Journal of
Marine Science and Engineering

MDPI

Review

A Survey of Underwater Acoustic Target Recognition Methods Based on Machine Learning

Xinwei Luo *, Lu Chen, Hanlu Zhou and Hongli Cao

Key Laboratory of Underwater Acoustic Signal Processing of Ministry of Education, Southeast University, Nanjing 210096, China
* Correspondence: luoxinweil@seu.edu.cn

Abstract: Underwater acoustic target recognition (UATR) technology has been implemented widely in the fields of marine biodiversity detection, marine search and rescue, and seabed mapping, providing an essential basis for human marine economic and military activities. With the rapid development of machine-learning-based technology in the acoustics field, these methods receive wide attention and display a potential impact on UATR problems. This paper reviews current UATR methods based on machine learning. We focus mostly, but not solely, on the recognition of target-radiated noise from passive sonar. First, we provide an overview of the underwater acoustic acquisition and recognition process and briefly introduce the classical acoustic signal feature extraction methods. In this paper, recognition methods for UATR are classified based on the machine learning algorithms used as UATR technologies using statistical learning methods, UATR methods based on deep learning models, and transfer learning and data augmentation technologies for UATR. Finally, the challenges of UATR based on the machine learning method are summarized and directions for UATR development in the future are put forward.

Keywords: machine learning; UATR; underwater acoustic dataset; classification and recognition

Citation: 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. https://doi.org/10.3390/jmse11020384

Academic Editor: Marco Cococcioni

Received: 16 December 2022
Revised: 3 February 2023
Accepted: 7 February 2023
Published: 9 February 2023

1. Introduction

Acoustic waves are the only energy form known to humans that can travel long distances in water and they are generally considered to be the best information carrier for sensing and recognizing underwater targets [1]. Exploring accurate UATR methods can better promote the development of related fields, such as seabed mapping [2], marine biodiversity detection [3,4], vessel target recognition [5], etc. The recognition and detection process based on underwater acoustic target signals is shown in Figure 1, mainly including target signal acquisition by a sonar system, array processing, data preprocessing, feature extraction, and target recognition [6]. This paper mainly reviews the target recognition methods based on sonar signals. Due to the complexity of the marine environment, the acoustic signals obtained by the spatiotemporal sampling of the sensor array in the ocean are not only the target signals but also the environmental noises and other target interference signals. Usually, the beam signals in the target direction are obtained by the spatial filtering characteristics of array processing, and then are combined with other preprocessing works to reduce the impact of noises. In the early stage, underwater acoustic targets are recognized by human ears. This method does not involve the feature extraction process shown in Figure 1. Generally, acquired original signals or the signals after preprocessing are recognized directly by human ears. The disadvantages of this recognition method are also apparent. When the amount of data increases, the time and energy required will increase proportionally, and it can hardly meet the requirements of real-time recognition. The most important thing is that the frequency distribution range of the underwater acoustic signals is broad. Listening to these sounds, which human ears cannot adapt to, for a long time is inevitably harmful to human health. In addition, researchers also used various

frequency domain and time-frequency analysis methods to identify targets based on the spectral features of underwater acoustic signals. Among them, power spectrum analysis, low-frequency analysis and recording (LOFAR), and detection of envelope modulation on noise (DEMON) are commonly used to extract underwater acoustic signal features [7,8]. Mel frequency cepstral coefficient (MFCC) and gammatone frequency cepstral coefficient (GFCC) are also commonly used features for underwater acoustic signal processing [9]. With the rise of machine learning methods, researchers in the underwater acoustic field have turned their attention to building automatic underwater acoustic target recognition methods using machine learning models [10]. Some studies combine the underwater acoustic signal feature extraction method with the machine learning model and use the signal frequency domain features as the input of the machine learning model for target recognition [11]. This paper introduces the classical underwater acoustic signal feature extraction methods in Section 2. Deep learning models have a strong learning ability and can extract the potential knowledge representation of signals. Therefore, some studies directly input the raw signals into the deep learning model for target recognition [12]. This paper reviews several studies on UATR using machine learning methods and divides them into three categories according to the different models used. The first category is the recognition methods based on the statistical learning model. These methods construct statistical probability models using labels and features of samples and classify test samples. The second category is recognition methods using deep learning models. These methods use the powerful learning ability of deep neural networks to extract features from input samples and perform recognition. The third one is transfer learning methods and data augmentation technologies, which are used to solve the problem of the lack of labeled underwater acoustic signal samples. This is a severe problem faced by UART.

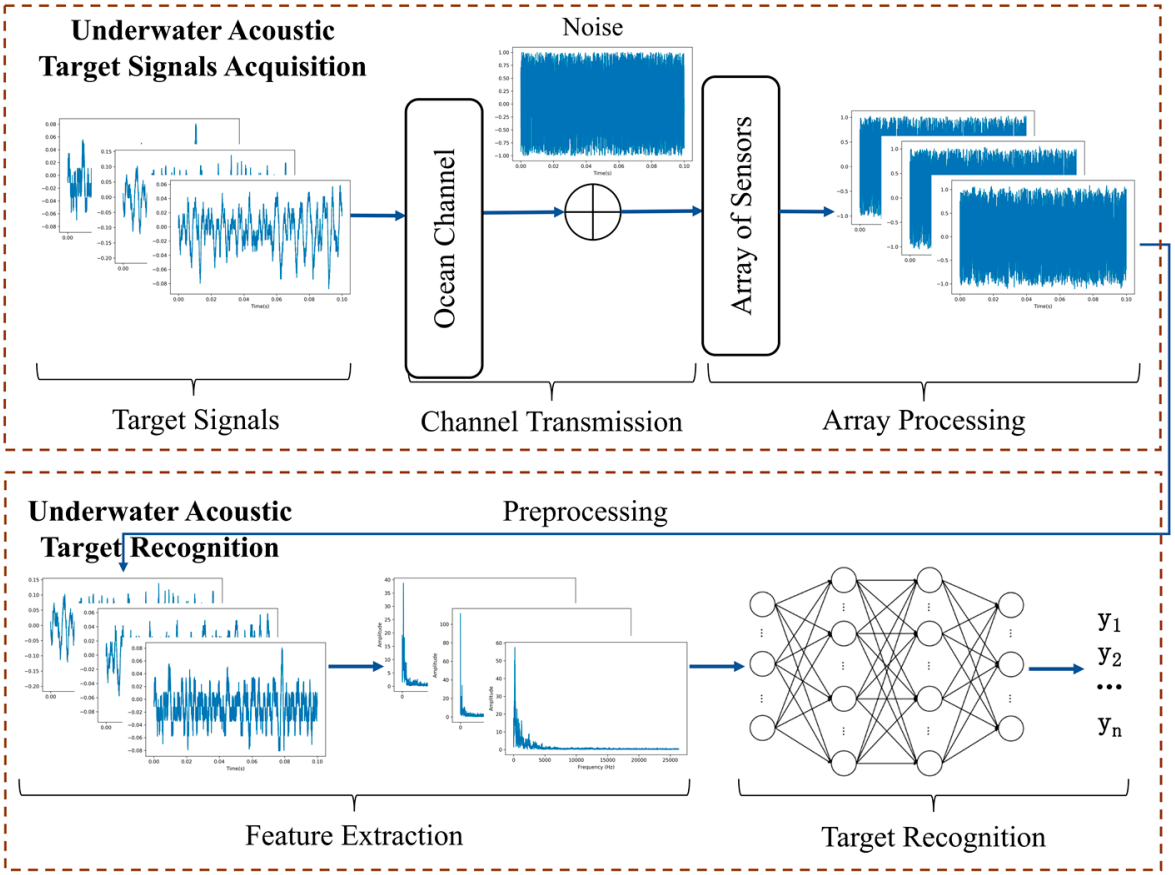

Figure 1. Flow chart of underwater acoustic signal acquisition and target recognition.

At present, machine learning methods have shown better performance than traditional methods in the field of underwater acoustics, including underwater acoustic target recognition [5], underwater acoustic source localization [13], single-channel source separation [14], and so on. However, they are data-driven methods, and underwater acoustic data acquisition faces enormous challenges. At the same time, traditional signal processing methods are more interpretable than many machine learning algorithms. Additionally, interpretability is necessary for some systems and application scenarios. In addition, machine learning models with good recognition performance on specific scenes and data may not apply to other datasets. Because the model only learns the potential feature representation of training data, it cannot be a universal method. However, even though machine learning methods still have many limitations, we can still see the development potential of such models in underwater acoustics fields [15–17].

This paper reviews the recent studies of UATR based on machine learning and analyzes the technical characteristics, performance, and challenges of these studies, which will provide a reference for researchers in the field of UATR technology. The next section describes the widely used feature extraction methods for UATR. Section 3 discusses the UATR technology based on machine learning models and related research works in detail. Section 4 analyzes the challenges encountered by the UATR methods based on machine learning. Section 5 gives the conclusion and discussion of this paper.

2. Data Preprocessing and Feature Extraction Methods

Acoustic signals are time sequence signals, but their frequency domain usually contains more information. Therefore, it is necessary to preprocess and extract features of the raw data to reduce data dimensions and suppress noise before inputting into recognition models.

Ship-radiated noise is one of the main research objects in the UATR field. It has the characteristics of being short-term stationary, and the frequency spectrum obtained by Fourier transform is relatively stable, which is more conducive to the extraction of target recognition features. LOFAR spectrums are generally used to characterize ship-radiated noise [8]. A LOFAR spectrum is a two-dimensional image of frequency and time obtained by a short time Fourier transformer (STFT). Narrowband components in ship-radiated noise can be found by extracting line spectral trajectories from the LOFAR spectrum [18]. Researchers can intuitively obtain the time-frequency distribution information of target signal energy through the LOFAR spectrum. This time–frequency spectrum analysis method is conducive to the detection and recognition of dynamic targets [19].

The periodic rotation of the propeller in the nonuniform flow field gives the ship-radiated noise a unique rhythm. Researchers often use the DEMON spectral analysis method to estimate the shaft frequency and blade number of the target propeller [20]. DEMON analysis is a wideband de-modulation technique, which can separate the modulation envelope caused by propeller cavitation from the ship-radiated noise signal, and estimate shaft number, shaft frequency, and blade number through spectrum analysis and line spectrum detection. These parameters related to target propellers are useful features for underwater target detection and recognition [21]. However, the DEMON analysis method has poor performance under low SNR. Some researchers simultaneously use LOFAR and DEOMON methods to extract ship-radiated noise signal features [22,23].

Studies showed that the human ear's perception of sound frequencies is not linear. To make the extracted sound features more consistent with the sound perception mechanism of human ears, researchers have proposed the extraction method of MFCC features. It uses the Mel filter bank to filter signals and then takes logarithm and inverse Fourier transform to obtain MFCC. This feature extraction method has become the basis of speech processing [24–26]. In 2007, Lim T et al. applied MFCC to UATR, and their research showed that this method has a good potential for application in UATR [27]. Tong et al.

proposed an effective UATR method. They first extracted three types of underwater targets MFCC features. Then, they classified and identified them using the K-Nearest Neighbor algorithm [28]. Similarly, GFCC is also an acoustic signal feature based on auditory perception, which is implemented based on Gammatone filter bank. Some studies have shown that GFCC features can better describe the targets than MFCC features under low SNR or interference conditions. Additionally, their classification results have better recognition accuracy and robustness [29–31]. In the field of UATR, many studies have shown the effectiveness and practicability of these two feature extraction methods [32–34].

In fact, the ship-radiated noise also contains components that change rapidly with time. Empirical mode decomposition (EMD) [35] can decompose non-stationary complex signals into various signal components called intrinsic mode function (IMF). Huang proposed the Hilbert–Huang transform (HHT) method [35] to divided the signals into the sum of several IMFs. In contrast to the frequency definition of traditional time-frequency analysis methods, HHT uses phase derivatives to obtain frequencies and accurately describes the instantaneous frequency components of signals [36]. HHT can characterize local instantaneous characteristics, so it has good adaptability to non-stationary signals. In 2014, Zeng and Wang et al. applied HHT to underwater acoustic target recognition and achieved better recognition results than MFCC [37]. In 2010, Bao et al. proposed a ship classification approach based on EMD, which approved the effectiveness of recognition by analyzing nonlinear features of radiated sound [38]. Other feature extraction methods in UATR include formant analysis, wavelet transform, linear predictive cepstral coefficient (LPCC), etc. [39–41].

The preprocessing of raw data and feature extraction are very important to improve the accuracy of UATR. Therefore, it is necessary to make sufficient preparation for the preprocessing and feature extraction of the raw data, extract the features of the target, and reduce the redundant information, so that the recognition model can have a good performance.

3. UATR Methods Based on Machine Learning

This section reviews machine-learning-based UATR techniques, which use a machine learning model [42,43] to conduct the mapping between underwater acoustic signals and their labels [44,45]. UATR methods based on machine learning are divided into three main categories: (1) methods based on statistical learning models, such as support vector machine (SVM), gaussian mixed model (GMM), hidden Markov model (HMM), etc.; (2) methods based on deep learning algorithms, such as convolution neural network (CNN), recurrent neural network (RNN), attention mechanism, etc.; (3) methods based on transfer learning and data augmentation strategies, which are proposed to solve the problem of insufficient data caused by the difficulty of data acquisition, storing, and labeling in the UATR field. In this section, the relevant studies on UATR based on machine learning methods are summarized in the form of the table listed, and the feature extraction methods, datasets used, recognition effects, and main contributions of these recognition methods are briefly explained.

3.1. UATR Methods Based on Statistical Learning

Statistical learning methods are based on traditional statistical methods to establish probability and statistical models for analysis and prediction. They are relatively simple and have easy-to-understand parameters. Statistical learning models can achieve good results on small datasets and are less prone to overfitting. Therefore, in the UATR domain, many studies are based on statistical learning models [46,47]. Table 1 summarizes studies that apply statistical learning models for UATR, and briefly describes the model used, feature extraction methods, datasets, recognition performance, and the main contributions.

Table 1. An overview of UATR methods using statistical learning.

Method	Feature	Dataset	Performance	Main Contributions
Single-class SVM [48] Moura et al., 2015	LOFAR Spectrum	* Ship-Radiated Noise	77.9% SP	Applying single-class SVM to passive sonar system detection to solve the classification problem of sparse negative samples.
SVM+BAT [49] Sherin et al., 2015	MFCC	* Sound Signals of Ships and Marine Animals	75% Acc	The BAT algorithm is used to optimize kernel parameters and achieves higher classification accuracy.
WSFSelect-SVME [50] Yang et al., 2016	Wavelet analysis features; waveform structure features; MFCC; auditory spectrum features.	UCI sonar dataset [1]; * Real-world underwater acoustic target dataset	81% CR1 99% CR2	Proposing a novel AdaBoost SVM model based on weighted sample and feature selection method to improve the accuracy of UATR and reduce extra computational and storage costs.
GMM+MUSIC [51] Peso et al., 2014	Cepstral Coefficients, Features	Sound Signals of Cetacean Species [2]	90.3% DR 18.1% ER	The unpredictability measure and MUSIC algorithm [52] are used to extract features for improving the recognition performance of GMM.
HMM [53] Kim et al., 2011	Features extracted by matching pursuit algorithm [54]	* Synthesized Active sonar Signals	91% Acc	Using synthesized sonar signals as input to avoid the problem of data acquisition, and applying a multi-aspect target classification scheme based on a hidden Markov model for classification.
HMM [55] Mohammed et al., 2018	GTCC	* Ship and Marine Species Acoustic Signals	93% SR	Investigating the performance of the GTCC-based HMM classifier with self-noise conditions and under Rayleigh fading environment.

* Dataset is proprietary. Acc: Accuracy. FR: False classification rate. SP: An index that comprehensively considers the recognition efficiency of known classes and test samples. DR: Detection Rate. ER: Error Rate. SR: Success Rate. CR1: Correct classification rate on sonar system. CR2: Correct classification on real-word dataset.

One of the statistical learning methods widely used in UATR is SVM [48,49]. Its core idea is to find the decision surface between different classes of data, make the two classes of samples fall on both sides of the decision surface, and make the samples far enough away from the decision surface [50]. The original SVM is based on the plane decision, which requires the samples to be linearly separable, but this condition usually cannot be satisfied

in practical cases. The solution of SVM is to map the samples to a new space, usually a higher dimensional space, using a kernel function, and then find a linear decision surface in the new space for classification [49]. Statistical learning theory shows that SVM has two advantages. First, it is a convex optimization problem, so the solution obtained must be a global optimum rather than a local optimum. Second, this algorithm is suitable for both linear and nonlinear problems. The computational complexity of SVM only depends on the dimension of support vectors rather than the size of datasets, which avoids the curse of dimensionality in a sense. Hence, it is suitable for datasets with a high-dimensional sample space.

Ref. [48] uses single-class SVM [56] for target detection in passive sonar systems. The single-class SVM is used to solve the problem that there is only one kind of training data. Additionally, the target data are expected to have the same characteristics as the training data. The reason single-class SVM is proposed is that in some specific scenarios, it is hard to obtain negative samples or to define the range of negative samples accurately. At this time, the model needs to recognize the type of unknown samples according to the characteristics of one class of known samples. De Moura and de Seixas [48] use the ship dataset acquired from a real marine environment to train the single-class SVM model, and the SP index [57] is 73.18%. The classification performance of SVM largely depends on the kernel function. Ref. [49] uses the BAT algorithm [58] to optimize the kernel parameters of SVM. Compared with other parameter optimization algorithms, such as genetic algorithms (GA) and particle swarm optimization (PSO), the BAT algorithm has the advantage that it can conduct global and local searches simultaneously to avoid falling into local optimum. The results reported show that the accuracy of the classifier using the BAT optimization algorithm is six percentage points higher than that using PSO algorithm [49]. An ensemble of SVM can improve the recognition accuracy of a UATR system. Yang et al. [50] proposes a novel SVM ensemble algorithm combined with sample selection and feature selection methods (WSFSelect-SVME). The proposed model solved the two limitations of traditional ensemble SVM methods. (1) The training data with poor quality will result in errors between actual and theoretical results. (2) Ensemble recognition systems usually have higher complexity and computational costs. The experimental results on the UCI sonar dataset and real-world underwater acoustic target dataset show that the WSFSelect-SVME model obtains better recognition performance and robustness than Adaboost SVM ensemble algorithm.

A set of studies applied the SVM model to UATR and achieved record-breaking results. The process of solving the support vectors involves the calculation of the N-order matrix (N is the number of samples). It requires a lot of memory and computing time when N is large. At the same time, the conventional SVM algorithm only supports binary classification. When dealing with multi-classification problems, the problem needs to be transformed into multiple binary classification problems, which reduces the classification efficiency.

The GMM is an extension of a single Gaussian probability density function, which is composed of multiple Gaussian distributions. GMM can approximate the density distribution of arbitrary shapes and can be used for multi-class target recognition. According to the different parameters of Gaussian probability density function (PDF), each Gaussian model can be regarded as a class. The GMM model first calculates the probability value of input samples. Additionally, whether the sample belongs to a Gaussian distribution can be judged according to the set threshold [59]. Research shows that GMM is suitable for modeling complex samples [60]. Parada and Cardénal-Lopez [51] proposed a method based on the GMM model to identify the two main sounds emitted by dolphins, whistle and pulse, as well as background noise. By introducing the uncertainty measure and MUSIC algorithm [52] for feature extraction, the detection rate of GMM is increased from 87.5% to 90.3%, and the classification error rate is reduced from 23.6% to 18.1%.

The GMM model can only approximate the Gaussian distribution of the calculated data and cannot extract the deep abstract features of the acoustic signal. Although GMM fits existing samples well, the fitting to unknown samples is unstable. Most importantly,

the recognition results of GMM in multidimensional features are not ideal [61]. Therefore, GMM is generally combined with other models to build a reliable UATR system.

Compared to GMMs, the advantage of HMMs is that they usually have a better prediction performance, instead of only focusing on fitting observed values [62]. During target recognition, HMM obtains the state transition probability matrix and observation probability matrix through training and makes decisions according to the maximum probability in the process of state transition [63]. In Kim et al. [53], a multi-direction target classification method based on HMM is proposed and applied to the classification of synthesized active sonar signals. Mohammed et al. [55] researched the efficiency and reliability of underwater acoustic target methods and proposed an HMM model based on Gammatone cepstral coefficient (GTCC). The experiment results on a dataset including ten types of ship and marine species show that the GTCC-based HMM model achieves an average accuracy of 89% under different SNR, which is 5 and 8 percentage points higher than ANN and statistical Euclidean distance classification, respectively.

Statistical learning methods build and train models based on traditional statistical analysis, which can only roughly fit the distribution of samples and have limited ability to extract features. Statistical learning methods struggle to handle the recognition tasks with large samples due to their limited model capacity. Moreover, both GMMs and HMMs have default assumptions, but the underwater acoustic data, in reality, struggle to meet these assumptions, which affects the generalization of the model. To better extract and use the features of underwater acoustic signals for UATR, deep learning models with strong feature extraction ability have been applied in this field.

3.2. UATR Methods Based on Deep Learning

In recent years, with the improvement in the computing ability of computers, the research of deep learning (DL) based on neural networks [64] has developed rapidly. A deep learning model can be composed of network modules with multiple processing layers. These layers extract features with different levels of abstraction and automatically adjust parameters through back propagation until suitable data features are extracted for downstream tasks. Deep learning models are widely used in speech recognition, image processing, intelligent control, expert systems, and other fields with their powerful feature extraction ability [65]. Researchers in the UATR field have also turned their attention to deep learning algorithms [66]. Deep learning-based UATR models are generally supervised, which train deep neural networks on datasets with labels, and then the network can predict the type of unknown samples. Table 2 lists the relevant studies on UATR using deep learning models.

Table 2. An overview of deep learning methods for UATR.

Method	Feature	Dataset	Performance	Main Contributions
Dense CNN [34] Doan et al., 2020	Original audio signal	* Real-world dataset	98.85% Acc	Using a dense CNN model for UATR, which reuses former feature maps. The proposed model achieves high recognition accuracy under low computational cost.
CNN, LSTM [67] Song et al., 2021	One-Dimensional Time Domain Signals, LOFAR Spectrum	* Underwater Targets and Ship Targets	90.1% Acc	Compared the recognition performance of CNN and LSTM models when the inputs are time domain signals and LOFAR spectrums, respectively.

Table 2. *Cont.*

Method	Feature	Dataset	Performance	Main Contributions
Depthwise Separable Convolution and Time-Dilated Convolution [68] Hu et al., 2021	One-dimensional Time domain Signals	* Ship-radiated Noise	90.9% Acc	Automatically extract the features of the one-dimensional time domain raw signals, and visualize the clustering performance of the proposed method.
Bi-GRU+GRU [69] Wang et al., 2020	Time domain Signals	* Shallow Sea Data	# 91% Acc	The proposed model can effectively tackle the changing input signal length.
CNN+Bi-LSTM+ Attention [70] Kamal et al., 2021	Features Extracted by Learnable Filterbank	* Indian Ocean Shallow Data	95.2% Acc	CNN and bidirectional LSTM model are used to extract features from multiple scales.
camResNet [71] Xue et al., 2022	Time domain Signals; Frequency domain signals	* Real-word dataset	98.2% Acc	Introduce channel attention mechanism to enhance the energy of signal features extracted by ResNet.
UATR-Transformer [72] Feng et al., 2022	Mel-spectrogram	Shipsear [73] DeepShip [74]	96.9% S_Acc 95.3% D_Acc	Taking the transformer architecture as the backbone for UATR for the first time.

* Dataset is proprietary. # The value is not given in the paper. It is approximate value according to the resulting diagram provided. Acc: Accuracy. S_ACC: Accuracy on Shipsear dataset. D_Acc: Accuracy on DeepShip dataset.

CNN is one of the mainstream deep learning architectures, which has been widely used in natural language processing, speech recognition, medical diagnosis, and other fields [75,76]. A basic convolutional neural network consists of the convolutional layer, activation function, and pooling layer. The convolutional layer is the core part of the network, and the convolutional kernel can be regarded as a feature recognizer. The training process of CNNs for UATR is to adjust the weights of the convolution kernel and make it suitable for target recognition. After the convolutional layer, the activation function enhances the generalization ability of the network by non-linear mapping between the input and output. Pooling can be regarded as a down-sampling operation, the main purpose of which is to reduce the resolution of the feature map. Common pooling methods include maximum pooling and average pooling [77]. The pooling operation is helpful to prevent overfitting of neural networks. When processing the target recognition tasks, CNNs send the output sample feature vectors to a fully connected layer to map the samples and labels [78]. In studies of UATR, some researchers use the acquired sonar image data as the CNN input for target recognition [79–81]. Others directly input time domain signals into CNN models to identify ship types [34,68]. In general, researchers first transform time domain signals into various spectrums and then use the CNN model to extract abstract features of spectrums and recognize underwater targets [8,82].

Doan et al. [34] use dense CNN to extract time domain signal features for UATR. The proposed target recognition network with the skip-connection technique could reuse former feature maps, which prevents the gradient vanishing problem. Experimental accuracy on a real-world dataset with 0 dB achieves 98.85%. Xiaoping et al. [67] compared the recognition ability of CNN and LSTM models for complex underwater acoustic signals. Experimental results show that when the classifier takes the time domain signal as input, the accuracy of CNN on the dataset containing eight types of underwater targets and six types of ships is five percentage points higher than that of LSTM. Hu et al. [68] used depth-separable

convolution and dilated convolution for passive UATR for the first time. The dilated convolution enlarges the receptive field of the model without increasing the parameters so that the features extracted by the model have better intra-class aggregation and inter-class separation characteristics. The proposed model achieves better recognition performance than traditional CNN model.

RNNs are a kind of neural network that is good at processing sequence data. The input of RNNs at each time step contains the output of the previous time step, and this structure makes the model capable of memory [83]. Underwater acoustic signals are complex time-varying signals with some correlation between each frame. Additionally, the memory ability of RNNs makes them suitable for learning the features of underwater acoustic signals. In recent years, RNNs have also become one of the major solutions for UATR. Wang et al. [69] proposed a hybrid time-series network, i.e., the combination of bi-directional gated recurrent unit (Bi-GRU) and multi-layer gated recurrent unit (GRU), for acoustic signal modulation identification in harsh underwater communication environments. The network optimizes the internal network structure by cascade order to obtain more hidden signal features. The experimental results show that the combined network of 4-layer Bi-GRU and 4-layer GRU have good recognition accuracy and robustness in an environment with serious interference. CNNs are effective local feature extractors. Additionally, the combination of CNN and LSTM can extract features from samples better. Kamal et al. [70] proposed a combination model of CNN and LSTM for target recognition based on shallow sea acoustic data. First, the standardized data are convolved with the filter to generate a learnable time-frequency representation. Then, the abstract features of the time-frequency representation are further extracted using a three-layer two-dimensional convolution. Bi-LSTM is used to capture the temporal features of the sequence from the front and back directions. Finally, the selective attention layer is used to select the most useful features for recognition. Experimental results on acoustic datasets collected in Indian Ocean shoals show that the recognition accuracy of this end-to-end deep learning model reaches 95.2%.

Attention is a kind of information selection and resource allocation method, which devotes limited resources to processing important information [84]. Generally, in order to select the information that is more important to the downstream task in the input set of vectors, the input information is represented in the form of key-value pairs. At the same time, query vectors are introduced, and the correlation between each input vector and the query vector is calculated by a scoring function [84]. To effectively extract the low-frequency spectrum under Doppler shift, Xue et al. designed a ResNet with channel attention mechanism model [71]. The target deep abstraction spectral features are extracted by ResNet. Then, the channel attention mechanism model is used to weigh the signal channels and complete information points in each channel. The targets are recognized by one-dimensional convolution. The recognition accuracy on a real-world dataset containing four kinds of underwater acoustic targets reaches 98.2%. Transformer [85] is an attention-based architecture. It is widely used in image processing and natural language processing fields. The transformer model was introduced to the UATR field for the first time by Feng et al. and achieves good recognition performance [72]. Compared with the CNN-based model, transformer architecture can consider both global and local information.

Deep learning is proving to be a potential tool for UATR. However, its application in this field is limited to a few methods. There are a set of deep learning methods and applications with excellent performance waiting to be explored.

3.3. Transfer Learning and Data Augmentation Strategies for UATR

Even though deep learning methods have achieved good performance in the UATR field, it is incredibly to train a reliable enough deep learning model if there is not exist a large amount of labeled data [86]. Many studies have shown that transfer learning (TL) and data augmentation methods are effective ways to solve the problem of model training in the case of insufficient data [87]. TL methods first train a network model on a large and

related dataset called the source domain, and then use a small target domain dataset to fine-tune the parameters to make the network model adapt to the new task requirements. These methods not only release the training pressure on an insufficient dataset but also reduce the training time on the target source and obtain robust models. Furthermore, the data augmentation methods such as generative adversarial networks (GANs) [88] also provide a solution for model training in the case of insufficient data, which expands the dataset by generating new samples. For training the UATR network, it is hard to construct a standard dataset of sufficient scale. Therefore, many researchers use transfer learning or data augmentation techniques for UATR. Table 3 lists the applications of TL and data augmentation methods in UATR.

Table 3. An overview of transfer learning and data augmentation strategies for UATR.

Method	Feature	Dataset	Performance	Main Contributions
VGG-19 [83] Huo et al., 2020	\	Seabed Objects-KLSG [3]	97.76% Acc	Combined semisynthetic data generation with deep transfer learning to improve the recognition accuracy.
GoogleNet [89] Nguyen et al., 2019	\	* CKI, TDI-2017, TDI-2018	91.6% Acc	Retrain the model to improve the recognition performance.
ResNet50 [90] Fuchs et al., 2018	\	ARACATI [4]	90% Acc	Transfer learning and pretrained CNN are used to extract image features which can replace manual features for FLS recognition system.
RSSD+CNN [82] Ke et al., 2018	Resonance-based Sparsity Signal Decomposition (RSSD)	ShipsEar [73]	93.28% Acc	The model designed is pretrained in an unsupervised manner using one-dimensional convolution and fine-tuned in a supervised manner.
VGG-19 [91] Korkmaz et al., 2022	Spectrograms	Dolphin whistles dataset [5]	92.3% Acc	The mean recognition accuracy of VGG model implementing the transfer learning method is 25.9 percentage points higher than the baseline and 11.7 percentage points higher than vanilla CNN model.
RBM+BP [92] Luo et al., 2021	Combining Power Spectrum and Demodulation Spectrum	ShipsEar [73]	92.6% Acc	Designed a data augmentation method using an RBM auto-encoder and improved the performance of the underwater acoustic target recognition system.
cDCGAN+ResNet [93] Luo et al., 2021	Multi-window Spectral Analysis	ShipsEar [73]	96.32% Acc	The proposed conditional deep convolutional GAN model (cDCGAN) has achieved good results using data augmentation method.
DCGAN+S-ResNet [94] Jiang et al., 2022	STFT	* Five Different Types of Underwater Targets	92% Acc	The improved DCGAN is used to solve the problem of data insufficient. The S-ResNet is proposed to reduce the model parameters and computational complexity.
WGAN-GP+CNN, LSTM [67] Song et al., 2021	One-Dimensional Time Domain Signals, LOFAR spectrums	* Underwater Target and Ship Targets	90.1% Acc	Using WGAN-GP to augment the time domain signals and LOFAR spectrums, respectively, then comparing the recognition performance of CNN and LSTM models.

* Dataset is proprietary. Acc: Accuracy. \ No related content.

In recent years, many deep CNN models with remarkable effects on image recognition have been proposed, and many researchers are trying to transfer the pretrained deep CNN model to the UATR tasks. ImageNet, a large image database published by Google, provides a good dataset to pretrain these CNN models [95]. The most direct application of transfer learning in the underwater target recognition field is to transfer the pre-trained model on the ImageNet dataset to the underwater sonar image dataset. In the paper by Lipton et al. [83], TL technology is applied to sonar seabed image classification. The proposed method pretrains a VGG19 model using the ImageNet dataset. Then, the parameters of VGG19 are fine-tuned using the dataset acquired in a real scenario and semi-generated data. Experimental results show that the VGG19 network transferred from the ImageNet dataset achieves 97.76% accuracy, which is better than the results of SVM and shallow CNN networks. With the support of pretrained CNN models and ImageNet dataset, many studies have chosen to transfer pretrained deep CNN models to underwater target recognition tasks. For example, a pretrained GoogleNet [89] is used for underwater human body automatic detection based on sonar images. In Fuchs et al. [90], ResNet50 pretrained on ImageNet is transferred to forward-looking sonar (FLS) image data classification, and the accuracy reaches 95%. Underwater acoustic spectrums have a similar format to images. Therefore, the TL strategies mentioned above have immense potential in UATR. Ke et al. [82] proposed a one-dimensional convolution automatic encoding–decoding model to recognize ship-radiated noise. It is combined with the feature extraction method based on resonant sparse signal decomposition. The model is trained on a large unlabeled dataset and then fine-tuned using a small, labeled dataset. The recognition accuracy of this model on the ShipsEar dataset [73] reaches 93.28%. Korkmaz et al. [91] compared the recognition performance of dolphin whistles using PamGuard [96], a software that automatically identifies marine mammals, vanilla CNN, and VGG models using the transfer learning approach. The results showed that the mean recognition accuracy of the CNN model was much higher than that of the PamGuard software, while the VGG model using the migration learning technique had an additional 11.7 percentage points higher recognition accuracy than the vanilla CNN. This study offers great potential for the deployment of marine biological detection systems using deep learning techniques.

Data augmentation methods are kinds of technology that can build synthetic data by transforming the existing labeled data using various transformations. Due to the difficulty of acquiring ship radiation signals, it is difficult to construct a sufficient number of labeled training data. Researchers have tried to apply various data augmentation techniques for UATR tasks. Luo et al. [92] used a restricted Boltzmann machine (RBM) autoencoder to augment the ship-radiated noise signal dataset for training the UATR system. In this method, RBM is used to encode the combined data of the power spectrum and demodulation spectrum of ship-radiated noise automatically without supervision. Then, reconstructed samples are obtained by decoding feature vectors layer by layer. After the above data augmentation processing, the recognition accuracy of a 4-layer Back Propagation (BP) classifier is improved from 91.4% to 92.6%. Luo et al. [93] proposed a conditional deep convolution generative adversarial network (cDCGAN) model for data augmentation. The cDCGAN uses CNN to build the generator and discriminator and introduces the label information to the training process. It increases the number of ship-radiated noise samples. A ResNet-based classifier is used to recognize ship type. The test accuracy is improved from 90.94% to 96.32% after the data augmentation processing. In the work by Jiang et al. [94], an improved DCGAN [97] architecture is used to augment the training data of ship-radiated noise targets, and then the proposed S-ResNet is used as a classifier. The recognition accuracy of the S-ResNet classifier improved by about six percentage points after using data augmentation technology. Schmidhuber [64] used WGAN-GP [98] to expand the time domain signal and the LOFAR spectrum and uses CNN and LSTM to classify underwater targets and ship signals. The experimental results show that the recognition accuracy of the CNN model increased by 3.7 percentage points when training with the dataset was augmented.

4. Challenges

The real marine environment is complex and changeable, and acquiring data from it poses various problems. For example, performing realistic underwater experiments has a high cost. The propagation of the acoustic signals process exits expansion loss, absorption loss, and boundary loss, so acquiring high-quality underwater sound signals is time and energy consuming. At the same time, due to the limitations of underwater communication hardware equipment, it is difficult to guarantee the quality of the acquired data. There may be problems such as non-homogenous resolution, a too-weak target signal, non-uniform intensity, and reverberation [98]. High-resolution sensors can improve the system of sonar systems, but they are expensive. Therefore, there is currently a lack of publicly available datasets for UATR. Again, data management, labeling, and storage also consume a lot of time and energy. Most UATR methods are tested on the recordings collected in relatively simple sea areas of the environment, and they do not apply to signals from complex sea areas. On top of that, the underwater acoustic data collected for military purposes are mostly highly classified and hard to use for academic studies. The dataset imbalance is one of the main problems for machine learning-based UATR.

Due to the lack of public datasets, most current studies on UATR based on machine learning use self-constructed datasets to evaluate the model recognition performance. The published literature does not provide detailed information about their dataset, so it is hard to compare the performance of various methods in the same dimension. In the complex marine environment, underwater acoustic signals are affected by various factors, such as time, temperature, depth, salinity, geographic location, and sensor type [99]. In the research, a set of factors should be comprehensively considered and should design appropriate models, which also brings daunting challenges to the underwater acoustic target recognition works.

In response to the problem of insufficient training data, some researchers have used transfer learning and data augmentation techniques in UATR. However, data augmentation technology has certain limitations. It is often a simple transformation based on raw data. Even the data generated through the neural network model is similar to the known samples in its distribution. Whether it can represent the data characteristics in the real environment remains to be verified. Transfer learning requires pretraining on a large number of source domain sample data. Whether it is possible to find a source domain close to the target domain and how to determine the appropriate source domain size are problems that are currently faced. Moreover, the pretraining process of the model in the source domain also requires a lot of computing resources and time.

In addition, it is a grant challenge to explore the model architecture and parameters suitable for underwater acoustic signals and improve the efficiency of network training because there is a set of problems in the training process, such as the model failing to convergence caused by the gradient vanishing or gradient explosion. Many machine learning models, and more so deep learning models based on neural networks, are black-box models with little interpretability. However, in many practical application scenarios, the predicted basis of the model is required. It limits the application scope of such methods to a large extent.

Each method has its limitations. Statistical learning models have a small number of parameters and compute faster but are only suitable for small datasets. Deep learning models with more complex structures usually provide better recognition accuracy than statistical learning methods but require more computational resources and training time. In addition, deep learning models are less computationally efficient as they are usually programmed based on the highly integrated python language. Deep learning models require data with high quality and poor generalization, which makes it hard to deploy current deep learning-based UATR methods directly to underwater acoustic monitoring systems.

5. Conclusions and Discussion

This paper reviews the recent studies of underwater acoustic target recognition based on machine learning, gives the flow chart from underwater acoustics signals acquisition to data processing and target recognition, analyzes the pros and cons of the machine learning framework used in relevant studies as well as the recognition performance, and summarizes the challenges faced by UATR based on machine learning. Due to the lack of training data, the current development trend of UATR is to combine manual features with machine learning methods. This paper introduced the feature extraction methods and their respective characteristics in underwater acoustic target recognition in Section 2. Then, we summarized the UATR methods based on machine learning and analyze the applicable scene of different machine learning models by citing the papers in Section 3. According to the different machine learning methods used, this paper organized this part into three categories: (1) UATR methods based on statistical learning; (2) UATR technologies using CNNs, RNNs, and other deep learning methods; (3) the application of transfer learning and data augmentation technology for UATR. By surveying the relevant literature, we found some problems and challenges facing UATR, including data acquisition, management, storage, labeling, computing resource problems, and some disadvantages of machine learning methods. The details were given in Section 4.

Machine learning methods have been widely used in natural language processing and computer vision fields and have achieved breakthroughs. On the contrary, the development of the application of machine learning in UATR both nationally and internationally is slow. Hence, there is still a set of work to be completed in the future. First, the dataset problem is one of the main problems faced by underwater acoustic target recognition methods based on machine learning. Effective data acquisition and preprocessing methods are the focus of the current progress in automatic UATR. At present, what needs to be solved urgently is the establishment of an open and standardized dataset for model training and testing, facilitating the comparative analysis of the studies, and promoting the development of UATR. At the same time, future works should pay attention to the basic research on the underwater acoustics, and try to use the fusion of multiple spectral features methods to describe the features of underwater targets from multiple dimensions. For the insufficient and unbalanced training data issue, we need to select appropriate transfer learning and data augmentation techniques according to the actual situation to improve the performance of UATR. More studies should also be conducted to understand and evaluate existing state-of-the-art deep learning architectures and their application in underwater acoustic signal classification. In addition, interpretability is an important issue that needs to be tackled in the future development of UATR based on machine learning. Only when the model has certain interpretability can it be applied to more actual scenarios and play a significant role. In the complex underwater environment, it is hard to solve all the challenges faced by UATR using a single form of data. Multimodal learning is one of the current research hotspots, aiming to learn information from multiple modalities in various modalities and to achieve the communication and transformation of information from different modalities [100]. Many researchers have tried to use multimodal learning for studies such as underwater navigation [101] and underwater communication [102]. Using multimodal data such as acoustic, optical, and imagery for UATR offers new ideas for future research.

Author Contributions: Conceptualization, X.L. and L.C.; methodology, X.L., L.C. and H.Z.; validation, L.C., H.Z. and H.C.; formal analysis, X.L. and L.C.; investigation, L.C.; resources, X.L and H.C; writing—original draft preparation, X.L., L.C. and H.Z.; writing—review and editing, X.L. and L.C.; visualization, X.L., L.C. and H.Z.; supervision, X.L.; project administration, X.L.; funding acquisition, X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by the National Natural Science Foundation of China under Grant 12174053, Grant 91938203; and in part by the Fundamental Research Funds for Central Universities No.2242022K30016.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

Notes

[1] https://archive-beta.ics.uci.edu/dataset/151/connectionist+bench+sonar+mines+vs+rocks, accessed on 15 December 2022.

[2] http://www.cemma.org, accessed on 15 December 2022.

[3] https://github.com/HHUCzCz/-SeabedObjects-KLSG--II, accessed on 16 November 2021.

[4] https://goo.gl/mwd4gj, accessed on 29 May 2017.

[5] https://csms-acoustic.haifa.ac.il/index.php/s/2UmUoK80Izt0Roe, accessed on 15 December 2022.

References

1. Urick, R.J. *Principles of Underwater Sound*; McGraw-Hill Book, Co.: Los Angeles, CA, USA, 1983; 423p.
2. Kenny, A.J.; Cato, I.; Desprez, M.; Fader, G.; Schüttenhelm, R.T.E.; Side, J. An Overview of Seabed-Mapping Technologies in the Context of Marine Habitat Classification\ding73. *ICES J. Mar. Sci.* **2003**, *60*, 411–418. [CrossRef]
3. Testolin, A.; Kipnis, D.; Diamant, R. Detecting Submerged Objects Using Active Acoustics and Deep Neural Networks: A Test Case for Pelagic Fish. *IEEE Trans. Mob. Comput.* **2022**, *21*, 2776–2788. [CrossRef]
4. Testolin, A.; Diamant, R. Combining Denoising Autoencoders and Dynamic Programming for Acoustic Detection and Tracking of Underwater Moving Targets. *Sensors* **2020**, *20*, 2945. [CrossRef] [PubMed]
5. Yuan, F.; Ke, X.; Cheng, E. Joint Representation and Recognition for Ship-Radiated Noise Based on Multimodal Deep Learning. *J. Mar. Sci. Eng.* **2019**, *7*, 380. [CrossRef]
6. Shiliang, F.; Shuanping, D.; Xinwei, L.; Ning, H.; Xiaonan, X. Development of Underwater Acoustic Target Feature Analysis and Recognition Technology. *Bull. Chin. Acad. Sci. Chin. Version* **2019**, *34*, 297–305.
7. Aksuren, I.G.; Hocaoglu, A.K. Automatic Target Classification Using Underwater Acoustic Signals. In Proceedings of the 2022 30th Signal Processing and Communications Applications Conference (SIU), Safranbolu, Turkey, 15–18 May 2022.
8. 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.
9. Chao-xion, S. Application of the Lifting Wavelet Transform Based MFCC in Target Identification. *Tech. Acoust.* **2014**, *33*, 372–375.
10. Teng, B.; Zhao, H. Underwater Target Recognition Methods Based on the Framework of Deep Learning: A Survey. *Int. J. Adv. Robot. Syst.* **2020**, *17*, 172988142097630. [CrossRef]
11. 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]
12. Tian, S.; Chen, D.; Wang, H.; Liu, J. Deep Convolution Stack for Waveform in Underwater Acoustic Target Recognition. *Sci. Rep.* **2021**, *11*, 9614. [CrossRef]
13. Wang, Y.; Peng, H. Underwater Acoustic Source Localization Using Generalized Regression Neural Network. *J. Acoust. Soc. Am.* **2018**, *143*, 2321–2331. [CrossRef]
14. Chen, J.; Liu, C.; Xie, J.; An, J.; Huang, N. Time–Frequency Mask-Aware Bidirectional LSTM: A Deep Learning Approach for Underwater Acoustic Signal Separation. *Sensors* **2022**, *22*, 5598. [CrossRef]
15. Bianco, M.J.; Gerstoft, P.; Traer, J.; Ozanich, E.; Roch, M.A.; Gannot, S.; Deledalle, C.-A. Machine Learning in Acoustics: Theory and Applications. *J. Acoust. Soc. Am.* **2019**, *146*, 3590–3628. [CrossRef]
16. Yang, H.; Lee, K.; Choo, Y.; Kim, K. Underwater Acoustic Research Trends with Machine Learning: Passive SONAR Applications. *J. Ocean Eng. Technol.* **2020**, *34*, 227–236. [CrossRef]
17. Yang, H.; Lee, K.; Choo, Y.; Kim, K. Underwater Acoustic Research Trends with Machine Learning: General Background. *J. Ocean Eng. Technol.* **2020**, *34*, 147–154. [CrossRef]
18. Pan, Y.; Zhao, A.; Yu, F.; Zhang, X. Line Spectrum Comparison Method Based on LOFAR Spectrum Feature. In Proceedings of the 2015 International Industrial Informatics and Computer Engineering Conference, Shaanxi, China, 10–11 January 2015; Atlantis Press: Paris, France, 2015.
19. Jin, G.; Liu, F.; Wu, H.; Song, Q. Deep Learning-Based Framework for Expansion, Recognition and Classification of Underwater Acoustic Signal. *J. Exp. Ampmathsemicolon Theor. Artif. Intell.* **2019**, *32*, 205–218. [CrossRef]
20. Sichun, L.; Desen, Y. DEMON Feature Extraction of Acoustic Vector Signal Based on 3/2-D Spectrum. In Proceedings of the 2007 2nd IEEE Conference on Industrial Electronics and Applications, Harbin, China, 23–25 May 2007.
21. 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, Monterey, CA, USA, 19–23 September 2016.
22. De Moura, N.N.; de Seixas, J.M.; Ramos, R. Passive Sonar Signal Detection and Classification Based on Independent Component Analysis. In *Sonar Systems*; InTech: London, UK, 2011.

23. 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]
24. Jankowski, C.R.; Quatieri, T.F.; Reynolds, D.A. Measuring Fine Structure in Speech: Application to Speaker Identification. In Proceedings of the 1995 International Conference on Acoustics, Speech, and Signal Processing, Detroit, MI, USA, 9–12 May 1995.
25. On, C.K.; Pandiyan, P.M.; Yaacob, S.; Saudi, A. Mel-Frequency Cepstral Coefficient Analysis in Speech Recognition. In Proceedings of the 2006 International Conference on Computing & Informatics, Kuala Lumpur, Malaysia, 6–8 June 2006.
26. Tazi, E.B. A Robust Speaker Identification System Based on the Combination of GFCC and MFCC Methods. In Proceedings of the 2016 5th International Conference on Multimedia Computing and Systems (ICMCS), Marrakech, Morocco, 29 September–1 October 2016.
27. Lim, T.; Bae, K.; Hwang, C.; Lee, H. Classification of Underwater Transient Signals Using MFCC Feature Vector. In Proceedings of the 2007 9th International Symposium on Signal Processing and Its Applications, Sharjah, United Arab Emirates, 12–15 February 2007.
28. 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.
29. 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.
30. Zhengchon, M. Speaker Recognition Algorithm Based on Gammatone Filter Bank. *Comput. Eng. Appl.* **2015**, *51*, 200–203.
31. Grimaldi, M.; Cummins, F. Speaker Identification Using Instantaneous Frequencies. *IEEE Trans. Audio Speech Lang. Process.* **2008**, *16*, 1097–1111. [CrossRef]
32. Wang, X.; Liu, A.; Zhang, Y.; Xue, F. Underwater Acoustic Target Recognition: A Combination of Multi-Dimensional Fusion Features and Modified Deep Neural Network. *Remote Sens.* **2019**, *11*, 1888. [CrossRef]
33. Wei, Z.; Ju, Y.; Song, M. A Method of Underwater Acoustic Signal Classification Based on Deep Neural Network. In Proceedings of the 2018 5th International Conference on Information Science and Control Engineering (ICISCE), Zhengzhou, China, 20–22 July 2018; pp. 46–50.
34. 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*, 1500905. [CrossRef]
35. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.-C.; Tung, C.C.; Liu, H.H. The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis. *Proc. R. Soc. Lond. Ser. Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [CrossRef]
36. Dätig, M.; Schlurmann, T. Performance and Limitations of the Hilbert–Huang Transformation (HHT) with an Application to Irregular Water Waves. *Ocean Eng.* **2004**, *31*, 1783–1834. [CrossRef]
37. Zeng, X.; Wang, S. Underwater Sound Classification Based on Gammatone Filter Bank and Hilbert-Huang Transform. In Proceedings of the 2014 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), Guilin, China, 5–8 August 2014.
38. Bao, F.; Li, C.; Wang, X.; Wang, Q.; Du, S. Ship Classification Using Nonlinear Features of Radiated Sound: An Approach Based on Empirical Mode Decomposition. *J. Acoust. Soc. Am.* **2010**, *128*, 206–214. [CrossRef] [PubMed]
39. 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.
40. Azimi-Sadjadi, M.R.; Yao, D.; Huang, Q.; Dobeck, G.J. Underwater Target Classification Using Wavelet Packets and Neural Networks. *IEEE Trans. Neural Netw.* **2000**, *11*, 784–794. [CrossRef]
41. Wang, N.; He, M.; Sun, J.; Wang, H.; Zhou, L.; Chu, C.; Chen, L. Ia-PNCC: Noise Processing Method for Underwater Target Recognition Convolutional Neural Network. *Comput. Mater. Ampmathsemicolon Contin.* **2019**, *58*, 169–181. [CrossRef]
42. Jordan, M.I.; Mitchell, T.M. Machine Learning: Trends, Perspectives, and Prospects. *Science* **2015**, *349*, 255–260. [CrossRef]
43. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
44. Cao, X.; Zhang, X.; Yu, Y.; Niu, L. Deep Learning-Based Recognition of Underwater Target. In Proceedings of the 2016 IEEE International Conference on Digital Signal Processing (DSP), Beijing, China, 16–18 October 2016; pp. 89–93.
45. 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; Atlantis Press: Paris, France, 2017.
46. Xu, F.; Zou, Z.-J.; Yin, J.-C.; Cao, J. Identification Modeling of Underwater Vehicles' Nonlinear Dynamics Based on Support Vector Machines. *Ocean Eng.* **2013**, *67*, 68–76. [CrossRef]
47. Vieira, M.; Amorim, M.C.P.; Sundelöf, A.; Prista, N.; Fonseca, P.J. Underwater Noise Recognition of Marine Vessels Passages: Two Case Studies Using Hidden Markov Models. *ICES J. Mar. Sci.* **2019**, *77*, 2157–2170. [CrossRef]
48. 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.
49. 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.

50. 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.
51. Parada, P.P.; Cardenal-López, A. Using Gaussian Mixture Models to Detect and Classify Dolphin Whistles and Pulses. *J. Acoust. Soc. Am.* **2014**, *135*, 3371–3380. [CrossRef]
52. Schmidt, R. Multiple Emitter Location and Signal Parameter Estimation. *IEEE Trans. Antennas Propag.* **1986**, *34*, 276–280. [CrossRef]
53. Kim, T.; Bae, K. HMM-Based Underwater Target Classification with Synthesized Active Sonar Signals. *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.* **2011**, *E94-A*, 2039–2042. [CrossRef]
54. Mallat, S.G.; Zhang, Z. Matching Pursuits with Time-Frequency Dictionaries. *IEEE Trans. Signal Process.* **1993**, *41*, 3397–3415. [CrossRef]
55. Mohammed, S.K.; Hariharan, S.M.; Kamal, S. A GTCC-Based Underwater HMM Target Classifier with Fading Channel Compensation. *J. Sens.* **2018**, *2018*, 1–14. [CrossRef]
56. Smola, A. *Learning with Kernels*; GMD-Forschungszentrum Informationstechnik: Berlin, Germany, 1998.
57. dos Anjos, A.; Torres, R.C.; Seixas, J.M.; Ferreira, B.C.; Xavier, T.C. Neural Triggering System Operating on High Resolution Calorimetry Information. *Nucl. Instrum. Methods Phys. Res. Sect. Accel. Spectrometers Detect. Assoc. Equip.* **2006**, *559*, 134–138. [CrossRef]
58. Yang, X.-S. A New Metaheuristic Bat-Inspired Algorithm. In *Nature Inspired Cooperative Strategies for Optimization (NICSO 2010)*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 65–74.
59. Lee, K.C. Underwater Acoustic Localisation by GMM Fingerprinting with Noise Reduction. *Int. J. Sens. Netw.* **2019**, *31*, 1. [CrossRef]
60. Kannan, S. Intelligent Object Recognition in Underwater Images Using Evolutionary-Based Gaussian Mixture Model and Shape Matching. *Signal Image Video Process.* **2020**, *14*, 877–885. [CrossRef]
61. Tęgowski, J.; Koza, R.; Pawliczka Vel Pawlik, I.; Skóra, K.; Trzcińska, K.; Zdroik, J. Statistical, Spectral and Wavelet Features of the Ambient Noise Detected in the Southern Baltic Sea. In Proceedings of the 23rd International Congress on Sound and Vibration, Athens, Greece, 10–14 July 2016; International Institute of Acoustics and Vibration (IIAV): Auburn, AL, USA, 2016.
62. Pujol, P.; Pol, S.; Nadeu, C.; Hagen, A.; Bourlard, H. Comparison and Combination of Features in a Hybrid HMM/MLP and a HMM/GMM Speech Recognition System. *IEEE Trans. Speech Audio Process.* **2005**, *13*, 14–22. [CrossRef]
63. Binesh, T.; Supriya, M.H.; Pillai, P.R.S. Discrete Sine Transform Based HMM Underwater Signal Classifier. In Proceedings of the 2011 International Symposium on Ocean Electronics, Kochi, India, 16–18 November 2011.
64. Schmidhuber, J. Deep Learning in Neural Networks: An Overview. *Neural Netw.* **2015**, *61*, 85–117. [CrossRef]
65. Gevrey, M.; Dimopoulos, I.; Lek, S. Review and Comparison of Methods to Study the Contribution of Variables in Artificial Neural Network Models. *Ecol. Model.* **2003**, *160*, 249–264. [CrossRef]
66. Hu, G.; Wang, K.; Peng, Y.; Qiu, M.; Shi, J.; Liu, L. Deep Learning Methods for Underwater Target Feature Extraction and Recognition. *Comput. Intell. Neurosci.* **2018**, *2018*, 1–10. [CrossRef] [PubMed]
67. 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.
68. Hu, G.; Wang, K.; Liu, L. Underwater Acoustic Target Recognition Based on Depthwise Separable Convolution Neural Networks. *Sensors* **2021**, *21*, 1429. [CrossRef] [PubMed]
69. 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]
70. Kamal, S.; Satheesh Chandran, C.; Supriya, M.H. 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]
71. Xue, L.; Zeng, X.; Jin, A. A Novel Deep-Learning Method with Channel Attention Mechanism for Underwater Target Recognition. *Sensors* **2022**, *22*, 5492. [CrossRef]
72. Feng, S.; Zhu, X. A Transformer-Based Deep Learning Network for Underwater Acoustic Target Recognition. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1505805. [CrossRef]
73. 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]
74. Irfan, M. DeepShip: An Underwater Acoustic Benchmark Dataset and a Separable Convolution Based Autoencoder for Classification. *Expert Syst. Appl.* **2021**, *12*, 115270. [CrossRef]
75. Albawi, S.; Mohammed, T.A.; Al-Zawi, S. Understanding of a Convolutional Neural Network. In Proceedings of the 2017 International Conference on Engineering and Technology (ICET), Antalya, Turkey, 21–23 August 2017.
76. Hershey, S.; Chaudhuri, S.; Ellis, D.P.W.; Gemmeke, J.F.; Jansen, A.; Moore, R.C.; Plakal, M.; Platt, D.; Saurous, R.A.; Seybold, B.; et al. CNN Architectures for Large-Scale Audio Classification. In Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017.
77. Akhtar, N.; Ragavendran, U. Interpretation of Intelligence in CNN-Pooling Processes: A Methodological Survey. *Neural Comput. Appl.* **2019**, *32*, 879–898. [CrossRef]

78. Scherer, D.; Müller, A.; Behnke, S. Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition. In *Artificial Neural Networks—ICANN 2010*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 92–101.
79. Valdenegro-Toro, M. Object Recognition in Forward-Looking Sonar Images with Convolutional Neural Networks. In Proceedings of the OCEANS 2016 MTS/IEEE Monterey, Monterey, CA, USA, 19–23 September 2016.
80. Nguyen, H.T.; Lee, E.-H.; Bae, C.H.; Lee, S. Multiple Object Detection Based on Clustering and Deep Learning Methods. *Sensors* **2020**, *20*, 4424. [CrossRef]
81. Galusha, A.P.; Dale, J.; Keller, J.; Zare, A. Deep Convolutional Neural Network Target Classification for Underwater Synthetic Aperture Sonar Imagery. In *Proceedings of the Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXIV, Baltimore, MD, USA, 10 May 2019*; Isaacs, J.C., Bishop, S.S., Eds.; SPIE: Bellingham, WA, USA, 2019.
82. Ke, X.; Yuan, F.; Cheng, E. Underwater Acoustic Target Recognition Based on Supervised Feature-Separation Algorithm. *Sensors* **2018**, *18*, 4318. [CrossRef]
83. Lipton, Z.C.; Berkowitz, J.; Elkan, C. A Critical Review of Recurrent Neural Networks for Sequence Learning. *arXiv* **2015**, arXiv:150600019.
84. Niu, Z.; Zhong, G.; Yu, H. A Review on the Attention Mechanism of Deep Learning. *Neurocomputing* **2021**, *452*, 48–62. [CrossRef]
85. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 11.
86. Chen, X.-W.; Lin, X. Big Data Deep Learning: Challenges and Perspectives. *IEEE Access* **2014**, *2*, 514–525. [CrossRef]
87. Ying, J.J.-C.; Lin, B.-H.; Tseng, V.S.; Hsieh, S.-Y. Transfer Learning on High Variety Domains for Activity Recognition. *Proc. ASE BigData SocialInform.* **2015**, *2015*, 1–6.
88. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. *Commun. ACM* **2014**, *63*, 139–144. [CrossRef]
89. Nguyen, H.-T.; Lee, E.-H.; Lee, S. Study on the Classification Performance of Underwater Sonar Image Classification Based on Convolutional Neural Networks for Detecting a Submerged Human Body. *Sensors* **2019**, *20*, 94. [CrossRef]
90. Fuchs, L.R.; Gallstrom, A.; Folkesson, J. Object Recognition in Forward Looking Sonar Images Using Transfer Learning. In Proceedings of the 2018 IEEE/OES Autonomous Underwater Vehicle Workshop (AUV), Porto, Portugal, 6–9 November 2018; pp. 1–6.
91. Korkmaz, B.N.; Diamant, R.; Danino, G.; Testolin, A. Automated Detection of Dolphin Whistles with Convolutional Networks and Transfer Learning. *Front. Artif. Intell.* **2023**, *6*, 1099022. [CrossRef]
92. 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]
93. 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]
94. Jiang, Z.; Zhao, C.; Wang, H. Classification of Underwater Target Based on S-ResNet and Modified DCGAN Models. *Sensors* **2022**, *22*, 2293. [CrossRef] [PubMed]
95. Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Fei-Fei, L. ImageNet: A Large-Scale Hierarchical Image Database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009.
96. Gillespie, D.; Caillat, M.; Gordon, J.; White, P. Automatic Detection and Classification of Odontocete Whistles. *J. Acoust. Soc. Am.* **2013**, *134*, 2427–2437. [CrossRef]
97. Radford, A.; Metz, L.; Chintala, S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv* **2016**, arXiv:1511.06434.
98. Neupane, D.; Seok, J. A Review on Deep Learning-Based Approaches for Automatic Sonar Target Recognition. *Electronics* **2020**, *9*, 1972. [CrossRef]
99. Domingos, L.C.F.; Santos, P.E.; Skelton, P.S.M.; Brinkworth, R.S.A.; Sammut, K. A Survey of Underwater Acoustic Data Classification Methods Using Deep Learning for Shoreline Surveillance. *Sensors* **2022**, *22*, 2181. [CrossRef]
100. Baltrusaitis, T.; Ahuja, C.; Morency, L.-P. Multimodal Machine Learning: A Survey and Taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 423–443. [CrossRef]
101. Gjanci, P.; Petrioli, C.; Basagni, S.; Phillips, C.A.; Boloni, L.; Turgut, D. Path Finding for Maximum Value of Information in Multi-Modal Underwater Wireless Sensor Networks. *IEEE Trans. Mob. Comput.* **2018**, *17*, 404–418. [CrossRef]
102. Han, S.; Noh, Y.; Lee, U.; Gerla, M. Optical-Acoustic Hybrid Network toward Real-Time Video Streaming for Mobile Underwater Sensors. *Ad Hoc Netw.* **2019**, *83*, 1–7. [CrossRef]