remote sensing

MDPI

Review

Artificial Intelligence-Based Underwater Acoustic Target Recognition: A Survey

Sheng Feng [1], Shuqing Ma [2], Xiaoqian Zhu [2,*] and Ming Yan [2]

[1] College of Computer Science, National University of Defense Technology, Changsha 410073, China; fengsh14@lzu.edu.cn
[2] College of Meteorology and Oceanography, National University of Defense Technology, Changsha 410073, China; mashuqing@nudt.edu.cn (S.M.)
* Correspondence: zhu_xiaoqian@sina.com

Abstract: Underwater acoustic target recognition has always played a pivotal role in ocean remote sensing. By analyzing and processing ship-radiated signals, it is possible to determine the type and nature of a target. Historically, traditional signal processing techniques have been employed for target recognition in underwater environments, which often exhibit limitations in accuracy and efficiency. In response to these limitations, the integration of artificial intelligence (AI) methods, particularly those leveraging machine learning and deep learning, has attracted increasing attention in recent years. Compared to traditional methods, these intelligent recognition techniques can autonomously, efficiently, and accurately identify underwater targets. This paper comprehensively reviews the contributions of intelligent techniques in underwater acoustic target recognition and outlines potential future directions, offering a forward-looking perspective on how ongoing advancements in AI can further revolutionize underwater acoustic target recognition in ocean remote sensing.

Keywords: literature review; machine learning; deep learning; ocean remote sensing; underwater target recognition

check for updates

Citation: Feng, S.; Ma, S.; Zhu, X.; Yan, M. Artificial Intelligence-Based Underwater Acoustic Target Recognition: A Survey. *Remote Sens.* **2024**, *16*, 3333. https://doi.org/10.3390/rs16173333

Academic Editors: Haiyong Xu, Jingyue Lu, Lei Zhang, Marc Pinto, Farman Ali and Xuebo Zhang

Received: 7 August 2024
Revised: 5 September 2024
Accepted: 6 September 2024
Published: 8 September 2024

1. Introduction

Ships navigating in the ocean inevitably emit sound waves. These emitted underwater signals typically contain discriminative characteristics that are critical for classifying different ships, such as the ship structure, propeller configuration, engine power, and ship condition. Common UATR tasks include vessel type classification [1–3], sonar imagery recognition [4–8], and recognition of communication signal modulation [9,10]. As an important branch of ocean remote sensing, underwater acoustic target recognition (UATR) technology has a wide range of marine applications, including marine resource development, ocean rescue, biological protection, and military applications. Its main purpose is to process and analyze sonar-received signals in order to determine the target type based on discriminative characteristics. From the perspective of pattern recognition, the task of UATR fundamentally addresses the challenge of enhancing separability among different types of ships, that is, how to pull homologous samples closer and push heterologous samples apart to construct a compact set. As shown in Figure 1, it is always possible to find separable boundaries between different ship classes for compact sets, whereas for noncompact sets, it is difficult to find an obvious separable boundary.

In fact, the inherent complexity and uncertainty of the underwater environment, including the sound sources, sea surface, water column, and seabed sediments, dictates the unique characteristics of underwater sound propagation, which poses numerous challenges for UATR. First of all, the speed of underwater sound is influenced by factors such as water temperature, density, and depth, leading to refraction where sound paths bend towards regions of lower sound speed. As illustrated in Figure 2, underwater sound signals are characterized by multipath propagation, along with effects such as attenuation, seabed penetration, scattering, and reverberation [11].

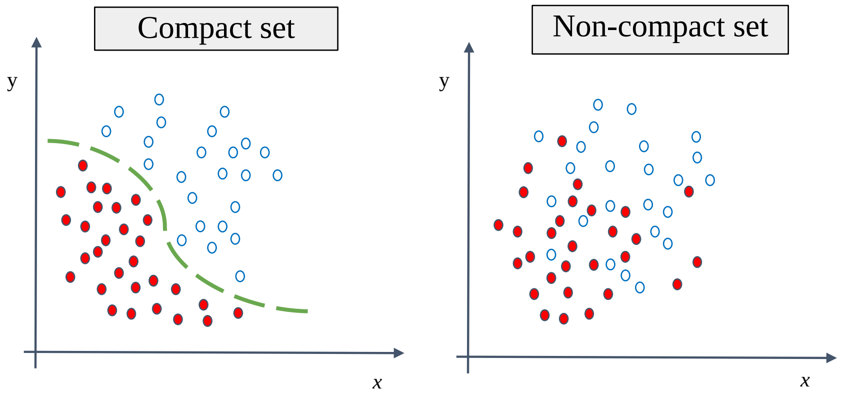

Figure 1. Schematic illustration of the separability between underwater signals from the perspective of pattern recognition.

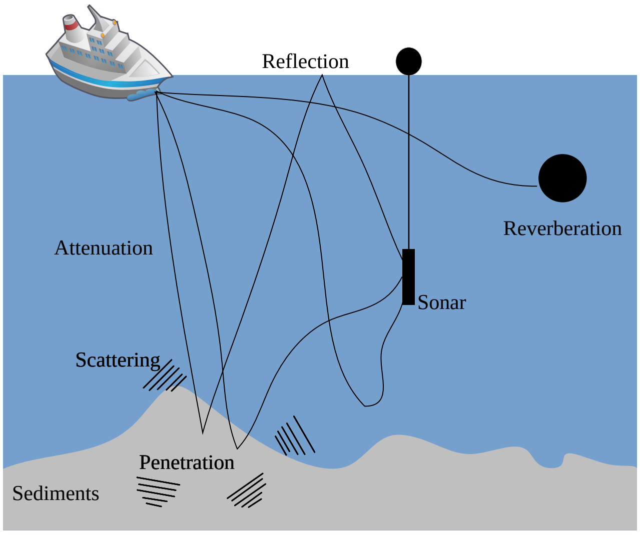

Figure 2. Schematic diagram of a typical underwater signal propagation channel.

Secondly, the unique acoustic characteristics of underwater targets are highly diverse, encompassing a variety of ship types and marine animals, each with distinctive acoustic signatures and complex sound generation mechanisms that pose challenges to recognition models. Coupled with the variability of these acoustic features against the ocean background noise, the task of UATR becomes exceptionally challenging.

In the early days, UATR methods primarily relied on experienced sonar operators to make manual determinations, including auditory and visual recognition. These operators were trained to recognize enemy submarines in ambient noise and even to distinguish marine animals from potential threats. This manual approach, combining auditory perception with visual interpretation of sonar displays, defined the early methods of UATR. However, the reliance on human interpretation not only limited the recognition speed but also introduced a significant error due to various psychological and physiological factors, which may greatly reduce the recognition accuracy. On the other hand, with the advancement of ship-radiated noise reduction, the increasing trend towards low-frequency characteristics further degrades the effectiveness of traditional methods. To address this issue, artificial intelligence (AI)-based UATR methods are rapidly developing [12], which attempt to extract meaningful knowledge from extensive training samples to find separable boundaries between different targets and subsequently recognize new samples based on the acquired knowledge. Generally, these intelligent recognition techniques have been implemented in sonar devices, demonstrating autonomous, efficient, and accurate performance in comparison to conventional methods [3].

Figure 3 illustrates the workflow of a typical intelligent UATR system, including data collection, data preprocessing, feature extraction, and classification algorithms. Among these processes, data collection is conducted by sonar operators using hydrophone arrays to acquire acoustic data. Subsequently, data preprocessing is implemented to mitigate clutter interference and to eliminate the effects of Doppler shift [13–15]. Typical preprocessing techniques include resampling, filtering, and denoising, which contribute to obtain high-quality signal data. Importantly, the process of feature extraction and the design of classification algorithms are considered to be crucial aspects that affect the recognition performance.

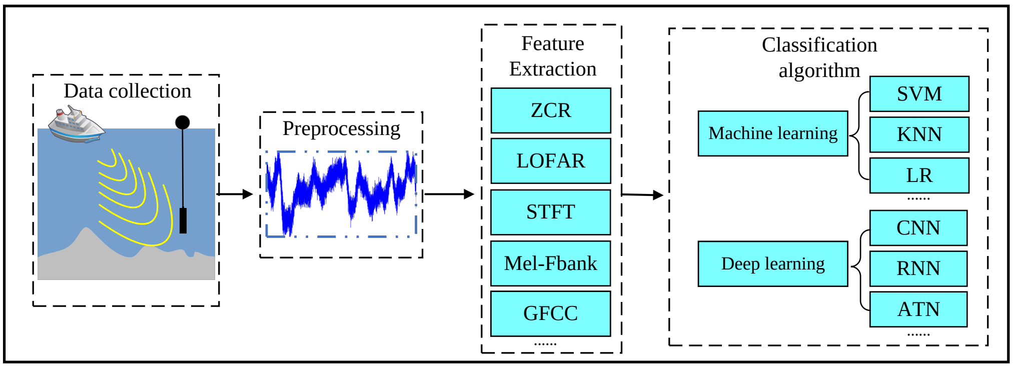

Figure 3. The workflow of a typical intelligent UATR system.

Feature extraction is typically associated with underwater acoustic signal processing techniques to extract the intrinsic physical information of the target. It is particularly crucial in addressing the second major challenge of target diversity. By employing a robust feature extraction method, it is possible to obtain discriminative features capable of distinguishing between various targets, thereby mitigating the complexity introduced by the wide range of underwater acoustic sources. In these methods, time-domain characteristics can offer straightforward acoustic representations, such as zero-crossing rate (ZCR), peak-to-peak amplitude, and waveform energy. Nonetheless, these features may not capture the time-varying characteristics that are crucial for achieving accurate UATR. To address this limitation, joint time–frequency techniques have been introduced into UATR. These techniques aim to extract discriminative features containing both temporal and spectral information, such as the short-time Fourier transform (STFT), low-frequency analysis and recording (LOFAR), Mel-filter bank (Mel-Fbank), and Gammatone Frequency Cepstral Coefficients (GFCCs).

For classification algorithms, machine learning (ML) methods were firstly introduced into the field of intelligent UATR, such as support vector machine (SVM), *K* nearest neighbor algorithm (*K*NN), hidden Markov model (HMM), and logistic regression (LR). Due to the relatively simple structure of ML models, these methods primarily utilize shallow signal features for intelligent recognition of simple targets, such as whale calls [16] and medium echoes [17]. As a subset of ML, deep learning (DL) methods are capable of learning deep representations from complex acoustic signals using deep neural network (DNNs). Composed of stacked layers, DNNs are particularly effective in automatically extracting hierarchical features from acoustic data. Recent research has extensively investigated DL-based UATR methods, demonstrating their superiority over shallow ML classifiers by leveraging high-level information inherent in acoustic signals [18,19]. Initially, early DL-based intelligent UATR systems were constrained by computational resources, particularly limited GPU memory, which made it difficult to perform fast and stable DNN training.

However, in the early 21st century, significant advancements in electronic technology were made, notably enhanced computational capabilities of GPUs. Consequently, the application of advanced DL techniques has made significant progress in intelligent UATR, thereby making substantial contributions to the field of ocean remote sensing.

In this paper, we provide a comprehensive review of two pivotal components within AI-based UATR systems: commonly utilized feature extraction methods and classification algorithms. These components represent the mainstream of ocean remote sensing due to their significant potential and impact. Furthermore, we highlight major challenges in the current field of intelligent UATR and propose potential solutions. The structure of this paper is as follows: Section 2 investigated widely utilized feature extraction methods to enhance separability in UATR. Section 3 discusses the shallow ML algorithms contributions to UATR. In Section 4, a wide range of DL applications in UATR are presented. Finally, Section 5 outlines prospective research directions aimed at proposing potential solutions to current challenges in AI-based UATR.

2. Feature Extraction Methods

As is well-known, each ship target exhibits a unique underwater acoustic pattern, with recognition performance largely dependent on feature selection. In this case, feature extraction techniques are pivotal for enhancing recognition performance. Ideally, the feature representation should be highly discriminative to provide sufficient information for classification algorithms. In such cases, it can effectively reduce intra-class variance, and increase inter-class variance, to find the compact set of acoustic signals. In particular, the waveform envelope of underwater acoustic signals, known as a kind of natural time series data, carries the inherent information of acoustic targets, providing direct insights into their temporal characteristics. Time-domain feature extraction methods from signal waveforms are closely related to the physical properties and dynamic behavior of underwater acoustic targets [20]. Deng et al. [21] proposed an integrated model-based feature extraction method that combines broadband–phase correlation and spatial distribution features within an embedded recognition system; real-time experiments demonstrated its effectiveness. In addition, Meng et al. [22] introduced a feature extraction method based on time-domain waveform structure, which achieved more than 89.5% recognition accuracy based on statistical zero wavelength and peak amplitude features using ML algorithms. Sun and Zhang [23] utilized time series chaotic dynamics to reconstruct the phase space of the ship-radiated noise. They utilized relevant dimensions and the maximum Lyapunov exponent features, thereby achieving successful UATR. While single-domain statistical feature extraction methods have shown promising results on small-scale datasets, they are also subject to limitations. Specifically, for underwater acoustic signals, time-domain analysis effectively captures time-varying characteristics but lacks frequency information. Conversely, frequency-domain analysis reveals the signal's frequency components but disregards temporal information. In AI-based UATR, researchers have increasingly focused on discriminative spectral features, which can be broadly categorized into three types: the first category is the interpretable spectral features that hold physical significance, which investigate the physical structure and motion state of the target ship. The second category deals with the problem of insufficient single-domain features by using time–frequency joint features, where time–frequency analysis techniques are used to integrate the temporal and spectral domain information of signals. The final category concerns autocoding features, which involve end-to-end learning of feature representations utilizing DNNs or other methods without manual feature engineering.

2.1. Physical Significance Feature

In marine environments, ship-radiated noise mainly includes three types: mechanical noise, hydrodynamic noise, and propeller noise. Mechanical noise is produced by the vibrations of the internal ship equipment, such as ventilation and pumps, which typically exhibit unique noise patterns with prominent line spectra. Hydrodynamic noise refers to noise generated by the interaction between fluid flow and ship hull, which is relatively smooth and characterized by a continuous spectrum. Propeller noise constitutes the main components of the ship-radiated noise, which has a wide frequency range from 5 Hz to 100 KHz. This type of noise is notable for its significant low-frequency line spectra and continuous spectrum characteristics. Let $s(t)$ represent ship noise, given by

$$\{s(t)\} = \{x(t)\} + \sum_{i=1}^{n}\{l_i(t)\}, \tag{1}$$

where $\{x(t)\}$ denotes wide-sense stationary random processes, $l_i(t)$ denotes initially random periodic signals, and $i = 1, 2, \ldots, n$, representing n periodic signals. Then we have the power spectrum:

$$S(f) = \lim_{T\to\infty} \frac{1}{T} E\left(\left|S_{k,T}(f)\right|^2\right), \tag{2}$$

where T denotes the length of the signal segment during Fourier transformation; E denotes ensemble averaging; and k represents the segment number of the signal.

In summary, the aforementioned physical components can reveal the physical structure and motion characteristics of the ship with robustness, providing useful information of significant physical relevance. To capture this meaningful information, the LOFAR [18,24] and the DEMON spectra [25] are commonly used in UATR studies. These two types of spectrograms are illustrated in Figure 4.

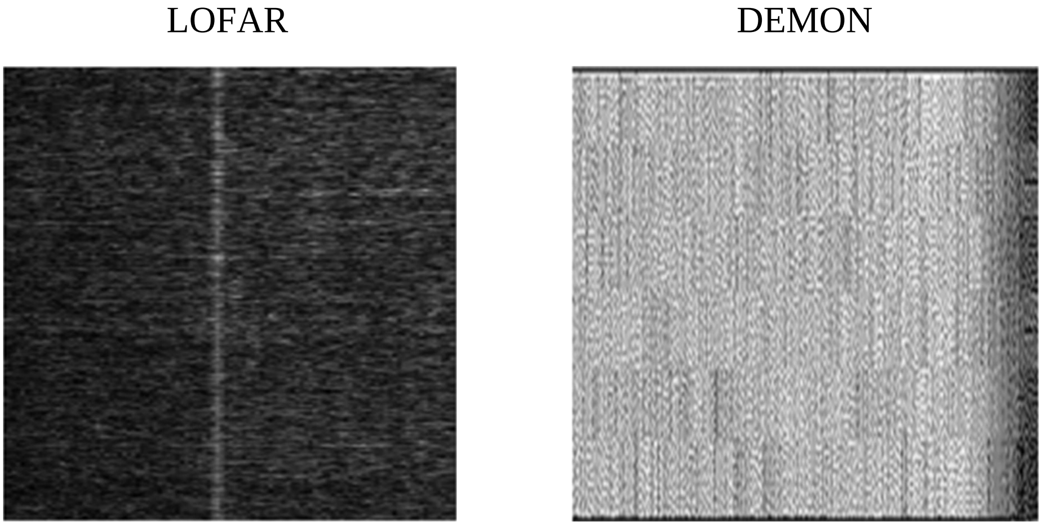

Figure 4. Physical significance feature of an underwater signal sample, including its LOFAR and DEMON [26] spectrograms.

2.1.1. LOFAR Physical Spectrum

Known for its capability to depict low-frequency line spectra, the LOFAR spectrum is particularly useful for recognizing underwater acoustic targets [26,27]. The main process to extract the LOFAR spectrum can be summarized as follows:

(1) The signal sequence is firstly framed to obtain L overlapped samples. Common framing functions include the Hanning or Hamming windows.

(2) Each framed signal undergoes normalization and centering:

$$u_j(n) = \frac{M_j(n)}{\max\left(M_j(n)\right)}, \quad 1 \leqslant j \leqslant L, \tag{3}$$

$$x_j(n) = u_j(n) - \frac{1}{L}\sum_{i=1}^{L} u_j(i). \tag{4}$$

(3) The short-time Fourier transform (STFT) is applied to signal $x_j(n)$ to obtain the LOFAR spectrum:

$$X_j(k) = FFT[x_j(n)]. \tag{5}$$

These spectra are arranged over time in a coordinate system to generate the complete LOFAR spectrogram. Due to the predominant low-frequency range of ship-radiated noise, the LOFAR spectrogram plays a critical role in analyzing the characteristics of underwater targets. The bright spectral lines in the spectrogram often correspond to the most important features of underwater acoustic targets. Jin et al. [26] utilized an adversarial generative network (GAN) to generate simulated samples based on their LOFAR spectrogram for data augmentation, which has proven its effectiveness to solve the few-shot UATR problem. In addition, Shi et al. [28] analyzed the line spectral traces on the LOFAR spectrogram and calculated the relative variance between adjacent time frames to enhance the detection sensitivity of low-intensity moving targets.

2.1.2. DEMON Physical Spectrum

In UATR, the DEMON spectrum also plays a crucial role due to its ability to extract detailed physical information from acoustic signals, such as propeller blade count and shaft frequency. This technique is specifically designed to detect modulations originating from propeller cavitation envelopes [29]. In practice, the DEMON spectrum entails the selection of the spectral band exhibiting the most pronounced time-varying modulation for envelope detection, which is subsequently followed by detailed spectral analysis to derive relevant parameters.

Specifically, the amplitude of broadband noise $n(t)$ generated by propeller cavitation is modulated by a set of sinusoidal signals. This broadband signal can be represented as

$$x(n) = \left(1 + \sum_{n=1}^{N} A_n \sin(2\pi n f_0 + \phi_n)\right) s(n) + g(n), \tag{6}$$

where A is the modulation amplitude, nf_0 is the modulation frequency, $s(n)$ denotes the broadband noise signal, and $g(t)$ represents background noise. After bandpass filtering using the coefficient $h_1(n)$, $y_1(n)$ is computed as

$$y_1(n) = \sum_{k=0}^{M-1} x(n-k)h_1(k), \tag{7}$$

Then, the Hilbert transform H is further utilized for envelope demodulation:

$$y_2(n) = |H(y_1(n))|. \tag{8}$$

After removing the DC component to obtain $y_3(n)$, subsequent downsampling and low-pass filtering are performed to yield $y_4(n)$:

$$y_4(n) = \sum_{k=0}^{P-1} y_3(dn-k)h_2(k), \tag{9}$$

where d denotes the downsampling factor and $h_2(k)$ represents the low-pass filter coefficients. Further spectral analysis using $y_4(n)$ provides power spectral density for extracting line spectra. These procedures facilitate the extraction of crucial physical parameters related to underwater targets. However, the complexity introduced by interference line spectra within DEMON spectrograms can degrade recognition performance under low signal-to-noise ratios (SNRs). To address this problem, Tong et al. [30] proposed the adaptive weighted envelope spectrum (AWES) method for better demodulation capability, which dynamically assigns weights to spectral components based on their modulation intensities. Furthermore, Li et al. emphasized that using LOFAR or DEMON alone is insufficient for estimating arbitrary rotation speeds in target detection tasks [31]. To address this issue, they combined the LOFAR and DEMON spectrograms to enhance ship estimation under various conditions. Figure 5 illustrates an example of a fused LOFAR and DEMON spectrogram after filtering. Experimental results demonstrate that feeding the comb-filtered combined feature (LOFAR + DEMON) into the DL model achieves the highest accuracy of 98%, outperforming the individual filtered LOFAR (83%) or DEMON (94%).

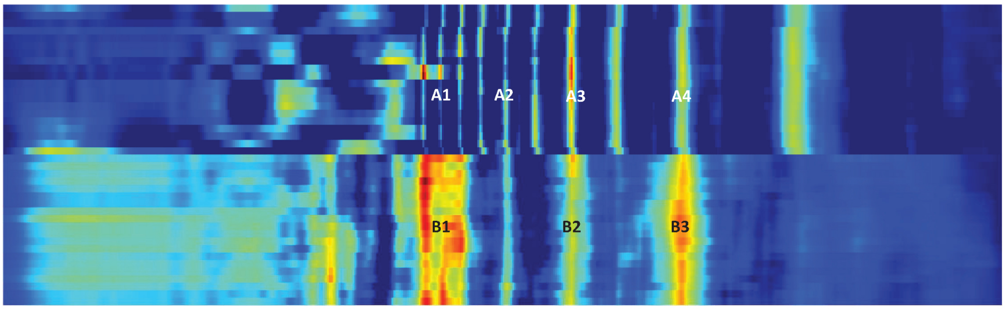

Figure 5. An example of a fused LOFAR and DEMON spectrogram with comb filtering [31]. A1–A4 and B1–B3 represent the primary spectral lines.

2.2. Joint Time–Frequency Feature

Numerous studies have shown that underwater acoustic signals exhibit nonstationary characteristics [32–34], with their frequencies varying significantly over time. Traditional Fourier transform methods only provide average information about the signal without revealing the temporal evolution of signal frequencies. To address this problem, researchers have attempted to represent the signal within a joint time–frequency domain, and have proposed a variety of time–frequency analysis techniques [35]. The objective of these techniques is to construct a joint time–frequency distribution function $F_x(t, f)$ that can well characterize the time-varying properties of the signal $x(t)$, such as the energy variation with time t and frequency f. This robust feature, used in the processing process, can provide dynamic time-varying characteristics for subsequent classification algorithms [36,37]. The commonly employed time–frequency features in UATR mainly include general time–frequency feature extraction, auditory perception features, and multidimensional features.

2.2.1. General Time–Frequency Feature

In acoustic processing, STFT is a fundamental method for analyzing underwater acoustic signals, providing a basis for the development of various time–frequency analysis methods, which assumes signal stationarity within a narrow window. Specifically, for a continuous nonstationary signal $x(t)$, given a narrow window function w sliding along the time axis, the STFT of $x(t)$ can be defined as

$$\mathrm{STFT}_x(t, \omega) = \int_{-\infty}^{\infty} x(\tau) w^*(\tau - t) e^{-j\omega\tau} d\tau, \tag{10}$$

where ω represents the angular frequency and $*$ denotes the complex conjugate. To further enrich information representation, Zhang et al. [38] proposed a multiscale STFT feature extraction method by expanding the spectrogram channel. Experimental results have illustrated its effectiveness in low-frequency information extraction, thereby improving recognition performance on real-world datasets. However, the fixed window size in STFT limits adaptive adjustment. If the window function w is too narrow, high time resolution captures snapshot characteristics, but frequency resolution is reduced due to short window segments. Conversely, if w is too wide, time domain precision is reduced, resulting in lower time resolution that fails to capture rapid signal changes. To address this issue, high-resolution time–frequency feature extraction methods have been applied to UATR [39,40]. In the early 1980s, wavelet transform was introduced as a mathematical tool for analyzing geophysical signals [41], which adapts its window size with frequency changes, making it suitable for analyzing nonstationary signals. In continuous wavelet transform (CWT), a time dilation operator is employed, defined as the time scale. CWT maps one-dimensional square integrable functions (energy-limited signals) into a two-dimensional function of time scale a and time shift b. For a square integrable function $y(t)$, its CWT with respect to the mother wavelet $\psi(t)$ can be defined as

$$W_{\psi}y(a,b) = \langle y, \psi_{a,b} \rangle = \int_{-\infty}^{+\infty} y(t) \frac{1}{\sqrt{|a|}} \psi^{*} \cdot \left(\frac{t-b}{a} \right) \mathrm{d}t, \tag{11}$$

and the mother wavelet function $\psi(t)$ is derived from $\psi_{a,b}$ through affine transformations:

$$\psi_{a,b}(t) = [U(a,b)\psi(t)] = \frac{1}{\sqrt{|a|}} \psi \left(\frac{t-b}{a} \right). \tag{12}$$

Here, $U(a,b)$ denotes affine transformations of the $\psi(t)$, while the unitary transformation ensures energy preservation with a normalization factor of $\frac{1}{\sqrt{|a|}}$. Clearly, when $a > 1$, the wavelet function expands, thereby reducing its frequency, enabling the measurement of lower-frequency components; when $a < 1$, the wavelet function compresses, increasing its frequency and facilitating the measurement of higher-frequency components using wavelet transform. Li et al. [42] proposed a wavelet packet transform-based algorithm for extracting features from underwater radiated noise. This method decomposed signals into four wavelet packets and used energy features from each frequency band as inputs to a Radial Basis Function (RBF) network classifier. Experimental results highlighted that wavelet packet transform can enhance signal fractal characteristics during feature reduction, thus effectively accomplishing diverse recognition tasks. Rademan et al. [43] applied continuous wavelet transform to extract discriminative features from whale calls, achieving superior detection accuracy and specificity compared to STFT methods on simulated datasets. Furthermore, other high-frequency analysis methods such as Hilbert–Huang transform (HHT) [44] and Wigner–Ville distribution (WVD) [45,46] have also been employed for feature extraction. These time–frequency methods hold potential for providing discriminative features, selected based on task requirements, resource constraints, and specific advantages in various recognition scenarios. STFT, valued for its simplicity and real-time capabilities, is commonly used in real-time sonar monitoring and rapid target detection. In contrast, the wavelet transform has multiscale analysis capabilities and performs well in noisy environments with reverberation effects or with target features at different scales, effectively revealing temporal variations. For scenarios involving rapid moving target resulting in frequency modulation or Doppler effects, HHT can provide detailed analysis of target motion dynamics. Moreover, WVD has demonstrated effectiveness in applications demanding high time–frequency resolution, revealing intricate details that are difficult to achieve with other methods by properly handling cross-terms.

2.2.2. Auditory Perceptual Features

In modern ocean engineering, considerable research focuses on auditory perceptual features that closely approximate human auditory perception [47–50]. Among these features, Mel frequency-based auditory perceptual features are mostly utilized in UATR [51]. This method utilizes f_{mel} for triangular filtering and compression to derive Mel-Fbank coefficients from radiated noise. Figure 6 illustrates the response function of equi-height Mel-Fbank, composed of triangular filters symmetrically responding around their central frequencies. Specifically, filters exhibit maximal response at their center frequencies and gradually attenuate symmetrically on either side.

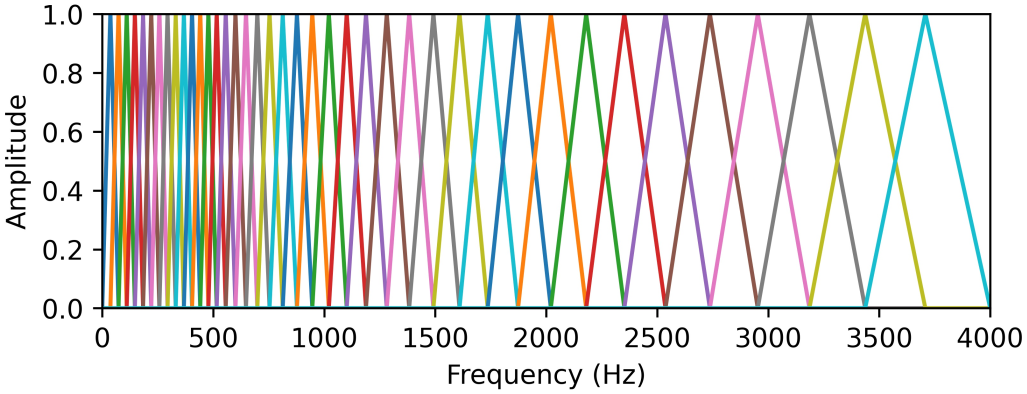

Figure 6. The Mel triangular filters implemented by Librosa package.

These frequency responses of Mel-Fbank can be expressed as

$$H(f) = \begin{cases} \frac{f_h - f}{f_h - f_0}, & f_0 < f \leqslant f_h \\ \frac{f_l - f}{f_l - f_0}, & f_l \leqslant f \leqslant f_0 \end{cases}. \tag{13}$$

Furthermore, the Mel-Fbank can be represented as

$$E(m) = \sum_{f=f_l(m)}^{f_h(m)} H_m(f)|X_n(f)|, \quad l = 1, 2, \cdots, L. \tag{14}$$

The resulting Mel-spectrogram, a two-dimensional representation of Mel-Fbank features, depicts f_{mel} frequency intensities across time frames. Due to overlapping frequency components, Mel-spectrograms are known to preserve intrinsic signal details, exhibit high feature inter-correlation, and demonstrate excellent capabilities for nonlinear feature extraction. Taking Mel-spectrograms as natural images, Liu et al. [49] integrated these auditory perceptual features of underwater targets into DNNs to learn frequency characteristics per hop group, highlighting their benefits in feature extraction under varying underwater acoustic communication scenarios. On the contrary, Tang et al. [50] argued that Mel-spectrograms differ significantly from natural images in that they contain rich time–frequency information. To efficiently handle this important information, they developed a three-dimensional spectrogram network, achieving an optimal balance between recognition performance and model complexity. Further application of Discrete Cosine Transform (DCT) on Mel-spectrograms yields Mel-Frequency Cepstral Coefficients (MFCCs) [52]. Utilizing these MFCC features extracted from ship-radiated noise, further classification of acoustic targets using ML [53] and DL techniques [54,55] has demonstrated substantial improvements in recognition performance.

The feature of Gammatone Frequency Cepstral Coefficients (GFCCs) is another widely used auditory perceptual representation in UATR [19,56,57]. Similar to MFCC features, the GFCC is obtained through DCT from a Gammatone filter bank, possessing excellent capabilities in representing signal spectral structures. The main difference between GFCCs and MFCCs lies in their design principles: GFCC models filter responses based on the cochlear frequency decomposition, whereas the MFCC relies on the f_{mel} scale. Due to their robust representation of nonlinearities and spectral envelopes, GFCC features have proven effective in describing sound signals, particularly in scenarios with low SNRs [58].

2.2.3. Multidimensional Fusion Features

To enrich feature diversity, fusion-based feature extraction methods have recently gained increasing attention in the field of UATR [59–61]. Following the concept of feature fusion, these approaches combine different time–frequency features to enrich signal representations, thereby providing discriminative information for classification algorithms. As shown in Figure 7, these methods typically involve two primary strategies. The first strategy adopts a single-network approach with frontend feature fusion during preprocessing. The frontend fusion integrates various features into multiple input channels, similar to the RGB channels in natural images, thereby constructing a multidimensional fused feature designed for UATR. Conversely, the second strategy employs ensemble models with backend feature fusion, leveraging multiple networks to process data and classify target labels using probability vectors across multiple model backends.

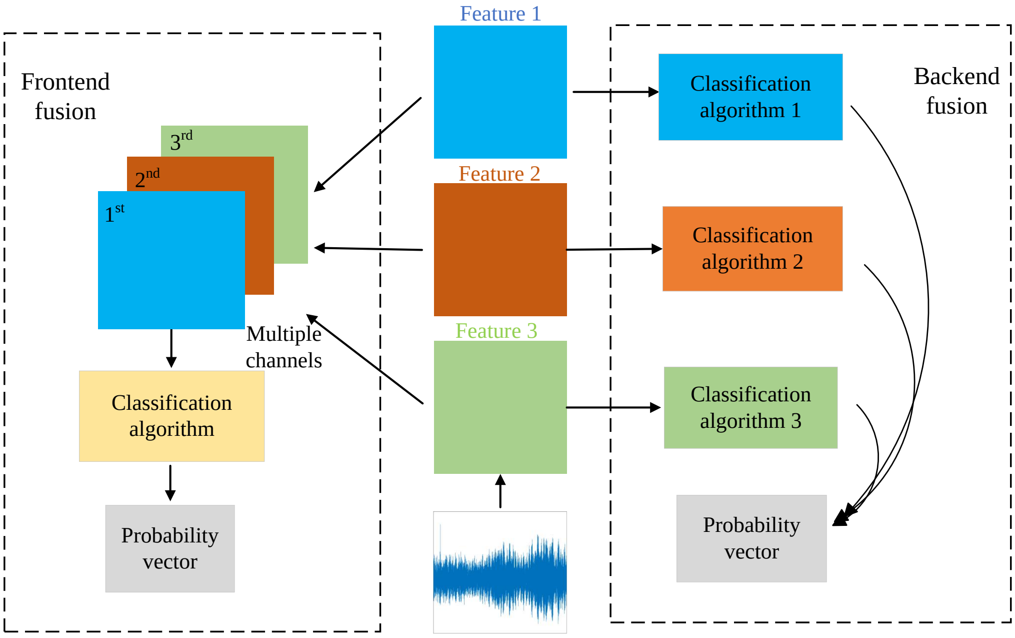

Figure 7. Two aspects of the multidimensional features.

In the frontend fusion approach, Tan et al. [62] extracted various features from underwater signals, including Amplitude Modulation Spectrogram, MFCC, Relative Spectral Transform–Perceptual Linear Prediction, GFCC, and Delta feature. These features are then fused to form AMCG-Delta features. Integration of various spectral data augmentation techniques resulted in a recognition accuracy 6.95% higher than that of the baseline method. Liu et al. [63] combined Mel-spectrograms with their first and second derivatives to form 3D Mel-spectrograms that were fed into a DNN architecture, achieving satisfactory recognition performance in multiple UATR tasks. Similarly, Wu et al. [64] also employed 3D Mel-spectrograms as input features, further combining transfer learning from ImageNet to enhance the feature applicability in underwater acoustic modalities. Their experimental findings on publicly available datasets showcased significant improvements in recognition performance, as can be seen in Figure 8.

On the other hand, backend fusion methods in underwater acoustic signal processing draw upon the principles of ensemble learning [65]. Zhang et al. [66] simultaneously extracted STFT magnitude spectrum, STFT phase spectrum, and bispectrum features of underwater signals, feeding each feature set into multiple networks and subsequently integrating their outputs to make decisions. This method takes full advantage of various time–frequency features to provide the integrated model with improved recognition accuracy and anti-noise robustness. Despite its capability to yield discriminative features, it is crucial to acknowledge that fusion-based feature extraction introduces considerable redundancy by combining multiple types of time–frequency representations. Moreover, these methods are sensitive to the selection of time–frequency feature combinations, as improper selection could result in suboptimal recognition performance.

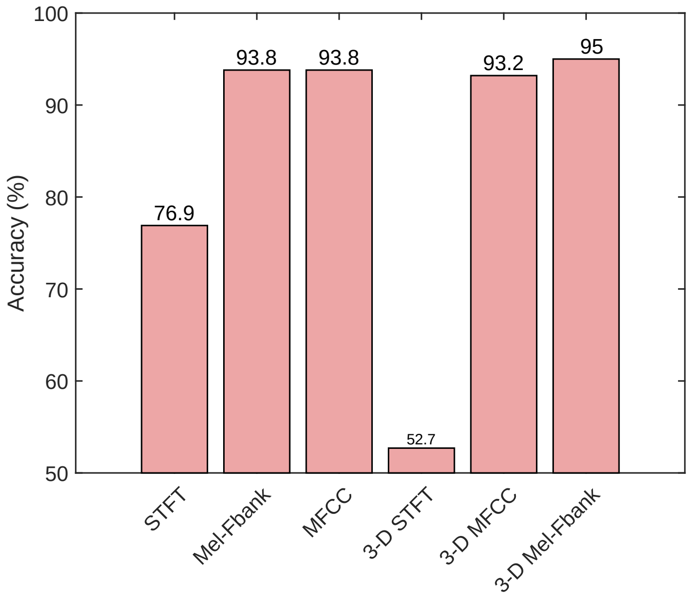

Figure 8. Accuracy comparison using various feature extraction methods with ResNet18 on the Shipsear dataset, as adapted from Wu et al. [64].

2.3. Autoencoding Feature

Manual feature extraction methods require complicated feature engineering, which involves numerous underwater acoustic processing techniques that are computationally intensive and inefficient. In contrast, autoencoding feature extraction methods have recently gained popularity in UATR due to their ability to automatically extract features [67,68], which is straightforward and efficient to achieve end-to-end UATR. AEs are the primary autoencoding feature extraction method in UATR [69,70]. In the unsupervised learning paradigm, AEs are originally designed for data compression, denoising, and sparse representation [71], which encode high-dimensional data into low-dimensional latent representations with the goal of reducing redundancy and capturing implicit patterns.

As shown in Figure 9, an AE consists of an encoder and a decoder, which together map an input x to an output x' while requiring that x' and x are sufficiently close, $x \approx x'$. The input can be a signal waveform or a spectrum. In the time domain, Dong et al. [72] designed a Bidirectional Denoising Autoencoder (BDAE) to compare the differences between original and denoised signals, thereby learning anti-noise robust characteristics of underwater acoustic signals. Experimental results demonstrated the effectiveness of BDAE features particularly in low SNRs. Li et al. [73] designed deep convolutional filter banks to decompose and merge ship-radiated noise, extracting deep time–frequency features through fully connected layers. Their research suggested that setting more filter banks can achieve better classification performance than advanced methods. From a spectral perspective, Luo et al. [74] used a Restricted Boltzmann Machine (RBM) for automatic encoding of power and demodulation spectra. Their method facilitated hierarchical feature extraction from underwater acoustic signals without the need for task-specific supervision. Experimental results showed that the proposed method can obtain discriminative signal representations to achieve effective UATR.

Moreover, self-supervised learning (SSL) methods can effectively extract automatic encoding features through well-designed upstream tasks. Generally, this learning paradigm can be categorized into generative and contrastive learning based on the upstream task. Contrastive learning focuses on learning meaningful representations by comparing differences between samples, whereas generative methods emphasize extracting features by leveraging the inherent structure of the data themselves. Combined with contrastive learning, Sun et al. [75] proposed the contrastive learning Underwater Acoustic Target Recognition (CCU) model for feature extraction. CCU dynamically extracts discriminative features from diverse data inputs, achieving superior recognition performance compared to AEs. Furthermore, Wang et al. [76] proposed an Acoustic-embedding Memory Unit Modified Space AE (ASAE) for both generative and contrastive learning. This enhancement of space AE serves as an upstream task to efficiently extract high-level semantic information from acoustic target signals. Experimental results demonstrated that ASAE spectrogram features can achieve better recognition performance than mainstream time–frequency features.

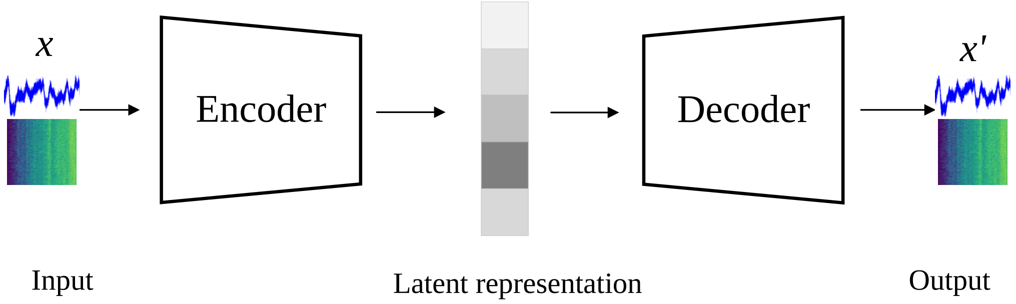

Figure 9. The basic architechture of AE.

While autoencoding features have shown success in UATR, these features essentially lack a close connection to physical mechanisms and target characteristics due to insufficient data preprocessing. Consequently, these features are considered difficult to provide intuitive physical explanations for, which are critical factors for model interpretability.

3. Machine Learning-Based Recognition Methods

Due to the simplicity, efficiency, and ease of implementation, ML methods have become widely utilized for intelligent UATR [77–79]. Their standard paradigm involves feeding extracted features into shallow ML algorithms, which then find classification boundaries based on feature similarity, effectively dividing compact sets in high-dimensional space. Table 1 provides an overview of representative ML-based UATR methodologies.

Among these ML methods, distance-based methods such as SVMs [80] and KNN [81] are well-known for their effectiveness in discovering decision boundaries within the feature space, which is crucial for recognizing underwater acoustic signals. As the most popular ML methods for UATR, SVMs employ kernel functions to map signal features into high-dimensional spaces, where separable boundaries between data points can be always found, ensuring different category samples lie on opposite sides of these boundaries while maintaining sufficient margin distance, thereby addressing the task of UATR. Many studies demonstrated that combining an SVM with individual features can achieve high classification results [82–86]. However, these studies did not explore the performance of multidimensional feature approaches. Liu et al. [87] demonstrated that using multiple features can lead to better results on real-world acoustic datasets.

Table 1. Representative works of ML-based UATR methods. Note that the default metric is recognition accuracy. Adjusted Rand index (ARI) is a widely used clustering evaluation metric.

Reference	Feature	Method	Dataset	Result
Li et al. [82]	wavelet analysis	SVM	three-category dataset	90.12%
Yang et al. [88]	feature selection	SVM	UCI sonar dataset [89]	81%
Moura et al. [83]	LOFAR	SVM	four-category dataset	76.73%
Sherin et al. [84]	MFCC	SVM	four-category dataset	74.28%
Wang et al. [85]	STFT	SVM	four-category dataset	88.64% (10 dB)

Table 1. *Cont.*

Reference	Feature	Method	Dataset	Result
Yao et al. [86]	STFT	SVM	two-category dataset	90%
Liu et al. [87]	multidimensional fusion feature	SVM	4-category dataset	97%
Choi et al. [90]	cross-spectral density matrix	SVM	simulation dataset	97.98%
Wei et al. [91]	MFCC	SVM	five-category dataset	93%
Chen et al. [92]	spectral ridge	SVM	four-category whale call dataset	99.415%
Yaman et al. [93]	512-dimensional feature vector	SVM	five-category propeller dataset	99.8%
Saffari et al. [94]	FFT	*K*NN	ten-category ship simulation dataset	98.26%
Li et al. [95]	complex multiscale diffuse entropy	*K*NN	four-category dataset	96.25%
Alvaro et al. [81]	-	*K*NN	five-category dataset	98.04%
Jin et al. [96]	eigenmode function	*K*NN	Shipsear dataset [1]	95%
Mohammed et al. [97]	GFCC	HMM	ten-category dataset	89%
You et al. [98]	GFCC	HMM	simulation dataset	90% (8 dB)
Yang et al. [17]	mutual information	LR	UCI sonar dataset [89]	94.7%
Seo et al. [99]	FFT	LR	simulation dataset	77.43%
Yang et al. [100]	auditory cortical representation	LR	3-category dataset	100%
K et al. [101]	mutual information	decision tree	UCI sonar dataset [89]	95%
Yaman et al. [93]	512-dimensional feature vector	decision tree	five-category dataset	99%
Yu et al. [102]	covariance matrix	decision tree	two-category dataset	99.2%
Zhou et al. [103]	multicorrelation coefficient	random forest	South China Sea dataset	93.83%
Choi et al. [90]	cross-spectral covariance matrix	random forest	two simulation datasets	96.83%
Chen et al. [92]	spectral ridge	random forest	four-category whale call dataset	99.69%
Wang et al. [104]	MFCC	GMM	four-category dataset	ARI 77.97
Sabara et al. [105]	-	GMM	SUBECO dataset [105]	74%
Yang et al. [106]	MFCC	GMM	five-category dataset	-
Yang et al. [106]	MFCC	fuzzy clustering	five-category dataset	-
Agersted et al. [107]	intensity spectrum	hierarchical clustering	Norwegian Institute of Marine Research	Best clusters = 7

*K*NN method predicts new samples by voting rule with the k nearest training data points. Common distance metrics $d(x, y)$ used to measure the similarity, such as Euclidean, Manhattan and Minkowski distances. Li et al. [95] introduced refined composite multiscale dispersion entropy (RCMDE) features for ship-radiated noise and utilized *K*NN for target recognition. Comparative experiments on 4-category ship dataset demonstrated that RCMDE-*K*NN can effectively capture the implicit pattern of underwater signals, which achieved a recognition accuracy at 96.25%. Jin et al. [96] extracted intrinsic mode function (IMF) feature based on target signal center frequency and energy intensity. This feature is then fed into *K*NN algorithm for nearshore ship recognition. Experimental results on the Shipsear dataset [1] demonstrated its effectiveness with a 95% recognition accuracy. However, the strength of these distance-based methods comes with the trade-off of potentially high computational demands due to the need to compute multiple distances [83,84]. To reduce spatial complexity, Yang et al. [88] proposed a novel ensemble SVM approach through weighted sampling and feature selection, named as the WSFSelect-SVME. This method integrates multiple SVMs classifiers to obtain robust recognition performance in real-world datasets. Additionally, Wang et al. [85] explored the use of DL algorithms for feature extraction, followed by SVM for chromatic feature classification. Experimental results on noisy datasets demonstrated its effectiveness, which achieved an average recognition rate of 94.92%.

On the other hand, HMM [97] and GMM [104] are two representative probability-based models derived from natural language processing (NLP), which can provide rich probabilistic explanations on the acoustic data. Specifically, HMM is utilized to characterize the temporal sequence of underwater acoustic signals, capturing dynamic sound variations over time through state transition and observation probabilities [97,98]. On the other hand, GMM models the static features of underwater acoustic signals, emphasizing the statistical properties of various frequency components using Probability Density Function modeling [106]. Despite their success in underwater acoustic target recognition, HMMs and GMMs are are acknowledged for their challenging training processes and sensitivity to parameter selection. These characteristics can significantly influence the practicality of deploying these models in ocean applications. To address this problem, various optimization methods, such as genetic algorithm [98], particle swarm algorithm [28] have been integrated into these models to reduce the sensitivity, which have achieved promising results.

Additionally, tree-based methods shine for their ability to provide a visual representation of the decision-making process. In the field of UATR, decision trees and random forest are commonly used tree-based methods. Specifically, decision trees work by recursively partitioning the data space based on individual acoustic features. Its architecture is intuitive: internal nodes represent features, branches signify decision rules, and leaf nodes indicate the outcomes, which has made certain achievements using various acoustic features [93,102]. Expanding on this foundation, random forests construct multiple trees, each contributing to a collective prediction, which can improve the generalization and reduce the risk of overfitting. The randomness introduced in the selection of data subsets and features for each tree is important to the robust UATR performance [90,103]. However, to avoid overfitting, a careful balance of model complexity is required for this ensemble method.

Lastly, regression methods, such as linear regression (LR) [17] has been proven its effectiveness in UATR. Yang et al. [17] conducted preprocessing of sonar-collected underwater signals to isolate critical attributes such as signal intensity, frequency, and reflection angles. By integrating feature selection techniques, they successfully performed binary classification tasks on rocks and mines using LR models. Seo et al. [99] also employed LR models for UATR. They initially used LR to filter target signals from strong noise environments and further trained the model using simulated data to identify real-world samples, which achieved efficient UATR under limited real data conditions. It is noteworthy that while this method alleviates data requirements with simulated data, the fidelity to actual marine environments remains a challenge, impacting performance and requiring substantial time investment for data generation. Despite the efficiency and simplicity of regression methods, their performance can be limited by the linear assumption, which may lead to suboptimal results in the presence of complex, nonlinear datasets where they might yield suboptimal results.

Overall, the successful application of various ML methods demonstrates their potential in handling underwater acoustic data, which further drives intelligent development in UATR. These advances have provided a strong theoretical foundation for DL-based methods, while also providing valuable insights into underwater data processing. Nonetheless, ML algorithms typically apply limited linear or nonlinear transformations on acoustic data, capturing shallow internal structures. In scenarios with limited samples and constrained computational resources, their ability to represent complex underwater signals is diminished, which limits their generalization and recognition accuracy.

4. Deep Learning-Based Recognition Methods

The DL-based intelligent UATR methods effectively utilize various neural networks to classify underwater targets. By mimicking the neural structure of the human brain, DL methods rely on layered architectures to learn deep abstract patterns within data. Based on learned universal principles, these methods analyze, reason, and predict outcomes for unseen samples, as illustrated in the standard prediction framework shown in Figure 10.

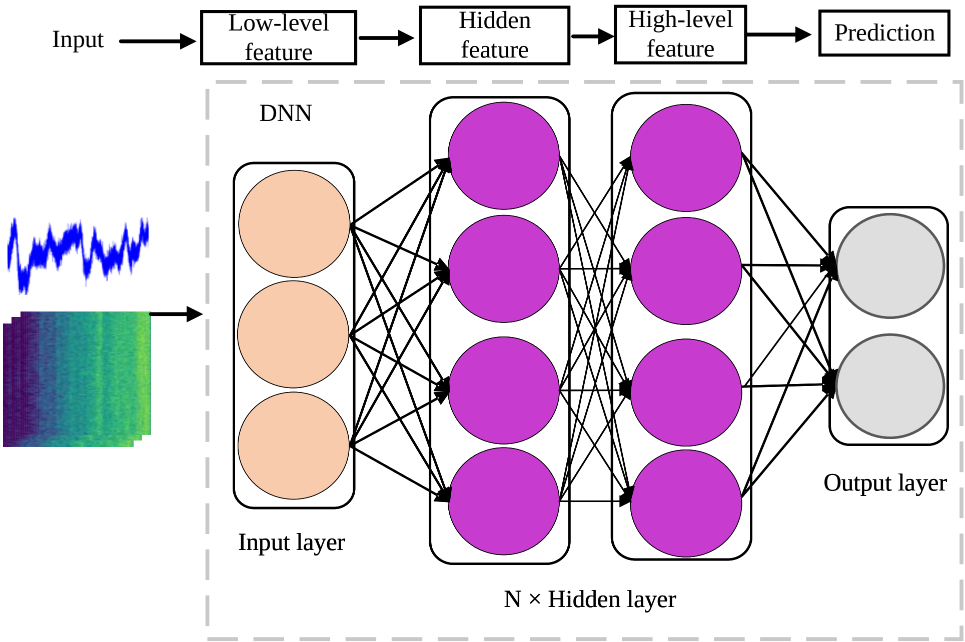

Figure 10. A standard prediction framework of DL-based UATR methods.

Within this paradigm, as DNN progress through layers, input features transform from low-level to high-level representations after passing through multiple hidden layers. These high-level features are finally integrated through the output layer to make the final prediction. Due to their superior signal processing capabilities, DL-based methodologies have emerged as a compelling alternative to conventional ML techniques, and have become a prominent area of current research [18,108].

Figure 11 depicts the advantages and disadvantages of representative DL-based UATR methods. Among these methods, shallow neural networks (SNNs) are firstly applied for UATR, such as multilayer perceptions (MLP) [109,110] and Restricted Boltzmann Machines (RBMs) [111,112], which achieved certain successes in underwater signal processing. However, their shallow architecture limited the ability to discern intricate patterns. The complexity of underwater environments and the instability of data necessitate the use of DNNs, particularly under conditions of few-shot and low SNRs.

Recurrent neural networks (RNNs), convolutional neural networks (CNNs), attention neural networks (ATNs), and Transformers have risen as pivotal DNN-based methodologies for intelligent UATR, demonstrating significant effectiveness in practical scenarios. Other DL technologies, such as generative adversarial networks (GANs) [26,113], transfer learning [114], and SSL [115,116], have opened new frontiers for intelligent UATR, showcasing enormous development potential. Importantly, the data learning processes in DL often involve substantial amounts of underwater acoustic signals, making the partitioning of datasets a critical factor that influences performance. Two of the most extensively studied datasets within this field are Shipsear [1] and Deepship [2]. The partitioning of these datasets is typically conducted using two principal methods. The first method is known as the random partitioning approach. This approach can potentially lead to a misleadingly optimistic assessment of model performance. This can occur if a data segment that is temporally subsequent is present in the training set, while an earlier segment is allocated to the test set. Such a scenario can lead to an overestimation of recognition performance, as the model may inadvertently benefit from future information that is not available in real-world conditions. The accuracy comparison of widely utilized DNNs employing random partitioning is depicted in Figure 12a.

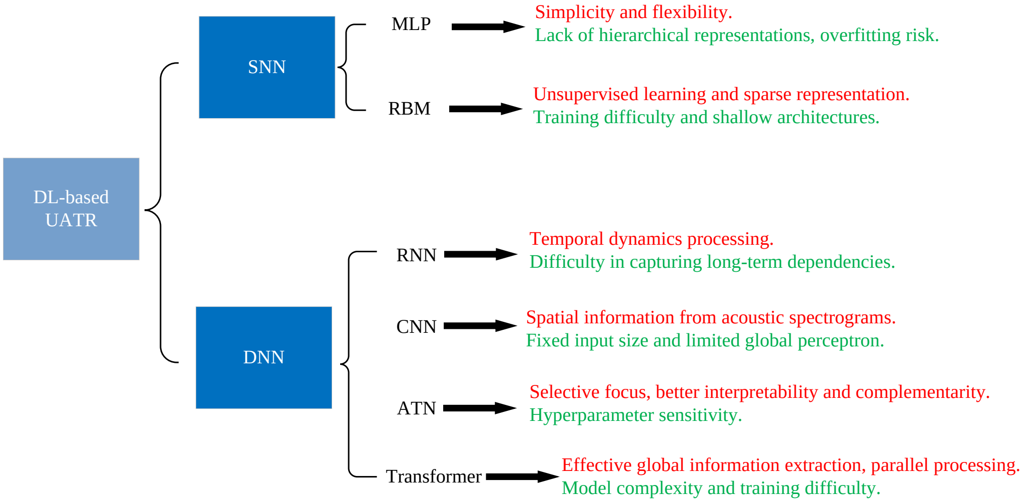
DL-based UATR
SNN
DNN
MLP
Simplicity and flexibility.
Lack of hierarchical representations, overfitting risk.
RBM
Unsupervised learning and sparse representation.
Training difficulty and shallow architectures.
RNN
Temporal dynamics processing.
Difficulty in capturing long-term dependencies.
CNN
Spatial information from acoustic spectrograms.
Fixed input size and limited global perceptron.
ATN
Selective focus, better interpretability and complementarity.
Hyperparameter sensitivity.
Transformer
Effective global information extraction, parallel processing.
Model complexity and training difficulty.

Figure 11. Representative DL neural networks in the field of intelligent UATR. The advantages and disadvantages of each method are marked in red and green, respectively.

The second method, known as causal partitioning, provides a more reasonable approach to dataset segmentation. Specifically, this technique segments audio files in accordance with their temporal sequence, ensuring that the model is not exposed to data from the future. In this way, it effectively prevents data leakage and provides a more accurate assessment of the prediction performance. The recognition accuracy comparison of DNNs leveraging causal partitioning is presented in Figure 12b. Causal partitioning respects the temporal integrity of the data, which is essential for training models that can generalize well to new, unseen data while maintaining the temporal causality inherent in time-series data. Overall, in the field of UATR, the mainstream recognition models are evolving from CNNs to Transformers. As can be seen from Figure 12, the advanced Transformer models have achieved superior recognition performance across both partitioning methods compared to CNN models. Moreover, the continuous integration of various cutting-edge computer techniques on Transformer architecture, including SSL [117,118], is expected to further improve recognition capabilities.

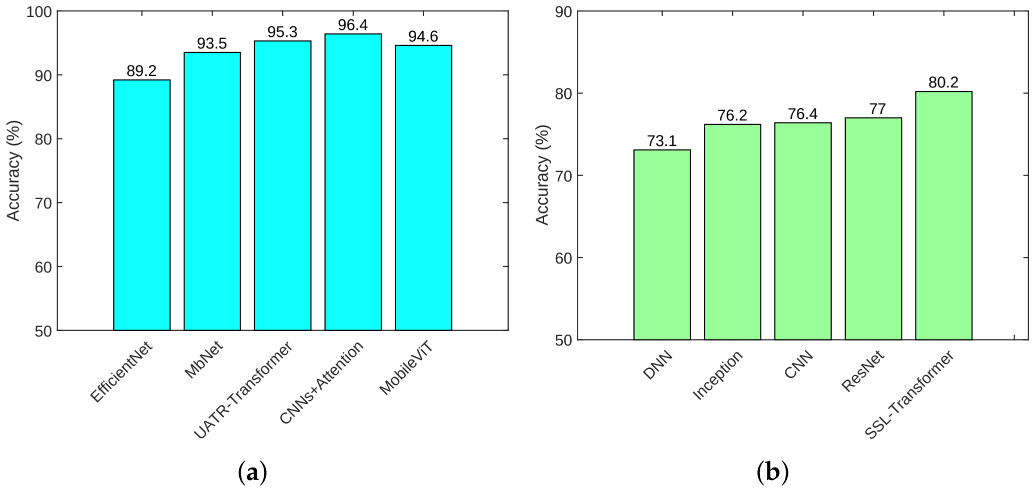

Figure 12. Accuracy comparison with different DNNs on the Deepship dataset. (**a**) Random partition, adapted from Zhou et al. [119], (**b**) causal partition, adapted from Irfan et al. [2] and Xu et al. [117].

4.1. RNN

In the field of DL, the RNN is a type of DNN capable of processing continuous data such as text [120] and speech [121]. Its fundamental concept involves utilizing internal recurrent connections to retain information from previous inputs, which are then integrated into current outputs. By leveraging these recursive connections, RNN-based UATR methods can directly handle temporal waveform data through memory functions that dynamically leverage temporal correlations, enabling accurate target classification [122–124]. In these methods, long short-term memory (LSTM) is one of the most commonly used RNN architectures [125,126]. Figure 13 presents the structure of the computational units in LSTM. LSTM leverages three gating mechanisms (input gate, forget gate, and output gate) that facilitate dynamic utilization of long-term dependencies to process underwater acoustic signals, which can address the problem of gradient vanishing and exploding commonly encountered in traditional RNNs. Specifically, within LSTM, gates controlled by activation functions σ (typically sigmoid functions in input and forget gates) regulate the information flow. Initially, the model input x_t is selectively processed by the forget gate in conjunction with the previous hidden state s_{t-1} to determine which information to retain or discard from the cell state C_{t-1}.

$$f_t = \sigma\left(W_f \cdot [s_{t-1}, x_t] + b_f\right). \tag{15}$$

Subsequently, the input gate determines update values i_t, creating a new candidate cell state C_t' via tanh function, updated alongside i_t.

$$\begin{aligned} i_t &= \sigma(W_i \cdot [s_{t-1}, x_t] + b_i) \\ C_t' &= \tanh(W_C \cdot [s_{t-1}, x_t] + b_C). \end{aligned} \tag{16}$$

Historic information is retained via multiplication by f_t to form new cell state C_t.

$$C_t = f_t * C_{t-1} + i_t * C_t'. \tag{17}$$

Finally, the output gate, through a sigmoid layer, determines which information within the cell state should be passed to the next hidden state. The processed cell state C_t, after tanh function, is multiplied by the output of the sigmoid output gate, ultimately determining information passed to the subsequent hidden state.

$$\begin{aligned} o_t &= \sigma(W_o[s_{t-1}, x_t] + b_o) \\ s_t &= o_t * \tanh(C_t). \end{aligned} \tag{18}$$

Yu et al. [127] leveraged the robust memory selection capability of LSTMs to process instantaneous features extracted from underwater communication signals, achieving recognition accuracy exceeding 80%, which demonstrates its effectiveness in modulation pattern recognition. Moreover, Yang et al. [128] employed pretrained LSTMs and softmax classifiers to classify ship-radiated noise in severe underwater communication environments. Furthermore, Zhang et al. [125] exploited backend fusion with multiple LSTMs to process various acoustic features; the final feature maps are integrated into softmax layers to make predictions, which achieved superior recognition performance over single-feature approaches.

Gated recurrent units (GRUs), another variant of RNNs, offer a simplified architecture compared to LSTM. Particularly, GRUs merge cell states and hidden states into a unified hidden state and incorporate two gates (update and reset gates) to process acoustic signals [129], which contain fewer parameters, making GRUs suitable for deployment in real-time recognition systems. Qi et al. [130] utilized GRUs to extract fine-grained features from acoustic spectrograms, achieving satisfactory recognition performance with real ocean data. Moreover, Wang et al. [131] proposed a hybrid time-series network for modulation signal recognition in complex underwater communication environments. By integrating bidirectional GRUs to enhance the hidden representations of acoustic signals, the proposed method demonstrated robust recognition capability under severe interference.

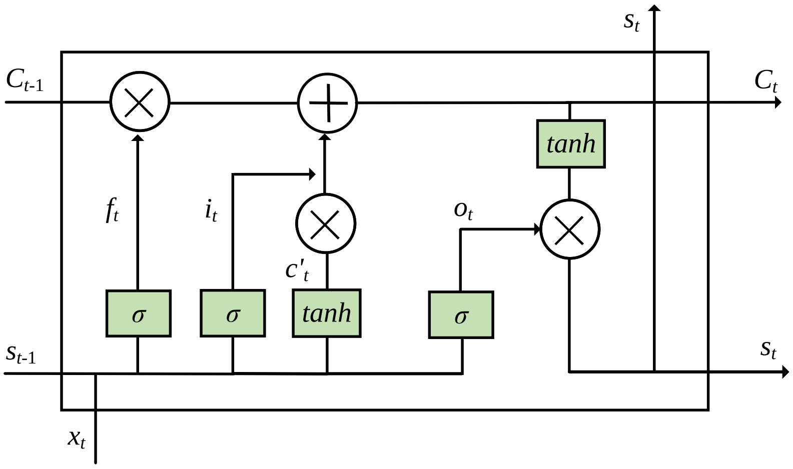

Figure 13. The computational units in LSTM.

To enhance feature representations, a widely adopted strategy integrates a CNN-based local feature extractor in front of RNNs [132]. Kamal et al. [133] proposed a hybrid CNN-LSTM model for recognizing underwater signals with real ocean data. The proposed model employed convolutional filters to learn sufficient time–frequency representations, subsequently extracting discriminative features via 2D convolutions, which achieved an impressive recognition accuracy of 95.2%. Moreover, Qi et al. [134] introduced an intelligent UATR method that combines LSTM with 1D CNN, where the LSTM module was utilized to capture phase and frequency characteristics from acoustic signals. Experimental results demonstrated its effectiveness to achieve superior recognition performance compared to a single network.

4.2. CNN

Inspired by the notable success of CNNs in handling natural images, early researchers directly applied side-scan and forward-scan sonar image data inputs to CNNs for UAT and have made certain achievements [135,136]. Subsequently, researchers began to explore the direct application of signal waveform to CNNs for UATR [19,137]. In comparison to processing time-domain waveforms, the sliding window in CNN is particularly suitable for extracting discriminative features from high-resolution time–frequency spectrograms [138]. In practice, the representation of time–frequency spectrograms can effectively capture stable energy distribution characteristics, which have shown superior recognition performance compared to waveform inputs [108,139]. In this case, the integration of CNNs with time–frequency spectrograms has been considered as the baseline method in the field of UATR [140,141]. ResNet [142] stands as a common CNN architecture for UATR, with its structure illustrated in Figure 14. Its main innovation, the residual block, consists of two consistently sized convolutional layers with a skip connection that adds input directly to the output of the block, formulated as $H(x) = F(x) + x$. This residual adding operation helps ResNet mitigate gradient vanishing because it allows gradients to propagate directly through to earlier layers, thus maintaining information as the network goes deeper. Additionally, hierarchical convolution and pooling layers are employed in standard ResNets to reduce resolution. Building upon the architecture of ResNet, researchers developed DenseNet [143], which incorporates dense connections between all layers for feature reuse and extraction efficiency. Both ResNet and DenseNet have been extensively researched for their advantages in extracting detailed texture features and facilitating spatial information flow, establishing their significance in UATR [144–146].

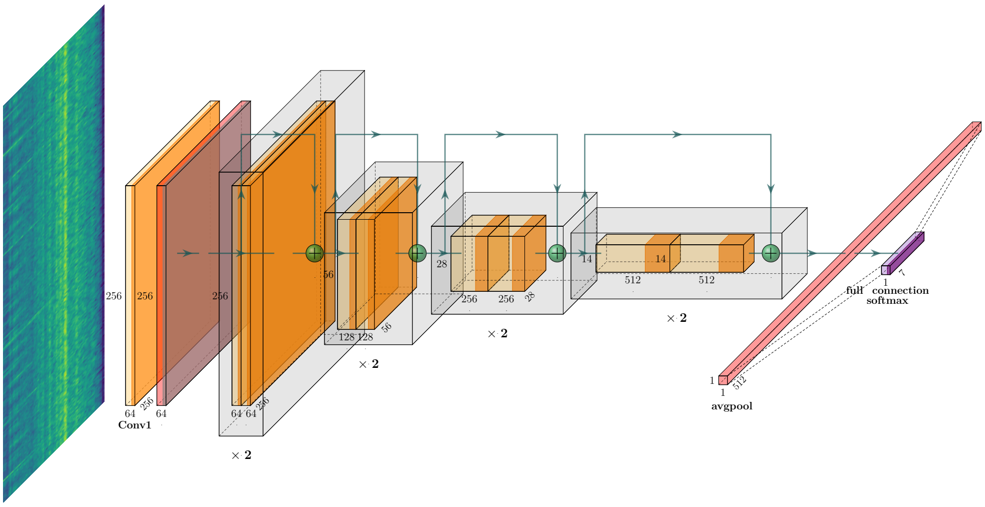

Figure 14. The UATR framework based on ResNet18, which commonly accepts the acoustic spectrograms as model input.

To improve efficiency, various convolution schemes have been further developed in CNNs. Depth-wise convolution (DW) [147] has been integrated into lightweight CNN architectures to increase feature discriminability while reducing computational costs [148]. Hu et al. [149] constructed CNNs using DW and dilated convolution mechanisms for passive UATR. The experimental results demonstrated that it outperformed traditional CNNs by extracting discriminative features. Using an Inception convolution block [150], Zheng et al. [151] applied the STFT features of underwater acoustic signals to GoogleNet [152] for shallow-water underwater target detection, demonstrating improved detection performance over energy detectors. Moreover, Cai et al. [153] leveraged Xception, a variant of Inception networks [147] optimized with DW convolutions, to develop a multibranch DNN capable of recognizing underwater targets. The proposed approach integrated Xception for detecting line spectral pair features, alongside other branches including MobileNet [154] and DenseNet, which achieved average recognition accuracies exceeding 95% across six classes of marine biological sounds. Considering hardware constraints and real-time requirements, a variety of lightweight CNNs [155–157] have been introduced with robust feature processing capabilities in real-time UATR tasks.

4.3. ATN

Underwater target signals essentially represent natural time series in which critical frequency information is essential for target classification. Attention mechanisms can effectively focus on these critical frequencies, thereby potentially enhancing feature processing capabilities. Xiao et al. [158] developed an interpretable ATN-based UATR method using STFT spectrograms, demonstrating significant potential to apply attention into UATR. By combining attention mechanisms with multiscale CNNs [119], ATNs can obtain superior recognition performance compared to baseline CNN and RNN models. Particularly, scalable attention modules, such as channel attention mechanism (CAM) [159] and spatial attention mechanism (SAM) [160], have been effectively integrated with CNNs to improve the feature extraction from signal spectrograms [161,162].

Bidirectional attention [163], a special co-attention and self-attention fusion technique, is designed to comprehensively capture semantic and structural information between different regions within model inputs. Liu et al. [164] combined bidirectional attention with Inception, dynamically weighting the receptive field to eliminate irrelevant backgrounds, thereby achieving satisfactory recognition accuracy. In addition, Wang et al. [161] introduced AMNet, a model combining multibranch backbone networks with convolutional attention modules. These modules selectively weight global information, thus adaptively filtering key information within acoustic spectrograms to potentially enhance classification performance.

In summary, ATNs exhibit substantial potential for robust signal representation capabilities to enhance recognition accuracy in the field of UATR, which makes them promising research directions in intelligent UATR.

4.4. Transformer

The Transformer, originally proposed by Google for machine translation in natural language processing tasks [165], can be viewed as a specialized form of ATN. This model primarily utilizes a multihead self-attention (MHSA) to describe correlations across multiple subspaces. The MHSA is a novel attention mechanism that significantly improves model representation and computational efficiency. As depicted in Figure 15, MHSAs simultaneously compute and integrate semantic information from H distinct heads on query Q, key K, and value V sets, thereby outperforming conventional attention mechanisms.

To leverage its robust data processing capabilities, Feng et al. [141] introduced UATR-Transformer. Specifically, the proposed model employs hierarchical tokenization strategies to derive computational units from acoustic Mel-spectrograms. Further feature transformation through an MHSA demonstrates that UATR-Transformer can achieve recognition performance comparable to state-of-the-art CNNs due to its both global and local information perception. Inspired by Audio Spectrogram Transformer (AST) in speech recognition [166], Li et al. [167] adapted its structure into UATR, proposing the Spectrogram Transformer Model (STM), which also outperforms CNNs in UATR tasks. The Swin Transformer [168], a specialized variant utilizing shifted windows for hierarchical feature extraction, was introduced by Wu et al. [169] in UATR as the time–frequency Swin-Transformer (TFST). TFST exhibits superior performance over both CNNs and traditional Transformers in learning discriminative features from moving targets on two real ocean datasets. Despite their satisfactory recognition accuracy, a primary drawback of Transformers is the high complexity of MHSAs, which impacts their real-time performance. To address this problem, Yao et al. [170] combined MobileNet with Transformer architecture to design MobileViT, which effectively integrates global and local information perception of underwater acoustic signals with low model complexity, validated through empirical dataset evaluations. SSL-Transformers exhibit robust generalization capabilities, making them powerful feature learners for underwater acoustic signals. To this end, several studies have explored SSL-Transformer methods for UATR [118,171], focusing on upstream tasks to learn generalized representations of underwater acoustic signals without label information, thereby addressing the challenge of limited data in UATR. While Transformers have demonstrated promising results in UATR, challenges of complex training requirements necessitate the integration of diverse training methods, including data augmentation [172] and knowledge distillation [173].

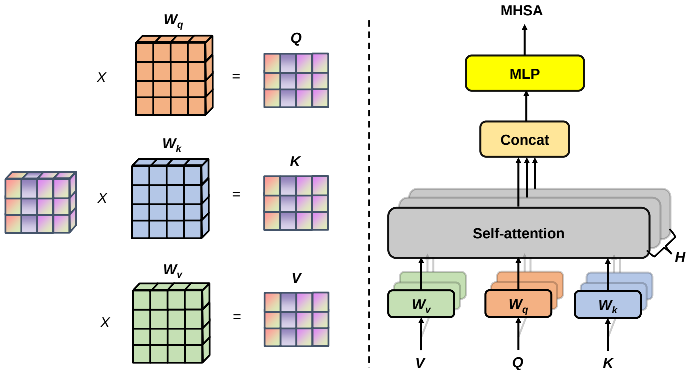

Figure 15. The MHSA mechanism in Transformer. The left side describes the self-attention mechanism, and the right side describes the MHSA. W_q, W_k, W_v comprise the learnable projection matrix.

5. Challenges and Future Prospects

As previously mentioned, the rapid development of AI has greatly promoted the application of data-driven methods in the field of UATR. In particular, the current DL methods have fully exploited the flexibility and scalability of DNNs to design advanced recognition systems. However, intelligent UATR still faces several serious challenges, which can be summarized as follows: the complex recognition condition issue, poor interpretability, weak generalization, and low robustness under adversarial attacks.

5.1. Complex Recognition Condition Issue

Recognizing underwater acoustic targets in conditions of low SNR and few-shot scenarios indeed poses a significant challenge in the field of underwater acoustics. As aforementioned, the multiple interferences and complex channels present in the ocean environment significantly increase the difficulty of UATR. Recognition tasks under low SNRs are further complicated by these challenges. The low SNR makes it extremely difficult to distinguish the weak signals from background noise, thereby complicating the model's ability to learn deep, discriminative features. To address this issue, many researchers have been dedicated to developing robust recognition methods, particularly through the use of well-designed DL models. Zhou et al. [174] introduced a cross-attention fusion joint training framework that synergizes denoising and recognition models for robust UATR. This framework has been shown to effectively tackle the UATR task in noisy environments by enhancing the SNR and improving feature extraction. Similarly, Li et al. [175] proposed a robust UATR method inspired by auditory passive attention mechanisms. Their approach considered three auditory passive attention loss functions to achieve robust UATR under low SNR conditions, achieving an average F1-score of 67.43% (SNR = −18 dB) and a notably high score of 90.79% (SNR = 6 dB). Additionally, several data preprocessing techniques, such as dual-path denoising [176] and adaptive filter techniques [177], can be also beneficial in addressing the challenges posed by low SNR conditions.

Due to the inherent challenges in acquiring high-quality underwater acoustic data, the task of few-shot UATR has become another prevalent research focus. Yang et al. [178] utilized a Siamese network that consists of 1D convolution and an LSTM for few-shot recognition. Experimental results based on sea trial data demonstrated its effectiveness. However, this method is prone to overfitting, potentially leading to a performance degradation. To address this issue, SSL methods have been increasingly adopted [179,180], aiming to enhance model generalizability and robustness with minimal reliance on labeled data. Common SSL techniques used in UATR can be broadly categorized into contrastive SSL [179,180] and generative SSL methods [117,171]. As depicted in Figure 16, contrastive SSL focuses on learning robust representations by comparing original data against its transformed versions, aiming to distinguish between positive and negative pairs to minimize the contrastive loss. On the other hand, generative SSL tasks are designed to capture the intrinsic structure of the data, enabling the model to generate new samples that closely resemble the original sample. In UATR, a generative model can be trained to reconstruct missing parts of a sonar signal or spectrogram and then optimized by minimizing the generative loss.

Figure 17 illustrated the accuracy comparison of few-shot models based on the Shipsear dataset. Among them, the contrastive SSL-based methods, CDCF achieves better results than conventional few-shot models, which demonstrates that SSL can help DL models to learn discriminative features from scarce acoustic data, thereby improving recognition accuracy in few-shot UATR tasks.

For future directions to address the issue of complex recognition conditions, further leveraging more powerful meta-learning approaches [181,182] with various feature extraction methods could be a promising strategy. This method is expected to help DL models to rapidly learn discriminative features in complex recognition environments.

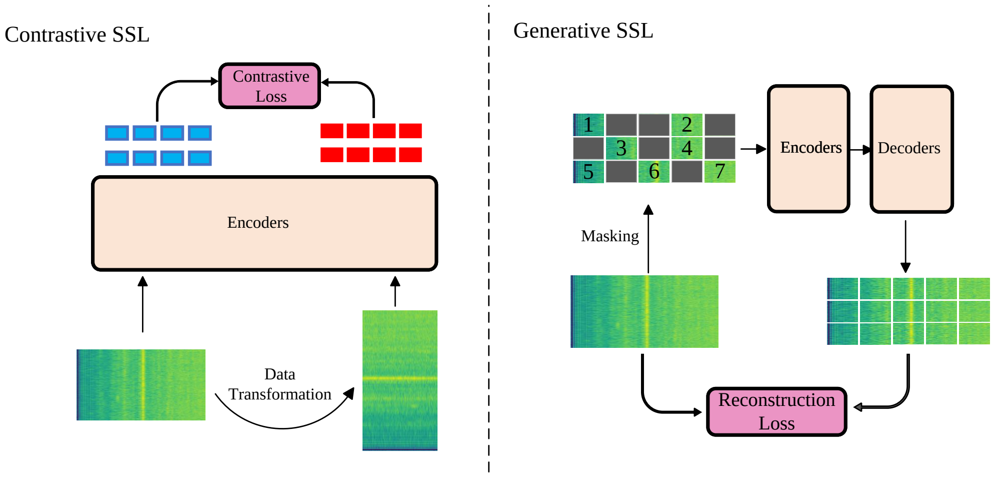

Figure 16. Contrastive and generative SSL methods for UATR.

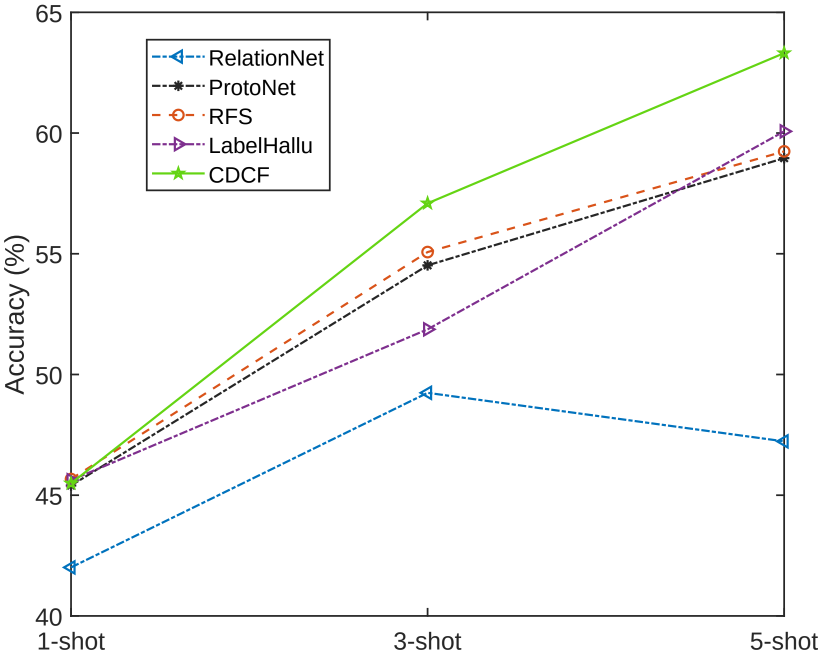

Figure 17. Accuracy comparison on three few-shot tasks based on Shipsear dataset, as adapted from Cui et al. [179].

5.2. Interpretability Problem

The challenge of the interpretability problem comes from the DL itself as a "black box" [183] and the complex mechanism of underwater acoustic signals. Currently, it remains unclear which discriminative features neural networks actually learn and whether temporal or frequency structures play crucial roles in the recognition process for all these models. In the field of computer vision, while visualizing feature maps partially explains how models recognize specific object contours by capturing texture and shape, these methods fail to fully reveal the inner workings of the models [38,139,184]. Even in the domain of speech recognition, which shares similarities with UATR tasks [185,186], existing interpretability studies have not provided a clear physical mechanism explanation due to lacking specific semantic information in acoustic signals.

On the other hand, UATR is closely related to underwater physics. The complex acoustic propagation channels and dynamic marine environments have substantial influence on acoustic signal transmission, which leads to nonstationary, non-Gaussianity, and nonlinearity characteristics. These complexities pose challenges in extracting crucial information within the underwater signals. As a result, DL models may encounter difficulties in effectively perceiving important physical information during decision making, which affects their interpretability.

To address this challenge, it is crucial to keep studying feature map visualization in UATR, such as Grad-CAM [146,187], t-distributed stochastic neighbor embedding (T-SNE) reduction [180,188,189], and attention [158] mechanisms. Among these methods, Grad-CAM and attention mechanisms are more inclined to explore interpretability from the underwater acoustic spectrogram, while t-SNE is more focused on learning the data distribution to enhance the interpretability, as can be seen in Figure 18. Generally speaking, these methods are mostly driven by image-based approaches without considering intrinsic properties. As a result, forthcoming interpretability research will focus on the perspective of features rather than simply treating acoustic features as natural images. The introduction of graph processing may be a promising solution because graphs are well suited to represent complex relationships [190]. By modeling signal features or signal samples as graphs, it is possible to form connected graphs that reveal the correlations between regions, thereby enhancing model interpretability.

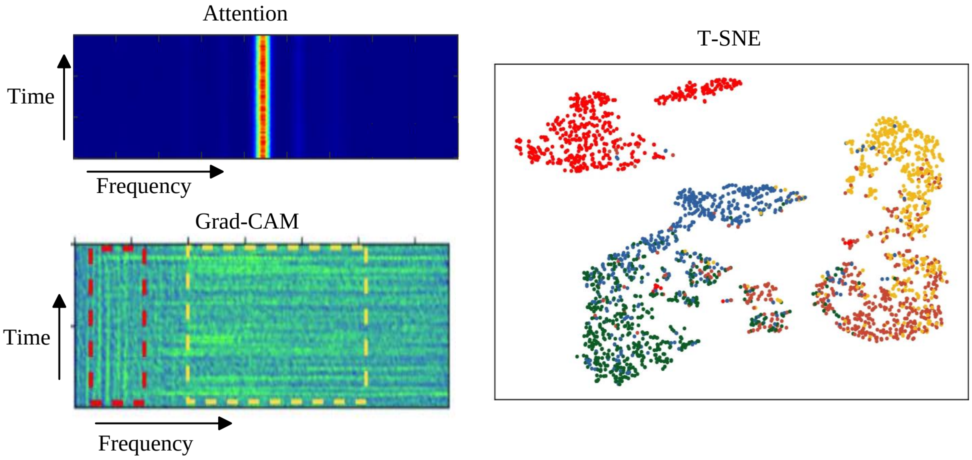

Figure 18. Interpretable methods in intelligent UATR.

5.3. Generalization Issue

In a standard intelligent UATR paradigm, both training and testing datasets typically come from the same sea areas and ship types. However, real ocean recognition scenarios often involve unseen sea areas and ships, leading to significant differences between data distributions known as environment mismatch. In such cases, the model's generalization ability to new data can be reduced, leading to a significant decrease in accuracy. To address this problem, data augmentation [63,114,191] and transfer learning [167,192] techniques are introduced into this field. However, due to the vast differences in actual marine environments, the volume of data used is still insufficient to leverage the advantages of data-driven learning. Consequently, these methods have not yet been able to consistently deliver high recognition accuracy across different sea areas.

The recent advancements of universal large models may further enhance the generalization of recognition models. Large DL models that have hundreds of millions of parameters, such as ChatGPT [193], Whisper [194], Pangu-Weather [195], and Xihe-ocean forecasting model [196], have achieved considerable success in their respective fields and demonstrate strong generalization capabilities even in zero-shot tasks. However, due to insufficient acoustic data in the field of underwater acoustic signals, it is impractical to directly train a universal recognition model to classify underwater signals. To address this challenge, advanced transfer learning techniques may benefit from universal model training [197], which involves aligning acoustic data with upstream tasks. Additionally, combining these techniques with efficient fine-tuning methods, such as Lora [198], QLora [199], and Adapter Tuning [200], holds potential for resolving the generalization problem arising from environmental mismatches.

5.4. Adversarial Robustness Challenge

Numerous studies have demonstrated the inherent fragility of AI methods. Specifically, well-designed perturbations have been shown to significantly degrade model performance, thereby reducing their reliability [201,202]. In practice, these perturbations create adversarial samples that cause neural networks to learn discontinuous mappings. Such adversarial samples can significantly confuse the backpropagation process of DL models, resulting in unexpected misclassifications. In safety-critical domains, the presence of adversarial samples poses a substantial threat to various intelligent applications, potentially leading to severe consequences [203]. Take autonomous driving as an example. Adversarial samples generated by altering light brightness or modifying traffic signal indicators can cause in-vehicle intelligent devices to malfunction, ultimately resulting in accidents [204]. UATR also belongs to the safety-critical area with significant military implications [205]. Adversarial attacks on intelligent recognition systems could lead to unexpected and severe outcomes with the misclassification of ships or inadequate detection of critical underwater targets.

The impact of adversarial attacks on intelligent recognition systems is illustrated in Figure 19. Taking the label data of motorboats as an example, both the waveform signals and the time–frequency spectrograms are subjected to small perturbations when under adversarial attacks, resulting in imperceptible adversarial data. These adversarial data are essentially indistinguishable from the original data, yet they lead to the classifier misclassifying motorboat samples as oil tankers. Such failures have the potential to impact strategic decisions, carrying serious implications for military operations and national security.

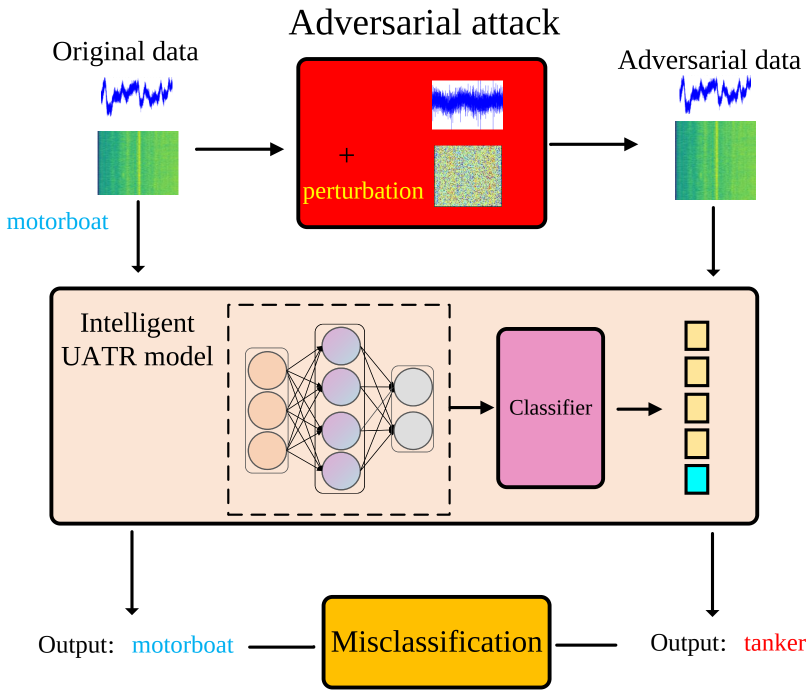

Figure 19. Misclassification caused by adversarial attacks on intelligent UATR systems.

In [135], Ma et al. investigated the Lambertian-based adversarial attacks in the field of underwater side-scan sonar image classification, which demonstrated higher attack success rate with other adversarial attacks. Moreover, Feng et al. [206] evaluated the adversarial robustness of the advanced intelligent UATR systems based on acoustic spectrogram inputs. Experimental findings indicate that both CNNs and Transformers significantly degrade the recognition performance in the presence of adversarial perturbations. However, it is worth noted that their study is based on data from the receiver side without considering the transmitter side. Since the UATR task is closely related to the transmitter, acoustic channel, and receiver, the best practice is to design the adversarial attack algorithm from the transmitter to attack the intelligent recognition system at the receiver. Such an approach is limited by the scarcity of underwater acoustic samples and the privacy of ship physical parameters.

For adversarial defense that shares insights, it is more appropriate to investigate it from the receiver side, since in practice the receiver may not clearly know the attack type and source. A major future direction in this area is to investigate intelligent recognition systems under strong adversarial attacks. Potential solutions include adversarial training [207], adversarial sample detection [208], and defensive distillation [209].

6. Conclusions

With the rapid progress in AI, AI-based UATR has seen significant maturation for effective ocean remote sensing. To further enhance the understanding of this field, this paper provides a comprehensive review of both traditional and state-of-the-art feature extraction and classification methods in AI-based UATR, highlighting their respective strengths and potential shortcomings. In addition, the current prevalent issues of the complex recognition condition, generalizability, interpretability, and adversarial robustness are thoroughly discussed. To address the above challenges, innovative research directions and suggested methodologies that hold promise are proposed for future research. These directions will address current challenges and are expected to become significant research areas in UATR. As these technologies evolve, they promise to revolutionize ocean remote sensing, offering new insights and capabilities for underwater exploration and monitoring.

Author Contributions: Conceptualization, S.F. and S.M.; investigation, S.F. and S.M.; project administration, S.M.; resources, M.Y.; supervision, S.M. and M.Y.; visualization, S.F.; writing—original draft, S.F.; writing—review and editing, S.F., S.M., X.Z. and M.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the National Defense Fundamental Scientific Research Program under Grant Number JCKY2020550C011.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Santos-Domínguez, D.; Torres-Guijarro, S.; Cardenal-López, A.; Pena-Gimenez, A. ShipsEar: An underwater vessel noise database. *Appl. Acoust.* **2016**, *113*, 64–69. [CrossRef]
2. Irfan, M.; Jiangbin, Z.; Ali, S.; Iqbal, M.; Masood, Z.; Hamid, U. DeepShip: An underwater acoustic benchmark dataset and a separable convolution based autoencoder for classification. *Expert Syst. Appl.* **2021**, *183*, 115270. [CrossRef]
3. Niu, H.; Li, X.; Zhang, Y.; Xu, J. Advances and applications of machine learning in underwater acoustics. *Intell. Mar. Technol. Syst.* **2023**, *1*, 8. [CrossRef]
4. Zhang, X.; Yang, P.; Wang, Y.; Shen, W.; Yang, J.; Wang, J.; Ye, K.; Zhou, M.; Sun, H. A Novel Multireceiver SAS RD Processor. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–11. [CrossRef]
5. Zhang, X.; Yang, P.; Wang, Y.; Shen, W.; Yang, J.; Ye, K.; Zhou, M.; Sun, H. LBF-Based CS Algorithm for Multireceiver SAS. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 1–5. [CrossRef]
6. Zhang, H.; Shafiq, M.O. Survey of transformers and towards ensemble learning using transformers for natural language processing. *J. Big Data* **2024**, *11*, 25. [CrossRef]
7. Zhang, X.; Yang, P.; Sun, H. Frequency-domain multireceiver synthetic aperture sonar imagery with Chebyshev polynomials. *Electron. Lett.* **2022**, *58*, 995–998. [CrossRef]
8. Zhang, X.; Yang, P.; Feng, X.; Sun, H. Efficient imaging method for multireceiver SAS. *IET Radar Sonar Navig.* **2022**, *16*, 1470–1483. [CrossRef]
9. Zhang, R.; He, C.; Jing, L.; Zhou, C.; Long, C.; Li, J. A Modulation Recognition System for Underwater Acoustic Communication Signals Based on Higher-Order Cumulants and Deep Learning. *J. Mar. Sci. Eng.* **2023**, *11*, 1632. [CrossRef]
10. Jiang, Z.; Zhang, J.; Wang, T.; Wang, H. Modulation recognition of underwater acoustic communication signals based on neural architecture search. *Appl. Acoust.* **2024**, *225*, 110155. [CrossRef]

11. Wu, J.; Chen, Y.; Jia, B.; Li, G.; Zhang, Y.; Yong, J. Optimal design of emission waveform for acoustic scattering test under multipath interference. In Proceedings of the 2020 5th International Conference on Communication, Image and Signal Processing (CCISP), Chengdu, China, 13–15 November 2020; pp. 102–106. [CrossRef]
12. Sumithra, G.; Ajay, N.; Neeraja, N.; Adityaraj, K. Hybrid Acoustic System for Underwater Target Detection and Tracking. *Int. J. Appl. Comput. Math.* **2023**, *9*, 149. [CrossRef]
13. Zhu, J.; Xie, Z.; Jiang, N.; Song, Y.; Han, S.; Liu, W.; Huang, X. Delay-Doppler Map Shaping through Oversampled Complementary Sets for High-Speed Target Detection. *Remote Sens.* **2024**, *16*, 2898. [CrossRef]
14. Zhu, J.; Song, Y.; Jiang, N.; Xie, Z.; Fan, C.; Huang, X. Enhanced Doppler Resolution and Sidelobe Suppression Performance for Golay Complementary Waveforms. *Remote Sens.* **2023**, *15*, 2452. [CrossRef]
15. Yoo, K.B.; Edelmann, G.F. Low complexity multipath and Doppler compensation for direct-sequence spread spectrum signals in underwater acoustic communication. *Appl. Acoust.* **2021**, *180*, 108094. [CrossRef]
16. Klionskii, D.M.; Kaplun, D.I.; Voznesensky, A.S.; Romanov, S.A.; Levina, A.B.; Bogaevskiy, D.V.; Geppener, V.V.; Razmochaeva, N.V. Solution of the Problem of Classification of Hydroacoustic Signals Based on Harmonious Wavelets and Machine Learning. *Pattern Recognit. Image Anal.* **2020**, *30*, 480–488. [CrossRef]
17. Quraishi, S.J.; Singh, M.; Prasad, S.K.; Arora, K.; Pathak, S.; Singh, A. A Machine Learning Approach to Rock and Mine Classification in Sonar Systems Using Logistic Regression. In Proceedings of the 2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan, 1–3 November 2023; pp. 462–468. [CrossRef]
18. Wang, P.; Peng, Y. Research on Feature Extraction and Recognition Method of Underwater Acoustic Target Based on Deep Convolutional Network. In Proceedings of the 2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA), Dalian, China, 25–27 August 2020; pp. 863–868. [CrossRef]
19. Doan, V.S.; Huynh-The, T.; Kim, D.S. Underwater Acoustic Target Classification Based on Dense Convolutional Neural Network. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
20. Meng, Q.; Yang, S. A wave structure based method for recognition of marine acoustic target signals. *J. Acoust. Soc. Am.* **2015**, *137*, 2242. [CrossRef]
21. Deng, J.; Yang, X.; Liu, L.; Shi, L.; Li, Y.; Yang, Y. Real-Time Underwater Acoustic Homing Weapon Target Recognition Based on a Stacking Technique of Ensemble Learning. *J. Mar. Sci. Eng.* **2023**, *11*, 2305. [CrossRef]
22. Meng, Q.; Yang, S.; Piao, S. The classification of underwater acoustic target signals based on wave structure and support vector machine. *J. Acoust. Soc. Am.* **2014**, *136*, 2265. [CrossRef]
23. Sun, Y.; Zhang, X. Analysis of Chaotic Characteristics of Ship Radiated Noise Signals with Different Data Lengths. In Proceedings of the OCEANS 2022, Chennai, VA, USA, 17–20 October 2022; pp. 1–7. [CrossRef]
24. van Haarlem, M. LOFAR: The Low Frequency Array. *Eas Publ. Ser.* **2005**, *15*, 431–444. [CrossRef]
25. Chung, K.W.; Sutin, A.; Sedunov, A.; Bruno, M.S. DEMON Acoustic Ship Signature Measurements in an Urban Harbor. *Adv. Acoust. Vib.* **2011**, *2011*, 952798. [CrossRef]
26. Jin, G.; Liu, F.; Wu, H.; Song, Q. Deep learning-based framework for expansion, recognition and classification of underwater acoustic signal. *J. Exp. Theor. Artif. Intell.* **2019**, *32*, 205–218. [CrossRef]
27. Chen, J.; Han, B.; Ma, X.; Zhang, J. Underwater Target Recognition Based on Multi-Decision LOFAR Spectrum Enhancement: A Deep-Learning Approach. *Future Internet* **2021**, *13*, 265. [CrossRef]
28. Shi, Y.; Piao, S.; Guo, J. Line spectrum detection and motion parameters estimation for underwater moving target. *J. Phys. Conf. Ser.* **2024**, *2718*, 012090. [CrossRef]
29. Pollara, A.; Sutin, A.; Salloum, H. Improvement of the Detection of Envelope Modulation on Noise (DEMON) and its application to small boats. In Proceedings of the OCEANS 2016 MTS/IEEE, Monterey, CA, USA, 19–23 September 2016; pp. 1–10. [CrossRef]
30. Tong, W.; Wu, K.; Wang, H.; Cao, L.; Huang, B.; Wu, D.; Antoni, J. Adaptive Weighted Envelope Spectrum: A robust spectral quantity for passive acoustic detection of underwater propeller based on spectral coherence. *Mech. Syst. Signal Process.* **2024**, *212*, 111265. [CrossRef]
31. Li, L.; Song, S.; Feng, X. Combined LOFAR and DEMON Spectrums for Simultaneous Underwater Acoustic Object Counting and F0 Estimation. *J. Mar. Sci. Eng.* **2022**, *10*, 1565. [CrossRef]
32. Yan, J.; Sun, H.; Chen, H.; Junejo, N.U.R.; Cheng, E. Resonance-Based Time-Frequency Manifold for Feature Extraction of Ship-Radiated Noise. *Sensors* **2018**, *18*, 936. [CrossRef] [PubMed]
33. Cao, X.; Togneri, R.; Zhang, X.; Yu, Y. Convolutional Neural Network with Second-Order Pooling for Underwater Target Classification. *IEEE Sens. J.* **2019**, *19*, 3058–3066. [CrossRef]
34. Li, J.; Wang, B.; Cui, X.; Li, S.; Liu, J. Underwater Acoustic Target Recognition Based on Attention Residual Network. *Entropy* **2022**, *24*, 1657. [CrossRef]
35. Gabor, D. Theory of communication. *J. Inst. Electr. Eng. Part I Gen.* **1946**, *94*, 58. [CrossRef]
36. Ioup, J.W. Time-frequency analysis for acoustics education and for listening to whales in the Gulf of Mexico. *J. Acoust. Soc. Am.* **2013**, *134*, 4124. [CrossRef]
37. Luo, X.; Zhang, M.; Liu, T.; Huang, M.; Xu, X. An Underwater Acoustic Target Recognition Method Based on Spectrograms with Different Resolutions. *J. Mar. Sci. Eng.* **2021**, *9*, 1246. [CrossRef]
38. Zhang, Y.; Zeng, Q. MSLEFC: A low-frequency focused underwater acoustic signal classification and analysis system. *Eng. Appl. Artif. Intell.* **2023**, *123*, 106333. [CrossRef]

39. Yang, M.; Li, X.; Yang, Y.; Meng, X. Characteristic analysis of underwater acoustic scattering echoes in the wavelet transform domain. *J. Mar. Sci. Appl.* **2017**, *16*, 93–101. [CrossRef]
40. Jing, L.; Zheng, T.; He, C.; Yin, H. Iterative adaptive frequency-domain equalization based on sliding window strategy over time-varying underwater acoustic channels. *JASA Express Lett.* **2021**, *1*, 076002. [CrossRef] [PubMed]
41. Morlet, J.; Arens, G.; Fourgeau, E.; Glard, D. Wave propagation and sampling theory—Part I: Complex signal and scattering in multilayered media. *Geophysics* **1982**, *47*, 203–221. [CrossRef]
42. Xin-xin, L.; Shi-e, Y.; Ming, Y. Feature extraction from underwater signals using wavelet packet transform. In Proceedings of the 2008 International Conference on Neural Networks and Signal Processing, Nanjing, China, 7–11 June 2008; pp. 400–405. [CrossRef]
43. Rademan, M.; Versfeld, D.; du Preez, J. Soft-output signal detection for cetacean vocalizations using spectral entropy, k-means clustering and the continuous wavelet transform. *Ecol. Inform.* **2023**, *74*, 101990. [CrossRef]
44. Han, Z.; Zhang, X.; Yan, B.; Qiao, L.; Wang, Z. The time-frequency analysis of the acoustic signal produced in underwater discharges based on Variational Mode Decomposition and Hilbert–Huang Transform. *Sci. Rep.* **2023**, *13*, 22. [CrossRef]
45. Choo, Y.S.; Byun, S.H.; Kim, S.M.; Lee, K. Target detection in pseudo Wigner-Ville distribution of underwater beamformed signals. *J. Acoust. Soc. Am.* **2019**, *146*, 2960. [CrossRef]
46. Wu, Y.; Li, X.; Wang, Y. Extraction and classification of acoustic scattering from underwater target based on Wigner-Ville distribution. *Appl. Acoust.* **2018**, *138*, 52–59. [CrossRef]
47. Wang, S.; Zeng, X. Robust underwater noise targets classification using auditory inspired time–frequency analysis. *Appl. Acoust.* **2014**, *78*, 68–76. [CrossRef]
48. Domingos, L.C.F.; Santos, P.E.; Skelton, P.S.M.; Brinkworth, R.S.A.; Sammut, K. An Investigation of Preprocessing Filters and Deep Learning Methods for Vessel Type Classification with Underwater Acoustic Data. *IEEE Access* **2022**, *10*, 117582–117596. [CrossRef]
49. Liu, Y.; Zhao, Y.; Gerstoft, P.; Zhou, F.; Qiao, G.; Yin, J. Deep transfer learning-based variable Doppler underwater acoustic communications. *J. Acoust. Soc. Am.* **2023**, *154*, 232–244. [CrossRef] [PubMed]
50. Tang, N.; Zhou, F.; Wang, Y.; Zhang, H.; Lyu, T.; Wang, Z.; Chang, L. Differential treatment for time and frequency dimensions in mel-spectrograms: An efficient 3D Spectrogram network for underwater acoustic target classification. *Ocean Eng.* **2023**, *287*, 115863. [CrossRef]
51. Abdul, Z.K.; Al-Talabani, A.K. Mel Frequency Cepstral Coefficient and its Applications: A Review. *IEEE Access* **2022**, *10*, 122136–122158. [CrossRef]
52. Zhang, Y.; Xu, K.; Wan, J. Rubost Feature for Underwater Targets Recognition Using Power-Normalized Cepstral Coefficients. In Proceedings of the 2018 14th IEEE International Conference on Signal Processing (ICSP), Beijing, China, 12–16 August 2018; pp. 90–93. [CrossRef]
53. Tong, Y.; Zhang, X.; Ge, Y. Classification and Recognition of Underwater Target Based on MFCC Feature Extraction. In Proceedings of the 2020 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), Macau, China, 21–24 August 2020; pp. 1–4. [CrossRef]
54. Ren, J.; Huang, Z.; Li, C.; Guo, X.; Xu, J. Feature Analysis of Passive Underwater Targets Recognition Based on Deep Neural Network. In Proceedings of the OCEANS 2019, Marseille, France, 17–20 June 2019; pp. 1–5. [CrossRef]
55. Hu, F.; Fan, J.; Kong, Y.; Zhang, L.; Guan, X.; Yu, Y. A Deep Learning Method for Ship-Radiated Noise Recognition Based on MFCC Feature. In Proceedings of the 2023 7th International Conference on Transportation Information and Safety (ICTIS), Xian, China, 4–6 August 2023; pp. 1328–1335. [CrossRef]
56. Lian, Z.; Xu, K.; Wan, J.; Li, G. Underwater acoustic target classification based on modified GFCC features. In Proceedings of the 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, China, 25–26 March 2017; pp. 258–262. [CrossRef]
57. Cao, Y.; Yan, J.; Sun, K.; Luo, X. Hydroacoustic Target Detection Based on Improved GFCC and Lightweight Neural Network. In Proceedings of the 2023 42nd Chinese Control Conference (CCC), Tianjin, China, 24–26 July 2023; pp. 6239–6243. [CrossRef]
58. Liu, G.K. Evaluating Gammatone Frequency Cepstral Coefficients with Neural Networks for Emotion Recognition from Speech. *arXiv* **2018**, arXiv:1806.09010.
59. Song, K.; Wang, N.; Zhang, Y. An Improved Deep Canonical Correlation Fusion Method for Underwater Multisource Data. *IEEE Access* **2020**, *8*, 146300–146307. [CrossRef]
60. Hong, F.; Liu, C.; Guo, L.; Chen, F.; Feng, H. Underwater Acoustic Target Recognition with a Residual Network and the Optimized Feature Extraction Method. *Appl. Sci.* **2021**, *11*, 1442. [CrossRef]
61. Chen, Z.; Tang, J.; Qiu, H.; Chen, M. MGFGNet: An automatic underwater acoustic target recognition method based on the multi-gradient flow global feature enhancement network. *Front. Mar. Sci.* **2023**, *10*, 1306229. [CrossRef]
62. Tan, J.; Pan, X. Underwater acoustic target recognition based on convolutional neural network and multi-feature fusion. In Proceedings of the Third International Conference on Computer Vision and Pattern Analysis (ICCPA 2023), Hangzhou, China, 31 March–2 April 2023; Shen, L., Zhong, G., Eds.; International Society for Optics and Photonics, SPIE; Volume 12754, p. 1275432. [CrossRef]
63. Liu, F.; Shen, T.; Luo, Z.; Zhao, D.; Guo, S. Underwater target recognition using convolutional recurrent neural networks with 3-D Mel-spectrogram and data augmentation. *Appl. Acoust.* **2021**, *178*, 107989. [CrossRef]

64. Wu, J.; Li, P.; Wang, Y.; Lan, Q.; Xiao, W.; Wang, Z. VFR: The Underwater Acoustic Target Recognition Using Cross-Domain Pre-Training with FBank Fusion Features. *J. Mar. Sci. Eng.* **2023**, *11*, 263. [CrossRef]
65. Yang, Y.; Lv, H.; Chen, N. A Survey on ensemble learning under the era of deep learning. *Artif. Intell. Rev.* **2023**, *56*, 5545–5589. [CrossRef]
66. Zhang, Q.; Da, L.; Zhang, Y.; Hu, Y. Integrated neural networks based on feature fusion for underwater target recognition. *Appl. Acoust.* **2021**, *182*, 108261. [CrossRef]
67. Luo, X.; Feng, Y. An Underwater Acoustic Target Recognition Method Based on Restricted Boltzmann Machine. *Sensors* **2020**, *20*, 5399. [CrossRef]
68. Nie, L.; Zhang, Y.; Wang, H. Classification of underwater soundscapes using raw hydroacoustic signals. *J. Acoust. Soc. Am.* **2023**, *154*, A304. [CrossRef]
69. Chen, Y.; Shang, J. Underwater Target Recognition Method Based on Convolution Autoencoder. In Proceedings of the 2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP), Chongqing, China, 11–13 December 2019; pp. 1–5. [CrossRef]
70. Khishe, M. DRW-AE: A Deep Recurrent-Wavelet Autoencoder for Underwater Target Recognition. *IEEE J. Ocean. Eng.* **2022**, *47*, 1083–1098. [CrossRef]
71. Berahmand, K.; Daneshfar, F.; Salehi, E.S.; Li, Y.; Xu, Y. Autoencoders and their applications in machine learning: A survey. *Artif. Intell. Rev.* **2024**, *57*, 28. [CrossRef]
72. Dong, Y.; Shen, X.; Wang, H. Bidirectional Denoising Autoencoders-Based Robust Representation Learning for Underwater Acoustic Target Signal Denoising. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–8. [CrossRef]
73. Li, J.; Yang, H.; Shen, S.; Xu, G. The Learned Multi-scale Deep Filters for Underwater Acoustic Target Modeling and Recognition. In Proceedings of the OCEANS 2019, Marseille, France, 17–20 June 2019; pp. 1–4. [CrossRef]
74. Luo, X.; Feng, Y.; Zhang, M. An Underwater Acoustic Target Recognition Method Based on Combined Feature with Automatic Coding and Reconstruction. *IEEE Access* **2021**, *9*, 63841–63854. [CrossRef]
75. Sun, B.; Luo, X. Underwater acoustic target recognition based on automatic feature and contrastive coding. *IET Radar Sonar Navig.* **2023**, *17*, 1277–1285. [CrossRef]
76. Wang, X.; Meng, J.; Liu, Y.; Zhan, G.; Tian, Z. Self-supervised acoustic representation learning via acoustic-embedding memory unit modified space autoencoder for underwater target recognition. *J. Acoust. Soc. Am.* **2022**, *152*, 2905–2915. [CrossRef] [PubMed]
77. Gomez, B.; Kadri, U. Earthquake source characterization by machine learning algorithms applied to acoustic signals. *Sci. Rep.* **2021**, *11*, 23062. [CrossRef] [PubMed]
78. Zelada Leon, A.; Huvenne, V.A.; Benoist, N.M.; Ferguson, M.; Bett, B.J.; Wynn, R.B. Assessing the Repeatability of Automated Seafloor Classification Algorithms, with Application in Marine Protected Area Monitoring. *Remote Sens.* **2020**, *12*, 1572. [CrossRef]
79. Harakawa, R.; Ogawa, T.; Haseyama, M.; Akamatsu, T. Automatic detection of fish sounds based on multi-stage classification including logistic regression via adaptive feature weighting. *J. Acoust. Soc. Am.* **2018**, *144*, 2709–2718. [CrossRef]
80. Xinhua, Z.; Zhenbo, L.; Chunyu, K. Underwater acoustic targets classification using support vector machine. In Proceedings of the International Conference on Neural Networks and Signal Processing, Nanjing, China, 14–17 December 2003; Volume 2, pp. 932–935. [CrossRef]
81. Alvaro, A.; Schwock, F.; Ragland, J.; Abadi, S. Ship detection from passive underwater acoustic recordings using machine learning. *J. Acoust. Soc. Am.* **2021**, *150*, A124. [CrossRef]
82. Li, H.; Cheng, Y.; Dai, W.; Li, Z. A method based on wavelet packets-fractal and SVM for underwater acoustic signals recognition. In Proceedings of the 2014 12th International Conference on Signal Processing (ICSP), HangZhou, China, 19–23 October 2014; pp. 2169–2173. [CrossRef]
83. de Moura, N.N.; de Seixas, J.M. Novelty detection in passive SONAR systems using support vector machines. In Proceedings of the 2015 Latin America Congress on Computational Intelligence (LA-CCI), Curitiba, Brazil, 13–16 October 2015; pp. 1–6. [CrossRef]
84. Sherin, B.M.; Supriya, M.H. Selection and parameter optimization of SVM kernel function for underwater target classification. In Proceedings of the 2015 IEEE Underwater Technology (UT), Chennai, India, 23–25 February 2015; pp. 1–5. [CrossRef]
85. Wang, B.; Wu, C.; Zhu, Y.; Zhang, M.; Li, H.; Zhang, W. Ship Radiated Noise Recognition Technology Based on ML-DS Decision Fusion. *Comput. Intell. Neurosci.* **2021**, *2021*, 8901565. [CrossRef]
86. Yao, Q.; Jiang, J.; Chen, G.; Li, Z.; Yao, Z.; Lu, Y.; Hou, X.; Fu, X.; Duan, F. Recognition method for underwater imitation whistle communication signals by slope distribution. *Appl. Acoust.* **2023**, *211*, 109531. [CrossRef]
87. Liu, F.; Li, G.; Yang, H. Application of multi-algorithm mixed feature extraction model in underwater acoustic signal. *Ocean Eng.* **2024**, *296*, 116959. [CrossRef]
88. Yang, H.; Gan, A.; Chen, H.; Pan, Y.; Tang, J.; Li, J. Underwater acoustic target recognition using SVM ensemble via weighted sample and feature selection. In Proceedings of the 2016 13th International Bhurban Conference on Applied Sciences and Technology (IBCAST), Islamabad, Pakistan, 12–16 January 2016; pp. 522–527. [CrossRef]
89. Fisher, R.A. *Iris*; UCI Machine Learning Repository: Espoo, Finland, 1988. [CrossRef]
90. Choi, J.; Choo, Y.; Lee, K. Acoustic Classification of Surface and Underwater Vessels in the Ocean Using Supervised Machine Learning. *Sensors* **2019**, *19*, 3492. [CrossRef]

91. Wei, M.; Chen, K.; Lin, Y.; Cheng, E. Recognition of behavior state of Penaeus vannamei based on passive acoustic technology. *Front. Mar. Sci.* **2022**, *9*, 973284. [CrossRef]
92. Chen, H.; Sun, H.; Junejo, N.U.R.; Yang, G.; Qi, J. Whale Vocalization Classification Using Feature Extraction with Resonance Sparse Signal Decomposition and Ridge Extraction. *IEEE Access* **2019**, *7*, 136358–136368. [CrossRef]
93. Yaman, O.; Tuncer, T.; Tasar, B. DES-Pat: A novel DES pattern-based propeller recognition method using underwater acoustical sounds. *Appl. Acoust.* **2021**, *175*, 107859. [CrossRef]
94. Saffari, A.; Zahiri, S.H.; Khishe, M. Automatic recognition of sonar targets using feature selection in micro-Doppler signature. *Def. Technol.* **2023**, *20*, 58–71. [CrossRef]
95. Li, Y.X.; Jiao, S.B.; Geng, B.; Zhang, Q.; Zhang, Y.M. A comparative study of four nonlinear dynamic methods and their applications in classification of ship-radiated noise. *Def. Technol.* **2022**, *18*, 183–193. [CrossRef]
96. Jin, S.Y.; Su, Y.; Guo, C.J.; Fan, Y.X.; Tao, Z.Y. Offshore ship recognition based on center frequency projection of improved EMD and KNN algorithm. *Mech. Syst. Signal Process.* **2023**, *189*, 110076. [CrossRef]
97. Mohammed, S.K.; Hariharan, S.M.; Kamal, S. A GTCC-Based Underwater HMM Target Classifier with Fading Channel Compensation. *J. Sens.* **2018**, *2018*, 6593037:1–6593037:14. [CrossRef]
98. You, H.; Byun, S.H.; Choo, Y. Underwater Acoustic Signal Detection Using Calibrated Hidden Markov Model with Multiple Measurements. *Sensors* **2022**, *22*, 5088. [CrossRef]
99. Seo, Y.; On, B.; Im, S.; Shim, T.; Seo, I. Underwater Cylindrical Object Detection Using the Spectral Features of Active Sonar Signals with Logistic Regression Models. *Appl. Sci.* **2018**, *8*, 116. [CrossRef]
100. Yang, L.; Chen, K. Performance and strategy comparisons of human listeners and logistic regression in discriminating underwater targets. *J. Acoust. Soc. Am.* **2015**, *138*, 3138–3147. [CrossRef]
101. K, S.; R, K.; Kumar, P.S.; V, R.; Lakshmi, G. Rock/Mine Classification Using Supervised Machine Learning Algorithms. In Proceedings of the 2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE), Bengaluru, India, 27–28 January 2023; pp. 177–184. [CrossRef]
102. Yu, Q.; Zhang, W.; Zhu, M.; Shi, J.; Liu, Y.; Liu, S. Surface and Underwater Acoustic Source Recognition Using Array Feature Extraction Based on Machine Learning. *J. Phys. Conf. Ser.* **2024**, *2718*, 012100. [CrossRef]
103. Zhou, X.; Yang, K. A denoising representation framework for underwater acoustic signal recognition. *J. Acoust. Soc. Am.* **2020**, *147*, EL377–EL383. [CrossRef] [PubMed]
104. Wang, Q.; Wang, L.; Zeng, X.; Zhao, L. An Improved Deep Clustering Model for Underwater Acoustical Targets. *Neural Process. Lett.* **2018**, *48*, 1633–1644. [CrossRef]
105. Sabara, R.; Soares, C.; Zabel, F.; Oliveira, J.V.; Jesus, S.M. Automatic Acoustic Target Detection and Classification off the Coast of Portugal. In Proceedings of the Global Oceans 2020: Singapore—U.S. Gulf Coast, Biloxi, MS, USA, 5–30 October 2020; pp. 1–9. [CrossRef]
106. Yang, K.; Zhou, X. Unsupervised Classification of Hydrophone Signals with an Improved Mel-Frequency Cepstral Coefficient Based on Measured Data Analysis. *IEEE Access* **2019**, *7*, 124937–124947. [CrossRef]
107. Agersted, M.D.; Khodabandeloo, B.; Liu, Y.; Melle, W.; Klevjer, T.A. Application of an unsupervised clustering algorithm on in situ broadband acoustic data to identify different mesopelagic target types. *Ices J. Mar. Sci.* **2021**, *78*, 2907–2921. [CrossRef]
108. Luo, X.; Chen, L.; Zhou, H.; Cao, H. A Survey of Underwater Acoustic Target Recognition Methods Based on Machine Learning. *J. Mar. Sci. Eng.* **2023**, *11*, 384. [CrossRef]
109. Baran, R.H.; Coughlan, J.M. Neural network for passive acoustic discrimination between surface and submarine targets. In Proceedings of the Automatic Object Recognition, San Francisco, CA, USA, 1 August 1991; Sadjadi, F.A., Ed.; International Society for Optics and Photonics, SPIE; Volume 1471, pp. 164–176. [CrossRef]
110. Khotanzad.; Lu.; Srinath. Target detection using a neural network based passive sonar system. In Proceedings of the International 1989 Joint Conference on Neural Networks, Washington, DC, USA, 18–22 June 1989; Volume 1, pp. 335–340. [CrossRef]
111. Filho, W. Preprocessing passive sonar signals for neural classification. *IET Radar Sonar Navig.* **2011**, *5*, 605–612. [CrossRef]
112. Yue, H.; Zhang, L.; Wang, D.; Wang, Y.; Lu, Z. The Classification of Underwater Acoustic Targets Based on Deep Learning Methods. In Proceedings of the 2017 2nd International Conference on Control, Automation and Artificial Intelligence (CAAI 2017), Sanya, China, 25–26 June 2017; pp. 526–529. [CrossRef]
113. Zhao, J.; Wang, S.; Jia, X.; Gao, Y.; Zhu, W.; Ma, F.; Liu, Q. Underwater target perception algorithm based on pressure sequence generative adversarial network. *Ocean Eng.* **2023**, *286*, 115547. [CrossRef]
114. Li, D.; Liu, F.; Shen, T.; Chen, L.; Zhao, D. Data augmentation method for underwater acoustic target recognition based on underwater acoustic channel modeling and transfer learning. *Appl. Acoust.* **2023**, *208*, 109344. [CrossRef]
115. Yang, J.; Yan, S.; Zeng, D.; Tan, G. Self-supervised learning minimax entropy domain adaptation for the underwater target recognition. *Appl. Acoust.* **2024**, *216*, 109725. [CrossRef]
116. Wang, X.; Wu, P.; Li, B.; Zhan, G.; Liu, J.; Liu, Z. A self-supervised dual-channel self-attention acoustic encoder for underwater acoustic target recognition. *Ocean Eng.* **2024**, *299*, 117305. [CrossRef]
117. Xu, K.; Xu, Q.; You, K.; Zhu, B.; Feng, M.; Feng, D.; Liu, B. Self-supervised learning–based underwater acoustical signal classification via mask modeling. *J. Acoust. Soc. Am.* **2023**, *154*, 5–15. [CrossRef] [PubMed]
118. You, K.; Xu, K.; Feng, M.; Zhu, B. Underwater acoustic classification using masked modeling-based swin transformer. *J. Acoust. Soc. Am.* **2022**, *152*, A296. [CrossRef]

119. Zhou, A.; Li, X.; Zhang, W.; Zhao, C.; Ren, K.; Ma, Y.; Song, J. An attention-based multi-scale convolution network for intelligent underwater acoustic signal recognition. *Ocean Eng.* **2023**, *287*, 115784. [CrossRef]
120. Schoene, A.M.; Turner, A.P.; De Mel, G.; Dethlefs, N. Hierarchical Multiscale Recurrent Neural Networks for Detecting Suicide Notes. *IEEE Trans. Affect. Comput.* **2023**, *14*, 153–164. [CrossRef]
121. Bansal, A.; Garg, N.K. Robust technique for environmental sound classification using convolutional recurrent neural network. *Multimed. Tools Appl.* **2023**, *83*, 54755–54772. [CrossRef]
122. Wang, Y.; Wu, H.; Zhang, J.; Gao, Z.; Wang, J.; Yu, P.S.; Long, M. PredRNN: A Recurrent Neural Network for Spatiotemporal Predictive Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 2208–2225. [CrossRef]
123. Zhang, X.; Zhong, C.; Zhang, J.; Wang, T.; Ng, W.W. Robust recurrent neural networks for time series forecasting. *Neurocomputing* **2023**, *526*, 143–157. [CrossRef]
124. Hewamalage, H.; Bergmeir, C.; Bandara, K. Recurrent Neural Networks for Time Series Forecasting: Current status and future directions. *Int. J. Forecast.* **2021**, *37*, 388–427. [CrossRef]
125. Zhang, S.; Xing, S. Intelligent Recognition of Underwater Acoustic Target Noise by Multi-Feature Fusion. In Proceedings of the 2018 11th International Symposium on Computational Intelligence and Design (ISCID), Hangzhou, China, 8–9 December 2018; Volume 1, pp. 212–215. [CrossRef]
126. Zhang, S.; Wang, C.; Sun, Q. Underwater Target Noise Recognition and Classification Technology based on Multi-Classes Feature Fusion. *JNWPU* **2020**, *38*, 366–376. [CrossRef]
127. Yu, X.; Li, L.; Yin, J.; Shao, M.; Han, X. Modulation Pattern Recognition of Non-cooperative Underwater Acoustic Communication Signals Based on LSTM Network. In Proceedings of the 2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP), Chongqing, China, 11–13 December 2019; pp. 1–5. [CrossRef]
128. Yang, H.; Xu, G.; Yi, S.; Li, Y. A New Cooperative Deep Learning Method for Underwater Acoustic Target Recognition. In Proceedings of the OCEANS 2019, Marseille, France, 17–20 June 2019; pp. 1–4. [CrossRef]
129. Cho, K.; van Merrienboer, B.; Gülçehre, Ç.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP, Doha, Qatar, 25–29 October 2014; A meeting of SIGDAT, a Special Interest Group of the ACL; Moschitti, A., Pang, B., Daelemans, W., Eds.; ACL: Edinburgh, UK, 2014; pp. 1724–1734. [CrossRef]
130. Qi, P.; Yin, G.; Zhang, L. Underwater acoustic target recognition using RCRNN and wavelet-auditory feature. *Multimed. Tools Appl.* **2023**, *83*, 47295–47317. [CrossRef]
131. Wang, Y.; Zhang, H.; Xu, L.; Cao, C.; Gulliver, T.A. Adoption of hybrid time series neural network in the underwater acoustic signal modulation identification. *J. Frankl. Inst.* **2020**, *357*, 13906–13922. [CrossRef]
132. Liu, Y.; Li, X.; Yang, L.; Bian, G.; Yu, H. A CNN-Transformer Hybrid Recognition Approach for sEMG-Based Dynamic Gesture Prediction. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 1–16. [CrossRef]
133. Kamal, S.; Satheesh Chandran, C.; Supriya, M. Passive sonar automated target classifier for shallow waters using end-to-end learnable deep convolutional LSTMs. *Eng. Sci. Technol. Int. J.* **2021**, *24*, 860–871. [CrossRef]
134. Qi, P.; Sun, J.; Long, Y.; Zhang, L.; Tianye. Underwater Acoustic Target Recognition with Fusion Feature. In *Neural Information Processing, Proceedings of the 28th International Conference, ICONIP 2021, Sanur, Bali, Indonesia, 8–12December 2021*; Proceedings, Part I 28; Mantoro, T., Lee, M., Ayu, M.A., Wong, K.W., Hidayanto, A.N., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 609–620.
135. Ma, Q.; Jiang, L.; Yu, W. Lambertian-based adversarial attacks on deep-learning-based underwater side-scan sonar image classification. *Pattern Recognit.* **2023**, *138*, 109363. [CrossRef]
136. Steiniger, Y.; Kraus, D.; Meisen, T. Survey on deep learning based computer vision for sonar imagery. *Eng. Appl. Artif. Intell.* **2022**, *114*, 105157. [CrossRef]
137. Wang, Y.; Jin, Y.; Zhang, H.; Lu, Q.; Cao, C.; Sang, Z.; Sun, M. Underwater Communication Signal Recognition Using Sequence Convolutional Network. *IEEE Access* **2021**, *9*, 46886–46899. [CrossRef]
138. Xiaoping, S.; Jinsheng, C.; Yuan, G. A New Deep Learning Method for Underwater Target Recognition Based on One-Dimensional Time-Domain Signals. In Proceedings of the 2021 OES China Ocean Acoustics (COA), Harbin, China, 14–17 July 2021; pp. 1048–1051. [CrossRef]
139. Zhu, P.; Zhang, Y.; Huang, Y.; Zhao, C.; Zhao, K.; Zhou, F. Underwater acoustic target recognition based on spectrum component analysis of ship radiated noise. *Appl. Acoust.* **2023**, *211*, 109552. [CrossRef]
140. Dong, Y.; Shen, X.; Jiang, Z.; Wang, H. Recognition of imbalanced underwater acoustic datasets with exponentially weighted cross-entropy loss. *Appl. Acoust.* **2021**, *174*, 107740. [CrossRef]
141. Feng, S.; Zhu, X. A Transformer-Based Deep Learning Network for Underwater Acoustic Target Recognition. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
142. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
143. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269. [CrossRef]

144. Yao, Y.; Zeng, X.; Wang, H.; Liu, J. Research on Underwater Acoustic Target Recognition Method Based on DenseNet. In Proceedings of the 2022 3rd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), Xi'an, China, 15–17 July 2022; pp. 114–118. [CrossRef]
145. Hong, F.; Liu, C.; Guo, L.; Chen, F.; Feng, H. Underwater Acoustic Target Recognition with ResNet18 on ShipsEar Dataset. In Proceedings of the 2021 IEEE 4th International Conference on Electronics Technology (ICET), Chengdu, China, 7–10 May 2021; pp. 1240–1244. [CrossRef]
146. Sun, Q.; Wang, K. Underwater single-channel acoustic signal multitarget recognition using convolutional neural networks. *J. Acoust. Soc. Am.* **2022**, *151*, 2245–2254. [CrossRef]
147. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1800–1807. [CrossRef]
148. Mehta, S.; Rastegari, M.; Shapiro, L.; Hajishirzi, H. ESPNetv2: A Light-Weight, Power Efficient, and General Purpose Convolutional Neural Network. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 9182–9192. [CrossRef]
149. Hu, G.; Wang, K.; Liu, L. Underwater Acoustic Target Recognition Based on Depthwise Separable Convolution Neural Networks. *Sensors* **2021**, *21*, 1429. [CrossRef] [PubMed]
150. Miao, Y.; Zakharov, Y.V.; Sun, H.; Li, J.; Wang, J. Underwater Acoustic Signal Classification Based on Sparse Time–Frequency Representation and Deep Learning. *IEEE J. Ocean. Eng.* **2021**, *46*, 952–962. [CrossRef]
151. Zheng, Y.; Gong, Q.; Zhang, S. Time-Frequency Feature-Based Underwater Target Detection with Deep Neural Network in Shallow Sea. *J. Phys. Conf. Ser.* **2021**, *1756*, 012006. [CrossRef]
152. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9. [CrossRef]
153. Cai, W.; Zhu, J.; Zhang, M.; Yang, Y. A Parallel Classification Model for Marine Mammal Sounds Based on Multi-Dimensional Feature Extraction and Data Augmentation. *Sensors* **2022**, *22*, 7443. [CrossRef]
154. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:abs/1704.04861.
155. Yan, C.; Yu, Y.; Yan, S.; Yao, T.; Yang, C.; Liu, L.; Pan, G. Underwater target recognition using a lightweight asymmetric convolutional neural network. In Proceedings of the 17th International Conference on Underwater Networks & Systems, WUWNet '23, Shenzhen, China, 24–26 November 2023; Association for Computing Machinery: New York, NY, USA, 2024. [CrossRef]
156. Tian, G.; Haiyang, Y.; Haiyan, W.; Fan, W.; Xiao, C. CA_MobileNetV2 for Underwater Acoustic Target Recognition. In Proceedings of the 2023 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), Zhengzhou, China, 14–17 November 2023; pp. 1–5. [CrossRef]
157. Jiang, Z.; Zhao, C.; Wang, H. Classification of Underwater Target Based on S-ResNet and Modified DCGAN Models. *Sensors* **2022**, *22*, 2293. [CrossRef]
158. Xiao, X.; Wang, W.; Ren, Q.; Gerstoft, P.; Ma, L. Underwater acoustic target recognition using attention-based deep neural network. *JASA Express Lett.* **2021**, *1*, 106001. [CrossRef]
159. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141. [CrossRef]
160. Zhu, X.; Cheng, D.; Zhang, Z.; Lin, S.; Dai, J. An Empirical Study of Spatial Attention Mechanisms in Deep Networks. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6687–6696. [CrossRef]
161. Wang, B.; Zhang, W.; Zhu, Y.; Wu, C.; Zhang, S. An Underwater Acoustic Target Recognition Method Based on AMNet. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 1–5. [CrossRef]
162. Zhu, M.; Zhang, X.; Jiang, Y.; Wang, K.; Su, B.; Wang, T. Hybrid Underwater Acoustic Signal Multi-Target Recognition Based on DenseNet-LSTM with Attention Mechanism. In Proceedings of the 2023 Chinese Intelligent Automation Conference, Nanjing, China, 2–5 October 2023; Deng, Z., Ed.; Springer Nature: Singapore, 2023; pp. 728–738.
163. Seo, M.J.; Kembhavi, A.; Farhadi, A.; Hajishirzi, H. Bidirectional Attention Flow for Machine Comprehension. In Proceedings of the 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, 24–26 April 2017.
164. Liu, C.; Hong, F.; Feng, H.; Hu, M. 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, CA, USA, 20–23 September 2021; pp. 1–5. [CrossRef]
165. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is All You Need. *arXiv* **2017**, arXiv:1706.03762v7.
166. Gong, Y.; Chung, Y.A.; Glass, J. AST: Audio Spectrogram Transformer. In Proceedings of the Proc. Interspeech 2021, Brno, Czechia, 30 August–3 September 2021; pp. 571–575. [CrossRef]
167. Li, P.; Wu, J.; Wang, Y.; Lan, Q.; Xiao, W. STM: Spectrogram Transformer Model for Underwater Acoustic Target Recognition. *J. Mar. Sci. Eng.* **2022**, *10*, 1428. [CrossRef]

168. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 9992–10002. [CrossRef]
169. Wu, F.; Yao, H.; Wang, H. Recognizing the State of Motion by Ship-Radiated Noise Using Time-Frequency Swin-Transformer. *IEEE J. Ocean. Eng.* **2024**, *49*, 1–12. [CrossRef]
170. Yao, H.; Gao, T.; Wang, Y.; Wang, H.; Chen, X. Mobile_ViT: Underwater Acoustic Target Recognition Method Based on Local–Global Feature Fusion. *J. Mar. Sci. Eng.* **2024**, *12*, 589. [CrossRef]
171. Feng, S.; Zhu, X.; Ma, S. Masking Hierarchical Tokens for Underwater Acoustic Target Recognition with Self-Supervised Learning. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2024**, *32*, 1365–1379. [CrossRef]
172. Park, D.S.; Chan, W.; Zhang, Y.; Chiu, C.; Zoph, B.; Cubuk, E.D.; Le, Q.V. SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition. In Proceedings of the Interspeech 2019, 20th Annual Conference of the International Speech Communication Association, Graz, Austria, 15–19 September 2019; Kubin, G., Kacic, Z., Eds.; ISCA: Copenhagen, Denmark, 2019; pp. 2613–2617. [CrossRef]
173. Gou, J.; Yu, B.; Maybank, S.J.; Tao, D. Knowledge Distillation: A Survey. *Int. J. Comput. Vis.* **2021**, *129*, 1789–1819. [CrossRef]
174. Zhou, A.; Li, X.; Zhang, W.; Li, D.; Deng, K.; Ren, K.; Song, J. A Novel Cross-Attention Fusion-Based Joint Training Framework for Robust Underwater Acoustic Signal Recognition. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–16. [CrossRef]
175. Li, J.; Yang, H. Deep learning method with auditory passive attention for underwater acoustic target recognition under the condition of ship interference. *Ocean Eng.* **2024**, *302*, 117674. [CrossRef]
176. Song, Y.; Liu, F.; Shen, T. A novel noise reduction technique for underwater acoustic signals based on dual-path recurrent neural network. *IET Commun.* **2023**, *17*, 135–144. [CrossRef]
177. Yang, H.; Li, L.; Li, G. A New Denoising Method for Underwater Acoustic Signal. *IEEE Access* **2020**, *8*, 201874–201888. [CrossRef]
178. Yang, H.; Liu, M.; Zhang, S.; Zheng, R.; Dong, S. Few-shot Underwater Acoustic Target Recognition Based on Siamese Network. In Proceedings of the 2023 42nd Chinese Control Conference (CCC), Tianjin, China, 24–26 July 2023; pp. 8252–8257. [CrossRef]
179. Cui, X.; He, Z.; Xue, Y.; Tang, K.; Zhu, P.; Han, J. Cross-Domain Contrastive Learning-Based Few-Shot Underwater Acoustic Target Recognition. *J. Mar. Sci. Eng.* **2024**, *12*, 264. [CrossRef]
180. Tian, S.; Bai, D.; Zhou, J.; Fu, Y.; Chen, D. Few-shot learning for joint model in underwater acoustic target recognition. *Sci. Rep.* **2023**, *13*, 17502. [CrossRef]
181. Zheng, S.; Mai, S.; Sun, Y.; Hu, H.; Yang, Y. Subgraph-Aware Few-Shot Inductive Link Prediction Via Meta-Learning. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 6512–6517. [CrossRef]
182. Ma, R.; Li, S.; Zhang, B.; Fang, L.; Li, Z. Flexible and Generalized Real Photograph Denoising Exploiting Dual Meta Attention. *IEEE Trans. Cybern.* **2023**, *53*, 6395–6407. [CrossRef] [PubMed]
183. Liang, Y.; Li, S.; Yan, C.; Li, M.; Jiang, C. Explaining the black-box model: A survey of local interpretation methods for deep neural networks. *Neurocomputing* **2021**, *419*, 168–182. [CrossRef]
184. Gao, Y.; Mosalam, K.M. Deep learning visual interpretation of structural damage images. *J. Build. Eng.* **2022**, *60*, 105144. [CrossRef]
185. Agrawal, P.; Ganapathy, S. Interpretable Representation Learning for Speech and Audio Signals Based on Relevance Weighting. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2020**, *28*, 2823–2836. [CrossRef]
186. Montavon, G.; Samek, W.; Müller, K.R. Methods for interpreting and understanding deep neural networks. *Digit. Signal Process.* **2018**, *73*, 1–15. [CrossRef]
187. Xie, Y.; Ren, J.; Xu, J. Guiding the underwater acoustic target recognition with interpretable contrastive learning. In Proceedings of the OCEANS 2023, Limerick, Ireland, 5–8 June 2023; pp. 1–6. [CrossRef]
188. Chen, Y.; Du, S.; Quan, H. Feature Analysis and Optimization of Underwater Target Radiated Noise Based on t-SNE. In Proceedings of the 2018 10th International Conference on Wireless Communications and Signal Processing (WCSP), Hangzhou, China, 18–20 October 2018; pp. 1–5. [CrossRef]
189. Xu, Y.; Kong, X.; Cai, Z. Cross-validation strategy for performance evaluation of machine learning algorithms in underwater acoustic target recognition. *Ocean Eng.* **2024**, *299*, 117236. [CrossRef]
190. Han, K.; Wang, Y.; Guo, J.; Tang, Y.; Wu, E. Vision GNN: An Image is Worth Graph of Nodes. In Proceedings of the 36th International Conference on Neural Information Processing Systems, New Orleans, LA, USA, 28 November–9 December 2022; pp. 8291–8303.
191. Xu, J.; Xie, Y.; Wang, W. Underwater acoustic target recognition based on smoothness-inducing regularization and spectrogram-based data augmentation. *Ocean Eng.* **2023**, *281*, 114926. [CrossRef]
192. Li, D.; Liu, F.; Shen, T.; Chen, L.; Yang, X.; Zhao, D. Generalizable Underwater Acoustic Target Recognition Using Feature Extraction Module of Neural Network. *Appl. Sci.* **2022**, *12*, 10804. [CrossRef]
193. Brown, T.B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language Models are Few-Shot Learners. In Proceedings of the Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, Virtual Conference, 6–12 December 2020; Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H., Eds.

194. Radford, A.; Kim, J.W.; Xu, T.; Brockman, G.; McLeavey, C.; Sutskever, I. Robust speech recognition via large-scale weak supervision. In Proceedings of the 40th International Conference on Machine Learning. JMLR.org, Honolulu, HI, USA, 23–29 July 2023. ICML'23.
195. Bi, K.; Xie, L.; Zhang, H.; Chen, X.; Gu, X.; Tian, Q. Accurate medium-range global weather forecasting with 3D neural networks. *Nature* **2023**, *619*, 533–538. [CrossRef]
196. Wang, X.; Wang, R.; Hu, N.; Wang, P.; Huo, P.; Wang, G.; Wang, H.; Wang, S.; Zhu, J.; Xu, J.; et al. XiHe: A Data-Driven Model for Global Ocean Eddy-Resolving Forecasting. *arXiv* **2024**, arXiv:2402.02995.
197. Sun, W.; Yan, R.; Jin, R.; Zhao, R.; Chen, Z. FedAlign: Federated Model Alignment via Data-Free Knowledge Distillation for Machine Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 1–12. [CrossRef]
198. Hu, E.J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Chen, W. LoRA: Low-Rank Adaptation of Large Language Models. *arXiv* **2021**, arXiv:2106.09685.
199. Dettmers, T.; Pagnoni, A.; Holtzman, A.; Zettlemoyer, L. QLoRA: Efficient Finetuning of Quantized LLMs. In Proceedings of the Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, 10–16 December 2023; Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., Levine, S., Eds.
200. Lu, J.; Jin, F.; Zhang, J. Adapter Tuning with Task-Aware Attention Mechanism. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2023, Rhodes Island, Greece, 4–10 June 2023; IEEE: New York, NY, USA, 2023; pp. 1–5. [CrossRef]
201. Goodfellow, I.J.; Shlens, J.; Szegedy, C. Explaining and Harnessing Adversarial Examples. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, 7–9 May 2015
202. Madry, A.; Makelov, A.; Schmidt, L.; Tsipras, D.; Vladu, A. Towards Deep Learning Models Resistant to Adversarial Attacks. In Proceedings of the 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, 30 April–3 May 2018.
203. Liu, X.; Xie, L.; Wang, Y.; Zou, J.; Xiong, J.; Ying, Z.; Vasilakos, A.V. Privacy and security issues in deep learning: A survey. *IEEE Access* **2020**, *9*, 4566–4593. [CrossRef]
204. Modas, A.; Sanchez-Matilla, R.; Frossard, P.; Cavallaro, A. Toward Robust Sensing for Autonomous Vehicles: An Adversarial Perspective. *IEEE Signal Process. Mag.* **2020**, *37*, 14–23. [CrossRef]
205. Svenmarck, P.; Luotsinen, L.; Nilsson, M.; Schubert, J. Possibilities and challenges for artificial intelligence in military applications. In Proceedings of the NATO Big Data and Artificial Intelligence for Military Decision Making Specialists' Meeting, Bordeaux, France, 30 May–1 June 2018; pp. 1–16.
206. Feng, S.; Zhu, X.; Ma, S.; Lan, Q. Adversarial Attacks in Underwater Acoustic Target Recognition with Deep Learning Models. *Remote Sens.* **2023**, *15*, 5386. [CrossRef]
207. Bai, T.; Luo, J.; Zhao, J.; Wen, B.; Wang, Q. Recent Advances in Adversarial Training for Adversarial Robustness. In Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21; International Joint Conferences on Artificial Intelligence Organization, Montreal, QC, Canada, 19–27 August 2021; pp. 4312–4321. [CrossRef]
208. Aldahdooh, A.; Hamidouche, W.; Fezza, S.A.; Déforges, O. Adversarial example detection for DNN models: A review and experimental comparison. *Artif. Intell. Rev.* **2022**, *55*, 4403–4462. [CrossRef]
209. Çatak, F.Ö.; Kuzlu, M.; Çatak, E.; Cali, U.; Guler, O. Defensive Distillation-Based Adversarial Attack Mitigation Method for Channel Estimation Using Deep Learning Models in Next-Generation Wireless Networks. *IEEE Access* **2022**, *10*, 98191–98203. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.