Ocean Engineering 298 (2024) 117252

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng

Review

A survey on machine learning in ship radiated noise

Hilde I. Hummel [a,*], Rob van der Mei [a,b], Sandjai Bhulai [b]

[a] Centrum Wiskunde & Informatica, Department of Stochastics, Science Park 123, Amsterdam, 1098 XG, Netherlands
[b] Vrije Universiteit, Department Mathematics, De Boelelaan 1111, Amsterdam, 1081 HV, Netherlands

ARTICLE INFO

Keywords:
Ship radiated noise
Machine learning
Survey
Deep learning
Underwater sound

ABSTRACT

The utilization of machine learning in analyzing ship radiated noise (SR-N) is undergoing rapid evolution. Because the omnipresent background noise strongly depends on the highly variable environment, the application of such techniques poses challenges. Furthermore, publicly available labeled datasets are scarce. Motivated by this, there has been a surge in the number of publications regarding the implementation of machine learning in the monitoring of SR-N within the past few years. This comprehensive survey delineates the state-of-the-art machine learning techniques applied to SR-N, with a specific focus on passive measurements. Recent developments are categorized into several sub-areas, namely; publicly available datasets, data augmentation, signal denoising, feature extraction, detection, localization, and recognition of SR-N. Additionally, future research directions are explored.

1. Introduction

The health of our ocean environment is endangered by human-induced sound pollution. The main pollution source originates from noise generated by ships, called *Ship Radiated Noise* (SR-N). This raises the need to measure and analyze the SR-N levels. This quantifies the amount of pollution and identifies or locates the ships producing noise. The measurement of underwater noise is utilized using so-called sonar systems. These systems can be categorized into three groups: (1) active sonar, (2) side scan sonar, and (3) passive sonar. The active sonar system emits an acoustic pulse and listens to the returning echo. Similarly, the side scan sonar can generate an acoustic image based on the measured echo. This survey will focus on passive sonar systems. These systems do not emit any sound but quietly listen to the noise in the ocean. It passively detects sound waves coming towards the hydrophone(s). This makes the passive system the ultimate system to monitor SR-N. From passive measurements, the SR-N can be analyzed in more detail. This analysis is challenged by multiple factors as expressed in the passive sonar equation:

$$SNR(\mathrm{dB}) = -\big(NL(\mathrm{dB}) - AG(\mathrm{dB})\big) - TL(\mathrm{dB}) + SL(\mathrm{dB}). \quad (1)$$

In this equation, all terms have the specific underwater sound unit dB relative to 1 μPa. *SNR* is the signal-to-noise ratio, which is equivalent to the measured sound by the hydrophone(s). The Noise Level (*NL*) represents the background noise produced by the ocean environment. The Array Gain (*AG*) reduces the *NL*. This value is set to 0 dB for a single hydrophone. Besides the ocean's background noise, the measurement is also distorted by Transmission Loss (*TL*). This is the energy loss of the acoustic source during travel from source to receiver. The total amount of energy loss depends on multiple environmental factors such as water temperature, depth, multi-path distortions, and sea bed type. The *TL* and *NL* are highly variable and may change over time since they depend on the dynamic and complex ocean environment. The final element in the sonar equation is the Source Level (*SL*), which is the acoustic energy level of the SR-N. The acoustic energy level is composed of three elements (Smith and Rigby, 2022; Slamnoiu et al., 2016; Veirs et al., 2016; Urick and United States. Naval Sea Systems Command. Undersea Warfare Technology Office, 1984; Liu et al., 2023c):

- machinery noise(20–1000 Hz);
- flow noise (5–10 Hz);
- propeller noise and cavitation (50–150 Hz).

From these elements, cavitation is the primary noise source, accounting for 80%–85% of the total generated noise. The strength of the total SR-N varies between around 140 dB @1 m for small fishing vessels and around 195 dB @1 m for maritime oil tankers (Slamnoiu et al., 2016).

The combination of these noise-elements generates a unique sound profile of the ship. Altogether, it can be concluded that the generated sound profile depends on the maintenance state of the ship and the ocean environment. This introduces variations and challenges the SR-N analysis. Traditionally, the SR-N analysis is performed manually by experienced sonar operators. However, this is a slow and costly process and is prone to human errors. Automation by *Machine Learning* (ML)

* Corresponding author.
E-mail address: h.i.hummel@cwi.nl (H.I. Hummel).

https://doi.org/10.1016/j.oceaneng.2024.117252
Received 30 November 2023; Received in revised form 7 February 2024; Accepted 20 February 2024
Available online 23 February 2024

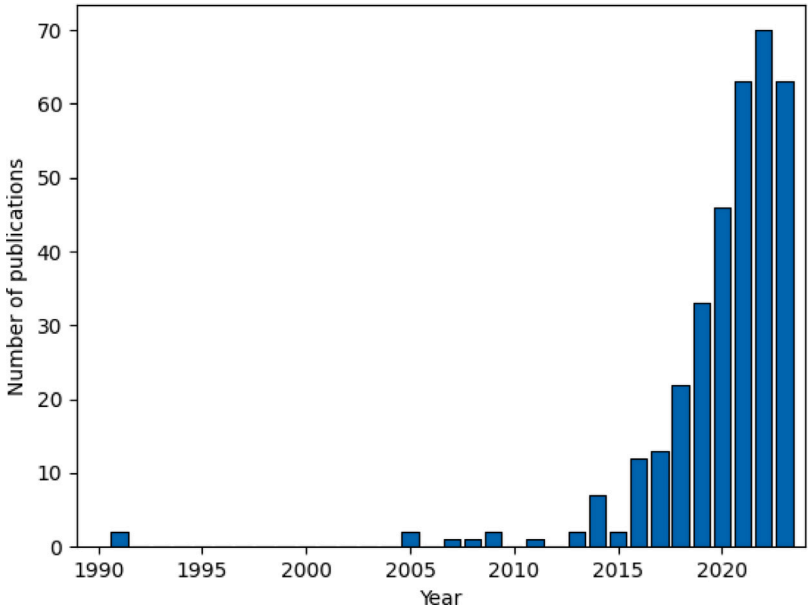

Fig. 1. Number of publications about the application of machine learning on SR-N. Papers derived from Google Scholar, last accessed on 30 December 2023.

could overcome these limitations. Recently, the potential of ML in underwater applications has been shown by outperforming traditional methods (Yuan et al., 2019; Wang and Peng, 2018; Chen et al., 2022c). For ML applications in automatic SR-N analysis, there has been a rise in the number of publications within the past few years (see Fig. 1). This graph has been created by selecting papers derived from Google Scholar. Here, the search criteria were 'Underwater Acoustic Target Recognition', 'Ship Radiated Noise', and 'Underwater Acoustic Signal Recognition'. From these search criteria, papers were selected if they applied an ML algorithm specified on SR-N.

In addition to other research initiatives (Neupane and Seok, 2020; Smith and Rigby, 2022; Huang et al., 2022; Niu et al., 2023; Luo et al., 2023), this survey offers a complete overview of the progressive state-of-the-art ML developments in automatic SR-N analysis. It summarizes the progress and critically examines the limitations and capabilities of recent ML advancements. This survey is categorized into three main segments: (1) an examination of the SR-N data in Section 2, (2) an in-depth exploration of data preprocessing methods, and (3) an assessment of the ultimate applications of automatic SR-N analysis (see Fig. 2). Within the preprocessing segment, various feature extraction methods are discussed in Section 3, followed by augmentation techniques in Section 4 and denoising methods in Section 5. Here, the denoising methods are discussed after the feature extraction methods, since some of the proposed techniques have been proposed both in feature extraction and denoising of SR-N recordings. The automatic analysis segment reviews the automatic detection in Section 6, automatic localization of the noise source in Section 7, and recognition of ships separately in Section 8. The survey culminates with a prospective outlook on potential future research directions in Section 9.

2. Datasets

Before any ML application can be developed, numerous SR-N measurements need to be gathered to create a dataset. For this purpose, hydrophones are deployed to record all types of underwater sounds. These hydrophones can be deployed in various settings. One of these settings is attached to a moving object, like a ship or an *Autonomous Underwater Vehicle* (AUV). This setting introduces self-noise by the moving object, complicating the implementation of the subsequent ML application. Consequently, a more commonly employed setting for data gathering is the utilization of stationary hydrophones. In this setup, individual hydrophones can be deployed, or an array can be formed by stacking hydrophones either vertically or horizontally. The utilization of arrays has shown to be less sensitive to noise compared to single hydrophones (Zhang et al., 2022b). Following the recording process, the collected data is combined to create a dataset of underwater sounds, encompassing SR-N. A substantial amount of underwater acoustic data is publicly available and can be employed to train various ML models for automatic SR-N analysis. Unfortunately, only a small fraction of this data is labeled. These labels can either be the identification or the relative location of the sound source. A description of the labeled dataset is given within this section, followed by a brief mention of available unlabeled datasets. The overview of the labeled datasets is given in Table 1.

Table 1
Specifications of publicly available labeled datasets called ShipsEar and Deepship.

Name	Sample rate	# Classes	Total duration	Citation
ShipsEar	22,050 Hz	12	3 h 8 m	Santos-Domínguez et al. (2016)
Deepship	32,000 Hz	4	47 h 4 m	Irfan et al. (2021)

2.1. Labeled datasets

For publicly available datasets, the labels are limited to the corresponding ship type that is recorded. This section describes two publicly available labeled datasets, called ShipsEar and Deepship.

2.1.1. ShipsEar

One of the publicly available labeled datasets is called ShipsEar. It contains recordings of background noise and multiple ship types (Santos-Domínguez et al., 2016). The dataset consists of 90 recordings, varying in duration between 15 s and 10 min. All recordings were made between the Autumn of 2012 and the Summer of 2013 on the Atlantic Coast near Port Ria de Vigo. This region has a maximum water depth of only 45 m. Nevertheless, there is abundant fishery activity within this shallow-water region, making it an attractive place for SR-N monitoring using hydrophones. Each hydrophone covers a wide frequency range of 1 Hz–28 kHz. Multiple hydrophones were stacked to create a vertical array. This scheme allows the recording of noise at different depths. When multiple recordings of a single vessel were available, the highest-quality recording was selected for the database. The ultimate labels were assigned using five categories:

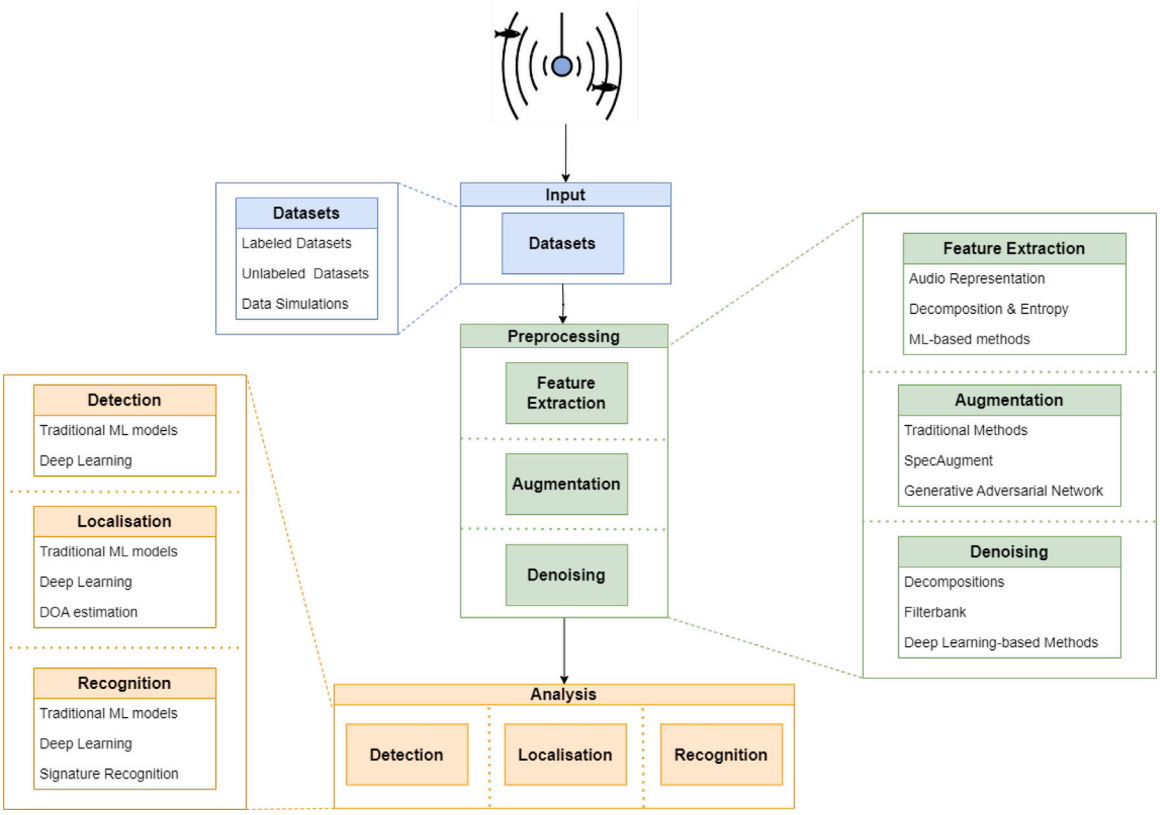

Fig. 2. Flowchart of ML applications in SR-N.

- Fishery;
- Motor boats/sailboats;
- Ferries;
- Liners;
- Background noise.

The target vessels were visually identified and verified during labeling. These labels and recordings were then used to create a basic classifier. The corresponding Cepstral Coefficients were extracted, and a (see Section 3.1.3) a *Gaussian Mixture Model* was optimized.

2.1.2. Deepship

Another publicly available labeled dataset is called Deepship. It is published as a benchmark dataset (Irfan et al., 2021). Compared to ShipsEar, Deepship contains more data. It has around 47 h of recordings of 265 different types of vessels and contains multiple types of underwater sounds. The recordings were made between 2 May 2016 and 4 October 2018, making it a diverse dataset. Due to this diversity, the strongly varying circumstances of the ocean environment are partly covered within the recordings. The recordings were made using a hydrophone with a 1 Hz–12 kHz frequency range. It was placed in the Strait of Georgia, which is a crowded strait with soil of sand and silt. The recordings were labeled using *Automatic Identification System* (AIS) data and were assigned to either of these four categories:

- Tanker;
- Tug;
- Passengers ship;
- Cargo ship.

An inclusion zone of two km was set to assign the corresponding label (see Fig. 3). When ships were out of this range, the recording was stopped. This dataset was utilized to create a baseline classifier. For this classifier, a separable *Convolutional Autoencoder*, with blocks based on Xception, was suggested.

2.2. Unlabeled datasets

Besides the limited quantity of publicly available labeled datasets, there is a massive amount of unlabeled data accessible. These datasets are generated for multiple purposes, such as marine mammal monitoring. Due to this multi-purpose, diversity is maintained within the recordings. This property makes these datasets a powerful data source to train ML models for automatic SR-N analysis. This section briefly mentions two commonly utilized unlabeled data sources.

2.2.1. Ocean Network Canada

Ocean Network Canada (ONC) is a massive publicly available unlabeled dataset and, therefore, a commonly utilized data source for ML applications. The dataset contains acoustic, physical, and biological recordings of the Atlantic, Pacific, and Arctic Ocean (Canada, 2007). From this data source, multiple labeled datasets, such as Deepship, have been created. Another subset has been utilized to create another labeled dataset in Domingos et al. (2022). Within the study, the labels have been assigned in the same way as Deepship.

2.2.2. National Oceanic And Atmospheric Administration

The *National Oceanic and Atmospheric Administration* (NOAA) is an underwater acoustic data source. This platform was originally intended to characterize sounds from fish and marine mammals and to monitor ambient noise and human-made sounds (Oceanic and Administration, 2017). They collaborate with the U.S. Navy and support research that examines the impact of human-generated sounds on marine mammals. The hydrophones are placed in Alaska, the Pacific, California Current, Northeast of the USA, the Gulf of Mexico, and the Pacific islands. These hydrophones are placed in shallow water and deep ocean, totaling a diverse publicly available data source.

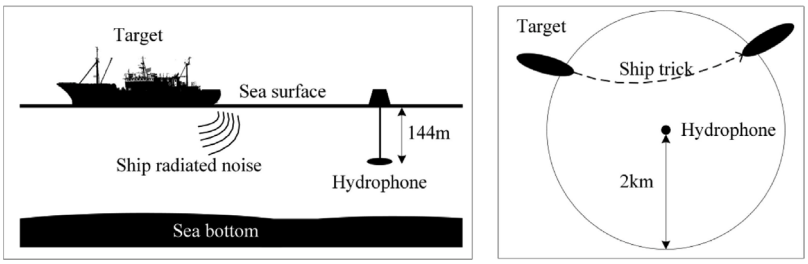

Fig. 3. An example visualization of the inclusion zone of the ONC dataset. The image on the left is a front view of the setting, and the image on the right is a helicopter view of the setting (Honghui et al., 2022).

2.3. Data simulation

To cope with the limited amount of labeled datasets, various studies proposed to generate simulated underwater sound data. Numerous underwater sound applications have been developed for this purpose. For instance, KRAKEN (Porter, 1992) has been utilized to simulate various types of ocean environments for automatic SR-N analysis (Cao and Ren, 2022; Li et al., 2018a; Zhang et al., 2022b). Additional simulators, such as ORCA.13 (Van Komen et al., 2019) and Bellhop (Li et al., 2022a), have also been suggested.

3. Feature extraction

From the SR-N recordings, information can be withdrawn by feature extraction methods. Unfortunately, the recorded SR-N is affected by environmental circumstances and the maintenance state of the ship itself. This results in a noisy audio signal with great variation within the same class of ship type. Therefore, the performance of the model relies on high-quality feature extraction. The discriminative capability of these types of features has not been explored thoroughly, making the feature extraction process extra important. As a result, many studies have focused on developing informative SR-N feature extraction techniques. An overview of the development of feature extraction from SR-N is given in Table 2.

3.1. Audio representations

Audio can be represented in either the *time domain*, the *frequency domain*, or the *time–frequency domain*. From these audio representations, different types of features can be extracted to represent SR-N. Below, multiple variants of audio representations as feature extraction methods are discussed.

3.1.1. Spectrograms

The time–frequency domain of the audio can be represented by so-called spectrograms. These are visual audio representations showing the signal strength over time at various frequencies by applying the *Short-Time Fourier Transformation* (STFT) on the raw audio. This SR-N audio representation is commonly utilized as an image input for a CNN to automatically recognize SR-N (Pan et al., 2021). A special type of spectrogram is the Mel-Spectrogram. This spectrogram is generated by taking the STFT of the original audio and converting the frequencies to the Mel scale. This scale is based on the human perception of sound, simulating human hearing. This type of spectrogram is commonly utilized for automatic SR-N recognition (Liu et al., 2021f; Vaz et al., 2022; Khalilabadi, 2023). Another study suggested creating a three-dimensional feature input by combining the original spectrogram with its delta and delta-delta spectrogram (Liu et al., 2021f). In addition to the STFT, other processing methods have been suggested to generate a spectrogram. For instance, the *Constant Q Transform* (CQT) (Kuzin et al., 2022; Chen et al., 2022b), which increases the time resolution at higher frequencies (Brown, 1991), and the wavelet transform (Chen et al., 2019a).

3.1.2. Low-frequency analysis and recording & detection of envelope modulation on noise

Within the specialization of traditional manual SR-N analysis, the audio is represented using the *Low-Frequency Analysis and Recording* (LOFAR) and *Detection of Envelope Modulation on Noise* (DEMON) processing algorithms. From the DEMON representation of the underwater audio, the propeller characteristics of the ship can be extracted. An optimized processing algorithm for DEMON spectrum generation was proposed to differentiate divers from SR-N (Slamnoiu et al., 2016). Another study proposed a DEMON-based feature extraction method to represent the SR-N audio for a CNN for final SR-N recognition (Bach et al., 2021). The LOFAR spectrum is a broadband representation to estimate the vibration noise of the ship's machinery (De Moura et al., 2011). It has been presented in seven different frequency ranges for target detection and recognition (Aksüren and Hocaoğlu, 2022). The recognition performance of multiple ship types has been evaluated, showing a discriminative performance. The discriminative ability of the DEMON algorithm and the LOFAR algorithm has been compared using three types of CNN networks (Wu et al., 2020). It was concluded that the LOFAR spectrogram has the best ability to discriminate, and a modified LeNet has the best recognition performance.

3.1.3. Cepstral coefficients

Another type of time–frequency representation of audio is cepstral coefficients. They are an extensively utilized representation of SR-N audio for ML applications. A prevalent type of coefficient in SR-N representation is called MFCC (*Mel Frequency Cepstral Coefficients*) (Zeng et al., 2013; Yao et al., 2023; Tong et al., 2020). These coefficients originate from speech recognition applications, showing promising results. The generation of MFCC starts by applying the *Discrete Fourier Transform* (DFT) on the raw audio. Next, the log transformation and Mel scale are applied. Finally, the signal is transformed by the *Discrete Cosine Transform* to generate the ultimate coefficients. Besides the conventional MFCCs, another type of coefficient based on an auditory filter and cubic-log compression has been suggested (Wu et al., 2014). Compared to the Mel-scale, these coefficients should give a better acoustic representation. A variant of the commonly suggested MFCC is called the *Power-Normalized Cepstral Coeffiecients* (PNCC). It has been stated that these coefficients are more robust to noise and, therefore, more suitable for automatic SR-N recognition than MFCCs (Wang et al., 2019a).

3.1.4. Filter banks

A widespread method to extract features from signals is the application of filter banks. These are a set of band-pass filters that separate the signal into multiple components. Several variants of filter banks are developed, like the gammatone filter bank. The design of these filters is based on a cochlear model (Slaney et al., 1993) and universally applied to extract informative features from SR-N. A combination of the gammatone filter bank and the Hilbert–Huang transform (see Section 3.2.2) is proposed to extract features from SR-N (Zeng and Wang, 2014). Other modifications of the gammatone filter bank have been suggested to optimize the features for SR-N information retrieval (Ma et al., 2021;

Table 2
Overview of the developments to extract informative features from SR-N.

Feature extraction	Technique	Description	Citation
Audio Representations	Spectrograms	Short Time Fourier Transformation and Mel-Spectrograms	Pan et al. (2021), Liu et al. (2021f)[a]
	LOFAR & DEMON	LOFAR and DEMON processing	Bach et al. (2021), Aksüren and Hocaoğlu (2022)[a]
	Cepstral Coefficients	MFCC and PNCC	Wu et al. (2014), Wang et al. (2019a)[a]
	Filter banks	Gammatone Filter banks and Mel filter banks	Lian and Wu (2022), Wu et al. (2023)[a]
	Combinations	Combinations of audio representations by feature fusion	Kuzin et al. (2022), Chen et al. (2022b)[a]
Decompositions & Entropy	Entropy as Complexity Measure	Diverse Entropy measures to estimate the complexity of SR-N	Li et al. (2022c), Wang et al. (2022a)[a]
	Wavelet decomposition & Mode decompositions	Decomposition algorithms to decompose SN-R in time-domain	Li et al. (2022g), Chen et al. (2023)[a]
	Hilbert–Huang Transform	Decomposed SR-N into IMFs and obtain the Hilbert spectrum	Ju et al. (2020), Yan et al. (2017)
ML based Feature Extraction	Distance based	increase inter-class feature distance and minimize the intra-class distance of SR-N	Li et al. (2020a)
	Sparse Bayesian Learning	extract discriminative features from time–frequency representations of SR-N	Zeng et al. (2020b)
	Deep Learning	MLP for signal processing to automatically extract features	Li et al. (2023c)

[a] : for more citations, see text.

Lian and Wu, 2022). Besides gammatone filters, other variants of filter banks have also been proposed, like the Mel-filter bank. This filter bank consists of triangular filters based on the Mel scale, representing the same features as the Mel spectrogram. The Mel scale is developed to represent the human perception of sound. Inspired by Mel-filter banks, variants of this method have been suggested for eventual underwater target recognition (Wu et al., 2023).

3.1.5. Combinations

Beyond the previously discussed methods, numerous combinations of feature extraction methods have been proposed to recognize SR-N automatically. For example, a multi-scale spectral feature set has been presented (Jiang et al., 2020), where multiple features were derived from the time domain at different detail levels. Another set of features was created by combining *Gammatone Frequency Cepstral Coefficients* (GFCC), log-Mel spectrogram, Chroma features, spectral bandwidth, spectral centroid, and the *Constant Q Transform* (CQT) (Kuzin et al., 2022; Chen et al., 2022b). The CQT is closely related to the Fast Fourier Transformation, but it increases the time resolution at higher frequencies (Brown, 1991). This combination should discriminate between several types of SR-N, even in the presence of noise.

3.2. Decompositions and entropy

In addition to the different types of audio representations, audio decomposition can be exploited as a feature extraction method for SR-N. These methods can decompose the audio into multiple components, called *Intrinsic Mode Functions* (IMFs). These decompositions lack a physical interpretation, and therefore they are combined with an entropy measure. Such a measure defines the uncertainty associated with the underlying random process and compresses the information load (Uruba, 2019). The discriminating performance of these measures is evaluated using simple machine learning classifiers, like *k-Nearest Neighbors* (kNN).

3.2.1. Entropy

As mentioned above, an entropy measure can be combined with a decomposition technique to extract informative features from SR-N. This section describes multiple types of entropy measures proposed to discriminate SR-N of different types of ships.

Dispersion entropy. The complexity and irregularity of time series is measured by dispersion entropy. This measure is sensitive to variations in frequency, amplitude, and even time series' bandwidth (Li et al., 2022c). Unfortunately, this measure suffers from a long computation time (Xie et al., 2023a).

Permutation entropy. The randomness and dynamic changes of a time series are measured by permutation entropy. Moreover, this measure has a low computation time since it compares neighboring values (Xie et al., 2023a). However, permutation entropy is not amplitude aware and is single scale. These are both considerable disadvantages of this measure.

Multi-scale analysis. Some of these entropy measures are combined with a multi-scale analysis. This analysis estimates the complexity of the data at different scales (Chen et al., 2019b). Multi-scale reverse dispersion entropy is a variant of dispersion entropy and has been suggested multiple times as a feature extraction method for SR-N (Li et al., 2021d,c,b). Even with a simple kNN, this method shows satisfactory results. Furthermore, a multi-scale permutation entropy has been suggested to overcome the limitation of the permutation entropy measure (Li et al., 2021a; Wang et al., 2022a). The classification performance of this modified permutation entropy was compared with other adjustments of permutation entropy. This comparison showed that the multi-scale permutation entropy has the highest recognition rate.

Hierarchical entropy. The complexity of a time series is measured by hierarchical entropy (Jiang et al., 2011). It considers both low and high-frequency information and has therefore been presented as a feature extractor for SR-N (Li et al., 2019d; Chen et al., 2018). This extraction measure is compared to multi-scale sample entropy, using

a probabilistic neural network. This comparison concluded that the hierarchical entropy outperforms the multi-scale sample entropy (Chen et al., 2018).

Combinations. To describe the SR-N in more detail and to make the entropy measures more robust to the background noise, some studies proposed a combination of multiple entropy measures (Siddagangaiah et al., 2016; Li et al., 2022b; Xiao, 2022).

3.2.2. Decompositions

The entropy measures were combined with multiple types of decomposition techniques to describe the SR-N. Various types of decompositions have been suggested to extract useful features. This section describes the wavelet packet decomposition and multiple types of mode decompositions that were utilized in SR-N feature extraction techniques.

Wavelet packet. The traditional *Fast Fourier Transform* (FFT) is not satisfactory for the analysis of non-stationary, non-Gaussian, and non-linear signals (Frei and Osorio, 2007). To overcome these limitations, an alternative decomposition is presented called *Wavelet Packet Decomposition* (WPD). This decomposition approximates the *Discrete Wavelet Transform* (DWT) using filter banks. The performance of the WPD on SR-N has been compared with filter banks, wavelet packet component energy, and MFCC (Ren et al., 2019). The experimental comparisons imply better recognition accuracy with wavelet decomposition component spectrum than the other methods. Another study combined WPD with energy entropy to classify SR-N using a simple kNN (Li et al., 2022g). Their proposed method showed superiority. compared with other methods.

Empirical mode decomposition. The *Empirical Mode Decomposition* (EMD) decomposes the original signal into individual IMFs. The EMD is similar to FFT, but FFT transposes the original signal to the frequency domain while EMD remains in the time domain after decomposition. Moreover, FFT assumes that the signals comprise multiple simple sine waves. EMD overcomes this limitation by generating data-based IMFs. Therefore, EMD has been effectively employed in underwater acoustic recordings to extract informative features from SR-N (Niu et al., 2018; Li et al., 2016)(Li et al., 2016; Wang et al., 2019b). Unfortunately, EMD suffers from a long computation time, a large recovery error, and mode mixing. Mode mixing, in this case, means that a single part of information can be separated over multiple IMFs. To reduce these limitations, an adjusted EMD called *Selective Noise-assisted EMD* has been presented (Niu et al., 2018).

Ensemble of empirical mode decomposition. As previously stated, EMD is sensitive to noise and may suffer from mode mixing. This makes the physical meaning of IMFs hard to interpret. This problem has been taken into account in the *Ensemble of Empirical Mode Decomposition* (EEMD) (Wu and Huang, 2009). It defines IMF components by the mean of an ensemble of trials, and each trial consists of a signal with added white noise. In SR-N, EEMD has been implemented multiple times for feature extraction (Chen et al., 2023). This decomposition has been combined with multiply entropy measures like slope entropy (Li et al., 2022h), permutation entropy (Li and Li, 2018), or energy entropy (Li et al., 2019a). However, they do not consider noise reduction, which makes them less reliable under noisy conditions. Therefore, an *improved complementary ensemble empirical mode decomposition* with adaptive noise has been proposed (Chen et al., 2019b). After decomposition, permutation entropy is calculated from the signal dominant IMFs and weighted by its normalized mutual information. Their experiments indicate a higher performance of their proposed method than other entropy-based feature extraction techniques.

Variational mode decomposition. The *variational mode decomposition* (VMD) overcomes the mode mixing problem of EMD (Dragomiretskiy and Zosso, 2013). The proposed method looks for an optimal ensemble of IMFs, with their corresponding center frequency. This method shows promising results in dealing with noise in signals. This makes VMD a useful feature method for SR-N. This decomposition has commonly been combined with a type of permutation entropy (Li et al., 2017; Xie et al., 2020a,b, 2021; Li et al., 2022d,f; Zare and Nouri, 2023; Zhang et al., 2020a). Some methods added more information about the IMFs, like the normalized maximal information coefficient (Xie et al., 2020b) or a correlation coefficient (Xie et al., 2020a, 2021). Most of these methods are evaluated using an SVM classifier to recognize multiple ship types with high accuracy. As previously stated, permutation entropy is not amplitude-aware. To cope with the loss of information, VMD was combined with slope entropy (Li et al., 2022i), a complexity measure that considers the amplitude information. However, VMD and slope entropy need parameter selection. This selection can be optimized using an optimizer (Li et al., 2023d; Yi and Tian, 2022). After optimization, feature selection is applied to extract the most informative IMF combinations. This approach has been shown to result in a high recognition rate.

Intrinsic time-scale decomposition. The *intrinsic time-scale decomposition* (ITD) is a non-linear method for time–frequency representations of signals. It was developed to overcome the limitations of FFT and EMD (Frei and Osorio, 2007). The method decomposes the input signal into 'proper rotation' components while preserving temporal information. It can be applied in real-time, which is a major benefit of this method. This decomposition has been combined with multiply entropy measures as a feature extraction method for SR-N (Li et al., 2019b,c; Wang and Chen, 2019).

Hilbert-huang transform. Next to the decomposition techniques, SR-N has been presented in the time–frequency domain as a feature using the so-called*Hilbert–Huang transform* (HHT). This transform is a signal processing technique that applies to non-stationary and non-linear signals (Chaitanya et al., 2021). It consists of two steps, starting with decomposing the signal using EMD. Next, the Hilbert Spectral Analysis is applied to the IMFs to obtain the final Hilbert spectrum (see Fig. 4). This type of audio representation has been suggested as an informative feature for SR-N (Ju et al., 2020; Yan et al., 2017).

3.3. Machine learning-based feature extraction

Besides traditional feature extraction based on conventional audio processing techniques, ML models can be applied to automatically extract features from SR-N. These features are utilized to train the following ML model application. Some traditional methods have been suggested, like a distance-based method (Li et al., 2020a) and the application of Bayesian learning (Zeng et al., 2020b). Alongside these traditional ML-based feature extraction methods, numerous deep-learning techniques have been suggested. A *Multi-Layer Perceptron* (MLP)-based method utilizes hard-parameter sharing for optimal feature extraction (Li et al., 2023c). Likewise, a *Neural Network* scheme has been suggested to extract correlation-deep features from the raw audio.

3.4. Summary

This section described various feature extraction methods that have been suggested for SR-N data. However, these approaches are constrained by their underlying assumptions. For example, DFT assumes signal stationarity, and methods like EMD assume that the signal can be accurately represented by IMFs. Unfortunately, the assumptions made by these methods may not always align with the characteristics of SR-N recordings. In alternative research approaches, some studies attempted to directly translate speech-based feature extraction methods to SR-N. However, the distinctive characteristics of SR-N, divergent from those

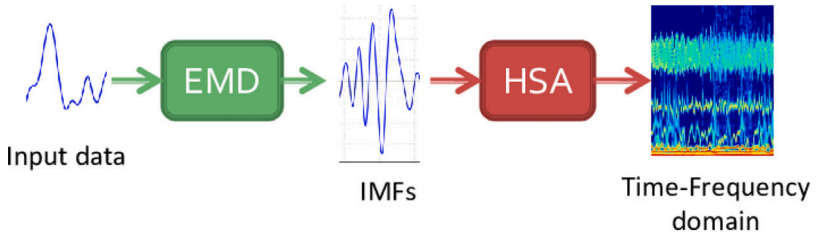

Fig. 4. Overview of the Hilbert–Huang Transform (Waskito et al., 2010).

Table 3
Overview of augmentation techniques to enhance SR-N datasets.

Augmentation	Technique	Description	Citation
Traditional Methods	Add pitch variation	Increase sample diversity by implementing a pitch shifting method	Yuanchao et al. (2023)
	Ray propagation model	Increase the number of samples by creating new samples with an underwater sound propagation model	Liu et al. (2021d)
SpecAugment	Spectrogram augmentation	Add time and frequency masks on spectrograms	Park et al. (2019), Li et al. (2022k)
Generative Adversarial Network	GAN	Apply a simple GAN using different audio representations	Liu et al. (2019b), Jin et al. (2020)
	InfoGAN	Generate underwater acoustic spectrograms	Yang et al. (2023a)
	TangGAN	Modified GAN architecture specialized for underwater acoustics	Pfau (2020)
	DCGAN	GAN model containing convolutional layers	Xie et al. (2023a), Yao et al. (2023)
	cDCGAN	Additional label information during the training process of a DCGAN	Luo et al. (2021b), Tian et al. (2021)[a]

[a] : for more citations, see text.

of speech, make such direct translations suboptimal for representing this type of audio. Conversely, a subset of studies proposed an ML-based feature extraction method directly from raw audio. These methods do not rely on these assumptions, which shows the potential of ML-based methods. Despite this potential, it is essential to recognize that this area of research is currently limited and necessitates further development.

4. Augmentation

Due to the complexity of SR-N analysis, ML models with many parameters have been developed. To reduce the chance of overfitting these big models during training, they need a big and diverse dataset. This diversity of the data is limited by the frequency stability of the harmonic line spectra of SR-N. Furthermore, only a limited amount of labeled SR-N data is available. To overcome these problems, several augmentation techniques have been published to enhance the training set. One such technique adds variation in the pitch of SR-N (Yuanchao et al., 2023). Also, physics-based underwater sound propagation models have been applied to generate more data samples (Liu et al., 2021d). An overview of augmentation techniques to enhance SR-N signals is provided in Table 3.

4.1. Spectrogram augmentation

A specialized technique to augment spectrograms, called SpecAugment, is presented in Park et al. (2019) (see Fig. 5). The spectrograms are augmented by applying frequency masks and time masks. This method has been suggested to enhance SR-N data samples (Li et al., 2022k; Hong et al., 2021b,a). Due to its simplicity, it is easy to implement. However, due to the complexity of the ocean environment, simplistic augmentation methods generate severely deviated synthetic data samples. This deviation may introduce a bias during the training process. To overcome this limitation, an updated spectrogram augmentation method is presented in Xu et al. (2023). This method uses smoothness-inducing regularization followed by local masking and replication is suggested to augment SR-N datasets.

4.2. Generative adversarial networks

Another suggested augmentation technique to augment SR-N data samples is the implementation of *Generative Adversarial Networks* (GANs). A GAN can generate underwater samples from gray-scale spectrum images (Liu et al., 2019b). The generation performance is evaluated using an independent classification network. Likewise, another GAN model was applied to expand the number of data samples by generating spectrograms instead of gray-scale images (Jin et al., 2020; Yang et al., 2023a). Besides spectrograms, MFCCs have also been suggested as input features for a *Deep Convolutional GAN* (DCGAN) (Yao et al., 2023). This same type of GAN has been applied to enhance data samples using wavelet time–frequency graphs as input features (Gao et al., 2020). Notably, all previously presented methods still apply manual feature extraction. This may limit the eventual augmentation performance. To overcome this, a one-dimensional DCGAN is applied directly on the audio waveform to enhance the dataset (Xie et al., 2023a). Unfortunately, simple GANs are not sufficient to capture complex underwater sounds (Luo et al., 2021b; Jiang et al., 2022). The acoustic data generated by GANs contain artifacts that result in machinery-like sounds (Thiem, 2020). To reduce this effect, another

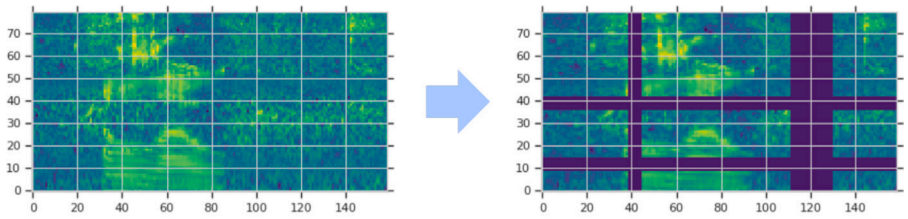

Fig. 5. Visualization of SpecAugment (Park et al., 2019).

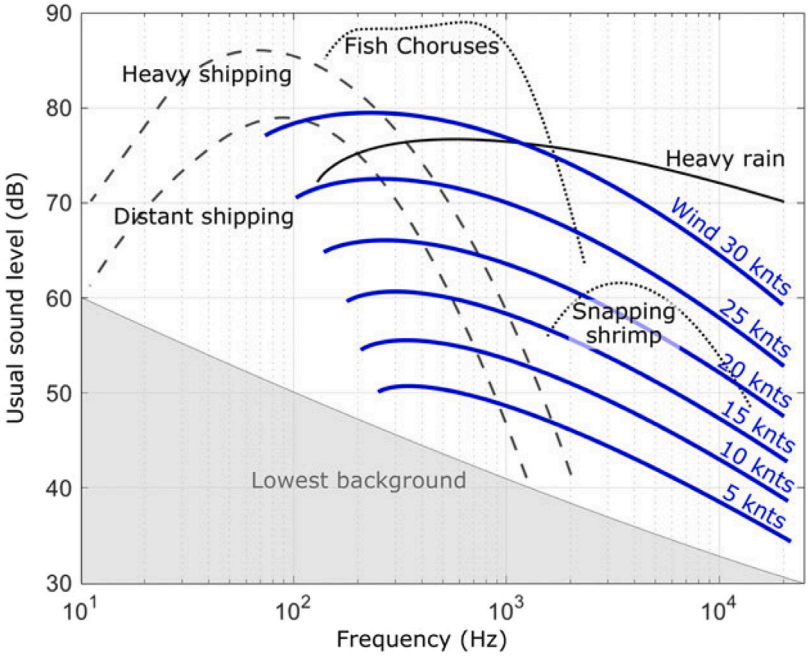

Fig. 6. Typical contributions to underwater ambient noise in the open ocean (Cauchy et al., 2018).

modified version of a GAN, named TangGAN, is proposed to generate underwater sounds (Thiem, 2020) and to enhance SR-N data (Pfau, 2020). Another study suggested incorporating label information during training to optimize the generator (Luo et al., 2021b; Tian et al., 2021).

4.3. Summary

Due to the varying and complex ocean environment, it is challenging to augment SR-N data. Simplistic techniques like SpecAugment and GANs are insufficient. Both methods suffer from multiple artifacts, which introduces a bias during the training of the following ML model. Luckily, several modifications of the GAN have been suggested to overcome these artifacts. This will preserve the quality of the augmented data. Eventually, this will lead to ML models with better performance in automatic SR-N analysis.

5. Denoising

So far, multiple feature extraction methods, types of input data, and how to enhance the data have been discussed. However, all of these methods suffer from the highly variable background noise. This background noise, referred to as ambient noise, lowers the eventual SNR (see Eq. (1)). This ambient noise is composed of numerous sources, like the weather and marine life (see Fig. 6). Eventually, the presence of ambient noise will weaken the performance of the subsequent ML application. For this reason, reducing this background noise during preprocessing of the data is important. Multiple techniques have been proposed to separate the background noise from SR-N, an overview of these techniques is given in Table 4.

5.1. Decompositions and filter banks

The application of decompositions has been suggested to distinguish the background noise from the original SR-N signal. Various decompositions techniques have been proposed, like optimization decomposition (Li et al., 2023a), a modified version of EEMD (Li et al., 2019f), and VMD (Yang et al., 2020; Li et al., 2022l; Ma et al., 2023; Fang et al., 2023). After the decomposition of the input signal, the generated IMFs were divided into three categories: noise IMF, noisy IMF, and signal IMF (Ma et al., 2023). Noise IMFs were discarded, and the noisy IMFs were denoised using Wavelet soft threshold. The denoised IMFs and signal IMFs were combined to reconstruct the original signal. The reconstructed signal is clearer after denoising in different simulated marine environments with variable SNR levels. Besides the discussed decompositions, the Gammatone Filter bank and dyadic discrete wavelet transform have been suggested to remove the background noise (Sonz and Zhang, 2021). The performance was tested with a simple kNN, outperforming the conventional MFCCs.

5.2. Deep learning-based denoising algorithms

Apart from the previously mentioned techniques for denoising, various Deep Learning techniques have been presented to denoise SR-N automatically. An overview of these methods is described in this section.

Table 4
Overview of denoising techniques applied to underwater acoustic data samples.

Denoising	Technique	Description	Citation
Decompositions	Optimization Decomposition	Secondary optimization decomposition model	Li et al. (2023a)
	Ensemble of Empirical Mode Decomposition	Modified EEMD	Li et al. (2019f)
	Variational Mode Decomposition	Mutual Information VMD	Yang et al. (2020)
		Snake Optimization VMD	Li et al. (2022l)
		VMD with IMF categorization	Ma et al. (2023)
Filter bank	Gammatone Filter bank	Gammatone Discrete Wavelet Coefficient	Sonz and Zhang (2021)
Deep Learning	Recurrent Neural Network	Bi-directional LSTM, Dual-Path RNN	Zhang et al. (2021b), Song et al. (2022)
	Convolution-based Models	WaveN2N, Generalized Channel-Invariant Network	Koh et al. (2020), Zeng et al. (2020a)
	Autoencoders	Denoising Autoencoder, Convolutional Autoencoder	Zhou and Yang (2020), Song et al. (2023c)[a]
	NAFSA-Net	Fullband-Subband attention blocks	Zhou et al. (2023)

[a] : for more citations, see text.

5.2.1. Recurrent neural networks

A *Long Short-Term Memory* (LSTM) model is a special type of a *Recurrent Neural Network* (RNN) that has been suggested to denoise SR-N recordings. This model type is specially designed to handle sequential data by taking long-term dependencies into account (Hochreiter and Schmidhuber, 1997). These characteristics make the LSTM model applicable for time series data, like SR-N. A deep bidirectional LSTM, with STFT magnitude features, is utilized to estimate the amplitude mask of an acoustic target (Zhang et al., 2021b). The completed pipeline is evaluated using the ShipsEar dataset, where the classifier achieved recognition accuracy of 60.67% on the denoised dataset. This is a reduction of 34% compared to the recognition accuracy on the original dataset. Alongside the LSTM, a dual-path RNN has been proposed to denoise SR-N recordings (Song et al., 2023a). Here, it was shown that this model could improve the SRN ratio by 12.02 dB and 9.48 dB, respectively.

5.2.2. Convolution-based model structures

A convolution-based network, called WaveN2N, is presented to denoise multi-channels array data (Koh et al., 2020). This technique is based on the computer vision method called Noise2Noise (Lehtinen et al., 2018), which has been shown to restore corrupted images from solely corrupted samples. WaveN2N is constructed in a self-supervised setting. The results were visually evaluated, showing more structure in the denoised spectrograms, implying noise level reduction. Apart from optimizing the separation between signal and noise, the background noise can be modeled on its own (Zeng et al., 2020a). A *Generalized Channel-Invariant Network* has been suggested to model the background noise of different marine environments, outperforming other *Convolutional Neural Network* (CNN)-based models.

5.2.3. Autoencoders

Beyond the convolution-based neural networks, a variety of autoencoders have been suggested to denoise SR-N recordings. A special type of autoencoder is the denoising autoencoder. These types of autoencoders try to reconstruct the undestroyed data using partially destroyed data samples as input. Similarly, a practical denoising and recognition method has been proposed (Zhou and Yang, 2020). First, the data was split into target and noise data by generating multi-images between marine noise and target signal using correlation and a dropout process. The denoising features were acquired by convolutional denoising autoencoder, which has been trained on the segmented multi-images. The denoising performance was evaluated using *fuzzy C-means* method. Here, the proposed denoising method showed the least overlap between noise and SR-N, compared to HHT, STFT, and MFCC. Another extension to the regular denoising autoencoder called a bidirectional autoencoder, has been proposed to denoise underwater audio (Dong et al., 2022). Due to a lack of clean underwater sound data, pseudo-clean recordings were created by denoising the SR-N recordings from ShipsEar using decompositions. This proposed method was evaluated by recognizing the ship type by an *Support Vector Machine* (SVM). The denoising of the input data before classification resulted in a recognition accuracy increase of 9%. Besides the denoising autoencoders, a convolution-based autoencoder has been suggested for underwater acoustic noise reduction (Song et al., 2023c). Here, it is stated that this network can extract the structural and local information of the spectrum, resulting in an SNR increase of 10.02 dB and 9.5 dB, respectively.

5.2.4. NAFSA-Net

The final denoising method discussed within this section is called a noise-aware deep learning model with fullband-subband attention network (NAFSA-Net) (Zhou et al., 2023). The NAFSA-Net pipeline starts with an encoder, utilized for feature extraction from the raw audio. Subsequently, a noise subnet and a target subnet are designed to adopt stacked fullband-subband attention blocks. Employment of fullband-subband attention blocks has been presented in speech enhancement (Chen et al., 2022d). These attention blocks extract global and local dependencies to describe the noise and target characteristics. An interaction module is designed to transmit auxiliary information between the sub-nets.

5.3. Summary

Generally, the suggested denoising methods for SR-N can be categorized into two groups: those relying on manually extracted features and those employing deep learning methods. The manually extracted feature methods are constrained by their assumptions that may not meet the SR-N signals. Even though decompositions are presented as an effective denoising method, there is no solid standard to differentiate between signal-dominated IMFs and noise-dominated IMFs (Dong et al., 2022). This leads to subjective influences in the denoising process of SR-N. In addition to these traditional methods, a combination of an encoder and decoder model and the denoising autoencoder are introduced. Both methods are completely data-based and highlight the potential of data-driven methods for denoising SR-N recordings. The denoising performance of these methods is evaluated by their recognition accuracy due to a lack of ground truth. For this reason, the true denoising performance of the data-driven methods is still unexplored.

6. Detection

Upon preprocessing SR-N recordings, the development of ML algorithms becomes instrumental for the automation of SR-N analysis. This section focuses on the automatic detection of SR-N, specifically detecting a ship once it comes within the range of a deployed hydrophone. Automatic detection is critical for harbor security or further analysis of the ship's signature. Detecting SR-N poses a considerable challenge due

Table 5
Overview of detection techniques applied to detect SR-N automatically.

Detection	Technique	Description	Citation
Traditional Methods	SVM	A comparative study examining the performance of multiple traditional ML methods	Prasad and Gurugopinath (2022)
Deep Learning	CNN	CRNN for fish and engine sound detection	Kammegne et al. (2023)
		Real-time detection and localization	Scherrer et al. (2022)
		Scaled spectrograms as input features	Park and Kim (2022)

to the low SNR of SR-N and the highly dynamic ocean environment. For these reasons, traditional detection methods fall short (Chen and Zhang, 2011). Consequently, several ML models have been developed to detect SR-N automatically and accurately. An overview of these ML models is given within Table 5.

The detection performance of diverse traditional supervised ML techniques has been compared to provide a comprehensive analysis (Prasad and Gurugopinath, 2022). This comparison showed that the SVM outperforms other supervised methods. Apart from these traditional methods, a CNN is frequently suggested for the automatic detection of SR-N. Some of these suggested methods are the automatic detection of fish calls and motor engine sounds by a *convolutional recurrent neural network* (Kammegne et al., 2023) and the real-time detection and navigation direction using two bottom-moored hydrophones (Scherrer et al., 2022). This real-time detection was realized by a shallow LeNet-based CNN, yielding a *True Positive Rate* of 0.92 on a private dataset. In order to conduct a comprehensive assessment of the CNN architecture's performance, it is compared with other deep learning methods (Prasad and Gurugopinath, 2023). The automatic detection capabilities of a CNN, a *gate recurrent unit* (GRU), LSTM, a basic Deep Neural Network, and an RNN are assessed, utilizing diverse data sources for a comprehensive evaluation. Within this comparative study, the proposed CNN and LSTM outperform any other deep learning method.

7. Localization

It is not just the detection of SR-N that proves to be difficult, equally challenging is the localization of the source of SR-N. During this localization, the range and depth of the acoustic source are determined using a single sensor. To do this, several conventional methods are utilized, which are based on inversion techniques. A common method is the well-known matched field processing. However, the accuracy of these inversion techniques is affected by the highly fluctuating environment (Lefort et al., 2017). For this reason, several ML techniques are proposed to locate SR-N automatically. This poses challenges too, due to the lack of labeled data. The available labeled datasets only take the ship identity into account and do not consider accurate localization labeling. Nevertheless, several ML techniques have been developed to cope with the automatic localization of SR-N. An overview of these ML techniques is given within Table 6.

7.1. Traditional machine learning models

A limited number of traditional ML models are developed to localize SR-N automatically. For instance, Lefort et al. (2017) proposed a kernel regression to perform automatic localization of SR-N and outperform original inversion techniques. The potential of machine learning has been further explored in Niu et al. (2017b) and Hu et al. (2021a). Both studies preprocess the raw measurements of a vertical array into a normalized sample covariance matrix. This matrix has been employed as the input for several machine learning methods: feed-forward neural network, SVM, kernel-Extreme Learning Machine, and Random Forest. Here, the SVM outperformed other methods resulting in a *Mean Absolute Percentage Error* (MAPE) of 2% (Niu et al., 2017b). However, the performance of these traditional methods decreases once the SNR decreases.

7.2. Deep learning techniques

Recently, more deep-learning techniques have been suggested for SR-N localization than traditional ML techniques. The drawback of Deep Learning methods over traditional techniques is the need for a large amount of labeled data. Unfortunately, this is not publicly available. To cope with this problem, most of the techniques are developed using simulated data.

7.2.1. Feed-forward neural network

The first Deep Learning method that has been suggested for automatic SR-N localization is the application of a *Feed-Forward Neural Network* (FNN) (Ozard et al., 1991). An associative FNN without hidden layers was proposed to estimate the source and depth of the target using the data of a vertical array. The same approach is applied in Niu et al. (2017a), where the performance of the FNN is compared with an SVM. Here, it was stated that ML applications have a higher detection range than conventional matched field processing. Evaluating a FNN for acoustic target ranging is difficult due to the lack of annotated data. For this reason, Chi et al. (2019) proposed a fitting-based early stopping method to evaluate an FNN for underwater target localization on a test set without known locations. Their proposed method resulted in a *Mean Squared Error (MSE)* of 0.24 km.

7.2.2. Convolutional neural network

Nowadays, the application of FNN in SR-N localization is replaced by CNNs. These models have proven to be powerful models for computer vision tasks. Hence, this model has extensively been utilized in underwater acoustic scenarios. It has outperformed traditional localization methods (Chen and Schmidt, 2021) and can be used to enhance the reliability of other underwater source localization methods (Xiao et al., 2021b). Occasionally, CNNs are combined with other *Artificial Neural Networks* (ANNs) (Liu et al., 2019a). The performance of a CNN-FNN architecture, with raw audio as input, is compared with an ANN with manually extracted features for underwater source localization (Huang et al., 2018b). The comparison concluded a superior performance of manually extracted features combined with ANN. Despite their initial performance, both models significantly degraded once the ocean environment changed. While it is acknowledged that manual feature extraction methods share this limitation, several methods have been published combining CNNs with manual feature extraction techniques for automatic underwater source localization. Examples include a normalized acoustic matrix (Liu et al., 2020b) and a cepstrogram (Ferguson et al., 2016, 2019; Ferguson, 2021). It has been stated that combining CNNs with cepstograms can even deal with multi-path distortion of the incoming sound signal. A CNN, consisting of Inception blocks, has been suggested to automatically localize SR-N. This proposed method resulted in a *Mean Absolute Error* (MAE) of 0.3 km in range and 12.1 m in depth on a private dataset. Other variations of the CNN architecture have been published to optimize automatic localization even further. One such variation is a ray-based blind deconvolution algorithm (Durofchalk et al., 2021). This approach has been compared with matched field processing and showed similar range accuracy. Moreover, an attention-based CNN is proposed for automatic ship ranging (Xiao et al., 2021a). The attention mechanism visualized the inherent features of concern of ships and the effect of underwater acoustic channels. The examination of how a CNN model processes raw

Table 6
Overview of ML techniques applied to localize SR-N automatically.

Localization	Technique	Description	Citation
Traditional ML models	Kernel methods	Kernel regression, kernel Extreme-Learning machine	Lefort et al. (2017), Hu et al. (2021a)
	SVM and Random Forest	Sample covariance matrix as input features	Niu et al. (2017b)
Deep Learning	FNN	Without hidden layers or with fitting based early stopping	Ozard et al. (1991), Niu et al. (2017a), Chi et al. (2019)
	CNN	Combined with other ANN and trained using either manually extracted features or raw waveform data	Chen and Schmidt (2021), Xiao et al. (2021b),[a]
		50-layer ResNet and 18 layer ResNet	Niu et al. (2019), Lin et al. (2020)
	Other NN	ANN, GRU, LSTM, CDC, and MLP	Yangzhou et al. (2019), Wang and Peng (2018),[a]
	Semi Supervised Learning	A two-step framework, CAE for feature extraction and MLP for automatic localization	Zhu et al. (2020, 2021b)[a]
	Multi-task learning	Simultaneously predict sea bed type and sound source localization	Van Komen et al. (2019, 2020)[a]
ML for DOA estimation	Various Deep Learning methods	ANN, RNN, CNN, FNN	Cao et al. (2019a), Whitaker et al. (2021)[a]

[a] : for more citations, see text.

underwater sound measurements directly is investigated (Herchig et al., 2022). The performance of a CNN trained using either magnitude data (real-valued) or pressure data (complex-valued) has been compared. It was shown that the complex-valued CNN outperformed the real-valued CNN. Unfortunately, when the simulated dataset was generated using more Sound Speed profiles, the performance of the complex-valued CNN decreased. To create a more generalized model to cope with these variations, more diverse training data is needed. Instead of using more training data, a transfer learning approach has been suggested in Ge et al. (2022). Here, a label distribution-guided transfer learning method was presented with an improved performance using only a limited amount of experimental data.

Residual neural network. The original CNN architecture is sensitive to the vanishing or exploding gradient problem. The *Residual Neural Network* (ResNet) is proposed to overcome this. Such a network consists of stacked residual blocks with skip connections. These connections connect the activation of a previous layer to a future layer. This makes this architecture suitable for automatic SR-N localization. A ResNet18 has been suggested to automatically localize SR-N in a shallow-water ocean environment. The model was trained using simulated data generated by KRAKEN and tested on experimental data. This resulted in a MAPE of 1.5% in the range. However, it is uncertain how this model would perform in more complex ocean environments.

7.2.3. Other neural networks

Several other NN architectures have been proposed for automatic SR-N localization. The sample covariance matrix is a widely used feature in source localization (Yangzhou et al., 2019). Therefore, this feature has been used multiple times as the input for a generalized regression NN (Wang and Peng, 2018; Jia et al., 2021; Huang et al., 2018a) and a one-dimensional ANN (Yangzhou et al., 2019). The performance of the one-dimensional ANN has been compared with the conventional matched field processing. However, the generality of the proposed method in other ocean environments is questioned. Since ANNs have shown to be successful in single underwater sound source localization, the application of deep learning methods for multi-source sound localization has also been explored. A GRU network has been proposed in Liu et al. (2021e) for multiple sound source localization without knowing the number of sources. Similarly, Huang et al. (2019) suggested an LSTM model for this purpose. The potential of the LSTM in automatic underwater sound source localization has already been shown in Qin et al. (2020). An LSTM network, with features based on the covariance matrix, was applied as a supervised regression problem. This proposed method decreased the MAPE by 40% in low SNR simulated data samples compared to conventional methods. Solving supervised problems with deep learning techniques often requires a large amount of labeled data. Due to the lack of publicly available annotations, Zhu et al. (2021a, 2022) proposed a two-step self-supervised learning scheme for automatic underwater source localization using *contrastive predictive coding*. The encoder of the extractor is isolated and coupled to an MLP while the parameters of the encoder remain frozen. The MLP is optimized using a small labeled dataset. This method is compared with an autoencoder and a complete supervised learning scheme using real data. Here, it was stated that the proposed self-supervised model outperforms the other methods, resulting in an MSE of 0.11 km (Zhu et al., 2022).

7.2.4. Semi-supervised learning

Similar to self-supervised learning, the proposal of semi-supervised learning emerges as a solution to address the scarcity of publicly available labeled data. This learning technique combines supervised and unsupervised learning, aiming to use labeled and unlabeled datasets. A two-step framework has been presented for automatic localization of underwater sound sources (Zhu et al., 2021b). First, a *Convolutional AutoEncoder* is trained in an unsupervised manner. Second, the latent features are utilized as input for an MLP, which is trained to estimate the source range in a supervised manner using a limited amount of labeled data. This method resulted in an MSE between 0.4 km and 0.48 km. To reduce the training time of this framework, a feature selection method has been proposed in Zhu et al. (2020, 2021b) based on Principle Component Regression before the two-step mechanism (results are visualized in Fig. 7). This proposed framework is more robust to unseen data and outperforms supervised methods when the number of labeled data samples decreases. An integrated feature selection method has been proposed in Jin et al. (2022) and Li et al. (2023b). Both studies applied a self-attention mechanism within the *Convolutional Autoencoder* for picking more useful features. This method resulted in an MAE of 0.06 km.

7.2.5. Multi-task learning

An alternative approach to enhance the performance of machine learning techniques for the automatic localization of SR-N is through the implementation of multi-task learning. This learning technique is based on inductive transfer to improve the generalization of the model. It provides this by learning multiple tasks in parallel with a shared representation (Caruana, 1997). This may result in higher prediction accuracy than models trained individually per task. This technique has been suggested to predict the range and depth of an underwater

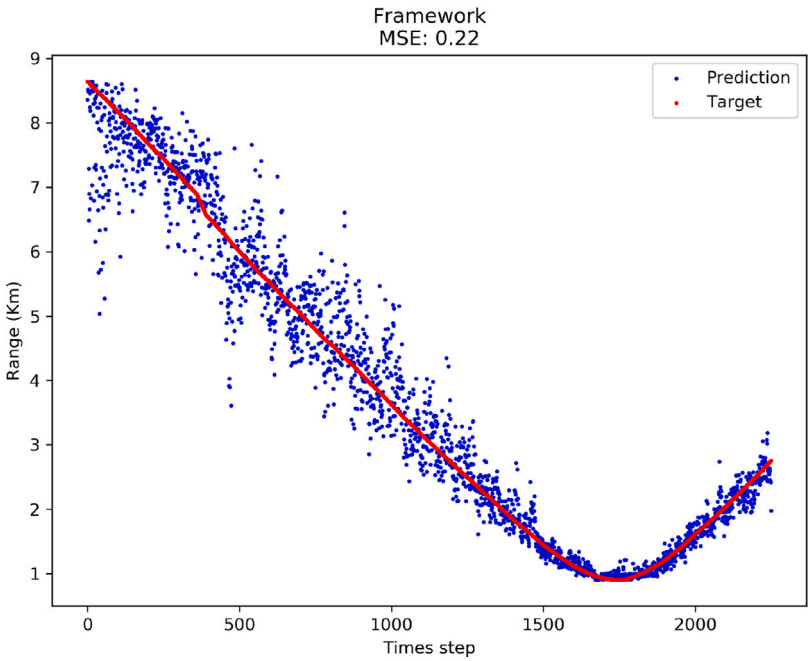

Fig. 7. Range estimation results from Zhu et al. (2021b) on Swellex-96 experiment.

acoustic source simultaneously (Liu et al., 2020a). Subsequently, this resulted in an MAE of 1.3 km in range and 9.5 m in depth on a private dataset. Additionally, various techniques have been presented showing the potential of simultaneous training of models for underwater sound source localization and sea bed classification (Van Komen et al., 2019; van Komen et al., 2019; Van Komen et al., 2020; Neilsen et al., 2021; Van Komen et al., 2021; Liu et al., 2023b). Underwater sound propagation paths are highly influenced by the type of sea bedding. Therefore, combining these tasks in a multi-task learning scheme may improve the final localization estimation accuracy. An FNN with manual feature extraction has been suggested to solve this task (Van Komen et al., 2019). However, the most commonly chosen model architecture is a CNN (van Komen et al., 2019; Van Komen et al., 2020; Neilsen et al., 2021; Liu et al., 2023b). Liu et al. (2023b) optimized the CNN architecture for automatic underwater sound source localization. An interpretable complex CNN based on the Barlett processor was suggested. This model creates a more physically relevant CNN output resulting in successful source localization. Another CNN architecture has been suggested in Van Komen et al. (2021) to predict seabed type, ship speed, and range using spectrograms as input features. Their proposed method achieved a seabed classification accuracy of 99% and a *Root Mean Squared Error* of 0.28 km in the Closest Point of Approach range prediction.

7.3. Machine learning for direction-of-arrival estimation

Thus far, only ML techniques that predict the range or depth of an acoustic source have been discussed. Beyond this, estimating the source azimuth of the target noise is also considered. This is called *Direction of Arrival* (DOA) estimation. This parameter is critical for signal processing of underwater sound, and as such, accurate estimation is necessary. Unfortunately, conventional DOA estimation methods degrade in performance in the complex time-variant ocean environment (Li et al., 2022a). The challenging conditions in this environment, characterized by substantial transmission loss and intense noise interference, result in significant distortion of the original acoustic signal, posing difficulties for accurate DOA estimation (Whitaker et al., 2021; Quan et al., 2021). Machine Learning models have the ability to learn and adapt to this time-variant ocean environment. For this reason, deep learning techniques have been developed. An ANN has been proposed to estimate the DOA using a single vector hydrophone (Cao et al., 2019a). The performance of this simple ANN was compared with the conventional complex sound intensity method, showing similar estimation accuracy. A more extensive comparison has been made in Whitaker et al. (2021). The DOA estimation performance of the frequency-masking average method was compared with a deep and shallow RNN architecture. The results show that the deep RNN outperforms both the shallow RNN and the conventional method and demonstrates the potential of deep neural networks. As previously mentioned, deep learning methods require a large amount of labeled data which is not publicly available. To cope with this problem, a deep transfer learning CNN framework is applied in Cao et al. (2021) where the synthetic data was combined with real data. Their proposed transfer learning technique utilized more accurate DOA estimations compared to a conventional CNN. Apart from transfer learning, other adjustments on a CNN have been proposed to optimize DOA estimation. For instance, Liu et al. (2021a) proposed a two-channel ResNet using both a real-valued channel and an imaginary input channel. Their proposed method outperformed the conventional MUSIC algorithm. It has been shown in various studies that CNNs outperform conventional methods and are able to extract useful features for accurate DOA estimation at different SNR levels (Cao and Ren, 2022; Li et al., 2022a). Likewise, other feature extraction techniques were combined with CNN models to improve DOA estimation. These extraction techniques are based on conventional methods like the wavelet transform (Quan et al., 2021) or local beam patterns (Nie et al., 2023b). Apart from CNN architectures, FNNs have been proposed for multi-target DOA estimation (Niu et al., 2017b; Ozanich et al., 2020). Both two-target and K-target, where K is unknown, problems are solved. Both studies demonstrate the potential of FNN for underwater sound source localization resulting in a mean error of 0.92°.

7.4. Summary

This section described the development of ML techniques in automatic range and depth estimation of the source of SR-N. Multiple Deep Learning methods have been described, where most techniques outperformed the conventional inversion methods. Particularly, the Deep Learning methods that can cope with sequential data show a great localization improvement. Even though these models have shown

Table 7
Overview of recognition ML techniques applied to automatically recognize various SR-N.

Recognition	Technique	Description	Citation
Traditional ML models	kNN	MFCC as input features	Tong et al. (2020)
	HMM	Cepstral Coefficients as input features	Küçükbayrak et al. (2009), Mohammed et al. (2018)[a]
	SVM	Combined with Cepstral coefficients or DL features	Lian et al. (2017), Can (2016)[a]
	SVDD	Combined with auditory based features	Li et al. (2018b)
Deep Learning	MLP & ANN	class-modular MLP, 3-layer ANN, 7-layer ANN	Li et al. (2018a), Axelsson and Rhen (2020)[a]
	RNNs	LSTM and GRU	de Souza et al. (2022), Yang et al. (2022)[a]
	CNN	Shallow 3-layered networks, VGGNet, and LeNet	Khalilabadi (2023), Yin et al. (2020), Wang et al. (2022b)[a]
		ResNet	Chen et al. (2019a), Domingos et al. (2022), Ren et al. (2022)[a]
		Multi Scale ResNet	Tian et al. (2021, 2023b)
		Depthwise Separable Convolutions	Zhang and Ding (2020), Hu et al. (2021b)[a]
		Auditory based	Shen et al. (2018, 2020)[a]
		Attention Module	Li et al. (2022e), Liu et al. (2021c)[a]
		Combinations with other networks	Hu et al. (2018), Liu et al. (2023a)[a]
	Contrastive Learning	supervised SimCLR	Chen et al. (2020)[a]
	Semi-Supervised Learning	Deep Belief Networks	Chen and Xu (2017), Luo and Feng (2020)[a]
		Variational Autoencoders	Satheesh et al. (2021), Bach et al. (2022)
		Stacked Autoencoders	Cao et al. (2019b), Haiyan et al. (2021)[a]
		Convolutional Autoencoders	Chen and Shang (2019), Lingzhi et al. (2023)[a]
	Transformers	Swin Transformer and Audio Spectrogram Transformer	Xu et al. (2022), Li et al. (2022k)
	Multi-target Recognition	Multi-target classification by CNNs	Pfau (2020), Sun and Wang (2022)
Signature Recognition	CNN	Automatic transient detection or tonal detection from LOFAR & MFCC	Tucker and Brown (2005), Park and Jung (2019)[a]
	Autoencoder	Enhance tonal signals instead of suppressing the background	Ju et al. (2022)

[a] : for more citations, see text.

their potential, to train these types of models, you need a large amount of labeled data. Since this is not publicly available, self-supervised learning and semi-supervised methods have been suggested. These types of learning techniques improve the training of the models by exploiting all the available SR-N data. Alongside the range and depth estimation, DOA estimation techniques have been discussed. These techniques have shown that a CNN can learn meaningful representation from the audio and have a robust performance under multiple SNR levels. This is exceptional given the severely complex and varying ocean environment.

8. Recognition

The final component of automated SR-N analysis pertains to the automatic recognition of SR-N. The unique acoustic signature of a ship's noise comprises specific transients and tonals. Transients are short peaks in amplitudes within the time domain, whereas tonals are characterized by a concentrated energy presence at a single frequency within a narrow section of the time–frequency spectrum. The combination of these measures is utilized to manually recognize the ship type based on its produced noise. As previously stated, manual SR-N recognition is a slow and costly process. To automate this process, multiple studies focus on automatically recognizing these signatures or the ship type directly using ML. An overview of these studies is given in Table 7.

8.1. Traditional machine learning models

Several traditional machine learning models have been suggested for automatic SR-N recognition. These traditional methods do not require the same large amount of data as deep learning and, therefore, have been represented for automatic SR-N recognition. However, the training process of these methods starts with manual feature extraction. Several traditional machine learning algorithms have been suggested for final recognition. Earlier studies suggested a *Hidden Markov Model* (HMM) (Yue et al., 2005; Küçükbayrak et al., 2009; Mohammed et al., 2018; Vieira et al., 2020) for automatic SR-N recognition. Other studies suggested implementing a simple kNN combined in conjunction with manually extracted features (Tong et al., 2020) or utilizing an SVM (Sherin and Supriya, 2015; Can, 2016; Lian et al., 2017). Notably, the SVM algorithm has been recurrently recommended and combined with manually extracted features from either the time-domain (Meng et al., 2014), frequency-domain (Leal et al., 2015), and time–frequency domain (Wang and Zeng, 2014). Additionally, the SVM has been combined with various wavelet-based techniques (Wang and Zeng, 2014; Li et al., 2014; Can et al., 2017). These methods outperformed traditional SVM and MFCC combinations with a recognition rate above 90% on private datasets. Related to the SVM, multiple *Support Vector Data Descriptors*, have been proposed to perform automatic recognition (Li et al., 2018b). The output of these classifiers is fused by a majority vote reaching a *True Negative Rate* of over 90% on a private dataset. Overall, traditional methods are still applied for automatic SR-N recognition. However, the late rise in the number of publications about automatic recognition of SR-N is dominated by deep learning applications.

8.2. Deep learning methods

In addition to the traditional machine learning methods, a wide variety of deep learning methods have been applied in automatic SR-N recognition. An overview of these methods is given in this section.

8.2.1. MultiLayer perceptron and shallow artificial neural networks

The application of NN in SR-N has shown to be less sensitive to noise compared to a traditional method (Yang and Chen, 2017). One of the many suggestions is the application of an MLP. This architecture was modified to create a class-modular MLP for automatic passive SR-N recognition (Souza Filho and de Seixas, 2016). Within this study, it was concluded that the design parameters have a significant impact on recognition accuracy. To optimize these parameters, several optimization algorithms have been proposed, like the Chimp Optimization Algorithm (Khishe and Mosavi, 2020), and the Whale Optimization

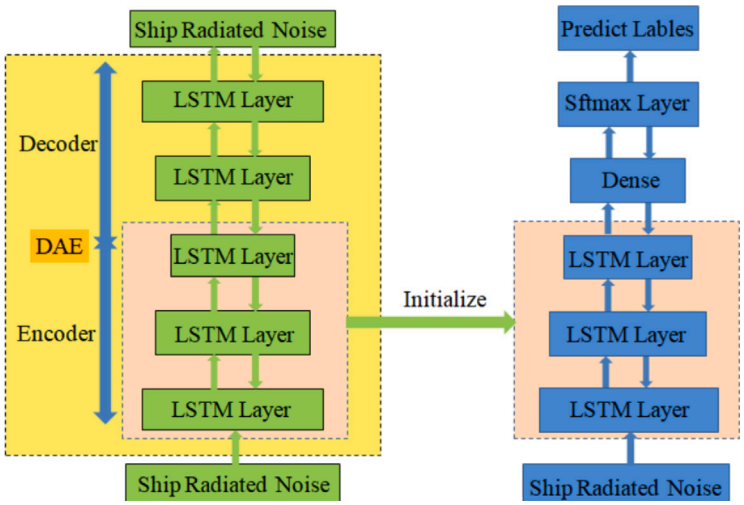

Fig. 8. DLSTM in DAE setting training and recognition. (Yang et al., 2019).

Algorithm (Qiao et al., 2021). Besides the MLP architecture, various studies suggested the application of a 3-layer ANN (Liu et al., 2014; Axelsson and Rhen, 2020; Jiang et al., 2021) or a 7-layer ANN (Li et al., 2018a). The performance of these neural networks still relies on manual feature extraction. These extraction techniques are sensitive to noise and degrade the ultimate recognition performance.

8.2.2. Recurrent neural network

For coping with sequential data, *Recurrent Neural Networks* (RNNs) have been suggested. The SR-N recordings are sequential; therefore, multiple types of RNNs have been employed for automatic SR-N recognition.

Long short-term memory network. The LSTM, a specific type of RNN, is a powerful model. This model is renowned for its capability to manage sequential data by considering long-term dependencies. This characteristic establishes the LSTM as a commonly chosen model for automatic SR-N recognition (Xu and Guo, 2021; Song et al., 2023b), especially with manually extracted features. Conversely, an LSTM incorporating time-domain information as input features was proposed (de Souza et al., 2022). Their results demonstrated a superior recognition rate compared to an MLP. In an effort to minimize dependence on manually extracted features, a bidirectional-LSTM for the automatic recognition of ships without prior feature extraction was introduced (Li et al., 2020b). Their experiments showcase the robust adaptability of the proposed method with a recognition rate above 80% on a private dataset. Additionally, a deep-LSTM has been proposed in a Deep Autoencoder network. This complete network was optimized to reconstruct the original SR-N in an unsupervised manner (Yang et al., 2019). Next, the deep LSTMs were isolated and reused for the final SR-N recognition (see Fig. 8). This framework achieved recognition accuracy of 90% on annotated data originating from the ONC dataset.

Gated recurrent unit. Another type of RNN utilized for automatic SR-N recognition is called a GRU. The architecture of a GRU is similar to an LSTM. However, a GRU has fewer parameters than an LSTM. This makes this model type less powerful and adaptable. On the other hand, it is less prone to overfit and less costly when it comes to training such a model. For these reasons, GRU has been employed for automatic SR-N recognition (Sun et al., 2020). Other studies suggest the combination of GRU with other types of NNs (Yang and Zeng, 2021). For instance, a GRU has been combined with either a 1D-CNN (Ashok and Latha, 2022) or a *Convolutional AutoEncoder* (Yang et al., 2022) showing an excellent recognition rate of over 82%.

8.2.3. Convolutional neural network

In the field of automatic SR-N recognition, CNNs dominate as the primary model architecture. Overall, the original audio is represented in the time–frequency domain, using various techniques. The time–frequency representation is then presented as an image for the CNN input layer. Numerous variations of the CNN architecture have been proposed, including deep linear CNNs as well as wide and shallow networks. Various simple and shallow CNN architectures have been proposed for automatic SR-N recognition (Yue et al., 2017; Applelid and Karlsson, 2019; Premus et al., 2020; Wang and Peng, 2020; Chen et al., 2021; Vaz et al., 2022; Khalilabadi, 2023; Zhang et al., 2021a). Even a shallow three-layered CNN achieved an accuracy between 88% and 95% on the ShipsEar dataset (Chen et al., 2021; Khalilabadi, 2023). The performance of another shallow three-layered CNN has been compared to the more conventional and complex LeNet and VGG architectures (Wu et al., 2018). Here, the shallow network outperformed these more complex architectures. Nevertheless, various complex and well-known CNN architectures have been suggested to accurately recognize SR-N. For instance, MobileNetV2 (de BA Barros and Ebecken, 2022), VGGNet (Choi et al., 2019; Yin et al., 2020), and some variations on the LeNet architecture (Wu et al., 2020; Jin et al., 2020; Bach et al., 2021; Wang et al., 2022b). Again high recognition rates are reported, achieving an accuracy between 90% and 98% on the ShipsEar dataset. Besides these well-known architectures, some modifications have been reported to further optimize the recognition rate. For instance, second-order pooling has been combined with a CNN architecture to capture the temporal correlations of the time–frequency representation (Cao et al., 2018). To exploit the spectral information from the input data, Pan et al. (2021) proposed to train a CNN per frequency band (see Fig. 9). Here, a positional encoding was added to retain the frequency information and a final global classifier fuses the output of the CNN models. Their proposed framework reached a recognition accuracy of 92%. Noteworthy, not all studies have incorporated a time–frequency algorithm as a precursor to automatic recognition. A CNN can be applied directly on the raw audio and learn distinguishable features (Doan et al., 2020; Mishachandar and Vairamuthu, 2021). A dense CNN has been proposed in Doan et al. (2020), within this architecture former feature maps are reused while maintaining low computational cost. Their proposed method achieved an astonishing recognition accuracy of over 98%. The following paragraphs will discuss additional variants of CNNs in automatic SR-N recognition.

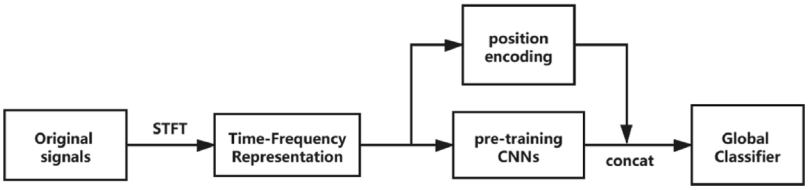

Fig. 9. Visualization of the proposed method in Pan et al. (2021). Here STFT stands for Short Time Fourier Transformation.

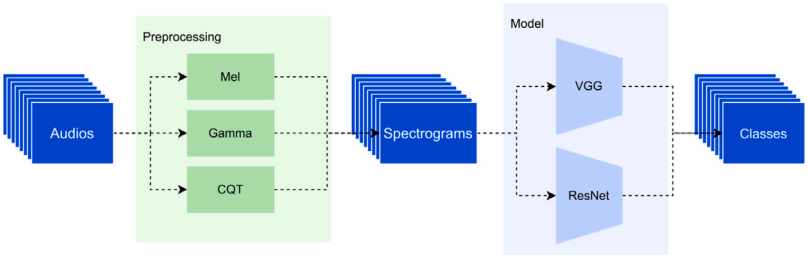

Fig. 10. Classification pipeline for comparing performance ResNet and VGGNet in Domingos et al. (2022).

ResNet. Apart from its recommendation for the automatic localization of SR-N, the ResNet architecture is also put forth for recognition applications (Yang et al., 2023b). Particularly, the ResNet18 has been proposed for automatic SR-N recognition (Hong et al., 2021a; Zhang et al., 2022b; Domingos et al., 2022; Yao et al., 2023). The performance of this architecture has been compared to the VGGNet architecture, outperforming the VGGNet by 14% accuracy on the ONC dataset (Domingos et al., 2022). The comparison pipeline is illustrated in Fig. 10. Apart from the commonly utilized ResNet18 architecture, deeper ResNet models have also been proposed. However, these deeper models require more data. For this reason, transfer learning has been suggested using models pre-trained on ImageNet (Song et al., 2020; Liu et al., 2021b). The impact of the number of hidden layers in ResNet models has been explored in Chen et al. (2019a), where the performance of ResNet models ranging from 50 layers to 152 layers was compared. These models resulted in a recognition accuracy between 93.1% to 95.9% on a private dataset. Noticeably, these suggested approaches involve separate feature extraction before model training, potentially impacting recognition performance. In contrast, an integrated system is proposed in Ren et al. (2022), where Gabor filters are applied to preprocess the incoming audio. The hyperparameters of both the Gabor filters and the ResNet50 model are optimized simultaneously. Their suggested framework reached a recognition accuracy of 80.7% on ShipsEar and 81.4% accuracy on Deepship.

Multi scale residual network. A deep convolutional stack network has been suggested in Tian et al. (2021), where the convolutional layers have been replaced by *Multi Scale Residual Units* (MSRU) (see Fig. 11). These units are designed based on the ResNet architecture. They are composed of multiscale convolutional kernels and a soft thresholding layer. This soft thresholding is inspired by the Deep residual shrinkage networks (Zhao et al., 2019). These MSRUs are stacked to form a *Multi Scale Residual Deep Network* (MSRDN). Within their recommended framework, features are automatically derived directly from the raw audio. Additionally, a large kernel function is suggested to capture many features. Nonetheless, it is worth noting that larger kernel sizes come with an increase in computation time. The follow-up study combined the raw audio with a time–frequency representation of the audio as input features for the MSRDN and moved the soft thresholding layer to the beginning of the MSRU (Tian et al., 2023b). Additionally, the MSRDN was modified by adding a ConvNext model to process the time–frequency representation, while the slightly modified MSRDN still processes the raw audio input. The models are trained using deep mutual learning, increasing the recognition rate of the model by 3% accuracy and decreasing the training time by 11.5%.

Depthwise separable convolutions. The previously mentioned types of CNNs only take the spatial dimension of the input into account. However, a depthwise separable CNN also takes the depth dimension, corresponding to the number of channels, into account. For this reason, a depthwise separable convolutional layer decomposes the original signal into multiple frequency components. A prominent CNN architecture leveraging depthwise separable convolutions is called MobileNet (Howard et al., 2017). Therefore, multiple studies applied a MobileNet architecture for automatic SR-N recognition (Zhang and Ding, 2020; de BA Barros and Ebecken, 2022). Additionally, a deep neural network incorporating depthwise separable convolutions and time-dilated convolutions has been introduced (Hu et al., 2020). This system is based on the human auditory system and decomposes the original audio into different frequency components using depthwise separable convolutions. Their proposed method is further optimized within Hu et al. (2021b), resulting in a recognition accuracy of 90.9% on a private dataset. Similarly, a deep learning method with hybrid routing is presented (Cheng and Zhang, 2021). This enables the network to exchange learned features and, therefore, improves the automatic recognition of SR-N. This method resulted in a recognition accuracy of 95% on a private dataset.

Auditory based. Various CNN architectures proposed in the literature are based on human sound perception, these are denoted as auditory-based CNNs. One such study applied deep convolutional filters to decompose the SR-N signal into multiple time-domain signals with different timescales (Li et al., 2019e). Here it was shown that the recognition accuracy increases when the number of filter groups is increased to some extent. Another study tried to mimic the human perception of sound (Shen et al., 2018, 2019). Human sound perception involves two main regions: the transmission of sound to the inner ear, where it is decomposed into frequency components at the cochlea, and the creation of auditory perception through neural signals in the auditory cortex. This system is mimicked by the auditory-based CNN models to automatically recognize SR-N. To mimic this complex system, a multi-component CNN network was proposed. First, a 1D CNN layer with Gammatone filter kernels decomposes the signal similar to the cochlea. The second part consists of a permute layer, an energy-pooling layer, a 2D frequency convolutional layer, and a fully connected layer to reconstruct the auditory cortex. A follow-up study applied dilated convolution in Gammatone initialed multi-scale convolution kernels (Shen et al., 2019). The final follow-up study suggested an integrated feature extractor optimized simultaneously with the corresponding classifier (Shen et al., 2020) with a recognition rate

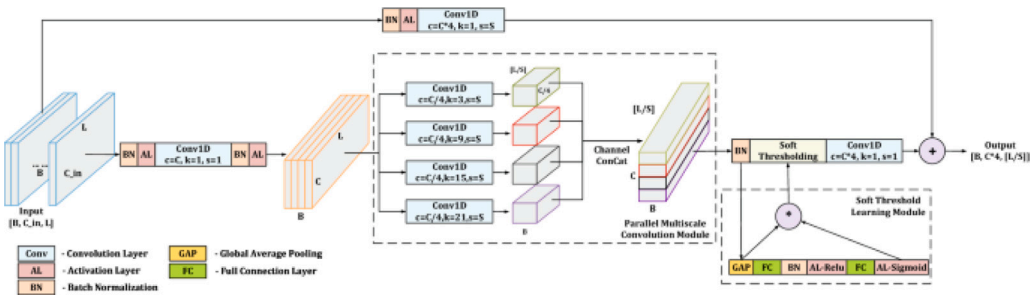

Fig. 11. The schematic visualization of a single Multi Scare Residual Unit which has been utilized as a building block for Multi-Scale Residual Deep Network (Tian et al., 2021).

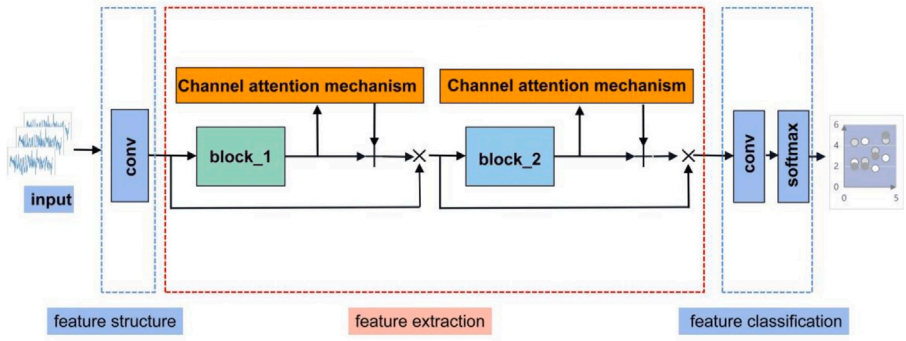

Fig. 12. The architecture of camResNet as proposed in Xue et al. (2022).

of 87.2% on data derived from ONC. Finally, another CNN architecture has been proposed to model the human timbre perception of SR-N (Li and Yang, 2021). The recognition accuracy has been compared to multiple control deep learning models, where their model showed an increase up to 13.5% compared to the other methods on a private dataset.

Attention module. Multiple CNN architectures applied for automatic SR-N recognition contain attention modules. These attention modules make the CNN focus on the SR-N signal instead of the non-informative background. One such architecture has been proposed in Li et al. (2022e). Here, an attempt was made to improve the generality of a Neural Network by learning a *Finite Impulse Response* filter incorporated in a 1D CNN layer. Next, an attention module that incorporated the STFT was proposed to extract features from the time–frequency domain. Their method achieved 84% accuracy on the ShipsEar dataset. Similarly, an Adaptive Generalized Network has been proposed to adaptively learn the wavelet parameters to extract the underwater characteristics at different frequencies (Xie et al., 2022a). In addition to this method, Channel Attention Modules have been suggested for automatic feature extraction as well. Here, these attention modules are integrated into a ResNet architecture(see Fig. 12) (Xue et al., 2022; Li et al., 2022j). The final classifier has a recognition accuracy of 99% on ShipsEar. Alternatively, a Dual Attention Network and a Multiresolution CNN have been suggested for automatic SR-N recognition (Liu et al., 2021c). The Multiresolution CNN has been employed to extract the aggregated characteristics of the manually extracted features. Next, the *Dual Attention Network* incorporates a *Position Attention Module* and a *Channel Attention Module*. Their proposed method achieved an accuracy score of 95.6% on the ShipsEar dataset.

Combinations with other networks. To further optimize the performance of a CNN in automatic SR-N recognition, various studies have explored different combinations of neural network architectures. These combinations include the combination of convolutional layers with an *Extreme Learning Machine* (Hu et al., 2018), a fusion of ResNet and DenseNet (Jin and Zeng, 2023), and a combination of convolutional layers with LSTM blocks (Liu et al., 2021f; Wang et al., 2021; Zhang et al., 2022a). Especially, the combination of 1D convolutional layers with LSTM blocks is frequently reported. Noticeable, is that these suggested models are shallow networks using only two convolutional layers followed by a single LSTM layer (Han et al., 2022) or four convolutional layers and four LSTM layers (Liu et al., 2023a) achieving an accuracy of 98.9% on ShipsEar dataset. Likewise, a CNN has been proposed to reconstruct the STFT by convolutional layers and kernel functions (Kamal et al., 2021). This network is then followed by a Bidirectional LSTM for final recognition, resulting in a recognition accuracy of 93% on a private dataset. An opposite approach is presented in Qi et al. (2021), where an LSTM is applied for automatic feature extraction and a CNN for the final recognition. Here, the two separate LSTM blocks have been suggested to extract information from both the frequency component and the phase component of the SR-N signal. The outputs of these blocks are fused and a single 1D convolutional layer with sigmoid activation has been suggested for the final classification. The classification accuracy of this method is 89.8% on a private dataset. The networks are fused by a fully connected layer. Apart from the combination of CNN and LSTM, a two-stream network that integrates MFCCNet with SpecNet has been proposed for automatic SR-N recognition (Xing et al., 2020). In this setup, the MFCCNet is based on GRU and SpecNet is based on VGGNet. The outputs of these networks are fused by a fully connected layer. This proposed method achieved 98.8% accuracy on a private dataset. Within this study, it is stated that VGGNet outperforms ResNet50, which is a contradictory result to Domingos et al. (2022).

8.2.4. Contrastive learning

Alternatively, a contrastive learning approach has been suggested to cope with the limited amount of labeled data (Xie et al., 2022b; Nie et al., 2023a; Sun and Luo, 2023; Zhu et al., 2023; Xie et al., 2023b). This is an unsupervised representation learning approach, where the goal is to learn a discriminative representation of the data, with sufficient results. In this approach, the CNN models are optimized to maximize the similarity among recordings from the same vessel while minimizing the similarity between recordings of different vessels (Nie et al., 2023a). Similarly, a framework based on SimCLR (Chen et al., 2020) has been suggested (Sun and Luo, 2023; Tian et al., 2023a). The

supervised SimCLR-based method resulted in 98% accuracy on both ShipsEar and Deepship (Sun and Luo, 2023).

8.2.5. Semi-supervised learning algorithms

Up until now, the proposed neural networks for automatic SR-N recognition were optimized in a supervised manner for classification purposes. These deep-learning applications require big data with labels. This proves to be a considerable disadvantage, considering the shortage of labeled data for automatic SR-N recognition (Yang et al., 2019; Haiyan et al., 2021). To cope with this problem, numerous learning algorithms have been proposed to optimize the final automatic SR-N recognition using the combination of labeled and unlabeled data. This learning framework is referred to as semi-supervised learning. This framework requires a small portion of labeled data and a large quantity of unlabeled data. In the context of automatic SR-N recognition, the publicly available labeled data is limited. However, a great amount of unlabeled data is available to the public. This means that semi-supervised learning is a promising metric for the training of SR-N ML applications. This section discusses several semi-supervised learning algorithms, like *Deep Belief Networks* and various variations of *Autoencoders*.

Deep belief networks. A specific type of unsupervised neural network employed in SR-N analysis is the *Boltzmann Machine*. This model operates, unlike previously mentioned methods, without the need for labeled data. This type of model finds application in automatically extracting features from SR-N (Xie et al., 2018) or the automatic encoding of the power spectrum of SR-N (Luo and Feng, 2020; Luo et al., 2021a). Here, the *Boltzmann Machines* are optimized to reconstruct the power spectrum of the original SR-N in an unsupervised manner. Next, the model is fine-tuned in a supervised manner using an MLP to perform the final recognition. A disadvantage of the Boltzmann is that the connections grow exponentially since all the nodes are fully connected. This issue is mitigated by the *Restricted Boltzmann Machines*. When these machines are stacked to create a neural network, this is called a *Deep Belief Networks* (DBNs). This network has been proposed for automatic SR-N recognition (Chen and Xu, 2017). Again, the DBN is trained in an unsupervised manner and fine-tuned in a supervised way. This method achieved a recognition accuracy of 98.2% on a simulated dataset.

Variational autoencoders. Even though Autoencoders show promising results in SR-N recognition, these models have some limitations. Autoencoders are optimized to encode and decode the input data, without considering the latent space. These characteristics make Autoencoders prone to overfitting. To overcome this problem, the *variational auotencoder* is introduced for automatic SR-N recognition (Satheesh et al., 2021; Bach et al., 2022). This algorithm improved the recognition rate with 10% compared to VGG-19, even at low SNR levels (Bach et al., 2022).

Stacked autoencoders. Recognizing SR-N automatically from underwater sound recordings is a complex task. Therefore, a single autoencoder may not be sufficient to create a representative latent space. For this reason, several stacked autoencoders have been suggested to recognize SR-N automatically (Cao et al., 2016; dos Santos Mello et al., 2018; Cao et al., 2019b; Chen et al., 2019c). Here, it is shown that these types of models outperform a traditional SVM or shallow Neural Network when the amount of labeled data samples is limited (Haiyan et al., 2021). Some variations on the traditional Stacked autoencoder are reported. For instance, a stacked denoising autoencoder has been suggested for automatic SR-N recognition (Chen et al., 2019c). Additionally, multiple sparse autoencoders have been stacked to create a Stacked Sparse AutoEncoder (Cao et al., 2019b). This method achieved a classification accuracy of 94% on a private dataset using three model layers.

Convolutional autoencoders. The convolutional autoencoder benefits from both the unsupervised pre-training as the traditional autoencoder and the automatic feature extraction advantage of a traditional CNN (Chen and Shang, 2019). This type of autoencoder can be optimized in a semi-supervised way to automatically recognize SR-N (Ke et al., 2018; Chen and Shang, 2019; Kamalipour et al., 2023). Here, the autoencoder architecture consists of convolution/pooling layers and deconvolution/unpooling layers (Chen and Shang, 2019). The training process of this type of model was optimized layer by layer resulting in a recognition accuracy of 93.3% on ShipsEar (Ke et al., 2018). In addition to the convolutional Autoencoder, a Deep Recurrent Autoencoder is proposed (Kamalipour et al., 2023). The output of both types of autoencoders is fused and is followed by a classifier. This study stated that the reconstruction rate of the lower frequency bands is better than the higher frequency bands. Besides this traditional Convolutional Autoencoder, a ResNet-based encoder with attention modules has been proposed, called CALNet (Lingzhi et al., 2023). This framework is connected to a decoder for unsupervised reconstruction of the noisy data samples. Finally, the encoder is isolated and connected to a classifier, to be finetuned to automatically recognize the clean data samples (see Fig. 13). This method outperforms U-net, DenseNet, and a single SE_ResNet by 16% in accuracy.

8.2.6. Transformers

In addition to the previously discussed methods, the application of transformer models in SR-N recognition is rising. Unlike CNNs, which fail to capture global information implicated in the spectrogram due to the use of a small kernel (Feng and Zhu, 2022), transformer models are based on multi-head attention modules. A masked modeling-based self-supervised learning method using a Swin Transformer has been suggested to automatically recognize SR-N (Xu et al., 2022). This method resulted in 78.03% accuracy on Deepship, outperforming a separable convolutional autoencoder. Another specialized audio transformer, called Audio Spectrogram Transformer (Gong et al., 2021), has also been suggested for automatic SR-N recognition (Li et al., 2022k). This transformer is reconstructed from the Vision Transformer (Dosovitskiy et al., 2020). The input of this transformer is the spectrogram of the original audio. This is then split into patches and treated as a sequence of patches. After linear projection, positional and class encoding are added to complete the input for the transformer encoder. The final classification is performed by a linear layer. The complete architecture of the Audio Spectrogram Transformer is visualized in Fig. 14. Unfortunately, transformers require a great amount of training data. To cope with the limited amount of available labeled SR-N data, it has been proposed to pre-train the Audio Spectrogram Transformer on Audioset and fine-tune the limited labeled SR-N data. This method was between 82.8% and 91.8% accurate on ShipsEar (Feng and Zhu, 2022).

8.2.7. Multi-target recognition

The majority of the automatic recognition of SR-N studies focus solely on single-target recognition. Here, only a single ship generates the noise and is recognized. In multi-target recognition, the recordings contain noise generated by multiple ships, and the goal is to identify all of these ships. Only a few studies have tried to solve automatic multi-target SR-N recognition (Yu et al., 2014). One particular study presented a CNN to perform multi-label classification (Pfau, 2020). The performance of their proposed CNN outperformed VGGNet, Inception, and MobileNet V2 pre-trained on ImageNet. Additionally, a ResNet architecture has been suggested to create a multi-target recognition method for an unknown number of targets (Sun and Wang, 2022). Multiple spectral features are combined to recognize an unknown number of targets using real-valued and complex-valued ResNet. This study showed that the magnitude STFT spectrum, complex-valued spectrum, and log-Mel spectrum can effectively recognize synthetic multi-target ship signals.

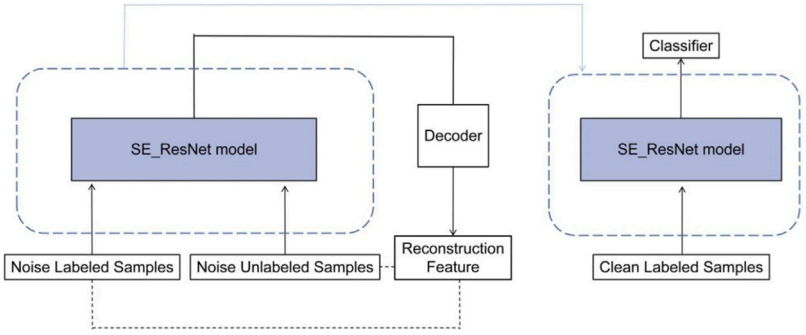

Fig. 13. The proposed training scheme of CALNet in Lingzhi et al. (2023).

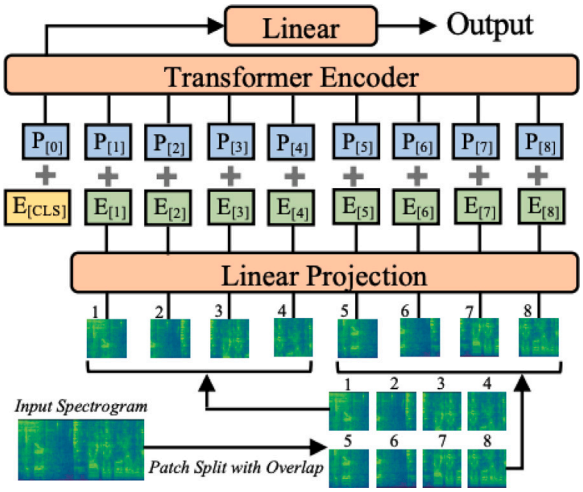

Fig. 14. The audio spectrogram transformer architecture (Gong et al., 2021).

8.3. Signature recognition

So far, the previously discussed methods focused solely on the recognition of the ship type considering the noise. This recognition is based on the sound signature of the ship. Instead of focusing on the whole signal of SR-N, various studies concentrate on the automatic detection of these individual sound signatures (Lim et al., 2007, 2008). The sound signatures of ships consist of tonals and transients. In the pursuit of automating transient detection, a model inspired by human perception of sound has been suggested (Tucker and Brown, 2005). On a different front, a CNN was introduced for the automatic detection of tonals in LOFAR spectrograms, demonstrating encouraging outcomes (Park and Jung, 2019). Besides the traditional DEMON and LOFAR representations, audio coefficients have been extracted to automatically recognize underwater transient signals using kNN (Guo and Gas, 2009) or an ANN (Can et al., 2016). Predominantly, deep learning techniques have been suggested to perform the final recognition. Alongside the manual feature extraction, the utilization of deep learning techniques is on the rise. An auto-associative NN is suggested to automatically extract the line spectrum directly from the raw audio (Huang et al., 2021). It was even stated that their proposed method can suppress background noise of simulated data. Instead of suppressing the background noise using an autoencoder neural network is suggested to enhance the tonal signals (Ju et al., 2022). This method has been compared with adaptive line enhancers, showing the superiority of their method. Finally, a different approach has been suggested in Honghui et al. (2022), to classify multiple attributes of the ship. To accomplish this, a group of neurons with learnable weights has been proposed to extract correlation-deep features. These features are utilized to create a multi-attribute correlation perception. Their proposed method achieves a stable correct recognition rate.

8.4. Summary

This section described various ML methods for automatic SR-N recognition. The majority of these methods still rely on conventional time–frequency representation algorithms. Even though suggestions were made to extract features in an unsupervised matter, these methods were still limited by the audio representation. To minimize the reliance on conventional audio processing methods, some suggestions were made to cope with the raw audio directly. This is still underexplored, but some studies concluded that a CNN can extract informative features directly from raw audio.

9. Outlook

In SR-N analysis applications, the success of ML models is evident. These methods offer a distinct advantage by not requiring prior information about the ocean environment like traditional processing methods. As a result of this independence, ML models manage to achieve comparable, and in some cases, superior results when compared to conventional methods. However, these models still cope with some limitations and challenges. This section gives a prospective outlook, delineating potential future research directions in the ML application for SR-N analysis.

9.1. Data input

Unfortunately, only a limited amount of labeled SR-N data is publicly available, and the existing databases vary in their annotation protocol. Additionally, the splitting of the datasets into training data and test data differs between studies. Some studies introduced a bias by windowing the data and randomly sampling the windowed data. Due to this bias, the resulting recognition accuracy levels are high. These variances create a challenge when attempting to create a fair comparison between ML applications in SR-N. A standardized approach in SR-N labeling would be beneficial to expand existing datasets and create more diverse datasets. This standardization could lead to the development of more trustworthy ML models applicable to SR-N and easier evaluation of these models.

9.2. Preprocessing

Numerous studies have focused on effective feature extraction from SR-N data. The manual feature extraction methods, aim to represent the complex non-stationary SR-N in a fixed format. Techniques such as mode decompositions and Gabor wavelets assume that SR-N is a composition of either IMFs or mother wavelets. However, it is crucial to note that these assumptions may not always be correct. Consequently, these

manual features exhibit sensitivity to noise. This sensitivity results in an ML model that degrades in performance once the ocean environment changes (Huang et al., 2018b). How to represent SR-N for ML models, insensitive to the variable ocean environment, remains an open question. However, the great amount of publicly available unlabeled data gives a potential for developing such a representation. Until now, to the best of our knowledge, no study of ML in SR-N analysis has ever conducted such an amount of data. The scarcity of publicly available labeled SR-N remains a challenge. Various preprocessing methods are discussed above to expand the limited amount of labeled SR-N data by generating synthetic samples. All of these methods are limited in synthesizing the complex imagery of underwater sound. Both semi-supervised learning and self-supervised learning have been suggested to overcome this limitation and have shown to increase the ML models performance in either denoising, localization, and recognition (Yang et al., 2019; Koh et al., 2020; Zhu et al., 2020, 2021a,b; Haiyan et al., 2021; Jin et al., 2022; Li et al., 2023b). These learning methods could be combined with the large unlabeled datasets, to leverage all available data and enhance the overall performance of automatic SR-N analysis application.

9.3. Analysis

Within this survey, the automatic SR-N analysis applications have been categorized into the detection, localization, and recognition of SR-N. The automatic analysis of SR-N is still dominated by CNNs, particularly those trained on time–frequency representations of audio. However, it is noteworthy that CNNs assume translational invariance, which does not apply to the frequency axis. Therefore, CNNs may not be the optimal solution to deal with time–frequency data. A convolution-free method based on the transformer has been suggested to overcome this limitation (Feng and Zhu, 2022). Several recent studies have also proposed the adoption of transformers in the automatic analysis of SR-N. Closely related fields, like audio event classification and speech recognition, have shown the potential of transformers applications to audio (Dong et al., 2018; Zhang et al., 2020b; Koutini et al., 2021; Chen et al., 2022a). These fields are out of the scope of this survey. Until now, the transformers applied to SR-N still represent the audio in the time–frequency domain. Future research needs to be conducted to explore the potential of transformers in automatic SR-N analysis.

9.4. Integrated system

Overall, the presented methods for ML applications in SR-N are partitioned in the preprocessing of the data and applied ML afterward for the final analysis. These processes are optimized separately. Incorporating an integrated system that optimizes both preprocessing and subsequent automatic analysis offers the potential for elevating the performance of the ML application. One such system is presented in Ren et al. (2022), where they optimized the Gabor filters simultaneously with the ResNet for automatic SR-N analysis. This study shows the strength of an integrated system for automatic SR-N analysis using ML.

CRediT authorship contribution statement

Hilde I. Hummel: Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Methodology, Investigation, Formal analysis, Conceptualization. **Rob van der Mei:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Funding acquisition. **Sandjai Bhulai:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Hilde Hummel reports financial support was provided by Dutch Government. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

Aksüren, İ.G., Hocaoğlu, A.K., 2022. Automatic target classification using underwater acoustic signals. In: 2022 30th Signal Processing and Communications Applications Conference. SIU, IEEE, pp. 1–4.

Applelid, G., Karlsson, M., 2019. Classify Different Types of Boat Engine Sounds with Machine Learning. (Bachelor's thesis). KTH Royal Institute of Technology.

Ashok, P., Latha, B., 2022. An improving recognition accuracy of underwater acoustic targets based on gated recurrent unit (GRU) neural network method. In: 2022 1st International Conference on Computational Science and Technology. ICCST, IEEE, pp. 1–6.

Axelsson, O., Rhen, C., 2020. Neural-network-based classification of commercial ships from multi-influence passive signatures. IEEE J. Ocean. Eng. 46 (2), 634–641.

Bach, N.H., Vu, L.H., Nguyen, V.D., 2021. Classification of surface vehicle propeller cavitation noise using spectrogram processing in combination with convolution neural network. Sensors 21 (10), 3353.

Bach, N.H., et al., 2022. Improving the classification of propeller ships using LOFAR and triple loss variational auto encoder. In: 2022 International Conference on Electrical, Computer and Energy Technologies. IEEE, pp. 1–5.

Brown, J.C., 1991. Calculation of a constant Q spectral transform. J. Acoust. Soc. Am. 89 (1), 425–434.

Can, G., 2016. Classification of vessel acoustic signatures using non-linear scattering based feature extraction (Ph.D. thesis). Bilkent Universitesi (Turkey).

Can, G., Akbaş, C.E., Çetin, A.E., 2016. Recognition of vessel acoustic signatures using non-linear teager energy based features. In: 2016 International Workshop on Computational Intelligence for Multimedia Understanding. IWCIM, IEEE, pp. 1–5.

Can, G., Akbaş, C.E., Çetin, A.E., 2017. Time-scale wavelet scattering using hyperbolic tangent function for vessel sound classification. In: 2017 25th European Signal Processing Conference. pp. 1794–1798.

Cao, H., Ren, Q., 2022. Distinguishing multiple surface ships using one acoustic vector sensor based on a convolutional neural network. JASA Expr. Lett. 2 (5), 054803.

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

Cao, H., Wang, W., Ni, H., Ren, Q., Ma, L., 2019a. Deep learning for DOA estimation using a vector hydrophone. In: Oceans 2019 MTS/IEEE Seattle. IEEE, pp. 1–4.

Cao, H., Wang, W., Su, L., Ni, H., Gerstoft, P., Ren, Q., Ma, L., 2021. Deep transfer learning for underwater direction of arrival using one vector sensor. J. Acoust. Soc. Am. 149 (3), 1699–1711.

Cao, X., Zhang, X., Togneri, R., Yu, Y., 2019b. Underwater target classification at greater depths using deep neural network with joint multiple-domain feature. IET Radar Sonar Nav. 13 (3), 484–491.

Cao, X., Zhang, X., Yu, Y., Niu, L., 2016. Deep learning-based recognition of underwater target. In: 2016 IEEE International Conference on Digital Signal Processing. IEEE, pp. 89–93.

Caruana, R., 1997. Multitask learning. Mach. Learn. 28, 41–75.

Cauchy, P., Heywood, K.J., Merchant, N.D., Queste, B.Y., Testor, P., 2018. Wind speed measured from underwater gliders using passive acoustics. J. Atmos. Ocean. Technol. 35 (12), 2305–2321.

Chaitanya, B., Yadav, A., Pazoki, M., Abdelaziz, A.Y., 2021. A comprehensive review of islanding detection methods. Uncertain. Modern Power Syst. 211–256.

Chen, Y., Du, S., Quan, H., Zhou, B., 2019a. Underwater target recognition method based on convolution residual network. In: MATEC Web of Conferences, vol. 283, EDP Sciences, p. 04011.

Chen, K., Du, X., Zhu, B., Ma, Z., Berg-Kirkpatrick, T., Dubnov, S., 2022a. HTS-AT: A hierarchical token-semantic audio transformer for sound classification and detection. In: ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, pp. 646–650.

Chen, S., Guan, S., Wang, H., Ye, N., Wei, Z., 2023. A new method of ship type identification based on underwater radiated noise signals. J. Mar. Sci. Eng. 11 (5), 963.

Chen, J., Han, B., Ma, X., Zhang, J., 2021. Underwater target recognition based on multi-decision lofar spectrum enhancement: A deep-learning approach. Fut. Internet 13 (10), 265.

Chen, T., Kornblith, S., Norouzi, M., Hinton, G., 2020. A simple framework for contrastive learning of visual representations. In: International Conference on Machine Learning. PMLR, pp. 1597–1607.
Chen, Z., Li, Y., Cao, R., Ali, W., Yu, J., Liang, H., 2019b. A new feature extraction method for ship-radiated noise based on improved CEEMDAN, normalized mutual information and multiscale improved permutation entropy. Entropy 21 (6), 624.
Chen, Z., Li, Y., Liang, H., Yu, J., 2018. Hierarchical cosine similarity entropy for feature extraction of ship-radiated noise. Entropy 20 (6), 425.
Chen, L., Liu, F., Li, D., Shen, T., Zhao, D., 2022b. Underwater acoustic target classification with joint learning framework and data augmentation. In: 2022 5th International Conference on Artificial Intelligence and Big Data. ICAIBD, IEEE, pp. 23–28.
Chen, J., Liu, C., Xie, J., An, J., Huang, N., 2022c. Time-frequency mask-aware bidirectional LSTM: A deep learning approach for underwater acoustic signal separation. Sensors 22 (15), 5598.
Chen, J., Rao, W., Wang, Z., Wu, Z., Wang, Y., Yu, T., Shang, S., Meng, H., 2022d. Speech enhancement with fullband-subband cross-attention network. arXiv:2211.05432.
Chen, R., Schmidt, H., 2021. Model-based convolutional neural network approach to underwater source-range estimation. J. Acoust. Soc. Am. 149 (1), 405–420.
Chen, Y., Shang, J., 2019. Underwater target recognition method based on convolution autoencoder. In: 2019 IEEE International Conference on Signal, Information and Data Processing. ICSIDP, IEEE, pp. 1–5.
Chen, Y., Xu, X., 2017. The research of underwater target recognition method based on deep learning. In: 2017 IEEE International Conference on Signal Processing, Communications and Computing. ICSPCC, IEEE, pp. 1–5.
Chen, Y., Xu, X., Zhou, B., Quan, H., 2019c. Underwater target recognition method based on t-SNE and stacked nonnegative constrained denoising autoencoder. Indian J. Geo-Mar. Sci. 1822–1832.
Chen, S., Zhang, H., 2011. Detection of underwater acoustic signal from ship noise based on WPT method. In: 2011 Fourth International Workshop on Chaos-Fractals Theories and Applications. IEEE, pp. 324–327.
Cheng, X., Zhang, H., 2021. Underwater target signal classification using the hybrid routing neural network. Sensors 21 (23), 7799.
Chi, J., Li, X., Wang, H., Gao, D., Gerstoft, P., 2019. Sound source ranging using a feed-forward neural network trained with fitting-based early stopping. J. Acoust. Soc. Am. 146 (3), EL258–EL264.
Choi, J., Choo, Y., Lee, K., 2019. Acoustic classification of surface and underwater vessels in the ocean using supervised machine learning. Sensors 19 (16), 3492.
de BA Barros, R.E., Ebecken, N.F., 2022. Development of a ship classification method based on convolutional neural network and cyclostationarity analysis. Mech. Syst. Signal Process. 170, 108778.
De Moura, N., De Seixas, J., Ramos, R., 2011. Passive sonar signal detection and classification based on independent component analysis. In: Sonar Systems. InTech Makati, Philippines, pp. 93–103.
de Souza, M.J., de Moura Júnior, N.N., de Seixas, J.M., 2022. Passive sonar classification using time-domain information and recurrent neural networks. In: 2022 IEEE Latin American Conference on Computational Intelligence. IEEE, pp. 1–6.
Doan, V.-S., Huynh-The, T., Kim, D.-S., 2020. Underwater acoustic target classification based on dense convolutional neural network. IEEE Geosci. Remote Sens. Lett. 19, 1–5.
Domingos, L.C., Santos, P.E., Skelton, P.S., Brinkworth, R.S., Sammut, K., 2022. An investigation of preprocessing filters and deep learning methods for vessel type classification with underwater acoustic data. IEEE Access 10, 117582–117596.
Dong, Y., Shen, X., Wang, H., 2022. Bidirectional denoising autoencoders-based robust representation learning for underwater acoustic target signal denoising. IEEE Trans. Instrum. Meas. 71, 1–8.
Dong, L., Xu, S., Xu, B., 2018. Speech-transformer: A no-recurrence sequence-to-sequence model for speech recognition. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, pp. 5884–5888.
dos Santos Mello, V., de Moura, N.N., de Seixas, J.M., 2018. Novelty detection in passive sonar systems using stacked autoencoders. In: 2018 International Joint Conference on Neural Networks. IJCNN, IEEE, pp. 1–7.
Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.
Dragomiretskiy, K., Zosso, D., 2013. Variational mode decomposition. IEEE Trans. Signal Process. 62 (3), 531–544.
Durofchalk, N.C., Jin, J., Vazquez, H.J., Gemba, K.L., Romberg, J., Sabra, K.G., 2021. Data driven source localization using a library of nearby shipping sources of opportunity. JASA Expr. Lett. 1 (12), 124802.
Fang, Y., Yu, H., He, Q., Bai, L., 2023. An underwater acoustic signal preprocessing method based on superposition variational mode decomposition residual. ISPP 2023, In: International Conference on Image, Signal Processing, and Pattern Recognition, vol. 12707, SPIE, pp. 373–380.
Feng, S., Zhu, X., 2022. A transformer-based deep learning network for underwater acoustic target recognition. IEEE Geosci. Remote Sens. Lett. 19, 1–5.
Ferguson, E.L., 2021. Multitask convolutional neural network for acoustic localization of a transiting broadband source using a hydrophone array. J. Acoust. Soc. Am. 150 (1), 248–256.
Ferguson, E.L., Ramakrishnan, R., Williams, S.B., Jin, C.T., 2016. Deep learning approach to passive monitoring of the underwater acoustic environment. J. Acoust. Soc. Am. 140 (4), 3351.
Ferguson, E.L., Williams, S.B., Jin, C.T., 2019. Convolutional neural network for single-sensor acoustic localization of a transiting broadband source in very shallow water. J. Acoust. Soc. Am. 146 (6), 4687–4698.
Frei, M.G., Osorio, I., 2007. Intrinsic time-scale decomposition: Time–frequency–energy analysis and real-time filtering of non-stationary signals. Proc. R. Soc. A: Math. Phys. Eng. Sci. 463 (2078), 321–342.
Gao, Y., Chen, Y., Wang, F., He, Y., 2020. Recognition method for underwater acoustic target based on DCGAN and DenseNet. In: 2020 IEEE 5th International Conference on Image, Vision and Computing. ICIVC, IEEE, pp. 215–221.
Ge, F.-X., Bai, Y., Li, M., Zhu, G., Yin, J., 2022. Label distribution-guided transfer learning for underwater source localization. J. Acoust. Soc. Am. 151 (6), 4140–4149.
Gong, Y., Chung, Y.-A., Glass, J., 2021. Ast: Audio spectrogram transformer. arXiv preprint arXiv:2104.01778.
Guo, Y., Gas, B., 2009. Underwater transient and non transient signals classification using predictive neural networks. In: 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, pp. 2283–2288.
Haiyan, N., Wenbo, W., Meng, Z., Qunyan, R., Li, M., 2021. Semi-supervised noise classification based on auto-encoder. In: 2021 OES China Ocean Acoustics. COA, IEEE, pp. 982–985.
Han, X.C., Ren, C., Wang, L., Bai, Y., 2022. Underwater acoustic target recognition method based on a joint neural network. Plos One 17 (4), e0266425.
Herchig, R., Palermo, N., Gerstoft, P., Daniel, T., Sternlicht, D., 2022. Comparing the performance of convolutional neural networks trained to localize underwater sound sources. In: Oceans 2022, Hampton Roads. IEEE, pp. 1–7.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.
Hong, F., Liu, C., Guo, L., Chen, F., Feng, H., 2021a. Underwater acoustic target recognition with a residual network and the optimized feature extraction method. Appl. Sci. 11 (4), 1442.
Hong, F., Liu, C., Guo, L., Chen, F., Feng, H., 2021b. Underwater acoustic target recognition with resnet18 on shipsear dataset. In: 2021 IEEE 4th International Conference on Electronics Technology. ICET, IEEE, pp. 1240–1244.
Honghui, Y., Junhao, L., Meiping, S., 2022. Underwater acoustic target multi-attribute correlation perception method based on deep learning. Appl. Acoust. 190, 108644.
Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861.
Hu, Z., Huang, J., Xu, P., Nan, M., Lou, K., Li, G., 2021a. Underwater acoustic source localization via kernel extreme learning machine. Front. Phys. 9, 653875.
Hu, G., Wang, K., Liu, L., 2020. An features extraction and recognition method for underwater acoustic target based on atcnn. arXiv:2011.14336.
Hu, G., Wang, K., Liu, L., 2021b. Underwater acoustic target recognition based on depthwise separable convolution neural networks. Sensors 21 (4), 1429.
Hu, G., Wang, K., Peng, Y., Qiu, M., Shi, J., Liu, L., 2018. Deep learning methods for underwater target feature extraction and recognition. Comput. Intell. Neurosci. 2018.
Huang, L., Pena, B., Liu, Y., Anderlini, E., 2022. Machine learning in sustainable ship design and operation: A review. Ocean Eng. 266, 112907.
Huang, Z., Xu, J., Gong, Z., Wang, H., Yan, Y., 2018a. A deep neural network based method of source localization in a shallow water environment. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, pp. 3499–3503.
Huang, Z., Xu, J., Gong, Z., Wang, H., Yan, Y., 2018b. Source localization using deep neural networks in a shallow water environment. J. Acoust. Soc. Am. 143 (5), 2922–2932.
Huang, Z., Xu, J., Gong, Z., Wang, H., Yan, Y., 2019. Multiple source localization in a shallow water waveguide exploiting subarray beamforming and deep neural networks. Sensors 19 (21), 4768.
Huang, C., Yang, K., Yang, Q., Zhang, H., 2021. Line spectrum extraction based on autoassociative neural networks. JASA Expr. Lett. 1 (1), 016003.
Irfan, M., Jiangbin, Z., Ali, S., Iqbal, M., Masood, Z., Hamid, U., 2021. DeepShip: An underwater acoustic benchmark dataset and a separable convolution based autoencoder for classification. Expert Syst. Appl. 183, 115270.
Jia, Y., Mo, Y., Wang, W., Guo, S., Ma, L., 2021. Deep-sea source ranging method using modified general regression neural network. In: 2021 OES China Ocean Acoustics. COA, IEEE, pp. 1032–1037.
Jiang, Y., Peng, C.-K., Xu, Y., 2011. Hierarchical entropy analysis for biological signals. J. Comput. Appl. Math. 236 (5), 728–742.
Jiang, J., Shi, T., Huang, M., Xiao, Z., 2020. Multi-scale spectral feature extraction for underwater acoustic target recognition. Measurement 166, 108227.
Jiang, J., Wu, Z., Lu, J., Huang, M., Xiao, Z., 2021. Interpretable features for underwater acoustic target recognition. Measurement 173, 108586.
Jiang, Z., Zhao, C., Wang, H., 2022. Classification of underwater target based on s-resnet and modified DCGAN models. Sensors 22 (6), 2293.
Jin, G., Liu, F., Wu, H., Song, Q., 2020. Deep learning-based framework for expansion, recognition and classification of underwater acoustic signal. J. Exper. Theor. Artif. Intell. 32 (2), 205–218.

Jin, P., Wang, B., Li, L., Chao, P., Xie, F., 2022. Semi-supervised underwater acoustic source localization based on residual convolutional autoencoder. EURASIP J. Adv. Signal Process. 2022 (1), 107.
Jin, A., Zeng, X., 2023. A novel deep learning method for underwater target recognition based on res-dense convolutional neural network with attention mechanism. J. Mar. Sci. Eng. 11 (1), 69.
Ju, D., Chi, C., Li, Z., Li, Y., Zhang, C., Huang, H., 2022. Deep-learning-based line enhancer for passive Sonar systems. IET Radar Sonar Nav. 16 (3), 589–601.
Ju, Y., Wei, Z., Huangfu, L., Xiao, F., 2020. A new low SNR underwater acoustic signal classification method based on intrinsic modal features maintaining dimensionality reduction. Polish Marit. Res. 27 (2), 187–198.
Kamal, S., Chandran, C.S., Supriya, M., 2021. Passive sonar automated target classifier for shallow waters using end-to-end learnable deep convolutional LSTMs. Eng. Sci. Technol. Int. J. 24 (4), 860–871.
Kamalipour, M., Agahi, H., Khishe, M., Mahmoodzadeh, A., 2023. Passive ship detection and classification using hybrid cepstrums and deep compound autoencoders. Neural Comput. Appl. 35 (10), 7833–7851.
Kammegne, C., Bayet, T., Brochier, T., Idy, D., Denis, C., Tremblay, Y., 2023. Detection and classification of underwater acoustic events. In: Pan-African Artificial Intelligence and Smart Systems: Second EAI International Conference, PAAISS 2022, Dakar, Senegal, November 2-4, 2022, Proceedings. Springer, pp. 251–269.
Ke, X., Yuan, F., Cheng, E., 2018. Underwater acoustic target recognition based on supervised feature-separation algorithm. Sensors 18 (12), 4318.
Khalilabadi, M.R., 2023. Underwater ship-radiated acoustic noise recognition based on mel-spectrogram and convolutional neural network. Int. J. Coast. Offsh. Environ. Eng. 8 (1), 10–15.
Khishe, M., Mosavi, M., 2020. Classification of underwater acoustical dataset using neural network trained by chimp optimization algorithm. Appl. Acoust. 157, 107005.
Koh, S., Chia, C.S., Tan, B.A., 2020. Underwater signal denoising using deep learning approach. In: Global Oceans 2020: Singapore–US Gulf Coast. IEEE, pp. 1–6.
Koutini, K., Schlüter, J., Eghbal-Zadeh, H., Widmer, G., 2021. Efficient training of audio transformers with patchout. arXiv preprint arXiv:2110.05069.
Küçükbayrak, M., Güneş, Ö., Arica, N., 2009. Underwater acoustic signal recognition methods. J. Naval Sci. Eng. 5 (3), 64–78.
Kuzin, D., Statsenko, L., Smirnova, M., 2022. Automated sea vehicle classification system based on neural network. In: 2022 International Conference on Ocean Studies. ICOS, IEEE, pp. 87–90.
Leal, N., Leal, E., Sanchez, G., 2015. Marine vessel recognition by acoustic signature. ARPN J. Eng. Appl. Sci. 10 (20), 9633–9639.
Lefort, R., Real, G., Drémeau, A., 2017. Direct regressions for underwater acoustic source localization in fluctuating oceans. Appl. Acoust. 116, 303–310.
Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M., Aila, T., 2018. Noise2Noise: Learning image restoration without clean data. arXiv:1803.04189.
Li, G., Bu, W., Yang, H., 2023a. Research on noise reduction method for ship radiate noise based on secondary decomposition. Ocean Eng. 268, 113412.
Li, X., Chen, J., Bai, J., Ayub, M.S., Zhang, D., Wang, M., Yan, Q., 2022a. Deep learning-based DOA estimation using CRNN for underwater acoustic arrays. Front. Mar. Sci. 9, 1027830.
Li, Y., Chen, X., Yu, J., 2019a. A hybrid energy feature extraction approach for ship-radiated noise based on CEEMDAN combined with energy difference and energy entropy. Processes 7 (2), 69.
Li, H., Cheng, Y., Dai, W., Li, Z., 2014. A method based on wavelet packets-fractal and SVM for underwater acoustic signals recognition. In: 2014 12th International Conference on Signal Processing. IEEE, pp. 2169–2173.
Li, Y., Gao, P., Tang, B., Yi, Y., Zhang, J., 2022b. Double feature extraction method of ship-radiated noise signal based on slope entropy and permutation entropy. Entropy 24 (1), 22.
Li, Y., Geng, B., Jiao, S., 2021a. Refined composite multi-scale reverse weighted permutation entropy and its applications in ship-radiated noise. Entropy 23 (4), 476.
Li, Y., Geng, B., Jiao, S., 2022c. Dispersion entropy-based Lempel-Ziv complexity: A new metric for signal analysis. Chaos Solitons Fractals 161, 112400.
Li, L., Gong, F., Liu, S., 2023b. Semi-supervised learning method for source range estimation in shallow water. J. Phys.: Conf. Ser. 2458, 012045.
Li, G., Hou, Y., Yang, H., 2022d. A novel method for frequency feature extraction of ship radiated noise based on variational mode decomposition, double coupled duffing chaotic oscillator and multivariate multiscale dispersion entropy. Alex. Eng. J. 61 (8), 6329–6347.
Li, C., Huang, Z., Xu, J., Yan, Y., 2018a. Underwater target classification using deep learning. In: Oceans 2018 MTS/IEEE Charleston. IEEE, pp. 1–5.
Li, Y., Jiao, S., Geng, B., 2021b. A comparative study of four multi-scale entropies combined with grey relational degree in classification of ship-radiated noise. Appl. Acoust. 176, 107865.
Li, Y., Jiao, S., Geng, B., Jiang, X., 2021c. Rcmfrde: Refined composite multiscale fluctuation-based reverse dispersion entropy for feature extraction of ship-radiated noise. Math. Probl. Eng. 2021, 1–18.
Li, Y., Jiao, S., Geng, B., Zhou, Y., 2021d. Research on feature extraction of ship-radiated noise based on multi-scale reverse dispersion entropy. Appl. Acoust. 173, 107737.
Li, Y., Li, Y., 2018. Feature extraction of underwater acoustic signal using mode decomposition and measuring complexity. In: 2018 15th International Bhurban Conference on Applied Sciences and Technology. IBCAST, IEEE, pp. 757–763.
Li, Y.-X., Li, Y.-A., Chen, Z., Chen, X., 2016. Feature extraction of ship-radiated noise based on permutation entropy of the intrinsic mode function with the highest energy. Entropy 18 (11), 393.
Li, Y., Li, Y., Chen, X., Yu, J., 2017. A novel feature extraction method for ship-radiated noise based on variational mode decomposition and multi-scale permutation entropy. Entropy 19 (7), 342.
Li, Z., Li, Y., Zhang, K., Guo, J., 2019b. A novel improved feature extraction technique for ship-radiated noise based on IITD and MDE. Entropy 21 (12), 1215.
Li, Z., Li, Y., Zhang, K., Guo, J., 2019c. A novel improved feature extraction technique for ship-radiated noise based on improved intrinsic time-scale decomposition and multiscale dispersion entropy. In: Proceedings, vol. 46, MDPI, p. 16.
Li, C., Liu, Z., Ren, J., Wang, W., Xu, J., 2020a. A feature optimization approach based on inter-class and intra-class distance for ship type classification. Sensors 20 (18), 5429.
Li, D., Liu, F., Shen, T., Chen, L., Yang, X., Zhao, D., 2022e. Generalizable underwater acoustic target recognition using feature extraction module of neural network. Appl. Sci. 12 (21), 10804.
Li, D., Liu, F., Shen, T., Chen, L., Zhao, D., 2023c. A robust feature extraction method for underwater acoustic target recognition based on multi-task learning. Electronics 12 (7), 1708.
Li, G., Liu, F., Yang, H., 2022f. Research on feature extraction method of ship radiated noise with K-nearest neighbor mutual information variational mode decomposition, neural network estimation time entropy and self-organizing map neural network. Measurement 199, 111446.
Li, Y., Ning, F., Jiang, X., Yi, Y., 2022g. Feature extraction of ship radiation signals based on wavelet packet decomposition and energy entropy. Math. Probl. Eng. 2022, 1–12.
Li, W., Shen, X., Li, Y., 2019d. A comparative study of multiscale sample entropy and hierarchical entropy and its application in feature extraction for ship-radiated noise. Entropy 21 (8), 793.
Li, Y., Tang, B., Jiao, S., 2022h. Optimized ship-radiated noise feature extraction approaches based on CEEMDAN and slope entropy. Entropy 24 (9), 1265.
Li, Y., Tang, B., Jiao, S., 2023d. SO-slope entropy coupled with SVMD: A novel adaptive feature extraction method for ship-radiated noise. Ocean Eng. 280, 114677.
Li, Y., Tang, B., Yi, Y., 2022i. A novel complexity-based mode feature representation for feature extraction of ship-radiated noise using VMD and slope entropy. Appl. Acoust. 196, 108899.
Li, J., Wang, B., Cui, X., Li, S., Liu, J., 2022j. Underwater acoustic target recognition based on attention residual network. Entropy 24 (11), 1657.
Li, P., Wu, J., Wang, Y., Lan, Q., Xiao, W., 2022k. STM: Spectrogram transformer model for underwater acoustic target recognition. J. Mar. Sci. Eng. 10 (10), 1428.
Li, Y., Xiao, L., Tang, B., Liang, L., Lou, Y., Guo, X., Xue, X., 2022l. A denoising method for ship-radiated noise based on optimized variational mode decomposition with snake optimization and dual-threshold criteria of correlation coefficient. Math. Probl. Eng. 2022.
Li, J., Yang, H., 2021. The underwater acoustic target timbre perception and recognition based on the auditory inspired deep convolutional neural network. Appl. Acoust. 182, 108210.
Li, S., Yang, S., Liang, J., 2020b. Recognition of ships based on vector sensor and bidirectional long short-term memory networks. Appl. Acoust. 164, 107248.
Li, J., Yang, H., Shen, S., Xu, G., 2019e. The learned multi-scale deep filters for underwater acoustic target modeling and recognition. In: OCEANS 2019-Marseille. IEEE, pp. 1–4.
Li, G., Yang, Z., Yang, H., 2019f. A denoising method of ship radiated noise signal based on modified CEEMDAN, dispersion entropy, and interval thresholding. Electronics 8 (6), 597.
Li, H., Yue, P., Jiangqiao, L., 2018b. Classification of underwater acoustic target using auditory spectrum feature and SVDD ensemble. In: 2018 Oceans-MTS/IEEE Kobe Techno-Oceans. OTO, IEEE, pp. 1–4.
Lian, Z., Wu, T., 2022. Feature extraction of underwater acoustic target signals using gammatone filterbank and subband instantaneous frequency. In: 2022 IEEE 6th Advanced Information Technology, Electronic and Automation Control Conference. IAEAC, IEEE, pp. 944–949.
Lian, Z., Xu, K., Wan, J., Li, G., 2017. Underwater acoustic target classification based on modified GFCC features. In: 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference. IAEAC, IEEE, pp. 258–262.
Lim, T., Bae, K., Hwang, C., Lee, H., 2007. Classification of underwater transient signals using mfcc feature vector. In: 2007 9th International Symposium on Signal Processing and Its Applications. IEEE, pp. 1–4.
Lim, T., Bae, K., Hwang, C., Lee, H., 2008. Underwater transient signal classification using binary pattern image of MFCC and neural network. IEICE Trans. Fundam. Electron. Commun. Comput. Sci. 91 (3), 772–774.
Lin, Y., Zhu, M., Wu, Y., Zhang, W., 2020. Passive source ranging using residual neural network with one hydrophone in shallow water. In: 2020 IEEE 3rd International Conference on Information Communication and Signal Processing. ICICSP, IEEE, pp. 122–125.

Lingzhi, X., Xiangyang, Z., Xiang, Y., Shuang, Y., 2023. Completion-attention ladder network for few-shot underwater acoustic recognition. Neural Process. Lett. 1–17.
Liu, Y., Chen, H., Wang, B., 2021a. DOA estimation based on CNN for underwater acoustic array. Appl. Acoust. 172, 107594.
Liu, F., Ding, H., Li, D., Wang, T., Luo, Z., Chen, L., 2021b. Few-shot learning with data enhancement and transfer learning for underwater target recognition. In: 2021 OES China Ocean Acoustics. COA, IEEE, pp. 992–994.
Liu, S., Fu, X., Xu, H., Zhang, J., Zhang, A., Zhou, Q., Zhang, H., 2023a. A fine-grained ship-radiated noise recognition system using deep hybrid neural networks with multi-scale features. Remote Sens. 15 (8), 2068.
Liu, C., Hong, F., Feng, H., Hu, M., 2021c. Underwater acoustic target recognition based on dual attention networks and multiresolution convolutional neural networks. In: Oceans 2021: San Diego–Porto. IEEE, pp. 1–5.
Liu, Y.-N., Niu, H.-Q., Li, Z.-L., 2019a. Source ranging using ensemble convolutional networks in the direct zone of deep water. Chin. Phys. Lett. 36 (4), 044302.
Liu, Y., Niu, H., Li, Z., 2020a. A multi-task learning convolutional neural network for source localization in deep ocean. J. Acoust. Soc. Am. 148 (2), 873–883.
Liu, M., Niu, H., Li, Z., 2023b. Implementation of bartlett matched-field processing using interpretable complex convolutional neural network. JASA Expr. Lett. 3 (2).
Liu, Y., Niu, H., Li, Z., Wang, M., 2021d. Deep-learning source localization using autocorrelation functions from a single hydrophone in deep ocean. JASA Expr. Lett. 1 (3), 036002.
Liu, Y., Niu, H., Yang, S., Li, Z., 2021e. Multiple source localization using learning-based sparse estimation in deep ocean. J. Acoust. Soc. Am. 150 (5), 3773–3786.
Liu, F., Shen, T., Luo, Z., Zhao, D., Guo, S., 2021f. Underwater target recognition using convolutional recurrent neural networks with 3-D Mel-spectrogram and data augmentation. Appl. Acoust. 178, 107989.
Liu, F., Song, Q., Jin, G., 2019b. Expansion of restricted sample for underwater acoustic signal based on generative adversarial networks. ICGIP 2018, In: Tenth International Conference on Graphics and Image Processing, vol. 11069, SPIE, pp. 1222–1229.
Liu, X.-A., Yang, D.-Q., Li, Q., 2023c. The load criteria for ship mechanical noise prediction in low frequencies and experimental validation. J. Ocean Eng. Sci. 8 (3), 278–293.
Liu, W., Yang, Y., Xu, M., Lü, L., Liu, Z., Shi, Y., 2020b. Source localization in the deep ocean using a convolutional neural network. J. Acoust. Soc. Am. 147 (4), EL314–EL319.
Liu, Y., Zhang, X., Shao, J., 2014. Quadratic time-frequency feature extraction and fusion for ship targets classification. In: 2014 12th International Conference on Signal Processing. ICSP, IEEE, pp. 194–198.
Luo, X., Chen, L., Zhou, H., Cao, H., 2023. A survey of underwater acoustic target recognition methods based on machine learning. J. Mar. Sci. Eng. 11 (2), 384.
Luo, X., Feng, Y., 2020. An underwater acoustic target recognition method based on restricted Boltzmann machine. Sensors 20 (18), 5399.
Luo, X., Feng, Y., Zhang, M., 2021a. An underwater acoustic target recognition method based on combined feature with automatic coding and reconstruction. Ieee Access 9, 63841–63854.
Luo, X., Zhang, M., Liu, T., Huang, M., Xu, X., 2021b. An underwater acoustic target recognition method based on spectrograms with different resolutions. J. Mar. Sci. Eng. 9 (11), 1246.
Ma, H., Xu, Y., Wang, J., Song, M., Zhang, S., 2023. SVMD coupled with dual-threshold criteria of correlation coefficient: A self-adaptive denoising method for ship-radiated noise signal. Ocean Eng. 281, 114931.
Ma, Y., Zhang, Y., Zhu, J., Xu, K., Cai, Y., 2021. A fast instantaneous frequency estimation for underwater acoustic target feature extraction. J. Phys.: Conf. Ser. 2031, 012018.
Meng, Q., Yang, S., Piao, S., 2014. The classification of underwater acoustic target signals based on wave structure and support vector machine. J. Acoust. Soc. Am. 136 (4), 2265.
Mishachandar, B., Vairamuthu, S., 2021. Diverse ocean noise classification using deep learning. Appl. Acoust. 181, 108141.
Mohammed, S.K., Hariharan, S.M., Kamal, S., 2018. A GTCC-based underwater HMM target classifier with fading channel compensation. J. Sens. 2018.
National Oceanic and Atmospheric Administration, 2017. Passive-acoustic-data. https://www.ncei.noaa.gov/products/passive-acoustic-data. (Accessed: 1 May 2023).
Neilsen, T., Escobar-Amado, C., Acree, M., Hodgkiss, W., Van Komen, D., Knobles, D., Badiey, M., Castro-Correa, J., 2021. Learning location and seabed type from a moving mid-frequency source. J. Acoust. Soc. Am. 149 (1), 692–705.
Neupane, D., Seok, J., 2020. A review on deep learning-based approaches for automatic Sonar target recognition. Electronics 9 (11), 1972.
Nie, L., Li, C., Wang, H., Wang, J., Zhang, Y., Yin, F., Marzani, F., Bozorg Grayeli, A., 2023a. A contrastive-learning-based method for the few-shot identification of ship-radiated noises. J. Mar. Sci. Eng. 11 (4), 782.
Nie, W., Zhang, X., Xu, J., Guo, L., Yan, Y., 2023b. Adaptive direction-of-arrival estimation using deep neural network in marine acoustic environment. IEEE Sens. J..
Niu, H., Gong, Z., Ozanich, E., Gerstoft, P., Wang, H., Li, Z., 2019. Deep learning for ocean acoustic source localization using one sensor. arXiv preprint arXiv:1903.12319.
Niu, F., Hui, J., Zhao, A., Cheng, Y., Chen, Y., 2018. Application of SN-EMD in mode feature extraction of ship radiated noise. Math. Probl. Eng. 2018, 1–16.
Niu, H., Li, X., Zhang, Y., Xu, J., 2023. Advances and applications of machine learning in underwater acoustics. Intell. Mar. Technol. Syst. 1 (1), 8.
Niu, H., Ozanich, E., Gerstoft, P., 2017a. Ship localization in Santa Barbara channel using machine learning classifiers. J. Acoust. Soc. Am. 142 (5), EL455–EL460.
Niu, H., Reeves, E., Gerstoft, P., 2017b. Source localization in an ocean waveguide using supervised machine learning. J. Acoust. Soc. Am. 142 (3), 1176–1188.
Ocean Network Canada, 2007. Dataportal. https://data.oceannetworks.ca/home. (Accessed: 1 May 2023).
Ozanich, E., Gerstoft, P., Niu, H., 2020. A feedforward neural network for direction-of-arrival estimation. J. Acoust. Soc. Am. 147 (3), 2035–2048.
Ozard, J.M., Zakarauskas, P., Ko, P., 1991. An artificial neural network for range and depth discrimination in matched field processing. J. Acoust. Soc. Am. 90 (5), 2658–2663.
Pan, A., Chen, X., Li, W., 2021. Recognition of underwater acoustic target using sub-pretrained convolutional neural networks. In: Proceedings of the 8th Conference on Sound and Music Technology: Selected Papers from CSMT. Springer, pp. 113–123.
Park, D.S., Chan, W., Zhang, Y., Chiu, C.-C., Zoph, B., Cubuk, E.D., Le, Q.V., 2019. Specaugment: A simple data augmentation method for automatic speech recognition. arXiv preprint arXiv:1904.08779.
Park, J., Jung, D.-J., 2019. Identifying tonal frequencies in a Lofargram with convolutional neural networks. In: 2019 19th International Conference on Control, Automation and Systems. ICCAS, IEEE, pp. 338–341.
Park, K.-M., Kim, D., 2022. Preprocessing performance of convolutional neural networks according to characteristic of underwater targets. J. Acoust. Soc. Korea 41 (6), 629–636.
Pfau, A.M., 2020. Multi-Label Classification of Underwater Soundscapes Using Deep Convolutional Neural Networks. Technical Report, Naval Postgraduate School.
Porter, M.B., 1992. The KRAKEN Normal Mode Program. Technical Report, Naval Research Lab Washington DC.
Prasad, S.K., Gurugopinath, S., 2022. Supervised learning techniques for detection of underwater acoustic sources. In: 2022 2nd International Conference on Intelligent Technologies. CONIT, IEEE, pp. 1–7.
Prasad, S.K., Gurugopinath, S., 2023. Deep learning techniques for detection of underwater acoustic sources. In: 2023 11th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks. IEMECON, IEEE, pp. 1–6.
Premus, V.E., Evans, M.E., Abbot, P.A., 2020. Machine learning-based classification of recreational fishing vessel kinematics from broadband striation patterns. J. Acoust. Soc. Am. 147 (2), EL184–EL188.
Qi, P., Sun, J., Long, Y., Zhang, L., 2021. Underwater acoustic target recognition with fusion feature. In: Neural Information Processing: 28th International Conference, ICONIP 2021, Sanur, Bali, Indonesia, December 8–12, 2021, Proceedings, Part I 28. Springer, pp. 609–620.
Qiao, W., Khishe, M., Ravakhah, S., 2021. Underwater targets classification using local wavelet acoustic pattern and multi-layer perceptron neural network optimized by modified whale optimization algorithm. Ocean Eng. 219, 108415.
Qin, D., Tang, J., Yan, Z., 2020. Underwater acoustic source localization using LSTM neural network. In: 2020 39th Chinese Control Conference. CCC, IEEE, pp. 7452–7457.
Quan, T., Yang, X., Jingjing, W., 2021. DOA estimation of underwater acoustic array signal based on wavelet transform with double branch convolutional neural network. In: Proceedings of the 15th International Conference on Underwater Networks & Systems. pp. 1–2.
Ren, J., Huang, Z., Li, C., Guo, X., Xu, J., 2019. Feature analysis of passive underwater targets recognition based on deep neural network. In: OCEANS 2019-Marseille. IEEE, pp. 1–5.
Ren, J., Xie, Y., Zhang, X., Xu, J., 2022. UALF: A learnable front-end for intelligent underwater acoustic classification system. Ocean Eng. 264, 112394.
Santos-Domínguez, D., Torres-Guijarro, S., Cardenal-López, A., Pena-Gimenez, A., 2016. ShipsEar: An underwater vessel noise database. Appl. Acoust. 113, 64–69.
Satheesh, C., Kamal, S., Mujeeb, A., Supriya, M., 2021. Passive sonar target classification using deep generative *beta*-VAE. IEEE Signal Process. Lett. 28, 808–812.
Scherrer, R., Aulnette, E., Quiniou, T., Kasarherou, J., Kolb, P., Selmaoui-Folcher, N., 2022. Boat detection in marina using time-delay analysis and deep learning. Int. J. Data Warehousing Min. (IJDWM) 18 (2), 1–16.
Shen, S., Yang, H., Li, J., 2019. Improved auditory inspired convolutional neural networks for ship type classification. In: OCEANS 2019-Marseille. IEEE, pp. 1–4.
Shen, S., Yang, H., Li, J., Xu, G., Sheng, M., 2018. Auditory inspired convolutional neural networks for ship type classification with raw hydrophone data. Entropy 20 (12), 990.
Shen, S., Yang, H., Yao, X., Li, J., Xu, G., Sheng, M., 2020. Ship type classification by convolutional neural networks with auditory-like mechanisms. Sensors 20 (1), 253.
Sherin, B., Supriya, M., 2015. Selection and parameter optimization of SVM kernel function for underwater target classification. In: 2015 IEEE Underwater Technology. IEEE, pp. 1–5.
Siddagangaiah, S., Li, Y., Guo, X., Chen, X., Zhang, Q., Yang, K., Yang, Y., 2016. A complexity-based approach for the detection of weak signals in ocean ambient noise. Entropy 18 (3), 101.

Slamnoiu, G., Radu, O., Rosca, V., Pascu, C., Damian, R., Surdu, G., Curca, E., Radulescu, A., 2016. DEMON-type algorithms for determination of hydro-acoustic signatures of surface ships and of divers. IOP Conf. Ser.: Mater. Sci. Eng. 145, 082013.

Slaney, M., et al., 1993. An efficient implementation of the patterson-holdsworth auditory filter bank. Apple Comput. Percept. Group Tech. Rep 35 (8).

Smith, T.A., Rigby, J., 2022. Underwater radiated noise from marine vessels: A review of noise reduction methods and technology. Ocean Eng. 266, 112863.

Song, Y., Liu, F., Shen, T., 2023a. A novel noise reduction technique for underwater acoustic signals based on dual-path recurrent neural network. IET Commun. 17 (2), 135–144.

Song, Y., Liu, F., Shen, T., 2023b. PLDA in i-vector based underwater acoustic signals classification. Ships Offshore Struct. 1–9.

Song, G., Liu, X., Zeng, X., Luo, H., Wang, D., Zhang, B., 2020. A deep-shallow network for passive underwater target recognition. In: 2020 IEEE 22nd International Conference on High Performance Computing and Communications; IEEE 18th International Conference on Smart City; IEEE 6th International Conference on Data Science and Systems. HPCC/SmartCity/DSS, IEEE, pp. 802–807.

Song, Y., Liur, F., Shen, T., 2022. Method of underwater acoustic signal denoising based on dual-path transformer network. IEEE Access.

Song, Y., Shen, T., Liu, F., 2023c. Underwater acoustic signal noise reduction based on fully convolutional time domain separation network. Available at SSRN 4349171.

Sonz, W., Zhang, X., 2021. Feature extraction and classification of ship targets based on gammatone filter bank. In: 2021 IEEE International Conference on Signal Processing, Communications and Computing. ICSPCC, IEEE, pp. 1–4.

Souza Filho, J.B., de Seixas, J.M., 2016. Class-modular multi-layer perceptron networks for supporting passive sonar signal classification. IET Radar Sonar Nav. 10 (2), 311–317.

Sun, B., Luo, X., 2023. Underwater acoustic target recognition based on automatic feature and contrastive coding. IET Radar Sonar Nav..

Sun, Q., Wang, K., 2022. Underwater single-channel acoustic signal multitarget recognition using convolutional neural networks. J. Acoust. Soc. Am. 151 (3), 2245–2254.

Sun, X., Yin, X., Yin, Y., Liu, P., Wang, L., Tang, R., 2020. Underwater acoustic target recognition based on ReLU gated recurrent unit. In: 2020 6th International Conference on Robotics and Artificial Intelligence. pp. 41–46.

Thiem, N., 2020. Creating Underwater Sounds Using Generative Adversarial Networks (Ph.D. thesis). Monterey, CA; Naval Postgraduate School.

Tian, S., Bai, D., Zhou, J., Fu, Y., Chen, D., 2023a. Few-shot learning for joint model in underwater acoustic target recognition. Sci. Rep. 13 (1), 17502.

Tian, S.-Z., Chen, D.-B., Fu, Y., Zhou, J.-L., 2023b. Joint learning model for underwater acoustic target recognition. Knowl.-Based Syst. 260, 110119.

Tian, S., Chen, D., Wang, H., Liu, J., 2021. Deep convolution stack for waveform in underwater acoustic target recognition. Sci. Rep. 11 (1), 9614.

Tong, Y., Zhang, X., Ge, Y., 2020. Classification and recognition of underwater target based on MFCC feature extraction. In: 2020 IEEE International Conference on Signal Processing, Communications and Computing. ICSPCC, IEEE, pp. 1–4.

Tucker, S., Brown, G.J., 2005. Classification of transient sonar sounds using perceptually motivated features. IEEE J. Ocean. Eng. 30 (3), 588–600.

Urick, R., United States. Naval Sea Systems Command. Undersea Warfare Technology Office, 1984. Ambient noise in the sea. In: AD-a460 546, Undersea Warfare Technology Office, Naval Sea Systems Command, Department of the Navy, URL https://books.google.nl/books?id=V79dvgEACAAJ.

Uruba, V., 2019. Energy and entropy in turbulence decompositions. Entropy 21 (2), 124.

Van Komen, D.F., Neilsen, T.B., Howarth, K., Knobles, D.P., Dahl, P.H., 2020. Seabed and range estimation of impulsive time series using a convolutional neural network. J. Acoust. Soc. Am. 147 (5), EL403–EL408.

van Komen, D., Neilsen, T.B., Knobles, D.P., Badiey, M., 2019. A convolutional neural network for source range and ocean seabed classification using pressure time-series. In: Proceedings of Meetings on Acoustics 177ASA, vol. 36, Acoustical Society of America, p. 070004.

Van Komen, D.F., Neilsen, T.B., Knobles, D.P., Badiey, M., 2019. A feedforward neural network for source range and ocean seabed classification using time-domain features. In: Proceedings of Meetings on Acoustics 177ASA, vol. 36, Acoustical Society of America, p. 070003.

Van Komen, D.F., Neilsen, T.B., Mortenson, D.B., Acree, M.C., Knobles, D.P., Badiey, M., Hodgkiss, W.S., 2021. Seabed type and source parameters predictions using ship spectrograms in convolutional neural networks. J. Acoust. Soc. Am. 149 (2), 1198–1210.

Vaz, G., Correia, A., Vicente, M., Sousa, J., Cruz, E., Dommergues, B., 2022. Marine acoustic signature recognition using convolutional neural networks. Available at SSRN 4119910.

Veirs, S., Veirs, V., Wood, J.D., 2016. Ship noise extends to frequencies used for echolocation by endangered killer whales. PeerJ 4, e1657.

Vieira, M., Amorim, M.C.P., Sundelöf, A., Prista, N., Fonseca, P.J., 2020. Underwater noise recognition of marine vessels passages: Two case studies using hidden Markov models. ICES J. Mar. Sci. 77 (6), 2157–2170.

Wang, J., Chen, Z., 2019. Feature extraction of ship-radiated noise based on intrinsic time-scale decomposition and a statistical complexity measure. Entropy 21 (11), 1079.

Wang, P., Chen, M., Wang, J., Deng, X., Chen, Z., 2022a. Auditory-based multi-scale amplitude-aware permutation entropy as a measure for feature extraction of ship radiated noise. In: 2022 IEEE 6th Advanced Information Technology, Electronic and Automation Control Conference. IAEAC, IEEE, pp. 1550–1555.

Wang, N., He, M., Sun, J., Wang, H., Zhou, L., Chu, C., Chen, L., 2019a. IA-PNCC: Noise processing method for underwater target recognition convolutional neural network. Comput. Mater. Continua 58 (1), 169–181.

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

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

Wang, P., Peng, Y., 2020. Research on feature extraction and recognition method of underwater acoustic target based on deep convolutional network. In: 2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications. AEECA, IEEE, pp. 863–868.

Wang, M., Qiu, B., Zhu, Z., Ma, L., Zhou, C., 2022b. Passive tracking of underwater acoustic targets based on multi-beam LOFAR and deep learning. Plos one 17 (12), e0273898.

Wang, B., Wu, C., Zhu, Y., Zhang, M., Li, H., Zhang, W., et al., 2021. Ship radiated noise recognition technology based on ML-DS decision fusion. Comput. Intell. Neurosci. 2021.

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

Waskito, P., Miwa, S., Mitsukura, Y., Nakajo, H., 2010. Parallelizing Hilbert-huang transform on a GPU. In: 2010 First International Conference on Networking and Computing. IEEE, pp. 184–190.

Whitaker, S., Barnard, A., Anderson, G.D., Havens, T.C., 2021. Recurrent networks for direction-of-arrival identification of an acoustic source in a shallow water channel using a vector sensor. J. Acoust. Soc. Am. 150 (1), 111–119.

Wu, Z., Huang, N.E., 2009. Ensemble empirical mode decomposition: A noise-assisted data analysis method. Adv. Adapt. Data Anal. 1 (01), 1–41.

Wu, J., Li, P., Wang, Y., Lan, Q., Xiao, W., Wang, Z., 2023. VFR: The underwater acoustic target recognition using cross-domain pre-training with fbank fusion features. J. Mar. Sci. Eng. 11 (2), 263.

Wu, H., Song, Q., Jin, G., 2018. Deep learning based framework for underwater acoustic signal recognition and classification. In: Proceedings of the 2018 2nd International Conference on Computer Science and Artificial Intelligence. pp. 385–388.

Wu, H., Song, Q., Jin, G., 2020. Underwater acoustic signal analysis: Preprocessing and classification by deep learning. Neural Netw. World 30 (2), 85–96.

Wu, Y., Yang, Y., Tao, C., Tian, F., Yang, L., 2014. Robust underwater target recognition using auditory cepstral coefficients. In: OCEANS 2014-TAIPEI. IEEE, pp. 1–4.

Xiao, L., 2022. Feature extraction of ship-radiated noise based on hierarchical dispersion entropy. Shock Vib..

Xiao, X., Wang, W., Ren, Q., Zhao, M., Ma, L., 2021a. Source ranging using attention-based convolutional neural network. In: 2021 OES China Ocean Acoustics. COA, IEEE, pp. 1038–1042.

Xiao, X., Wang, W., Su, L., Guo, X., Ma, L., Ren, Q., 2021b. Localization of immersed sources by modified convolutional neural network: Application to a deep-sea experiment. Sensors 21 (9), 3109.

Xie, J., Chen, J., Zhang, J., 2018. DBM-based underwater acoustic source recognition. In: 2018 IEEE International Conference on Communication Systems. ICCS, IEEE, pp. 366–371.

Xie, D., Esmaiel, H., Sun, H., Qi, J., Qasem, Z.A., 2020a. Feature extraction of ship-radiated noise based on enhanced variational mode decomposition, normalized correlation coefficient and permutation entropy. Entropy 22 (4), 468.

Xie, D., Hong, S., Yao, C., 2021. Optimized variational mode decomposition and permutation entropy with their application in feature extraction of ship-radiated noise. Entropy 23 (5), 503.

Xie, Z., Lin, R., Wang, L., Zhang, A., Lin, J., Tang, X., 2023a. Data augmentation and deep neural network classification based on ship radiated noise. Front. Mar. Sci..

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

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

Xie, Y., Ren, J., Xu, J., 2023b. Guiding the underwater acoustic target recognition with interpretable contrastive learning. In: OCEANS 2023-Limerick. IEEE, pp. 1–6.

Xie, D., Sun, H., Qi, J., 2020b. A new feature extraction method based on improved variational mode decomposition, normalized maximal information coefficient and permutation entropy for ship-radiated noise. Entropy 22 (6), 620.

Xing, G., Liu, P., Zhang, H., Tang, R., Yin, Y., 2020. A two-stream network for underwater acoustic target classification. In: 2020 6th International Conference on Robotics and Artificial Intelligence. pp. 248–252.

Xu, K., Feng, M., Zhu, B., et al., 2022. Underwater acoustic classification using masked modeling-based swin transformer. J. Acoust. Soc. Am. 152 (4), A296.

Xu, F., Guo, Y., 2021. Research on feature extraction method of underwater acoustic passive target. In: 2021 IEEE 4th International Conference on Automation, Electronics and Electrical Engineering. IEEE, pp. 668–671.

Xu, J., Xie, Y., Wang, W., 2023. Underwater acoustic target recognition based on smoothness-inducing regularization and spectrogram-based data augmentation. Ocean Eng. 281, 114926.
Xue, L., Zeng, X., Jin, A., 2022. A novel deep-learning method with channel attention mechanism for underwater target recognition. Sensors 22 (15), 5492.
Yan, J., Sun, H., Cheng, E., Kuai, X., Zhang, X., 2017. Ship radiated noise recognition using resonance-based sparse signal decomposition. Shock Vib..
Yang, L., Chen, K., 2017. Performance comparison of two types of auditory perceptual features in robust underwater target classification. Acta Acust. United Acust. 103 (1), 56–66.
Yang, H., Huang, X., Liu, Y., 2023a. Infogan-enhanced underwater acoustic target recognition method based on deep learning. In: Proceedings of 2022 International Conference on Autonomous Unmanned Systems. ICAUS 2022, Springer, pp. 2705–2714.
Yang, H., Li, L., Li, G., 2020. A new denoising method for underwater acoustic signal. IEEE Access 8, 201874–201888.
Yang, H., Xu, G., Yi, S., Li, Y., 2019. A new cooperative deep learning method for underwater acoustic target recognition. In: OCEANS 2019-Marseille. IEEE, pp. 1–4.
Yang, S., Xue, L., Hong, X., Zeng, X., 2023b. A lightweight network model based on an attention mechanism for ship-radiated noise classification. J. Mar. Sci. Eng. 11 (2), 432.
Yang, S., Zeng, X., 2021. Combination of gated recurrent unit and network in network for underwater acoustic target recognition. In: INTER-NOISE and NOISE-CON Congress and Conference Proceedings, vol. 263, Institute of Noise Control Engineering, pp. 486–492.
Yang, H., Zheng, K., Li, J., 2022. Open set recognition of underwater acoustic targets based on GRU-CAE collaborative deep learning network. Appl. Acoust. 193, 108774.
Yangzhou, J., Ma, Z., Huang, X., 2019. A deep neural network approach to acoustic source localization in a shallow water tank experiment. J. Acoust. Soc. Am. 146 (6), 4802–4811.
Yao, Q., Wang, Y., Yang, Y., 2023. Underwater acoustic target recognition based on data augmentation and residual CNN. Electronics 12 (5), 1206.
Yi, Y., Tian, G., 2022. Feature extraction method of ship radiated noise based on BOA-VMD and slope entropy. Front. Phys. 1044.
Yin, X., Sun, X., Liu, P., Wang, L., Tang, R., 2020. Underwater acoustic target classification based on LOFAR spectrum and convolutional neural network. In: Proceedings of the 2nd International Conference on Artificial Intelligence and Advanced Manufacture. pp. 59–63.
Yu, L., Cheng, Y.-m., Song, L., Liu, Z.-g., Chen, K.-z., 2014. Underwater acoustic multi-target recognition algorithm based on hierarchical information fusion structure. In: 17th International Conference on Information Fusion. FUSION, IEEE, pp. 1–8.
Yuan, F., Ke, X., Cheng, E., 2019. Joint representation and recognition for ship-radiated noise based on multimodal deep learning. J. Mar. Sci. Eng. 7 (11), 380.
Yuanchao, X., Zhiming, C., Xiaopeng, K., 2023. Improved pitch shifting data augmentation for ship-radiated noise classification. Appl. Acoust. 211, 109468.
Yue, Z., Wei, K., Qing, X., 2005. A novel modeling and recognition method for underwater sound based on HMT in wavelet domain. In: AI 2004: Advances in Artificial Intelligence: 17th Australian Joint Conference on Artificial Intelligence, Cairns, Australia, December 4-6, 2004. Proceedings 17. Springer, pp. 332–343.
Yue, H., Zhang, L., Wang, D., Wang, Y., Lu, Z., 2017. The classification of underwater acoustic targets based on deep learning methods. In: 2017 2nd International Conference on Control, Automation and Artificial Intelligence. CAAI 2017, Atlantis Press, pp. 526–529.
Zare, M., Nouri, N.M., 2023. A novel hybrid feature extraction approach of marine vessel signal via improved empirical mode decomposition and measuring complexity. Ocean Eng. 271, 113727.
Zeng, X., Liu, X., Song, G., Wang, D., Luo, H., Zhang, B., 2020a. Adversarial training for underwater target recognition in complex marine conditions. In: 2020 IEEE 22nd International Conference on High Performance Computing and Communications; IEEE 18th International Conference on Smart City; IEEE 6th International Conference on Data Science and Systems. HPCC/SmartCity/DSS, IEEE, pp. 1174–1179.
Zeng, X., Lu, C., Li, Y., 2020b. A multi-task sparse feature learning method for underwater acoustic target recognition based on two uniform linear hydrophone arrays. In: INTER-NOISE and NOISE-CON Congress and Conference Proceedings, vol. 261, Institute of Noise Control Engineering, pp. 4404–4411.
Zeng, X., Wang, S., 2014. Underwater sound classification based on Gammatone filter bank and Hilbert-Huang transform. In: 2014 IEEE International Conference on Signal Processing, Communications and Computing. ICSPCC, IEEE, pp. 707–710.
Zeng, X., Wang, Q., Zhang, C., Cai, H., 2013. Feature selection based on ReliefF and PCA for underwater sound classification. In: Proceedings of 2013 3rd International Conference on Computer Science and Network Technology. IEEE, pp. 442–445.
Zhang, Q., Da, L., Zhang, Y., Hu, Y., 2021a. Integrated neural networks based on feature fusion for underwater target recognition. Appl. Acoust. 182, 108261.
Zhang, J., Ding, Y., 2020. Underwater target recognition based on spectrum learning with convolutional neural network. In: 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference. ITOEC, IEEE, pp. 1520–1523.
Zhang, H., Junejo, N.U.R., Sun, W., Chen, H., Yan, J., 2020a. Adaptive variational mode time-frequency analysis of ship radiated noise. In: 2020 7th International Conference on Information Science and Control Engineering. ICISCE, IEEE, pp. 1652–1656.
Zhang, W., Li, X., Zhou, A., Ren, K., Song, J., 2021b. Underwater acoustic source separation with deep bi-LSTM networks. In: 2021 4th International Conference on Information Communication and Signal Processing. ICICSP, IEEE, pp. 254–258.
Zhang, W., Lin, B., Yan, Y., Zhou, A., Ye, Y., Zhu, X., 2022a. Multi-features fusion for underwater acoustic target recognition based on convolution recurrent neural networks. In: 2022 8th International Conference on Big Data and Information Analytics. BigDIA, IEEE, pp. 342–346.
Zhang, Q., Lu, H., Sak, H., Tripathi, A., McDermott, E., Koo, S., Kumar, S., 2020b. Transformer transducer: A streamable speech recognition model with transformer encoders and rnn-t loss. In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, pp. 7829–7833.
Zhang, W., Wu, Y., Shi, J., Leng, H., Zhao, Y., Guo, J., 2022b. Surface and underwater acoustic source discrimination based on machine learning using a single hydrophone. J. Mar. Sci. Eng. 10 (3), 321.
Zhao, M., Zhong, S., Fu, X., Tang, B., Pecht, M., 2019. Deep residual shrinkage networks for fault diagnosis. IEEE Trans. Ind. Inform. 16 (7), 4681–4690.
Zhou, X., Yang, K., 2020. A denoising representation framework for underwater acoustic signal recognition. J. Acoust. Soc. Am. 147 (4), EL377–EL383.
Zhou, A., Zhang, W., Li, X., Xu, G., Zhang, B., Ma, Y., Song, J., 2023. A novel noise-aware deep learning model for underwater acoustic denoising. IEEE Trans. Geosci. Remote Sens. 61, 1–13.
Zhu, X., Dong, H., Rossi, P.S., Landrø, M., 2020. Feature selection based on principal component analysis for underwater source localization by deep learning. arXiv preprint arXiv:2011.12754.
Zhu, X., Dong, H., Rossi, P.S., Landrø, M., 2021a. Self-supervised underwater source localization based on contrastive predictive coding. In: 2021 IEEE Sensors. IEEE, pp. 1–4.
Zhu, X., Dong, H., Rossi, P.S., Landrø, M., 2022. Time-frequency fused underwater acoustic source localization based on contrastive predictive coding. IEEE Sens. J. 22 (13), 13299–13308.
Zhu, X., Dong, H., Salvo Rossi, P., Landrø, M., 2021b. Feature selection based on principal component regression for underwater source localization by deep learning. Remote Sens. 13 (8), 1486.
Zhu, P., Zhang, Y., Huang, Y., Lin, B., Zhu, M., Zhao, K., Zhou, F., 2023. SFC-sup: Robust two-stage underwater acoustic target recognition method based on supervised contrastive learning. IEEE Trans. Geosci. Remote Sens..