sounds (Yang et al. 2020). This technology has broad utility in automating maritime traffic monitoring, identifying noise sources in ocean environmental monitoring systems, and enhancing security defense measures.

Underwater acoustic target recognition presents a formidable challenge, often accompanied by numerous practical obstacles (Dong et al. 2021; Xie et al. 2022a; Zhang et al. 2022b). Various factors, including intricate underwater environments, unpredictable transmission channels, and the volatile motion states of vessels, compound the complexity of analyzing underwater acoustic signals. The manual recognition of underwater acoustic features and targets requires significant human effort, which poses limitations in meeting practical demands (Xie et al. 2022a). Furthermore, discriminative patterns may exist in the data that are not easily discernible by human cognition (Bianco et al. 2019). Consequently, the emphasis of research has gradually shifted toward automatic underwater acoustic target recognition.

The automatic underwater acoustic target recognition system follows the paradigm of acoustic pattern recognition tasks and primarily comprises three key components: preprocessing, acoustic feature extraction, and the recognition module. Preprocessing strategies are employed to amplify target signals and mitigate irrelevant interference, thus enhancing the accuracy and robustness of the recognition system. Subsequently, acoustic feature extraction methods transform the processed signals into informative and low-dimensional acoustic features. Recognition models that leverage statistical methods, linear or nonlinear classifiers, or neural networks extract knowledge from these input features and predict potential underwater targets. Notably, ML techniques have ushered in significant advancements in
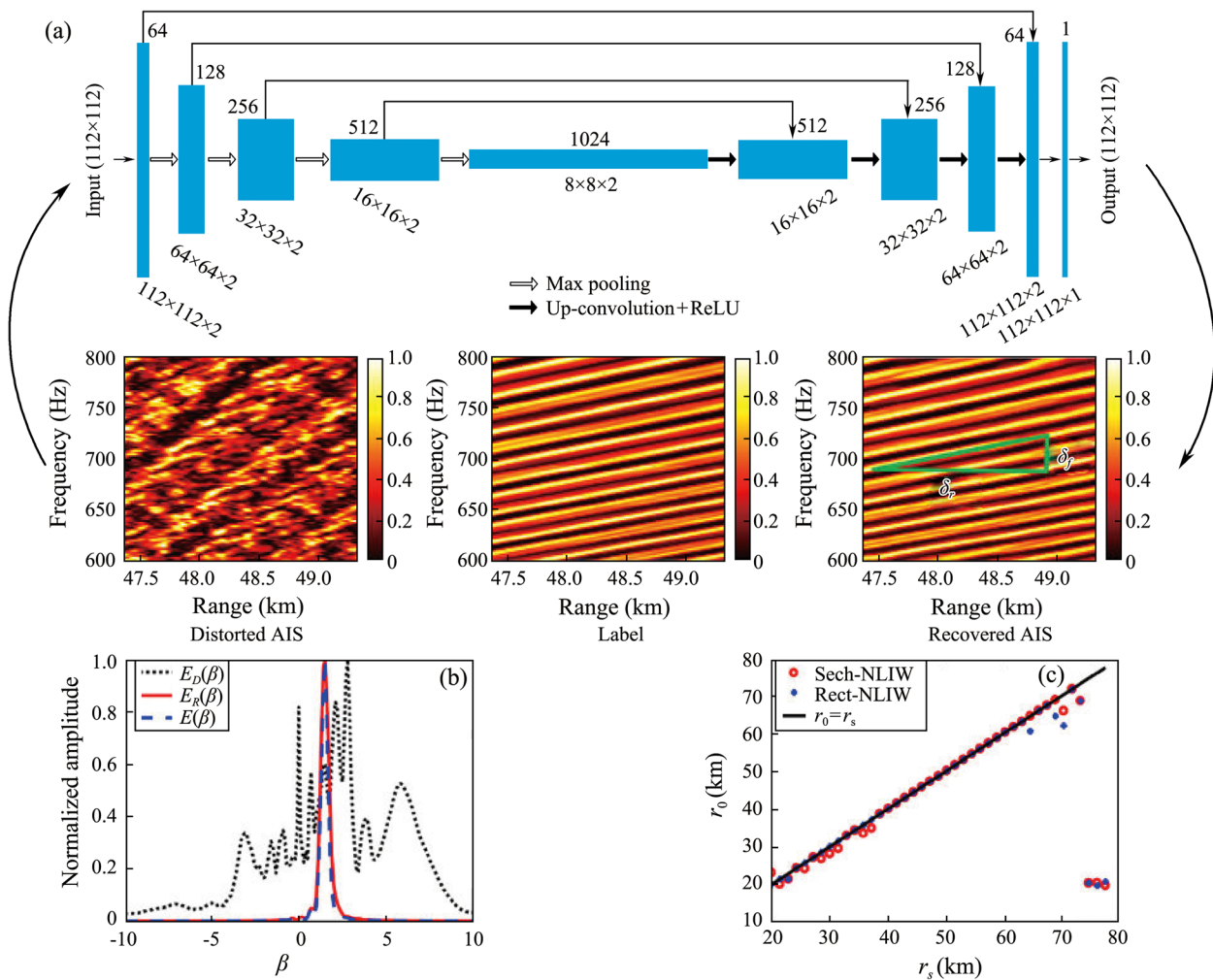
Fig. 5 **a** The input to the U-Net is a distorted AIS, while its output is the recovered AIS. The undistorted AIS serves as the corresponding label. Each box in the U-Net represents convolution layer(s), with the number of channels indicated at the top of the box. The *x-y*-size and the number of convolution layers are provided at the lower edge of the box. **b** Normalized b distributions of the distorted AIS, the recovered AIS, and the label. **c** Ranging results were obtained using the recovered AIS. Circles represent varying results for the Sech–NLIW case, while crosses represent varying results for the Rech–NLIW case. (Adapted from Li et al. 2020b)

automating preprocessing, intelligent feature extraction, and enhancement of pattern recognition capabilities.

In recent years, the development of ML algorithms and the accumulation of extensive databases have catalyzed a surge in research focused on automatic underwater acoustic target recognition. Researchers have dedicated significant efforts to creating automated systems that are both reliable and robust. Research in this field can be categorized into several directions. Some studies aim to optimize preprocessing algorithms to address background noise, signal interference, low signal-to-noise ratio (SNR), and limited data quantity (Zhou and Yang 2020; Dong et al. 2021). Others concentrate on developing intelligent feature extraction methods tailored to the unique characteristics of underwater acoustic signals (Jiang et al. 2020, 2021). Specific investigations are dedicated to constructing adaptive and accurate recognition models capable of effectively discerning underwater signals (Zhang et al. 2022b). In addition, some studies have focused on differentiating surface and underwater acoustic targets based on acoustic field characteristics rather than source features (Zhang et al. 2022a; Yu et al. 2023). In the following sections, we provide a comprehensive overview of the relevant scientific research in this domain.
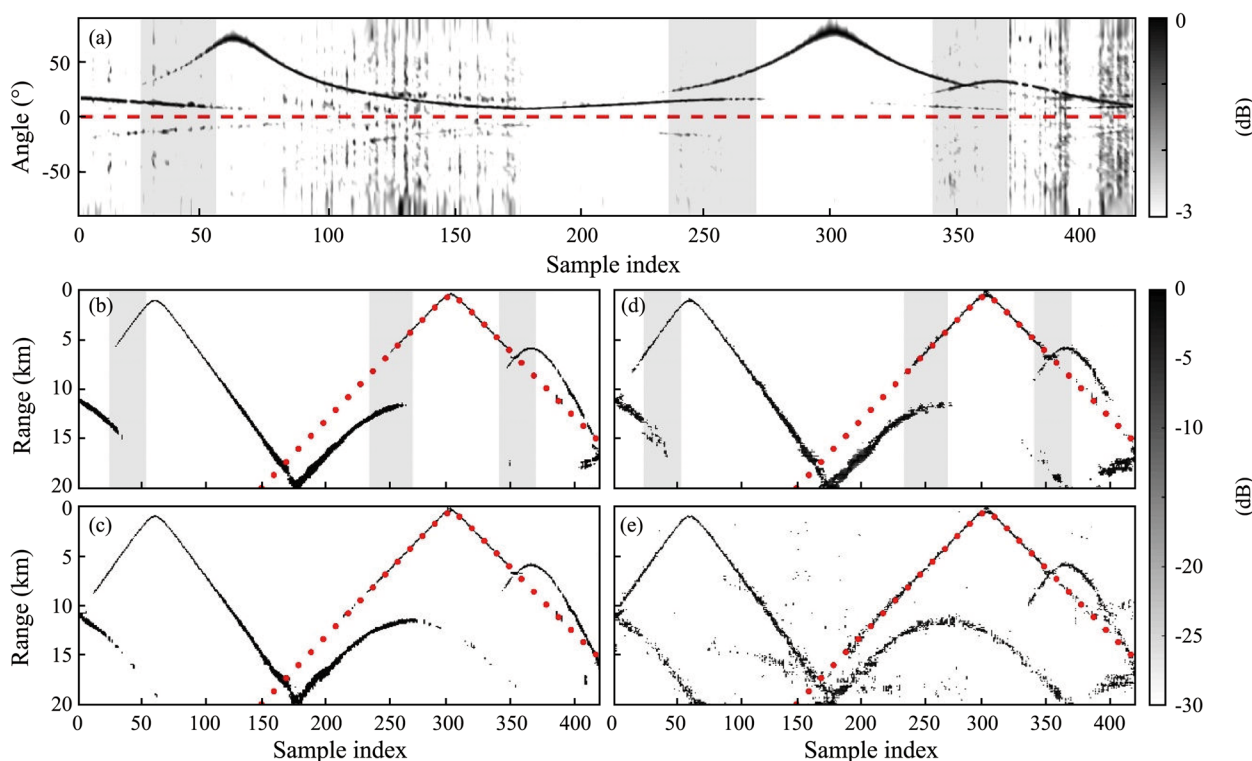
Fig. 6 a Conventional beamforming results. Shaded areas represent samples with an SSR below 3 dB. **b** Bartlett results without SVD preprocessing. **c** Bartlett results with SVD preprocessing. **d** GFGRU results without SVD preprocessing. **e** GFGRU results with SVD preprocessing. We have plotted the fifteen highest peaks as the number of sources is unknown. Circles indicate the actual ranges of the experimental ship. (Adapted from Liu et al. 2021d)

3.2 Preprocessing methods

Due to marine environments' complexity, underwater recognition systems often face challenges in achieving satisfactory generalization performance in real-world scenarios. To mitigate this issue, many researchers employ preprocessing techniques on signal records to minimize the impact of interference on recognition systems. For example, denoising algorithms are widely used to address ambient noise (Yang et al. 2022), pulse signals (Wang et al. 2022a), and self-noise in complex marine environments. Furthermore, filtering techniques, including band-pass and adaptive filtering, are commonly applied during the preprocessing stage. Currently, signal-processing algorithms continue to dominate in this domain. However, several preprocessing methods based on ML have emerged in recent years, showing promising performance. For instance, researchers have developed data-driven denoising encoders (Dong et al. 2022) to reduce noise interference adaptively. These machine-learning-based preprocessing methods can autonomously learn relevant parameters, thus alleviating the burdensome task of manual parameter adjustment and significantly reducing time costs.

3.3 Acoustic feature extraction

Acoustic feature extraction is pivotal in underwater acoustic target recognition because it transforms a time series of signals into representative features that encapsulate specific data attributes (Bianco et al. 2019). These features must effectively capture the intrinsic characteristics of underwater acoustic signals while remaining resilient to environmental variations such as ocean noise (Xie et al. 2022b), distortion, and variations in source-target distance (Xie et al. 2022a). Traditional feature extraction methods in this field encompass time-domain, frequency-domain, and time-frequency features. Furthermore, this study introduces feature extraction methods rooted in ML techniques.

Time-domain features in acoustic signals analysis are typically derived from the statistical properties of the signals. These features are crucial in quantifying various aspects of the signal's characteristics. For instance, energy-based features such as short-time average energy, peak energy, energy difference, and energy entropy have been widely applied to assessing acoustic signals' strength or power. Additionally, several other commonly utilized time-domain features, including zero crossing rate, autocorrelation, and amplitude envelope (Boashash

Table 1 Summary of source localization methods using ML

Category	Subcategory	Literature
Training data	Simulated data	Huang et al. 2018; Chi et al. 2019; Niu et al. 2019a, b; Liu et al. 2020b; Li et al. 2020b; Van Komen et al. 2020; Chen and Schmidt 2021; Liu et al. 2021c, d; Neilsen et al. 2021; Goldwater et al. 2023
	Real data	Lefort et al. 2017; Niu et al. 2017a, b; Wang and Peng 2018; Ferguson et al. 2019; Yangzhou et al. 2019; Ferguson 2021
	Mixed data	Wang et al. 2019a; Wang et al. 2020b; Liu et al. 2020a; Ge et al. 2022
Preprocessed features	Frequency-domain pressure	Lefort et al. 2017; Wang et al. 2019a
	Time-domain waveform	Huang et al. 2018; Van Komen et al. 2020; Goldwater et al. 2023
	Sample covariance matrix	Niu et al. 2017a, b; Wang and Peng 2018; Chi et al. 2019; Yangzhou et al. 2019; Chen and Schmidt 2021; Liu et al. 2021d; Ge et al. 2022
	Eigenvectors	Huang et al. 2018
	Time-domain autocorrelation function	Liu et al. 2021c
	Pressure magnitude	Niu et al. 2019a, b; Liu et al. 2020a
	Cepstrogram	Ferguson et al. 2019; Ferguson 2021
	Beamformed output	Wang et al. 2020b
	Time(range)-frequency spectrograms	Li et al. 2020b; Neilsen et al. 2021
Output	Classification	Niu et al. 2017a, b; Chi et al. 2019; Niu et al. 2019a, b; Wang et al. 2019a; Chen and Schmidt 2021; Wang et al. 2020b; Ge et al. 2022
	Regression	Lefort et al. 2017; Niu et al. 2017a, b; Huang et al. 2018; Wang and Peng 2018; Ferguson et al. 2019; Yangzhou et al. 2019; Liu et al. 2020a, b; Li et al. 2020b; Van Komen et al. 2020; Chen and Schmidt 2021; Ferguson 2021; Neilsen et al. 2021
ML backbone models	FNN/SVM/RF	Niu et al. 2017a, b; Chi et al. 2019; Yangzhou et al. 2019
	GRNN	Wang and Peng 2018
	ResNet	Niu et al. 2019a, b
	U-Net	Li et al. 2020b
	Xception	Liu et al. 2021c
	GFGRU	Liu et al. 2021d
	Other variants of convolutional networks	Huang et al. 2018; Ferguson et al. 2019; Liu et al. 2020a, b; Van Komen et al. 2020; Chen and Schmidt 2021; Ferguson 2021; Neilsen et al. 2021; Ge et al. 2022; Goldwater et al. 2023
Learning strategies	Early stopping with the additional constraint	Chi et al. 2019
	Multitask learning	Liu et al. 2020b; Ferguson 2021; Neilsen et al. 2021
	Transfer learning	Wang et al. 2019a; Wang et al. 2020b; Ge et al. 2022
	Data augmentation	Ferguson et al. 2019; Liu et al. 2021c; Neilsen et al. 2021
Data sets	Shallow water	Lefort et al. 2017; Niu et al. 2017a, b; Chi et al. 2019; Ferguson et al. 2019; Yangzhou et al. 2019; Li et al. 2020b; Van Komen et al. 2020; Chen and Schmidt 2021; Ferguson 2021; Neilsen et al. 2021; Goldwater et al. 2023
	Deep ocean	Wang et al. 2019a; Wang et al. 2020b; Liu et al. 2020b; Liu et al. 2020a; Liu et al. 2021c; Liu et al. 2021d
Receiver type	Vertical array	Lefort et al. 2017; Niu et al. 2017a, b; Huang et al. 2018; Chi et al. 2019; Wang et al. 2019a; Liu et al. 2020b; Wang et al. 2020b; Chen and Schmidt 2021; Neilsen et al. 2021; Ge et al. 2022
	Horizontal array	Li et al. 2020b; Ferguson 2021
	Single sensor	Niu et al. 2019a, b; Ferguson et al. 2019; Van Komen et al. 2020; Liu et al. 2021c; Goldwater et al. 2023

and O'shea 1990), as well as short-time mean amplitude difference (Jiang et al. 2020), are extensively employed in recognition of underwater targets. These features are invaluable for capturing signals' amplitude and temporal attributes, enabling the analysis of critical characteristics of underwater sound propagation. Such studies can offer insights into specific target identification or the differentiation of various noise types within underwater environments.

Frequency-domain features are derived by transforming acoustic signals into the frequency domain using short-time Fourier transform and wavelet transform.

Niu *et al. Intelligent Marine Technology and Systems*   (2023) 1:8

Page 9 of 23

These methods offer an efficient means of extracting spectral, harmonic, and phase characteristics from signals, often instrumental in distinguishing various underwater targets. Commonly employed frequency-domain features encompass the power spectrum (Hemminger and Pao 1994), Mel spectrum (Wang et al. 2019b; Liu et al. 2021a), Mel-frequency cepstral coefficients (MFCCs) (Wang et al. 2016), DEMON (detection of envelope modulation on noise) spectrum (Li et al. 2022b), spectral sub-band centroid (Chen and Xu 2017), and spectra based on LOFAR (low-frequency analysis and recording) (Li et al. 2022b), Hilbert Huang transform (comprising empirical mode decomposition and Hilbert spectral analysis) (Zeng and Wang 2014; Jin et al. 2023), wavelet transform (Khishe 2022; Xie et al. 2022a, b), and constant Q transform (Cao et al. 2018; Irfan et al. 2021). In addition, time-frequency spectrograms can be generated by concatenating framed frequency-domain features along the time dimension. Time-frequency spectrograms concurrently capture temporal and frequency information, rendering them potent tools for feature extraction in underwater acoustic target recognition (Liu et al. 2021a).

Furthermore, with the advancement of ML algorithms, data-driven neural networks have also found application in feature extraction. Numerous studies have used neural networks, such as CNNs (Irfan et al. 2021; Xie et al. 2022a), recurrent-wavelet architectures (Khishe 2022), and embedding memory units (Wang et al. 2022a, b), including autoencoders, to extract high-dimensional representations. These representations provide an enhanced characterization of the training data distribution and automatically capture profound semantic information as highly adaptable learners. Significantly, they demonstrate satisfactory recognition when dealing with abundant, high-quality data. However, they often lack explicit physical meaning and interpretability.

Recognizing the ongoing importance of traditional features in contemporary underwater acoustic recognition systems is crucial. Traditional features offer a more transparent physical interpretation and showcase robustness and generalization capabilities. While the ML-based approaches excel at capturing intricate patterns, traditional features remain indispensable components of the recognition framework. Their explicit physical interpretation enhances the system's comprehension and ensures resilience across various scenarios.

3.4 Recognition module
The recognition module automatically identifies underwater targets using the extracted features. This module essentially recognizes underwater acoustic targets by discerning patterns within the features. Underwater acoustic target recognition primarily comprises two principal paradigms: traditional ML-based and DL-based approaches.

Conventional traditional ML-based approaches typically involve an initial step of selecting discriminative features, followed by the use of ML algorithms such as Naïve Bayes, SVMs (Wang and Zeng 2014), *k*-nearest neighbor (KNN) (Ke et al. 2020; Jin et al. 2023), Gaussian mixture model (Wang et al. 2019b), or RFs (Wang et al. 2023) to make target class predictions. However, these traditional ML approaches heavily rely on manual engineering features that can effectively represent target information. This process is time-consuming and may introduce subjectivity.

DL-based approaches have demonstrated exceptional performance across various recognition tasks, including underwater acoustic target recognition. DL algorithms, such as DNNs (Irfan et al. 2021), CNNs (Cao et al. 2018; Irfan et al. 2021; Liu et al. 2021a; Ren et al. 2022; Xie et al. 2022a), recurrent neural networks (Liu et al. 2021a; Khishe 2022), transformers (Feng and Zhu 2022; Li et al. 2022a) and their variations, can automatically extract features from raw acoustic data, eliminating the need for manual feature engineering. DL-based approaches typically require substantial amounts of labeled data to train the models. However, they often achieve superior accuracy and robustness compared with traditional ML-based techniques. Moreover, DL methods rely less on prior knowledge and can recognize unseen data in real-world scenarios.

As depicted in Fig. 7, we present a visualization of the experimental results reported by Irfan et al. (2021). The figure displays the recognition accuracy of eight models: Naïve Bayes, KNN, SVM, RF, DNN, CNN, Inception Network, and Residual Network, across four distinct acoustic features: Mel spectrogram, Gammatone spectrogram, CQT spectrogram, and wavelet packets. The DNN-based methods demonstrate markedly superior recognition accuracy compared to traditional ML algorithms.

3.5 Optimization of training strategies
In addition to feature extraction methods and recognition models, a substantial portion of research has been dedicated to optimizing training strategies. The limited availability of data in underwater acoustic recognition tasks presents a significant challenge, as it renders recognition systems susceptible to overfitting and diminishes their capacity for generalization. Numerous advanced training strategies have been proposed to create more resilient recognition systems to tackle this issue.

These strategies encompass both manually designed and automatically generated augmentation techniques.
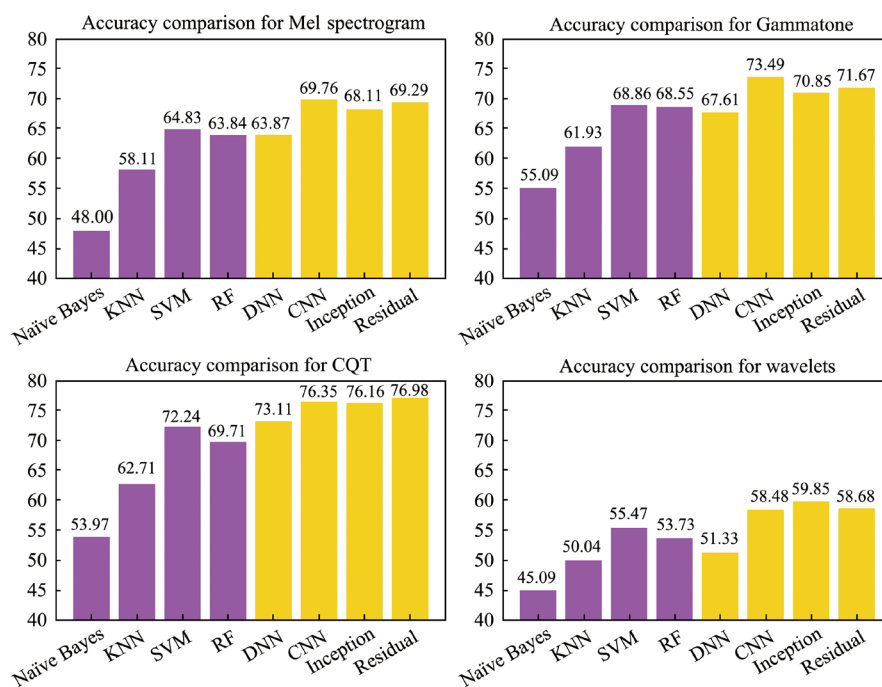
Fig. 7 Accuracy comparison between traditional ML algorithms (in purple) and DL-based methods (in yellow) on the DeepShip dataset. (Adapted from Irfan et al. 2021)

Manual-designed augmentation techniques involve modifying the training data, such as simulating channel modeling (Li et al. 2023a) and introducing simulated background noise (Kim et al. 2021). These techniques simulate various real-world conditions and augment the diversity of the training data, thus bolstering the system's generalization ability. In addition to manual-designed augmentation, automatic augmentation methods have garnered considerable attention. These techniques leverage ML algorithms to generate synthetic or perturb existing data. Examples of automatic augmentation techniques include spectrogram masking (Liu et al. 2021a), generative adversarial networks (Jin et al. 2020), and signal reconstruction (Luo et al. 2021). To further address data scarcity, some researchers use unlabeled data to construct self-supervised or unsupervised learning recognition systems (Wang et al. 2022b). In contrast, others incorporate additional data from different domains for transfer learning (Li et al. 2023a). Additionally, fusion methods, such as feature integration (Ke et al. 2020; Liu et al. 2021a) at the feature level and model ensemble at the model level, are widely employed to build robust recognition systems. These approaches enhance the model's generalization performance through additional data and mitigate overfitting in the recognition model.

3.6 Public databases and benchmarks

The challenges and high costs associated with underwater signal acquisition (Santos-Domínguez et al. 2016; Irfan et al. 2021), coupled with limited data availability and restrictions due to security and military applications, have contributed to the scarcity of real-world underwater signal data. Previous research relied heavily on simulated signals with predetermined characteristics like speed, direction, and distance. However, comprehensively simulating the exceedingly complex interference factors in underwater environments solely through simulated signals is nearly impractical. The disparity between simulated signals and real-world scenarios often leads to reduced generalization performance of recognition systems. With advancements in acquisition technology and growing demands from the research community, two publicly released real-world underwater acoustic databases have become fundamental resources for recent work in this field. These databases offer authentic and diverse datasets that better mirror actual scenarios. Consequently, much recent research has been based on these publicly available databases and has yielded promising results (Santos-Dominguez et al. 2016; Ke et al. 2020; Khishe 2022; Ren et al. 2022; Xie et al. 2022a, b). The details of the two databases are provided in Table 2. One of these, ShipsEar (Santos-Domínguez et al. 2016), comprises 90 records of ship and boat sounds from 12 different types (dredgers, fishing boats, trawlers, mussel boats,

tugboats, motorboats, pilot boats, sailboats, passenger ferries, ocean liners, Ro-Ro vessels, and background noise recordings), totaling 2.94 h of recordings. In addition to audio records, ShipsEar offers supplementary information, such as the target images, localization data, acquisition time, channel depth, wind, conditions, distance, atmospheric and oceanographic data, and notes. This additional information allows for a more comprehensive and detailed acoustic data analysis. The other database, DeepShip (Irfan et al. 2021), consists of 47.07 h of real-world underwater recordings of 265 ships categorized into four classes (tugboats, cargo ships, oil tankers, and passenger ships). The extensive scale of DeepShip effectively meets the data requirements of data-driven ML algorithms.

These two databases serve as a valuable benchmark for research in this field. Both databases must provide an official data division, such as training, validation, or test sets, for evaluating recognition tasks. Consequently, the reported results in the current studies are not directly comparable. Existing research has demonstrated that different division methods notably impact reported results (Liu et al. 2021a). A common practice involves dividing each audio record into multiple segments and randomly assigning them to training and test sets, which can result in some samples in the test set and training sets belonging to the same record. Given that the ship-radiated noise signals tend to be relatively stable over some time, DNN-based methods can easily achieve high performance through overfitting in such cases. Therefore, it is advisable to split the training and test sets based on entire audio records to prevent information leakage (Santos-Domínguez et al. 2016; Irfan et al. 2021; Liu et al. 2021c; Xie et al. 2022a; Xu et al. 2023). We hope further research efforts will standardize the division method to enable researchers to conduct more rigorous validations and comparisons.

4 Communication

Over the years, underwater acoustic (UWA) communication technology has evolved significantly, progressing from incoherent to coherent communication and from single-carrier to multicarrier communication, as exemplified by orthogonal frequency division multiplexing (OFDM). The demand for increased data rates and wider

bandwidth in underwater communication is steadily growing (Li et al. 2008). Concurrently, underwater networking technology is gaining popularity, and multimode networks can effectively facilitate information exchange and sharing (Sozer et al. 2000). However, the UWA channel encounters a series of challenges, including multipath effects, rapid fading, and significant background noise due to underwater sound propagation's intricate and dynamic nature (Qarabaqi and Stojanovic 2013). These challenges present substantial obstacles to reliable underwater information transmission.

In summary, given the substantial growth in the demands for underwater communication, traditional communication technology rooted in modular and model-driven approaches is encountering limitations. As underwater communication grapples with increasingly complex environmental dynamics and multiple dimensions of network resources demand precise configuration with fine-grained accuracy, traditional UWA communication technology will be subjected to rigorous assessments of accuracy and robustness. Implies that communication models built upon expert knowledge also exhibit certain limitations.

The application diagram of ML in UWA communication is shown in Fig. 8. Typical application scenarios that combine UWA communication with ML include the following: (a) The physical layer, primarily for communication between nodes, includes tasks such as underwater channel estimation and equalization (Chen et al. 2018; Zhang et al. 2019, 2021b, 2022d, 2022e), underwater adaptive modulation and coding (Fu and Song, 2018; Zhang et al. 2022g), communication quality prediction (Lucas and Wang 2020; Chen et al. 2021), and UWA communication signal detection (Chu et al. 2023). (b) The network layer, which encompasses aspects such as cluster-based routing protocols (Chen et al. 2022; Geng and Zheng 2022), optimal power allocation (Xiao et al. 2019; Wang et al. 2020a), and underwater network security (He et al. 2020; Mary et al. 2021). As research on underwater communication technology continues to advance, these research topics are accompanied by growing demands for intelligent and integrated underwater equipment. This trend presents new challenges related to the rapid increase in data volume, the dynamic nature of UWA application scenarios, and heightened security

Table 2 Information of ShipsEar and DeepShip

Dataset	Location	Record	Type	Duration (h)	Sample rate (Hz)	Extra info
ShipsEar	Georgia Delta node near Canada	90	12	2.94	52734	Yes
DeepShip	Spanish Atlantic coast in Northwest Spain	265	4	47.07	32000	None

requirements. ML offers new solutions to address the following challenges:

(a) Big Data versus DNNs: With the development of underwater information acquisition technology, a substantial volume of experimental data has been accumulated. This wealth of information requires further integration, distillation, and refinement. DL methods effectively consolidate and extract information from data (LeCun et al. 2015).

(b) Complex and Dynamic Environments versus Transfer Learning: The marine environment is complex and varied. It necessitates ML models with solid robustness to quickly adapt to unfamiliar surroundings, thereby enabling UWA communication in diverse scenarios. Transfer learning emphasizes using past knowledge and experience to guide learning in new tasks (Weiss et al. 2016). This ML approach is fundamental for achieving general artificial intelligence and is the primary method for breaking free from fixed-scene UWA communication.

(c) Multinode Network versus Reinforcement Learning: Underwater networking introduces challenges related to information fusion and intelligent interactions among multiple agents. With reinforcement learning, underwater multiagent systems learn habitual behaviors that maximize utility through direct interactions with the environment. They subsequently accomplish more complex tasks through
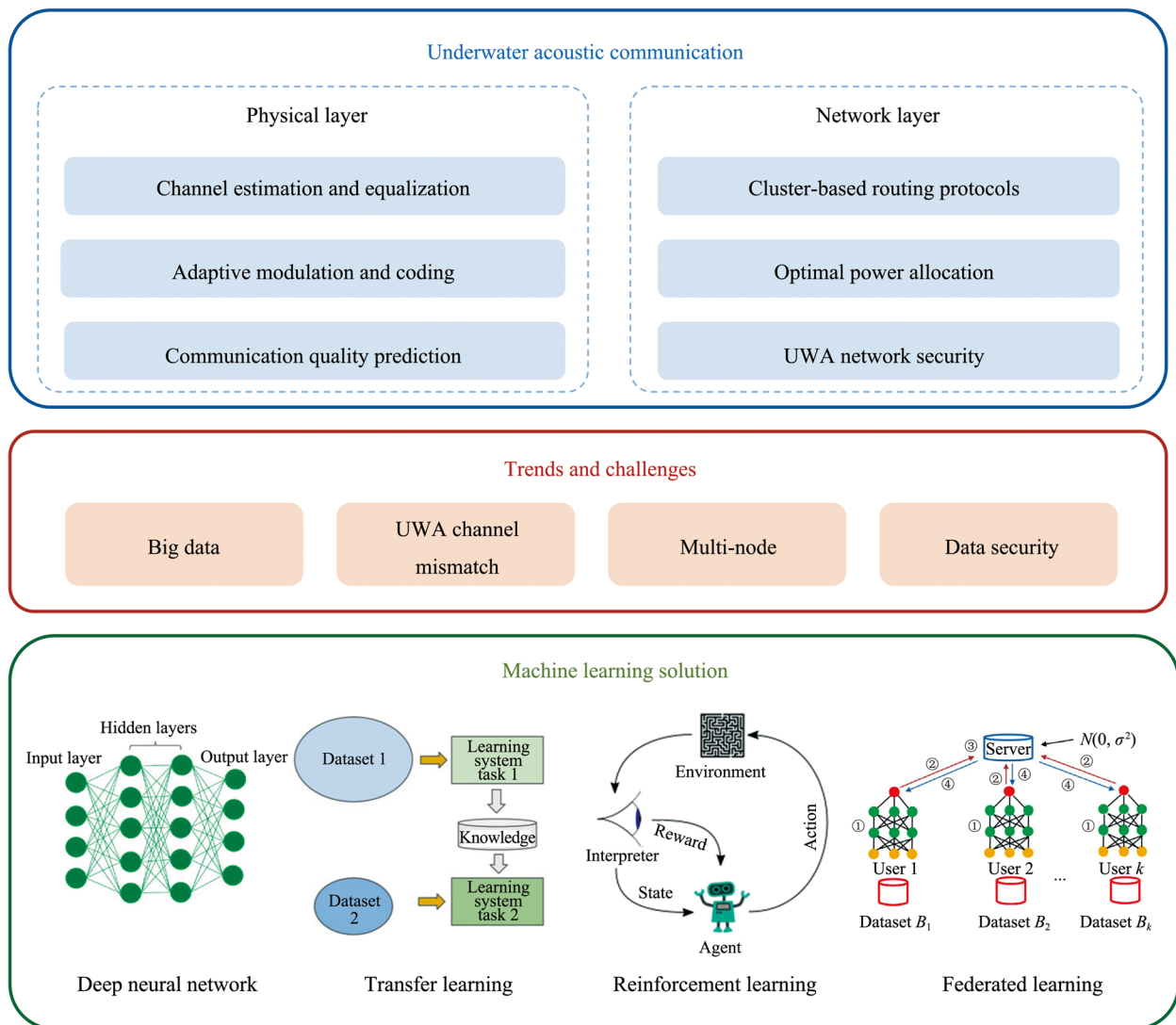
Fig. 8 ML-based UWA communications

interaction and decision-making in high-dimensional and dynamic real-world settings (Mnih et al. 2015).

(d) Data Security versus Federated Learning: Ensuring the privacy and security of UWA networking and communication data is paramount. Federated learning has emerged as an efficient method for preserving privacy (McMahan et al. 2017; Li et al. 2020a). This distributed ML approach can derive a comprehensive learning model through decentralized training and parameter sharing among participants without directly accessing the data sources. Minimizing the risk of data breaches while ensuring privacy and enabling model training on extensive datasets achieves these objectives.

As illustrated in Fig. 8, the physical layer of UWA communication is the foundation for the entire communication system. Numerous practical investigations have underscored the importance of advancing physical layer technology in enabling breakthroughs across the field. Currently, it stands as a pivotal research direction. In this context, channel estimation and equalization form the bedrock and nucleus of high-quality communication implementation within the physical layer. They are critical links connecting various modules within the physical layer. The subsequent sections provide an in-depth review of a particularly noteworthy application in communication architecture: UWA channel estimation and equalization.

Table 3 summarizes typical studies that apply ML models for UWA channel estimation and equalization, providing brief descriptions of the models used, communication systems involved, features employed, datasets utilized, model performance, and main contributions.

In early studies involving DNN-aided channel estimation, researchers aimed to replace traditional channel estimation and equalization modules with various depth network structures. This yielded improved performance results approaching the minimum mean square error (MMSE) solution. These studies employed typical network structures such as the multilayer perceptron (MLP) (Chen et al. 2018), which consisted of fully connected neural networks with five layers having 1024, 1500, 600, 128, and 32 neurons, respectively (Zhang et al. 2019). Additionally, relatively efficient lightweight DNN structures were explored (Jiang et al. 2019). However, a challenge arises when dealing with complex-valued UWA communication signals, often reshaped as two parallel real-valued tensors (with real and imaginary parts treated separately) for input into the network. This approach could waste memory resources and slow down the training process. Researchers designed a complex-valued network ($\mathbb{C}$-DNN) for UWA channel estimation to address

these challenges, as illustrated in Fig. 9 (Zhang et al. 2022c). Experiments conducted using the Watermark dataset measured at sea demonstrated that the complex-valued model can achieve nearly optimal channel tracking performance while conserving 50% of spatial resources compared with its real-valued counterparts.

Researchers have recently emphasized addressing practical issues in UWA communication through ML. Of notable interest are the challenges stemming from the scarcity of UWA data, which gives rise to the few-shot problem, and the intricate and dynamically challenging UWA environment, which leads to domain mismatch. This section delves into some noteworthy studies that have tackled these challenges.

4.1 Few-shot problem in UWA communications

The domain of UWA communication presents a few-shot problem that stems from the challenges associated with efficiently collecting UWA data. Factors such as demanding sea trial conditions lead to high acquisition costs, resulting in limited samples collected within a finite time frame. More data is needed to ensure effective model training, leading to overfitting. To address this issue, data augmentation, a widely employed technique in various ML domains, generates additional data from the limited dataset. By leveraging communication signal processing techniques, researchers incorporate perturbations and interferences commonly encountered in UWA communication scenarios, including timing errors, Doppler shift, and noise interference, into data augmentation to expand the dataset. One common approach involves the application of symbol timing offset $\widehat{y}(n) = y(n + \varepsilon)$ and Doppler shift $\widehat{y}(n) = y[(1 + \sigma)n]$ to the original data, as outlined in previous literature (Zhao et al. 2022). This study analyzes the performance enhancement achieved through data augmentation using simulated data, building upon this method. This analysis validates the effectiveness of the proposed approach.

In addition, Zhang et al. (2022f) identified the potential mechanism behind model performance degradation resulting from insufficient UWA samples. They emphasized the significance of fast-fading perturbations occupying the channel structure's high-frequency range. These components are crucial in enabling the model to attain sufficient training and acquire knowledge of channel distribution characteristics in specific UWA environments, thereby preventing overfitting. Building upon this theoretical analysis, the authors proposed an EMD-based data augmentation method that decomposes the channel and employs random replay to expand the channel samples (Zhang et al. 2022f), as depicted in Fig. 10. The feasibility of the data augmentation method was

Table 3 Typical study on ML-based UWA channel estimation

Method	System	Feature	Dataset	Performance	Contribution
MLP (Chen et al. 2018)	OFDM	Signal after channel estimation and equalization	Real-world dataset-Swan River, western Australia	Lower SER compared with traditional algorithms	First adopt an MLP-based algorithm to perform symbol detection
DNN (Zhang et al. 2019)	OFDM	Received signal in the frequency domain	Bellhop simulation dataset	Lower BER and more robust to the pilot number	The DNN-based receiver is designed to recover the transmitted symbols directly
Lightweight DNN (Jiang et al. 2019)	OFDM	Received signal in the frequency domain	Bellhop simulation dataset	Near MMSE performance, smaller storage resources, and running time	Lightweight DNN is designed to improve real-time performance
CsiPreNet (Liu et al. 2021b)	OFDMA	Historical CSI image	Measured time-varying channel dataset	Better BER performance than RLS predictor	CsiPreNet comprises CNN and LSTM networks and captures the temporal correlation and frequency correlation of CSIs
Meta-learning (Zhang et al. 2021a)	OFDM	Received signal in the frequency domain	Real-world dataset-Watermark (van Walree et al. 2017) and lake experiments	Saves 98% of gradient steps during the adaptation stage compared with conventional DNN	A meta-learning-based UWA-OFDM communication system is designed to alleviate the environment mismatch in real-world UWA applications
C-DNN (Zhang et al. 2022c)	OFDM	Received signal in the frequency domain	Real-world dataset-Watermark	Near-optimal channel tracking saves 50% of spatial resources	A complex-valued DNN aided channel tracking is proposed for UWA-OFDM
Data augmentation (Zhang et al. 2022f)	OFDM	Received signal in the frequency domain	Real-world dataset-Watermark	Eliminate overfitting and achieve 50%–80% BER performance improvement	EMD-based data augmentation is proposed to solve UWA data scarcity
SCABNet (Zhang et al. 2023)	OFDM	LS channel estimation results	Real-world dataset-Watermark	Better BER performance than existing methods, more robust and efficient than FC-DNN and ComNet	The proposed receiver combines DL with classical expert knowledge to improve system performance and interpretability
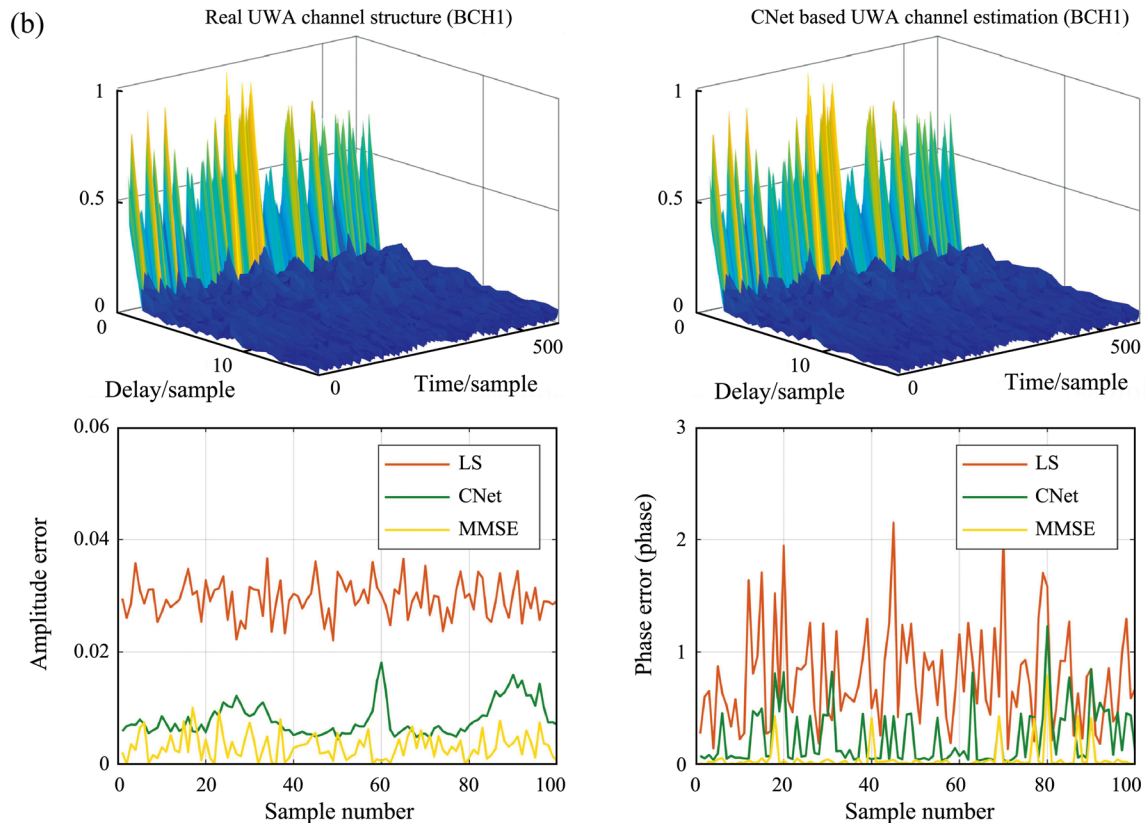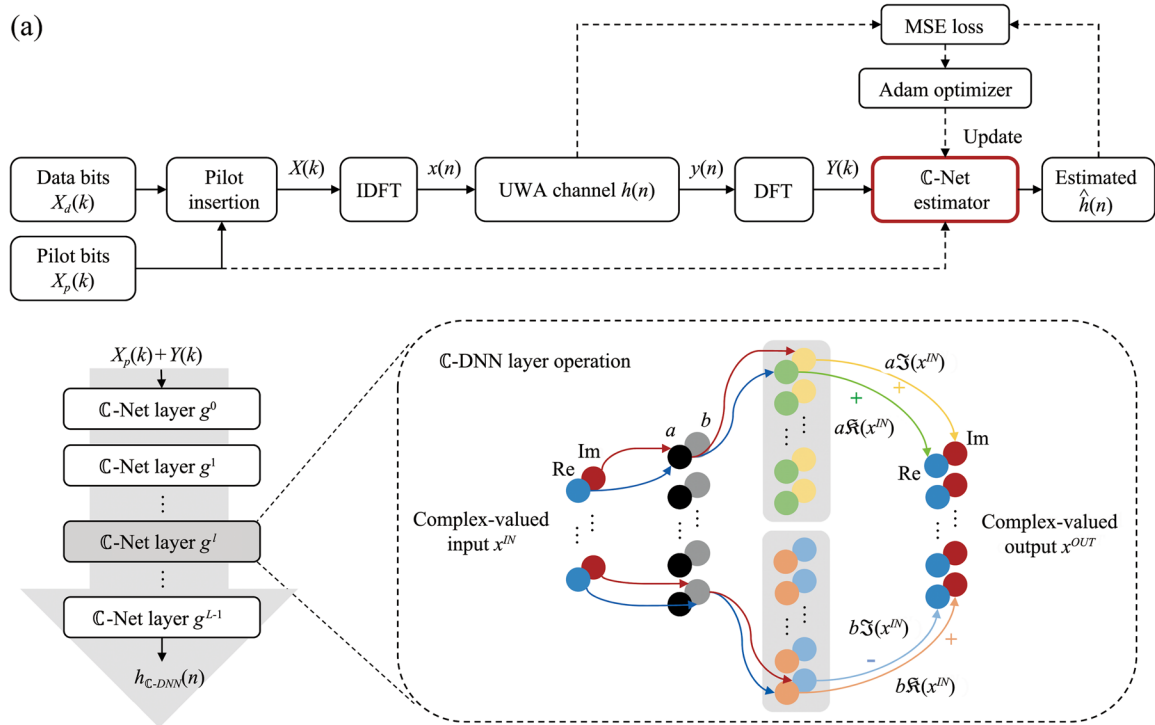
Fig. 9 **a** C-DNN for UWA communication system. **b** Time-varying UWA channel reconstruction and channel estimation error using the C-DNN estimator. (Adapted from Zhang et al. 2022c)

demonstrated through the experimental results shown in Fig. 11.

4.2 Environmental mismatch in UWA communications

In underwater acoustic (UWA) communication, the environmental mismatch problem arises because of the time–space-varying characteristics of the UWA channel. This variability poses a substantial challenge for the seamless transition of offline-trained models to online applications, particularly when environmental conditions change.

Currently, prevalent ML-based UWA communication system designs predominantly employ the conventional step-by-step iterative training approach, which unfortunately yields suboptimal model portability. Consequently, when the UWA communication environment undergoes alterations, a substantial volume of data from the new setting becomes necessary for retraining or fine-tuning purposes. This dependency on extensive retraining severely limits the model's generalizability.

To tackle the issue of source-target domain mismatch, researchers have suggested a meta-learning approach that incorporates meta-learning techniques into UWA channel estimation and equalization (Zhang et al. 2021a). This method enables swift adaptation to unfamiliar UWA environments in instances of environmental mismatch.

The researchers have developed a UWA-OFDM multi-task training platform based on a meta-learning training strategy, as illustrated in Fig. 12. The training tasks are drawn from known UWA communication task datasets (simulation or historical data) in various environments. In contrast, the target tasks are derived from communication sampling data in unknown environments. Through the meta-learning training process, the neural network model can rapidly discover parameter solutions that apply to unknown tasks within the parameter space. Consequently, it exhibits greater expressive power than traditional training methods for target tasks.

This study compared the meta-learning method's performance of transfer speed and error rate with conventional ML training methods. As depicted in Fig. 13, the experimental results demonstrate that the meta-learning-based model achieves convergence in unknown environments in just 100 iterations, whereas the model trained using traditional methods requires approximately 5000 iterations to reach convergence. The proposed method notably enhances response speed and effectively mitigates the impact of UWA mismatch on ML techniques. Overall, this approach constitutes a robust endeavor toward enabling UWA communication in diverse scenarios.
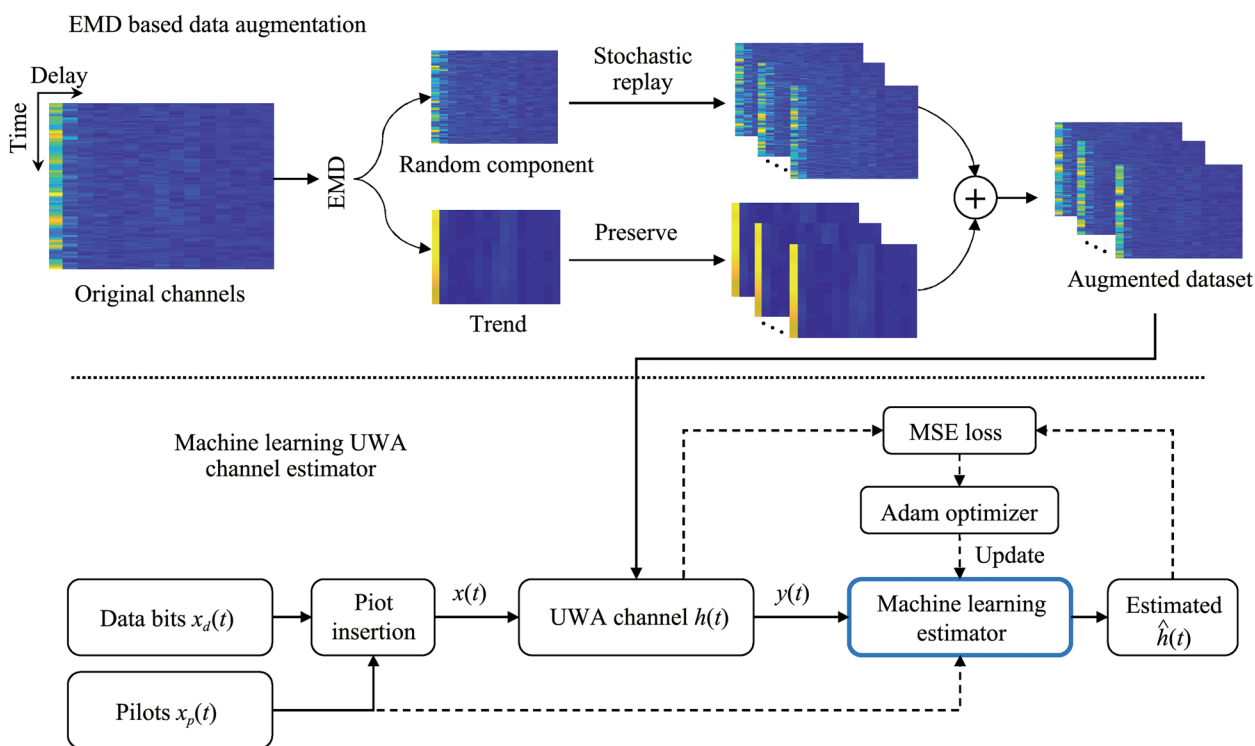
**Fig. 10** Data augmentation-aided UWA–OFDM. (Adapted from Zhang et al. 2022f)
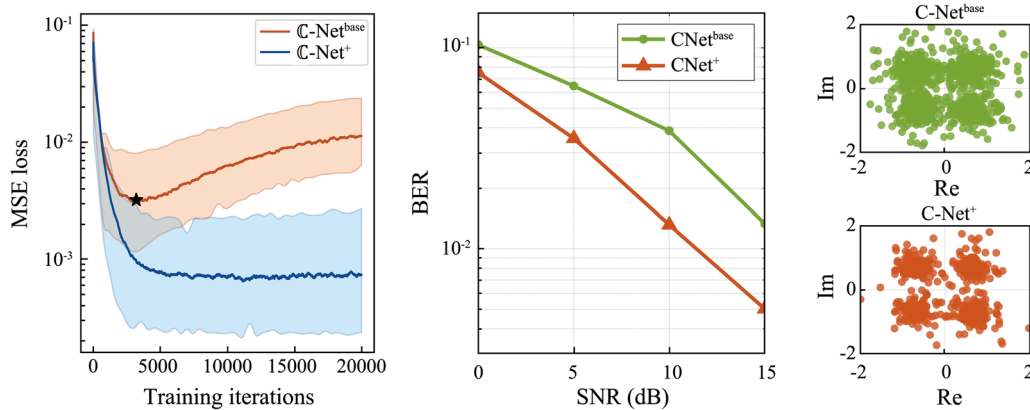
Fig. 11 Performance of data augmentation. From left to right: the loss curve before and after data augmentation, the BER performance, and constellation diagrams before and after data augmentation. (Adapted from Zhang et al. 2022f)

To further solve the problem that the training data at a single buoy may not be sufficient, a federated meta-learning (FML) scheme is proposed to train the DNN-based receiver by exploiting the model parameters gathered from multiple buoys within the ocean of things scenario (Zhao et al. 2022). This study analyzes the convergence performance of the FML. It describes a closed-form expression for the convergence rate, considering the effects of scheduling ratios, local epochs, and data volumes on an individual node. When trained with ample data, the simulation results demonstrate that the proposed C-DNN receiver outperforms classical MF-based detectors regarding BER performance and complexity.

5 Geoacoustic inversion

Geoacoustic inversion is a vital inverse problem in underwater acoustics. Its primary objective is to estimate the geoacoustic characteristics of the ocean floor based on recorded acoustic data. The most commonly employed technique is Matched-field inversion (MFI) (Collins et al. 1992). MFI deduces geoacoustic parameters by comparing acoustic measurements with replica
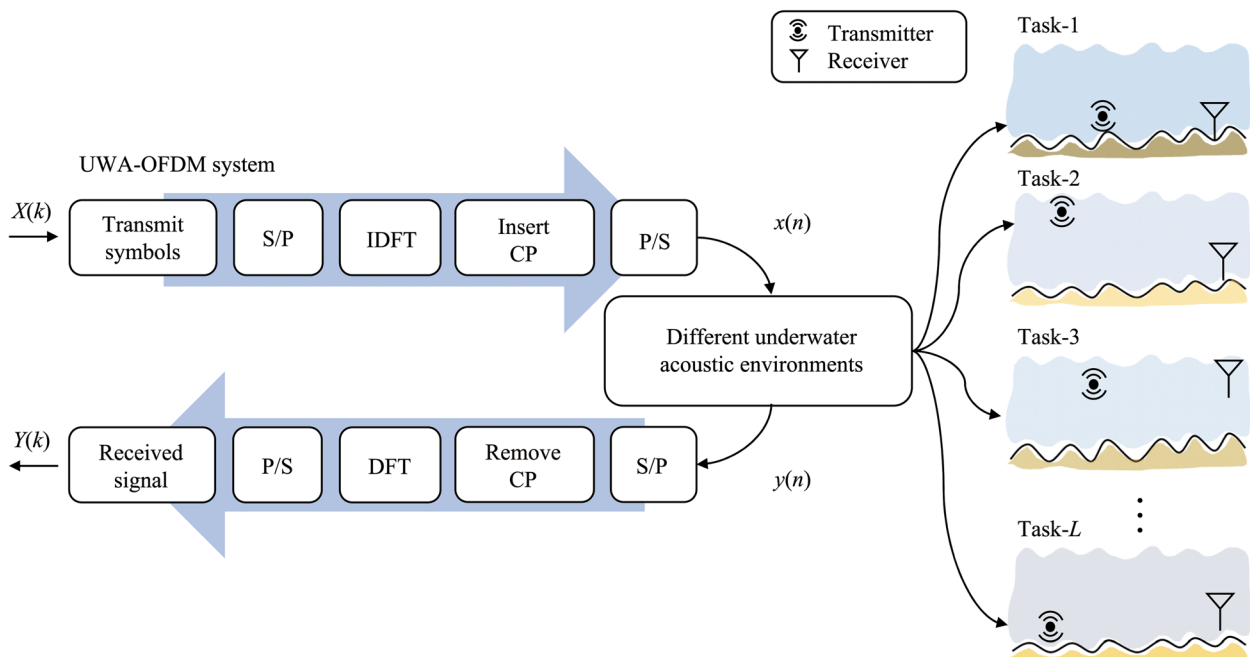
Fig. 12 Meta-learning based UWA-OFDM communications. (Adapted from Zhang et al. 2021a)
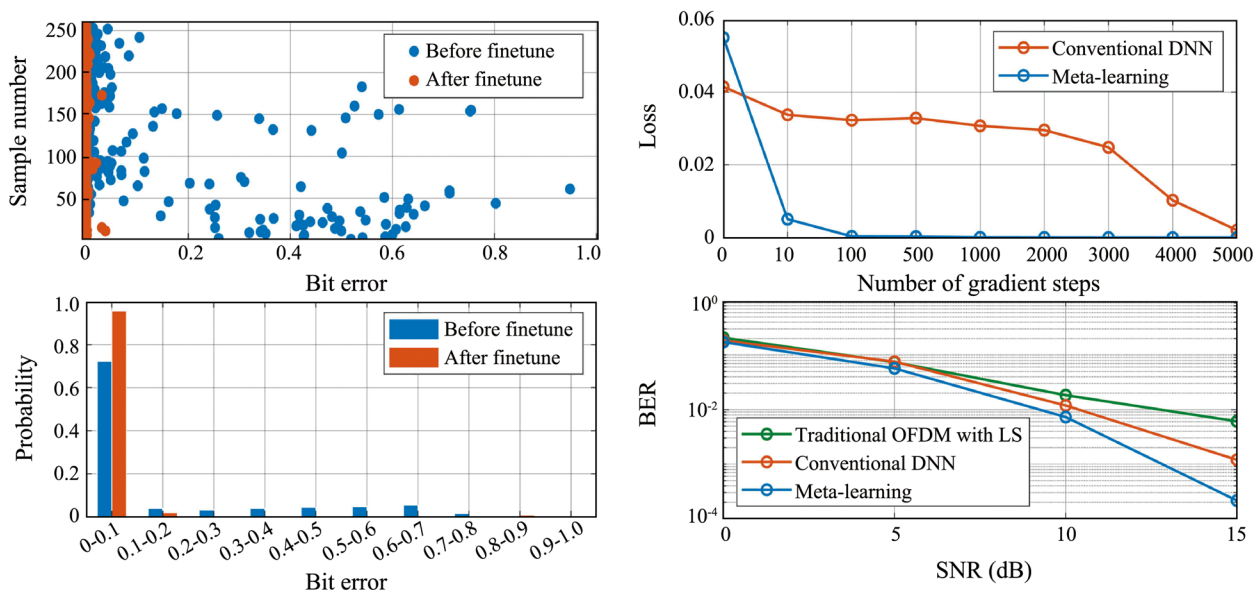
Fig. 13 Performance of meta-learning UWA channel estimation. From top to bottom, and left to right: The bit error distribution of the proposed method before and after meta-learning adaptation fine-tuning, convergence performance for an SNR of 15 dB with varying gradient steps during fine-tuning, and the BER performance using different approaches. (Adapted from Zhang et al. 2021a)

data, encompassing various unknown parameters computed through sound propagation models. However, MFIs face particular challenges. First, optimization methods such as the genetic algorithm (GA) and simulated annealing (SA) are time-consuming when multiple inversion parameters are involved. Second, these optimization techniques can be trapped in local minima because of the vast parameter search space and limited data. In contrast to MFI, ML methods directly learn a mapping from received data to geoacoustic parameters. This approach eliminates the need for explicit sound propagation models during testing and instead harnesses the capabilities of ML algorithms to infer the connection between measured data and desired parameters. ML offers a data-driven approach that enhances the accuracy and efficiency of the inversion process.

The application of ML to geoacoustic inversion began in the 1990s (Caiti and Parisini 1994; Michalopoulou et al. 1995; Caiti and Jesus 1996; Stephan et al. 1998; Benson et al. 2000). During that period, techniques such as radial basis function neural networks (RBFNNs) and other types of networks were employed to estimate geoacoustic parameters. Recently, features extracted from signals using a generalized additive model (Piccolo et al. 2019) have been used to estimate sound speed and attenuation. Integrating physical models with ML (Frederick et al. 2020) makes it feasible to classify ocean bottom sediments based on their acoustic characteristics. The

results demonstrate that ML methods surpass conventional MFI methods, particularly under low-frequency conditions.

Significant advancements in geoacoustic inversion were achieved by (Shen et al. 2020) using improved RBFNN incorporating the MFI kernel function. This approach yielded a performance comparable to that of conventional MFI techniques. Enhanced sensitivity of the objective functions to sediment density was attained by leveraging extensive datasets. In another application of ML techniques, a CNN was used to predict seabed types simultaneously, and source ranges from impulsive time series data (Van Komen et al. 2020). This application showcased the potential of ML methods for making simultaneous predictions of source ranges and seabed types.

In addition, a CNN (Neilsen et al. 2021) was employed to determine seabed types and source locations from a moving mid-frequency source. The power spectral levels of five tones (2, 2.5, 3, 3.5, and 4 kHz) served as input for the CNN, as depicted in Fig. 14. The performance of the trained CNN was analyzed under mismatched environments, highlighting the importance of accounting for environmental variability when using ML in ocean acoustics. These advancements underscore the promising capabilities of ML in geoacoustic inversion and its potential to enhance performance and accuracy in various ocean acoustic applications.

Motivated by the effectiveness of DL in handling multidimensional data, researchers introduced a CNN using

the multi-range vertical array data processing (MRP) method (Liu et al. 2022) for geoacoustic inversion. This approach enables exploiting a broad range of spatial diversity in the acoustic field. Unlike employing multiple separate networks for different geoacoustic parameters, a single CNN using the MTL method was proposed to estimate the geoacoustic parameters simultaneously. The combination of MTL with MRP (Liu et al. 2022) alleviates the coupling between the geoacoustic parameters.

From Fig. 15, it is evident that the distributions of the inversion results obtained from the MFI are not tightly concentrated around the ground truth. This observation highlights the increased complexity of the geoacoustic inversion problem when the test data are contaminated with noise, primarily because of the intricate coupling relationships. In contrast, the MRP-CNN produces more focused estimates that closely align with the ground truth. This enhanced performance can be attributed to the training process, in which the penalty factors in MTL are jointly optimized alongside the network parameters. The MRP-CNN effectively balances the influence of different geoacoustic parameters on the acoustic field. Consequently, the trained MRP-CNN demonstrates the capacity to mitigate the impact of parameter coupling during the inversion process (Liu et al. 2022).

One of the key advantages of employing ML for geoacoustic inversion is its ability to handle intricate and nonlinear relationships between the input data and seabed properties. Training the ML model on a substantial dataset of acoustic measurements and corresponding ground truth information can discern patterns and generate predictions based on the observed data. With an ample dataset, such as multi-range received data, the models can better understand the coupling between different geoacoustic parameters and leverage this understanding to enhance inversion outcomes. Additionally, ML techniques can significantly accelerate the inversion process. Traditional methods often involve time-consuming iterative or search-based algorithms that require extensive computational resources. Conversely, once the ML model is trained, it can swiftly generate predictions for new acoustic data within a relatively short timeframe. This efficiency is especially advantageous for real-time applications.

6 Limitations and prospects

Despite the significant progress in ML across various aspects of underwater acoustics, the practical application still needs to be improved. These limitations primarily include:

(1) Limited data availability: High-quality and labeled underwater acoustic datasets are often constrained, which poses challenges in training and validating ML models.

(2) Generalization: ML models trained on specific datasets may struggle to generalize effectively to unseen underwater acoustic scenarios, potentially leading to diminished performance.

(3) Robustness to noise and variability: Underwater acoustic environments are characterized by noise, signal distortions, and complex propagation phenomena. Developing ML models that exhibit robustness despite these challenges remains a significant research area.
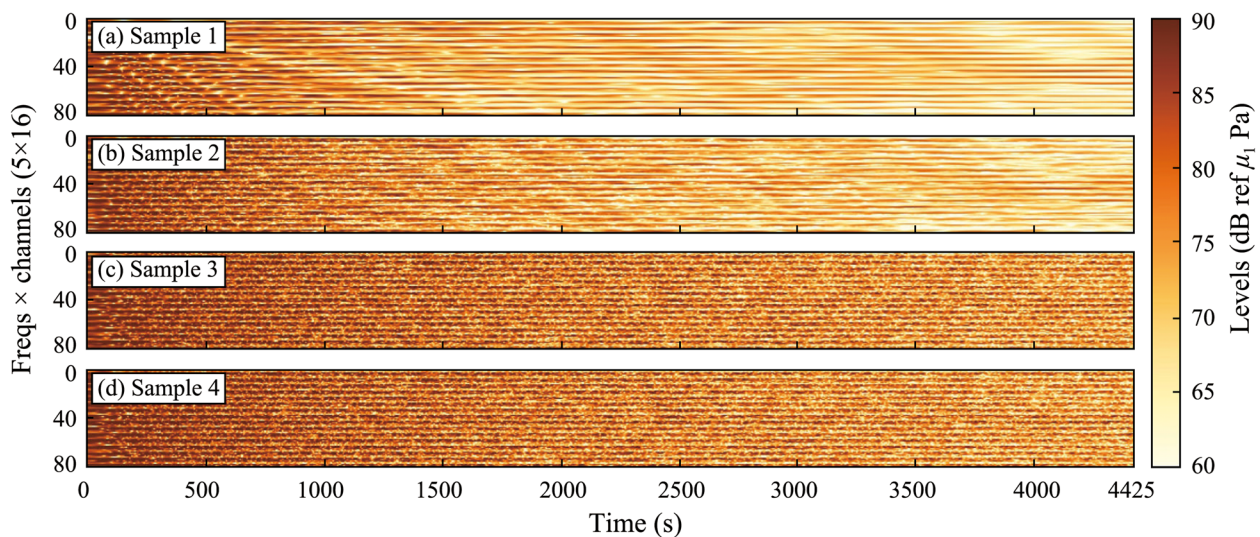
Fig. 14 The preprocessed input feature for ML inversion. (Adapted from Neilsen et al. 2021)
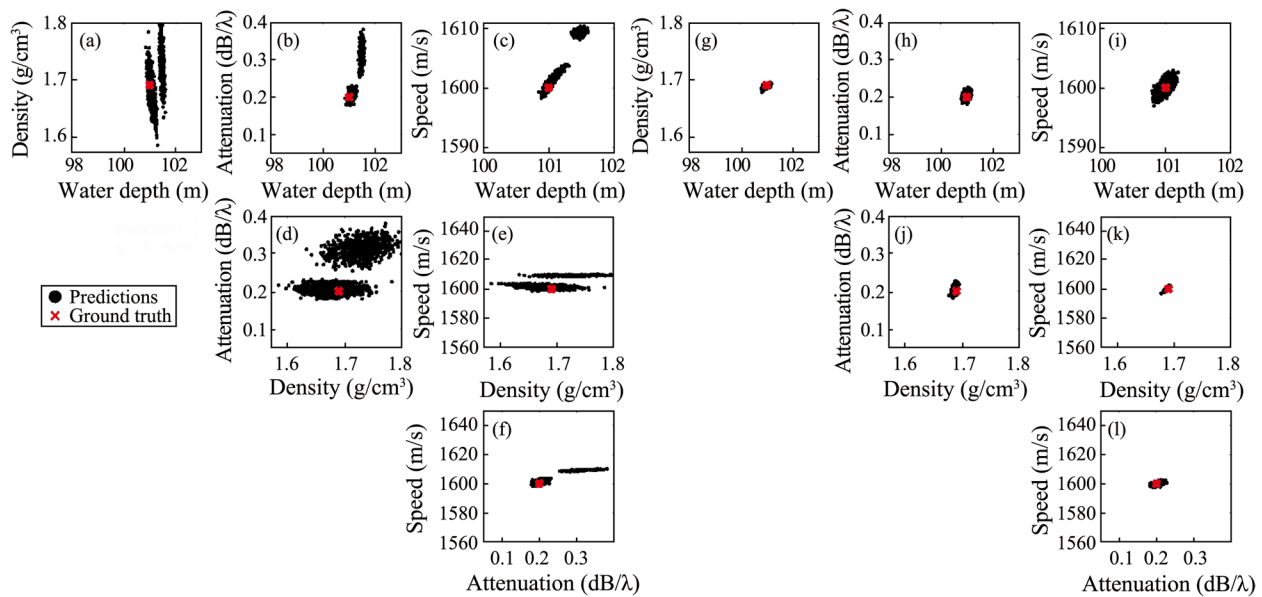
Fig. 15 2-D distributions of parameter estimates between **a**, **g** water depth and density; **b**, **h** water depth and attenuation; **c**, **i** water depth and speed; **d**, **j** density and attenuation; **e**, **k** density and speed; and **f**, **l** attenuation and speed. **a** through **f** correspond to the MFI estimation results, and **g** through **l** correspond to the MRP-CNN results. The red crosses represent the ground truth. (Adapted from Liu et al. 2022)

(4) Interpretable and explainable models: In specific applications, the ability to comprehend and elucidate the decision-making processes of ML models is crucial. Achieving the interpretability and explainability of underwater acoustic ML models is a noteworthy research pursuit.

Therefore, numerous research opportunities still exist in underwater acoustics using ML. Several future research directions include the following:

(1) Physics-Informed Neural Networks: Physics-informed neural networks (PINNs) can effectively generalize to unseen or sparse data points by incorporating physical laws. They can capture the underlying structure and dynamics of the system, leading to improved predictions even with limited training data. PINNs have potential in various underwater acoustic application scenarios.

(2) Transfer Learning and Domain Adaptation: Transfer learning techniques and domain adaptation methods can leverage knowledge from related domains and enhance the generalization ability of ML models in underwater acoustics.

(3) Ensemble and Hybrid Approaches: Exploring ensemble learning techniques and hybrid models that combine multiple ML algorithms or integrate physical models with ML to enhance performance and robustness.

(4) Active Learning and Data Augmentation: Developing strategies for active learning and data augmentation to address the limited availability of labeled underwater acoustic datasets and enhance the efficiency of model training.

(5) Explainable ML models in Underwater Acoustics: Developing interpretable and explainable ML models can provide insights into the decision-making process and enhance the trustworthiness of results in underwater acoustic applications.

By addressing these research directions and overcoming the associated challenges, underwater acoustic ML can advance further, leading to more accurate, efficient, and reliable solutions for various underwater acoustic tasks and applications.

Authors' contributions
Xiaolei Li prepared the section on source localization; Yonglin Zhang prepared the section on communication; Ji Xu prepared the section on target recognition; Haiqiang Niu prepared the remaining parts, refined all the sections, and finalized the paper. All authors contributed to the writing.

Availability of data and materials
No new dataset was generated or analyzed in the study.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests existing.

References

Benson J, Chapman NR, Antoniou A (2000) Geoacoustic model inversion using artificial neural networks. Inverse Probl 16(6):1627–1639

Bianco MJ, Gerstoft P, Traer J, Ozanich E, Roch MA, Gannot S et al (2019) Machine learning in acoustics: theory and applications. J Acoust Soc Am 146(5):3590–3628

Boashash B, O'shea P (1990) A methodology for detection and classification of some underwater acoustic signals using time-frequency analysis techniques. IEEE Trans Acoust Speech Signal Proc 38(11):1829–1841

Caiti A, Jesus SM (1996) Acoustic estimation of seafloor parameters: a radial basis functions approach. J Acoust Soc Am 100(3):1473–1481

Caiti A, Parisini T (1994) Mapping ocean sediments by RBF networks. IEEE J Ocean Eng 19(4):577–582

Cao X, Togneri R, Zhang XM, Yu Y (2018) Convolutional neural network with second-order pooling for underwater target classification. IEEE Sens J 19(8):3058–3066

Chen R, Schmidt H (2021) Model-based convolutional neural network approach to underwater source-range estimation. J Acoust Soc Am 149(1):405–420

Chen Y, Xu XN (2017) The research of underwater target recognition method based on deep learning. 2017 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), Xiamen, China, 22-25 October 2017, pp 1–5

Chen YG, Yu WJ, Sun X, Wan L, Tao Y, Xu XM (2021) Environment-aware communication channel quality prediction for underwater acoustic transmissions: a machine learning method. Appl Acoust 181:108128

Chen YG, Zhu JY, Wan L, Fang X, Tong F, Xu XM (2022) Routing failure prediction and repairing for AUV-assisted underwater acoustic sensor networks in uncertain ocean environments. Appl Acoust 186:108479

Chen ZP, He ZQ, Niu K, Rong Y (2018) Neural network-based symbol detection in high-speed OFDM underwater acoustic communication. In 2018 10th International Conference on Wireless Communications and Signal Processing (WCSP), Hangzhou, China, 18-20 October 2018, pp 1–5

Chi J, Li XL, Wang HZ, Gao DZ, Gerstoft P (2019) Sound source ranging using a feed-forward neural network trained with fitting-based early stopping. J Acoust Soc Am 146(3):EL258–EL264. https://doi.org/10.1121/1.5126115

Chu HL, Li C, Wang HB, Wang J, Tai YP, Zhang YL et al (2023) A deep-learning based high-gain method for underwater acoustic signal detection in intensity fluctuation environments. Appl Acoust 211:109513

Collins MD, Kuperman WA (1991) Focalization: environmental focusing and source localization. J Acoust Soc Am 90(3):1410–1422

Collins MD, Kuperman WA, Schmidt H (1992) Nonlinear inversion for ocean-bottom properties. J Acoust Soc Am 92(5):2770–2783

Dong YF, Shen XH, Jiang Z, Wang HY (2021) Recognition of imbalanced underwater acoustic datasets with exponentially weighted cross-entropy loss. Appl Acoust 174:107740

Dong YF, Shen XH, Wang HY (2022) Bidirectional denoising autoencoders-based robust representation learning for underwater acoustic target signal denoising. IEEE Trans Instrum Meas 71:1–8

Dosso SE, Wilmut MJ (2008) Uncertainty estimation in simultaneous Bayesian tracking and environmental inversion. J Acoust Soc Am 124(1):82–97

Dosso SE, Wilmut MJ (2009) Comparison of focalization and marginalization for Bayesian tracking in an uncertain ocean environment. J Acoust Soc Am 125(2):717–722

Feng S, Zhu XQ (2022) A transformer-based deep learning network for underwater acoustic target recognition. IEEE Geosci Remote Sens Lett 19:1–5

Ferguson EL (2021) Multitask convolutional neural network for acoustic localization of a transiting broadband source using a hydrophone array. J Acoust Soc Am 150(1):248–256

Ferguson EL, Williams SB, Jin CT (2019) Convolutional neural network for single-sensor acoustic localization of a transiting broadband source in very shallow water. J Acoust Soc Am 146(6):4687–4698

Finette S, Mignerey PC (2018) Stochastic matched-field localization of an acoustic source based on principles of Riemannian geometry. J Acoust Soc Am 143(3):3628–3638

Frederick C, Villar S, Michalopoulou ZH (2020) Seabed classification using physics-based modeling and machine learning. J Acoust Soc Am 148(2):859–872

Fu Q, Song AJ (2018) Adaptive modulation for underwater acoustic communications based on reinforcement learning. OCEANS 2018 MTS/IEEE Charleston, Charleston, SC, USA, 22-25 October 2018, pp 1–8

Ge FX, Bai YY, Li MJ, Zhu GP, Yin JW (2022) Label distribution-guided transfer learning for underwater source localization. J Acoust Soc Am 151(6):4140–4149. https://doi.org/10.1121/10.0011741

Geng X, Zheng YR (2022) Exploiting propagation delay in underwater acoustic communication networks via deep reinforcement learning. IEEE Trans Neural Netw Learn Syst 1–12. https://doi.org/10.1109/TNNLS.2022.3170050

Gerstoft P (1994) Inversion of seismoacoustic data using genetic algorithms and a posteriori probability distributions. J Acoust Soc Am 95(2):770–782

Gingras DF, Gerstoft P (1995) Inversion for geometric and geoacoustic parameters in shallow water: experimental results. J Acoust Soc Am 97(6):3589–3598

Goldwater M, Zitterbart DP, Wright D, Bonnel J (2023) Machine-learning-based simultaneous detection and ranging of impulsive baleen whale vocalizations using a single hydrophone. J Acoust Soc Am 153(2):1094–1107

He Y, Han GJ, Jiang JF, Wang H, Martinez-Garcia M (2020) A trust update mechanism based on reinforcement learning in underwater acoustic sensor networks. IEEE Trans Mob Comput 21(3):811–821

Hemminger TL, Pao YH (1994) Detection and classification of underwater acoustic transients using neural networks. IEEE Trans Neural Netw 5(5):712–718

Huang ZQ, Xu J, Gong ZX, Wang HB, Yan YH (2018) Source localization using deep neural networks in a shallow water environment. J Acoust Soc Am 143(5):2922–2932

Irfan M, Zheng JB, Ali S, Iqbal M, Masood Z, Hamid U (2021) DeepShip: an underwater acoustic benchmark dataset and a separable convolution based autoencoder for classification. Expert Syst Appl 183:115270

Jiang JJ, Shi T, Huang M, Xiao ZZ (2020) Multi-scale spectral feature extraction for underwater acoustic target recognition. Measurement 166:108227

Jiang JJ, Wu ZZ, Lu JN, Huang M, Xiao ZZ (2021) Interpretable features for underwater acoustic target recognition. Measurement 173:108586

Jiang RK, Wang XT, Cao S, Zhao JF, Li XR (2019) Deep neural networks for channel estimation in underwater acoustic OFDM systems. IEEE Access 7:23579–23594

Jin GH, Liu F, Wu H, Song QZ (2020) Deep learning-based framework for expansion, recognition and classification of underwater acoustic signal. J Exp Theor Artif Intell 32(2):205–218

Jin SY, Su Y, Guo CJ, Fan YX, Tao ZY (2023) Offshore ship recognition based on center frequency projection of improved EMD and KNN algorithm. Mech Syst Signal Proc 189:110076

Ke XQ, Yuan F, Cheng E (2020) Integrated optimization of underwater acoustic ship-radiated noise recognition based on twodimensional feature fusion. Appl Acoust 159:107057

Khishe M (2022) Drw-AE: a deep recurrent-wavelet autoencoder for underwater target recognition. IEEE J Ocean Eng 47(4):1083–1098

Kim KI, Pak MI, Chon BP, Ri CH (2021) A method for underwater acoustic signal classification using convolutional neural network combined with discrete wavelet transform. Int J Wavelets Multiresolut Inf Proc 19(04):2050092

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444

Lefort R, Real G, Dremeau A (2017) Direct regressions for underwater acoustic source localization in fluctuating oceans. Appl Acoust 116:303–310

Li BS, Zhou SL, Stojanovic M, Freitag L, Willett P (2008) Multicarrier communication over underwater acoustic channels with nonuniform Doppler shifts. IEEE J Ocean Eng 33(2):198–209

Li DH, Liu F, Shen TS, Chen L, Zhao DX (2023a) Data augmentation method for underwater acoustic target recognition based on underwater acoustic channel modeling and transfer learning. Appl Acoust 208:109344

Li L, Song S, Feng X (2022a) Combined LOFAR and DEMON spectrums for simultaneous underwater acoustic object counting and F_0 estimation. J Mar Sci Eng 10(10):1565

Li P, Wu J, Wang YX, Lan Q, Xiao WB (2022b) STM: spectrogram transformer model for underwater acoustic target recognition. J Mar Sci Eng 10(10):1428

Li T, Sahu AK, Talwalkar A, Smith V (2020a) Federated learning: challenges, methods, and future directions. IEEE Signal Proc Mag 37(3):50–60

Li XL, Song WH, Gao DZ, Gao W, Wang HZ (2020b) Training a U-Net based on a random mode-coupling matrix model to recover acoustic interference striations. J Acoust Soc Am 147(4): EL363–EL369. https://doi.org/10.1121/10.0001125

Li XL, Wang PY, Song WH, Gao W (2023b) Modal wavenumber estimation by combining physical informed neural network. J Acoust Soc Am 153(5):2637–2648

Liu F, Shen TS, Luo ZL, Zhao DX, Guo SJ (2021a) Underwater target recognition using convolutional recurrent neural networks with 3-D Mel-spectrogram and data augmentation. Appl Acoust 178:107989

Liu L, Cai L, Ma L, Qiao G (2021b) Channel state information prediction for adaptive underwater acoustic downlink OFDMA system: deep neural networks based approach. IEEE Trans Veh Technol 70(9):9063–9076

Liu MD, Niu HQ, Li ZL, Liu YN, Zhang QQ (2022) Deep-learning geoacoustic inversion using multi-range vertical array data in shallow water. J Acoust Soc Am 151(3):2101–2116

Liu WX, Yang YX, Xu MQ, Lü LG, Liu ZW, Shi Y (2020a) Source localization in the deep ocean using a convolutional neural network. J Acoust Soc Am 147(4):EL314–EL319

Liu YN, Niu HQ, Li ZL (2020b) A multi-task learning convolutional neural network for source localization in deep ocean. J Acoust Soc Am 148(2):873–883. https://doi.org/10.1121/10.0001762

Liu YN, Niu HQ, Li ZL, Wang MY (2021c) Deep-learning source localization using autocorrelation functions from a single hydrophone in deep ocean. JASA Express Lett 1(1):036002. https://doi.org/10.1121/10.0003647

Liu YN, Niu HQ, Yang SS, Li ZL (2021d) Multiple source localization using learning-based sparse estimation in deep ocean. J Acoust Soc Am 150(5):3773–3786. https://doi.org/10.1121/10.0007276

Lucas E, Wang ZH (2020) Supervised learning for performance prediction in underwater acoustic communications. Global Oceans 2020: Singapore–U.S. Gulf Coast, Biloxi, MS, USA, 05-30 October 2020, pp 1–6

Luo XW, Feng YL, Zhang MH (2021) An underwater acoustic target recognition method based on combined feature with automatic coding and reconstruction. IEEE Access 9:63841–63854

Mary DRK, Ko E, Kim SG, Yum SH, Shin SY, Park SH (2021) A systematic review on recent trends, challenges, privacy and security issues of underwater internet of things. Sensors 21(24):8262

McMahan HB, Moore E, Ramage D, Hampson S, Arcas BAY (2017) Communication-efficient learning of deep networks from decentralized data. Artificial intelligence and statistics, Fort Lauderdale, FL, USA, 20-22 April 2017, pp 1273–1282

Michalopoulou ZH, Alexandrou D, de Moustier C (1995) Application of neural and statistical classifiers to the problem of seafloor characterization. IEEE J Ocean Eng 20(3):190–197

Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG et al (2015) Human-level control through deep reinforcement learning. Nature 518(7540):529–533

Neilsen TB, Escobar-Amado CD, Acree MC, Hodgkiss WS, Van Konman DF, Knobles DP et al (2021) Learning location and seabed type from a moving mid-frequency source. J Acoust Soc Am 149(1):692–705

Niu HQ, Gerstoft P, Ozanich E, Li ZL, Zhang RH, Gong ZX et al (2020) Block sparse Bayesian learning for broadband mode extraction in shallow water from a vertical array. J Acoust Soc Am 147(6):3729–3739

Niu HQ, Gong ZX, Ozanich E, Gerstoft P, Wang HB, Li ZL (2019a) Deep-learning source localization using multi-frequency magnitude-only data. J Acoust Soc Am 146(1):211–222. https://doi.org/10.1121/1.5116016

Niu HQ, Li ZL, Wang HB, Gong ZX (2019b) Overview of machine learning methods in underwater source localization. J Signal Proc 35(9):1450–1459. https://doi.org/10.16798/j.issn.1003-0530.2019.09.002

Niu HQ, Ozanich E, Gerstoft P (2017a) Ship localization in Santa Barbara Channel using machine learning classifiers. J Acoust Soc Am 142(5): EL455–EL460. https://doi.org/10.1121/1.5010064

Niu HQ, Reeves E, Gerstoft P (2017b) Source localization in an ocean waveguide using supervised machine learning. J Acoust Soc Am 142(3):1176–1188. https://doi.org/10.1121/1.5000165

Ozard JM, Zakarauskas P, Ko P (1991) An artificial neural network for range and depth discrimination in matched field processing. J Acoust Soc Am 90:2658–2663

Piccolo J, Haramuniz G, Michalopoulou ZH (2019) Geoacoustic inversion with generalized additive models. J Acoust Soc Am 145(6):EL463–EL468

Qarabaqi P, Stojanovic M (2013) Statistical characterization and computationally efficient modeling of a class of underwater acoustic communication channels. IEEE J Ocean Eng 38(4):701–717

Ren JW, Xie Y, Zhang XW, Xu J (2022) UALF: A learnable front-end for intelligent underwater acoustic classification system. Ocean Eng 264:112394

Santos-Domínguez D, Torres-Guijarro S, Cardenal-Lopez A, Pena-Gimenez A (2016) ShipsEar: an underwater vessel noise database. Appl Acoust 113:64–69

Shen YN, Pan X, Zheng Z, Gerstoft P (2020) Matched-field geoacoustic inversion based on radial basis function neural network. J Acoust Soc Am 148(5):3279–3290

Sozer EM, Stojanovic M, Proakis JG (2000) Underwater acoustic networks. IEEE J Ocean Eng 25(1):72–83

Steinberg BZ, Beran MJ, Chin SH, Howard JH (1991) A neural network approach to source localization. J Acoust Soc Am 90:2081–2090

Stephan Y, Demoulin X, Sarzeaud O (1998) Neural direct approaches for geoacoustic inversion. J Comput Acoust 6(1–2):151–166

Van Komen DF, Neilsen TB, Howarth K, Knobles DP, Dahl PH (2020) Seabed and range estimation of impulsive time series using a convolutional neural network. J Acoust Soc Am 147(5):EL403–EL408

van Walree PA, Socheleau FX, Otnes R, Jenserud T (2017) The watermark benchmark for underwater acoustic modulation schemes. IEEE J Ocean Eng 42(4):1007–1018

Wang H, Wang B, Li Y (2022a) IAFNet: few-shot learning for modulation recognition in underwater impulsive noise. IEEE Commun Lett 26(5):1047–1051

Wang MF, Zhu ZJ, Qian GF (2023) Modulation signal recognition of underwater acoustic communication based on Archimedes Optimization Algorithm and Random Forest. Sensors 23(5):2764

Wang RN, Yadav A, Makled EA, Dobre OA, Zhao RQ, Varshney PK (2020a) Optimal power allocation for full-duplex underwater relay networks with energy harvesting: a reinforcement learning approach. IEEE Wirel Commun Lett 9(2):223–227

Wang S, Zeng X (2014) Robust underwater noise targets classification using auditory inspired time-frequency analysis. Appl Acoust 78:68–76

Wang WB, Li SC, Yang JS, Liu Z, Zhou WC (2016) Feature extraction of underwater target in auditory sensation area based on MFCC. IEEE/OES China Ocean Acoustics (COA), Harbin, China, 09-11 January 2016, pp 1–6

Wang WB, Ni HY, Su L, Hu T, Ren QY, Gerstoft P et al (2019a) Deep transfer learning for source ranging: deep-sea experiment results. J Acoust Soc Am 146(4): EL317–EL322. https://doi.org/10.1121/1.5126923

Wang WB, Wang Z, Su L, Hu T, Ren QY, Gerstoft P et al (2020b) Source depth estimation using spectral transformations and convolutional neural network in a deep-sea environment. J Acoust Soc Am 148(6):3633–3644

Wang XM, Liu AH, Zhang Y, Xue FZ (2019b) Underwater acoustic target recognition: a combination of multi-dimensional fusion features and modified deep neural network. Remote Sens 11(16):1888

Wang XM, Meng JX, Liu YT, Zhan G, Tian ZN (2022b) Self-supervised acoustic representation learning via acoustic-embedding memory unit modified space autoencoder for underwater target recognition. J Acoust Soc Am 152(5):2905–2915

Wang Y, Peng H (2018) Underwater acoustic source localization using generalized regression neural network. J Acoust Soc Am 143(4):2321–2331

Weiss K, Khoshgoftaar TM, Wang DD (2016) A survey of transfer learning. J Big Data 3(1):1–40

Xiao L, Jiang DH, Chen Y, Su W, Tang YL (2019) Reinforcement-learning-based relay mobility and power allocation for underwater sensor networks against jamming. IEEE J Ocean Eng 45(3):1148–1156

Xie Y, Ren JW, Xu J (2022a) Underwater-art: expanding information perspectives with text templates for underwater acoustic target recognition. J Acoust Soc Am 152(5):2641–2651

Xie Y, Ren JW, Xu J (2022b) Adaptive ship-radiated noise recognition with learnable fine-grained wavelet transform. Ocean Eng 265:112626

Xu J, Xie Y, Wang WC (2023) Underwater acoustic target recognition based on smoothness-inducing regularization and spectrogram-based data augmentation. Ocean Eng 281:114926

Yang H, Lee K, Choo Y, Kim K (2020) Underwater acoustic research trends with machine learning: passive SONAR applications. J Ocean Eng Technol 34(3):227–236

Yang ZD, Huo LS, Wang JK, Zhou J (2022) Denoising low SNR percussion acoustic signal in the marine environment based on the LMS algorithm. Measurement 202:111848

Yangzhou JY, Ma ZY, Huang X (2019) A deep neural network approach to acoustic source localization in a shallow water tank experiment. J Acoust Soc Am 146(6):4802–4811

Yu QK, Zhu M, Zhang W, Shi J, Liu Y (2023) Surface and underwater acoustic source recognition using multi-channel joint detection method based on machine learning. J Mar Sci Eng 11(8):1587

Zeng X, Wang S (2014) Underwater sound classification based on Gammatone filter bank and Hilbert-Huang transform. International Conference on Signal Processing, Communications and Computing (ICSPCC), Guilin, China, 05-08 August 2014, pp 707–710

Zhang W, Wu YQ, Shi J, Leng HZ, Zhao Y, Guo JZ (2022a) Surface and underwater acoustic source discrimination based on machine learning using a single hydrophone. J Mar Sci Eng 10(3):321

Zhang WL, Yang XH, Leng CL, Wang JJ, Mao SW (2022b) Modulation recognition of underwater acoustic signals using deep hybrid neural networks. IEEE Trans Wirel Commun 21(8):5977–5988

Zhang Y, Wang H, Li C, Meriaudeau F (2022c) Complex-valued deep network aided channel tracking for underwater acoustic communications. OCEANS Conference, Chennai, India, 21-24 February 2022, pp 1–5

Zhang YL, Li C, Wang HB, Wang J, Yang F, Meriaudeau F (2022d) Deep learning aided OFDM receiver for underwater acoustic communications. Appl Acoust 187:108515

Zhang YL, Wang HB, Li C, Chen DS, Meriaudeau F (2021a) Meta-learning-aided orthogonal frequency division multiplexing for underwater acoustic communications. J Acoust Soc Am 149(6):4596–4606

Zhang YL, Wang HB, Li C, Chen X, Meriaudeau F (2022e) On the performance of deep neural network aided channel estimation for underwater acoustic OFDM communications. Ocean Eng 259:111518

Zhang YL, Wang HB, Li C, Meriaudeau F (2022f) Data augmentation aided complex-valued network for channel estimation in underwater acoustic orthogonal frequency division multiplexing system. J Acoust Soc Am 151(6):4150–4164

Zhang YL, Wang HB, Tai YP, Li C, Meriaudeau F (2021b) A machine learning label-free method for underwater acoustic OFDM channel estimations. 15th International Conference on Underwater Networks Systems, Shenzhen, China, 23-26 November 2021, pp 1–5

Zhang YW, Li JX, Zakharov Y, Li X, Li JH (2019) Deep learning based underwater acoustic OFDM communications. Appl Acoust 154:53–58

Zhang YZ, Chang JZ, Liu Y, Xing LY, Shen XH (2023) Deep learning and expert knowledge based underwater acoustic OFDM receiver. Phys Commun 58:102041

Zhang YZ, Zhu JR, Wang HY, Shen XH, Wang B, Dong Y (2022g) Deep reinforcement learning-based adaptive modulation for underwater acoustic communication with outdated channel state information. Remote Sens 14(16):3947

Zhao H, Ji F, Li Q, Guan QS, Wang S, Wen MW (2022) Federated meta-learning enhanced acoustic radio cooperative framework for ocean of things. IEEE J Sel Top Signal Proc 16(3):474–486

Zhou XY, Yang KD (2020) A denoising representation framework for underwater acoustic signal recognition. J Acoust Soc Am 147(4):EL377–EL383

Publisher's Note