A Survey of Underwater Acoustic Target Recognition Using Deep Learning and Neural Networks

www.surveyx.cn

Abstract

The survey paper explores the transformative impact of deep learning and neural networks on underwater acoustic target recognition, emphasizing advancements in Convolutional Neural Networks (CNNs) and innovative frameworks such as the Convolution-based Mixture of Experts (CMoE). These technologies significantly enhance recognition accuracy by addressing challenges like intra-class diversity and inter-class similarity, achieving classification accuracies up to 96.32

1 Introduction

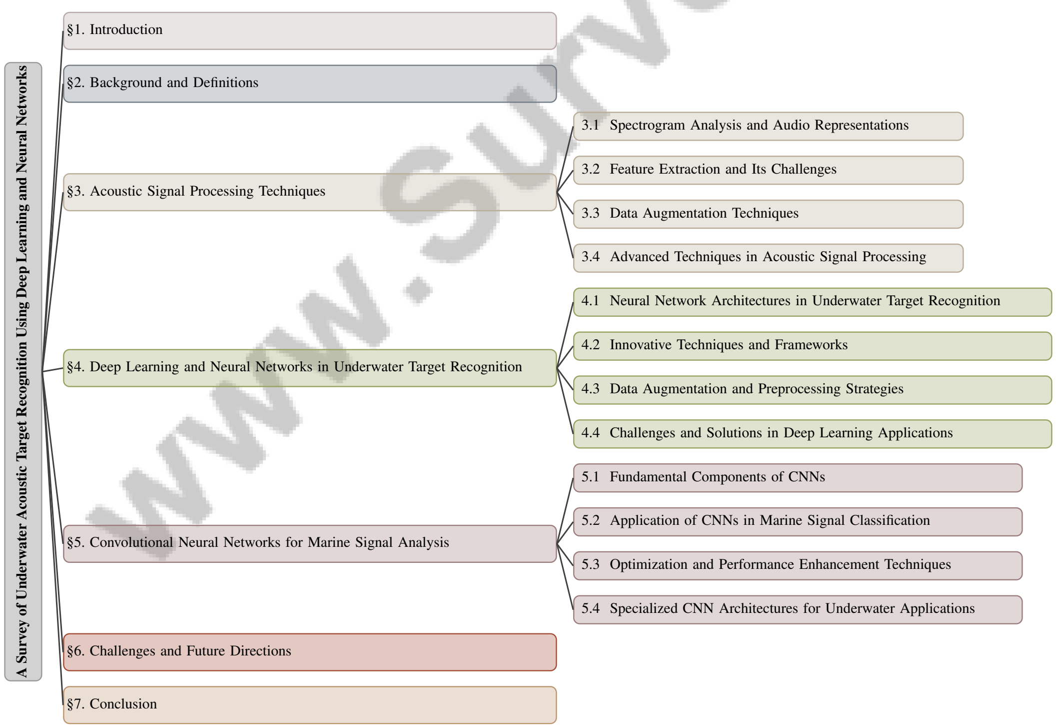


Figure 1: chapter structure

1.1 Importance of Underwater Acoustic Target Recognition

Underwater Acoustic Target Recognition (UATR) is pivotal for marine activities and military operations, facilitating the identification and classification of underwater objects [1]. Its applications span marine exploration, environmental monitoring, naval defense, and underwater navigation [2].

In ecological monitoring, UATR is essential for recognizing species such as fish and coral, which are critical for assessing biodiversity and marine ecosystem health [3]. Furthermore, recognizing ship-radiated signals enhances maritime safety and environmental monitoring [4].

Despite its significance, UATR presents challenges due to the complexities of underwater environments, including noisy signals, time-frequency interference, and the dynamic nature of acoustic signals [5]. The unpredictable transmission channels and motion states of underwater targets further complicate recognition accuracy and robustness [6], while the scarcity of publicly available datasets hinders the development of effective recognition models [7].

In military contexts, UATR is crucial for detecting evolving stealth technologies in underwater targets, which directly impacts national security through timely identification of underwater threats [1]. As advancements in UATR continue to support various scientific, operational, and defense-related activities, the demand for effective recognition systems remains a priority to address the complex challenges of the marine environment.

1.2 Advancements through Deep Learning and Neural Networks

Deep learning and neural networks have revolutionized underwater acoustic target recognition, enhancing feature extraction and classification accuracy. These advancements address challenges such as limited labeled data, variability in intrinsic characteristics, and noise interference. Techniques like Residual Networks and data augmentation methods, including SpecAugment, have achieved state-of-the-art recognition accuracy, exemplified by a 94.3

Convolutional Neural Networks (CNNs) are central to these advancements, effectively identifying complex patterns in acoustic data and enhancing recognition accuracy [3]. The integration of CNNs with frameworks such as convolution-based mixture of experts (CMoE) refines recognition processes by directing inputs to the most suitable expert layer, effectively addressing environmental variability challenges [4].

Additionally, deep neural networks (DNNs) have advanced audio source separation, although limitations related to specific sampling frequencies can restrict their applicability across diverse audio applications [8]. Innovative data augmentation techniques, such as multi-window spectral analysis (MWSA) combined with conditional deep convolutional generative adversarial networks (cDCGAN), enhance data diversity and improve classification accuracy with architectures like ResNet [7].

Transfer learning has emerged as a valuable strategy to mitigate the challenges of limited labeled data, allowing pre-trained models to adapt to specific underwater acoustic tasks, thus improving computational efficiency and accuracy [4]. Despite these advancements, challenges remain in enhancing the interpretability and generalization of deep learning models in practical applications, necessitating ongoing research to fully exploit the potential of these technologies in underwater target recognition [9].

1.3 Objectives of the Paper

This survey paper addresses the multifaceted challenges of underwater acoustic target recognition by exploring innovative methodologies and frameworks that enhance recognition accuracy in diverse environmental conditions [10]. A primary objective is to tackle the complexity of accurately recognizing underwater targets amidst variable acoustic environments, where noise and motion states contribute to classification inaccuracies. The proposed Convolution-based Mixture of Experts (CMoE) framework effectively manages high intra-class diversity and inter-class similarity in acoustic signals, improving recognition reliability [6].

The survey also aims to advance feature extraction and classification methodologies by introducing machine learning approaches leveraging CNNs for enhanced performance [3]. It facilitates model comparisons using benchmarks like QiandaoEAR22, crucial for identifying specific ships among multiple underwater targets [7]. The importance of adaptive data pruning and smoothness-inducing regularization is highlighted to overcome recognition challenges associated with ship-radiated signals, refining classification processes [4].

Additionally, the survey emphasizes developing a multimodal approach that integrates diverse data inputs to enhance recognition accuracy, addressing the inherent complexities of underwater acoustic environments [2]. By achieving these objectives, the paper aims to improve current methodologies and pave the way for future research and innovation in underwater acoustic target recognition.

1.4 Structure of the Survey

The survey is structured to provide a comprehensive exploration of underwater acoustic target recognition through deep learning and neural networks. It begins with an introduction outlining the field's importance and the advancements brought by these technologies. Following this, background information is presented, defining key concepts such as underwater acoustics, target recognition, and sonar technology [11].

Subsequent sections reflect the stages of signal acquisition, feature extraction, and recognition, offering insights into various techniques employed at each stage [12]. The survey discusses acoustic signal processing techniques, emphasizing spectrogram analysis, feature extraction challenges, and data augmentation methods. It then examines deep learning and neural network applications in underwater target recognition, highlighting innovative architectures and frameworks that enhance recognition capabilities [13].

A dedicated section focuses on the role of CNNs in marine signal analysis, exploring fundamental components, applications, and optimization techniques. The survey addresses challenges such as data scarcity, environmental variability, and computational limitations, proposing solutions and future research directions.

Finally, the paper concludes by summarizing key findings and reflecting on the transformative potential of deep learning and neural networks in underwater acoustic target recognition, underscoring the importance of continued research in this dynamic field [14].The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Introduction to Underwater Acoustics and Target Recognition

Underwater acoustics is foundational for understanding sound propagation in marine environments, critical for target recognition systems like sonar, which detect and classify underwater objects such as ships and submarines [4]. Environmental factors like temperature, salinity, and pressure significantly influence acoustic signal propagation, affecting the performance of recognition systems.

Recognizing underwater acoustic targets is challenging due to the marine environment's complexity and dynamism. Traditional methods often falter in noisy conditions, low signal-to-noise ratios, and high intra-class variability with inter-class similarity, complicating classification [4]. The development of robust recognition models is further impeded by the scarcity of comprehensive real-world datasets and the limitations of existing data augmentation techniques.

Incorporating machine learning into underwater acoustics shows promise for enhancing acoustic phenomena analysis and prediction. However, these methods are limited by the availability of labeled data and the inherent variability in acoustic signals. Real-world datasets, such as those capturing ship-radiated sounds, are crucial for training models to recognize specific underwater targets [4], but often lack the diversity to fully represent real-world variability, limiting model generalization.

Enhancing underwater target recognition accuracy and reliability requires integrating advanced recognition technologies with marine acoustic signal processing complexities. Considerations include sound propagation patterns, distance, channel depth, and the need for comprehensive data annotations. Recent advances, such as tri-modal contrastive learning frameworks and optimized feature extraction methods, improve recognition by leveraging diverse data perspectives and enriching model training with contextual information [15, 16, 5]. This integration is vital for addressing underwater acoustics' multifaceted challenges and enhancing sonar systems used in national defense, ocean exploration, and other critical applications.

2.2 Key Concepts in Acoustic Signal Processing

Underwater acoustic signal processing employs various techniques crucial for effective target recognition. Deep learning models, particularly Convolutional Neural Networks (CNNs) and Recurrent

Neural Networks (RNNs), are central to managing underwater soundscapes' complexities [11]. CNNs excel in identifying patterns and processing spectral representations, such as spectrograms, from raw audio samples [17], automating feature extraction for generative applications.

Long Short-Term Memory (LSTM) networks enhance temporal dependency capture in acoustic data, improving audio signal sequence processing [14]. Integrating multimodal features from diverse sources provides a comprehensive understanding of acoustic environments, surpassing unimodal approaches [2].

Hybrid models, like Mobile_ViT, which merges MobileNet's convolutional architecture with Transformers' self-attention mechanism, balance efficiency and performance in processing complex acoustic signals [18]. Techniques like the Dimension-wise Class Activation Map (dCAM) enhance model interpretability and accuracy by processing multivariate data dimensions simultaneously [19].

Machine learning and neural networks are pivotal in predicting environmental parameters and detecting anomalies in naval architecture, such as wave height predictions and offshore platform damage detection [20]. Recurrent neural networks provide robust solutions for sequence-based data processing, addressing training complexities in decoding convolutional codes [21].

The survey emphasizes self-supervised learning, semi-supervised fine-tuning, and few-shot learning as methodologies to enhance model performance with limited labeled samples, a common challenge in underwater acoustic signal processing [22]. Additionally, statistical models like Support Vector Machines (SVM) and Gaussian Mixture Models (GMM) complement deep learning approaches, particularly when high-quality underwater acoustic data is available [13]. These concepts collectively advance acoustic signal processing capabilities in underwater target recognition.

3 Acoustic Signal Processing Techniques

3.1 Spectrogram Analysis and Audio Representations

Method Name	Audio Representation	Methodological Integration	Application Versatility
ADP-SIR[4]	Spectrogram Embeddings	Spectrogram Analysis Integration	Underwater Target Recognition
UART[15]	Spectrograms	Tri-modal Contrastive	Underwater Target Recognition
SCC-CNN[3]	-	Advanced Feature Extraction	Coral Species Identification

Table 1: Overview of methodologies integrating spectrogram analysis and audio representations for underwater acoustic applications. The table compares different methods based on their audio representation techniques, methodological integration, and application versatility, highlighting their roles in enhancing underwater target recognition and species identification.

Spectrogram analysis is pivotal in underwater acoustic signal processing, providing a time-frequency representation critical for distinguishing underwater targets [4]. The Short-Time Fourier Transform (STFT) is frequently employed to transform time-domain signals into frequency-domain representations, facilitating the interpretation of acoustic signals. The selection of audio representation—whether raw waveforms, spectrograms, or Mel-frequency cepstral coefficients (MFCCs)—significantly impacts neural network performance, particularly in generating high-quality audio outputs. Spectrograms are favored for tasks like style transfer due to their ability to capture intricate audio patterns. Architectures such as SincNet illustrate that raw waveform modeling can achieve competitive results with fewer parameters, highlighting the importance of representation in optimizing neural network efficacy [17, 23, 24, 25, 26]. Table 1 presents a comparative analysis of various methodologies that utilize spectrogram analysis and audio representations in underwater acoustic signal processing, emphasizing their integration and application versatility.

Innovative methodologies like spectrogram embeddings enhance recognition by extracting meaningful features from training signals, thus improving underwater target recognition robustness [4]. Frameworks such as the tri-modal contrastive learning framework (UART) integrate spectrograms with audio and text inputs, enriching data perspectives and enhancing recognition capabilities [15]. The integration of spectrogram analysis with advanced preprocessing techniques, including Weber Local Descriptor, Phase Congruency, and ZCA Whitening, has been successful in applications like coral species identification, demonstrating these methods' versatility [3]. The QiandaoEAR22 dataset, with extensive records of ship-radiated and ambient noise, is crucial for developing spectrogram-based recognition models, emphasizing comprehensive datasets' role in advancing underwater acoustic research [7].

Combining spectrograms with advanced audio representations is essential for improving underwater acoustic target recognition systems' accuracy and effectiveness. These techniques enhance model generalization despite limited data and address underwater environments' complexities, which can lead to biased recognition outcomes. Innovative strategies such as smoothness-inducing regularization and specialized spectrogram-based data augmentation methods significantly advance marine signal analysis, facilitating more reliable maritime traffic monitoring and environmental assessments [27, 28, 29].

3.2 Feature Extraction and Its Challenges

Method Name	Feature Extraction Techniques	Challenges in Underwater Environments	Technological Advancements
JNN-UATR[14]	Convolution Pooling Layers	Environmental Noise Complexity	Joint Neural Network
CPECNN[30]	Cnns	Noise Interference	Transfer Learning
UATRM[31]	Mfcc Features	Noise Interference	Resnet18 Model
MSRDN[11]	Multiscale Convolution Layers	Interferential Noise	Multiscale Residual Units
UATR[1]	Multi-window Spectral	Environmental Noise Complexity	Deep Convolutional Gan

Table 2: Comparative analysis of various methods for feature extraction in underwater acoustic target recognition, highlighting the techniques employed, challenges faced in underwater environments, and technological advancements achieved. The table includes methods such as JNN-UATR, CPECNN, UATRM, MSRDN, and UATR, illustrating their distinct approaches to enhancing recognition performance.

Feature extraction is crucial in underwater acoustic target recognition, forming the foundation for accurate classification and analysis of acoustic signals. Capturing all necessary features from underwater acoustic signals using traditional neural networks often results in suboptimal recognition performance due to the complexity of underwater environments [14]. Convolutional Neural Networks (CNNs) address these challenges by identifying relevant characteristics from acoustic data to enhance classification accuracy. For example, CNNs have successfully extracted features from fish images, showcasing their potential in underwater applications [30]. In acoustic contexts, datasets are typically converted into spectrograms to facilitate classification, providing a visual representation of frequency components over time that aids in feature extraction [23]. Table 2 presents a comprehensive overview of feature extraction methods used in underwater acoustic target recognition, detailing the specific techniques, challenges encountered, and technological advancements associated with each approach.

Mel-frequency cepstral coefficients (MFCC) are another widely used technique for feature extraction, effectively capturing essential characteristics of underwater acoustic signals [31]. This method automates the extraction process, improving recognition systems' performance, particularly where manual feature design is common. The integration of multiscale residual units further enhances this process by capturing features at various scales, improving the model's adaptability to underwater complexities [11].

However, feature extraction is often hindered by the scarcity of ship-radiated noise samples available for training, limiting effective classifier development and leading to poor recognition performance [1]. Transfer learning has emerged as a promising strategy, allowing models trained on larger datasets to be adapted for specific underwater tasks, thereby enhancing feature extraction capabilities and improving classification outcomes.

Despite advancements in feature extraction techniques, including deep learning models and architectures like Residual Networks and Mobile_ViT, challenges persist in effectively capturing the diverse range of features required for accurate classification. These challenges arise from underwater complexities, limited labeled data, and noise interference, hindering the full leverage of advanced methods. Recent studies indicate that while some approaches have achieved impressive accuracy rates—such as 94.3

3.3 Data Augmentation Techniques

Data augmentation techniques are essential for enhancing model training in underwater acoustic target recognition by simulating diverse underwater signals and expanding data distribution [28]. These

AI-generated, for reference only.

techniques address data scarcity, a significant limitation in developing robust recognition systems due to the limited availability of labeled underwater acoustic datasets. By artificially expanding datasets, data augmentation improves deep learning models' generalization capabilities, enhancing their performance in real-world applications.

Primary data augmentation methods manipulate existing acoustic signals to create new training samples, employing techniques such as time-stretching, pitch-shifting, and noise addition. These introduce variability into the dataset, helping models become more resilient to diverse underwater conditions. Advanced strategies like smoothness-inducing regularization and specialized spectrogram-based data augmentation further address underwater environments' complexities and limited data availability, improving recognition models' generalization performance [32, 4, 16, 29, 28].

Generative models, such as Conditional Deep Convolutional Generative Adversarial Networks (cDCGAN), can generate realistic acoustic signals based on existing data, increasing both the volume and diversity of training data. This approach enhances robustness in recognition models, leading to improved performance in underwater acoustic target recognition tasks. Techniques like transformer-based style transfer and specialized data augmentation strategies leverage existing optical image datasets to generate high-fidelity pseudo-acoustic samples for training, addressing challenges posed by limited data availability [33, 27, 28, 18].

Multi-window spectral analysis (MWSA) significantly enhances the spectral representation of acoustic signals by providing diverse perspectives on the same dataset. This approach aids in meaningful feature extraction, improving recognition models' generalization capacity in complex underwater environments, where limited data availability poses challenges. By simulating various signal conditions, MWSA captures inter-class relationships, enabling models to learn robust patterns and perform effectively in real-world applications [17, 27, 28, 32]. This approach is particularly effective in scenarios with high intra-class diversity and inter-class similarity, allowing models to capture subtle differences and improve classification accuracy.

Data augmentation techniques are vital for enhancing underwater acoustic target recognition systems' performance, particularly in addressing limited and imbalanced datasets. Techniques such as smoothness-inducing regularization and specialized spectrogram-based methods, including local masking and replicating, simulate diverse underwater signals, improving recognition models' generalization capabilities. Given the complexities of underwater environments, where real data is often scarce, these strategies ensure models effectively learn from available data without being misled by simulated signals that may not accurately represent true conditions [16, 31, 27, 28]. By simulating diverse conditions and expanding the training dataset, these techniques enhance deep learning models' ability to generalize and perform accurately across various scenarios, advancing marine signal analysis.

3.4 Advanced Techniques in Acoustic Signal Processing

Method Name	Signal Processing Techniques	Innovative Models	Audio Source Separation
SMR-LMR[27]	Array Signal Processing	Smoothness-inducing Regularization	Local Masking Replicating
CNN-MR[34]	-	Convolutional Neural Networks	-
SN[26]	Convolutional Neural Network	Sincnet	-
DEMONet[10]	Cross-temporal Vae	Multi-expert Network	Demon Spectra
UART[16]	Data Preprocessing	Contrastive Learning Framework	Sfi Convolution Layer
DA[35]	-	Dourbal Algorithm	-
M3[32]	Multi-task Learning	Multi-expert Mechanisms	Multi-gate Mechanisms
SmoothReg-LMR[28]	Beamforming	Backbone Network	Sfi Convolution Layer
MSRDN[11]	Multiscale Convolution Layers	Deep Convolution Stack	Multiscale Residual Units

Table 3: This table categorizes advanced methodologies in acoustic signal processing, highlighting methods employed for signal processing, innovative modeling, and audio source separation. It provides a comparative overview of various techniques, including smoothness-inducing regularization, convolutional neural networks, and sampling-frequency-independent convolution layers, showcasing their application in underwater acoustic environments.

Advanced techniques in acoustic signal processing are crucial for overcoming challenges in underwater environments, characterized by complex signal propagation and significant background noise. Smoothness-inducing regularization is an innovative approach that mitigates overfitting by smoothing learned representations, enhancing models' robustness and generalization [27]. The

AI-generated, for reference only.

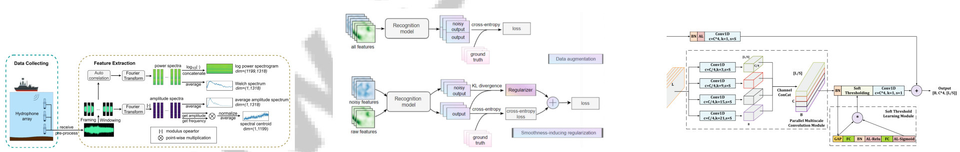
Local Masking and Replicating (LMR) strategy, a spectrogram-based data augmentation technique, diversifies training data and improves model performance [27].

Integrating convolutional neural networks (CNNs) with complex-valued signal processing marks a significant advancement, enabling direct processing of modulation patterns from complex-valued inputs without explicit feature extraction. This approach streamlines recognition and enhances accuracy, showcasing CNN adaptability to diverse signal processing tasks [34]. Additionally, the SincNet method exemplifies parameter-efficient models' potential, reducing parameters while maintaining performance in acoustic signal processing [26].

The DEMONet method, which integrates physical characteristics with discriminative time-frequency features, highlights domain-specific knowledge's importance in improving underwater acoustic recognition. This method underscores the necessity of combining advanced signal processing techniques with a deep understanding of the underlying physical phenomena [10]. The tri-modal contrastive learning approach utilized by UART enhances acoustic representation learning by incorporating descriptive natural language, enriching data perspectives available for model training [16].

The development of a sampling-frequency-independent (SFI) convolution layer facilitates audio source separation across various sampling frequencies. This innovation, achieved through the impulse invariant method, allows models to adapt to diverse sampling conditions, enhancing applicability in different acoustic scenarios [25]. The Dourbal algorithm contributes to computational efficiency by leveraging intermediate results for fast correlation vector calculations, streamlining recognition processes [35].

As illustrated in Figure 3, this figure categorizes advanced techniques in acoustic signal processing into three main areas: signal processing techniques, innovative models, and audio source separation methods. It highlights key methodologies such as smoothness regularization, complex-valued CNNs, and the SFI convolution layer. These advanced methodologies underscore innovative techniques' potential to address underwater acoustic environments' complex challenges, advancing signal processing systems' capabilities in marine applications. The continuous evolution and optimization of CNN architectures, as explored in various domains, including computer vision and natural language processing, further emphasize these technologies' transformative impact on acoustic signal processing [9]. Table 3 presents a comprehensive classification of advanced techniques in acoustic signal processing, emphasizing the integration of signal processing techniques, innovative models, and audio source separation methods in addressing underwater acoustic challenges.



(a) Hydrophone Array Data Processing Flowchart[32]

(b) Deep Learning Model for Image Recognition with Data Augmentation and Regularization[28]

(c) Parallel Multiscale Convolution Module with Soft Threshold Learning[11]

Figure 2: Examples of Advanced Techniques in Acoustic Signal Processing

Collectively, these examples underscore innovative approaches in acoustic signal processing, paving the way for more robust and precise applications [32, 28, 11].

4 Deep Learning and Neural Networks in Underwater Target Recognition

4.1 Neural Network Architectures in Underwater Target Recognition

Advancements in underwater target recognition have been significantly propelled by neural network architectures, especially Convolutional Neural Networks (CNNs), which excel in capturing spatial and temporal dependencies in noisy underwater settings [9, 8]. Architectures like the Multiscale Residual Deep Neural Network (MSRDN) utilize multiscale residual units to enhance feature capture across scales [5]. The Mobile_ViT architecture integrates CNNs for local feature extraction with Transformers for global analysis, optimizing efficiency in underwater contexts [18].

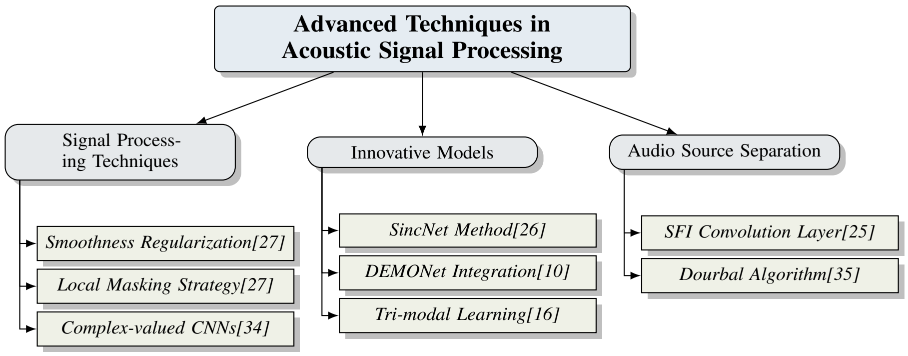
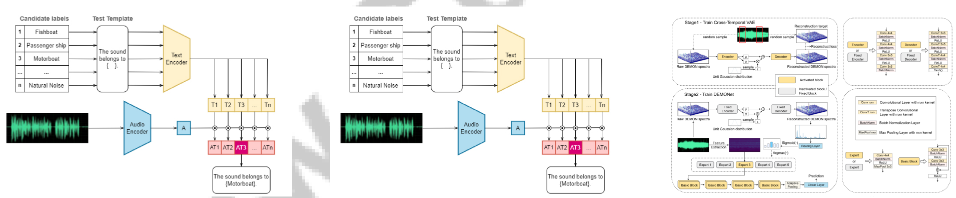


Figure 3: This figure illustrates the advanced techniques in acoustic signal processing, categorizing them into signal processing techniques, innovative models, and audio source separation methods, highlighting key methodologies like smoothness regularization, complex-valued CNNs, and SFI convolution layer.

SincNet employs CNNs with rectangular band-pass filters for raw waveform modeling, improving recognition accuracy in applications like speech recognition [26]. Integrated Contrastive Learning (ICL) enhances model robustness by extracting diverse spectrogram features [29]. Transfer learning, utilizing pre-trained Audio Neural Networks (PANNs) and ImageNet models, facilitates model adaptation for passive sonar tasks [23]. The UART framework replaces traditional paradigms with a contrastive learning approach, incorporating descriptive natural language for feature representation [15].

The SFI convolution layer adapts to varying sampling frequencies, enhancing model adaptability [25]. The CMoE framework dynamically assigns inputs to expert layers, improving classification accuracy in complex environments [6]. The evolution of neural network architectures, particularly CNNs and their integration with other networks, underscores their transformative role in advancing underwater target recognition [9].



(a) Audio Classification System[16]

(b) Audio Classification System[15]

(c) Deep Learning Framework for Spectral Analysis and Prediction[10]

Figure 4: Examples of Neural Network Architectures in Underwater Target Recognition

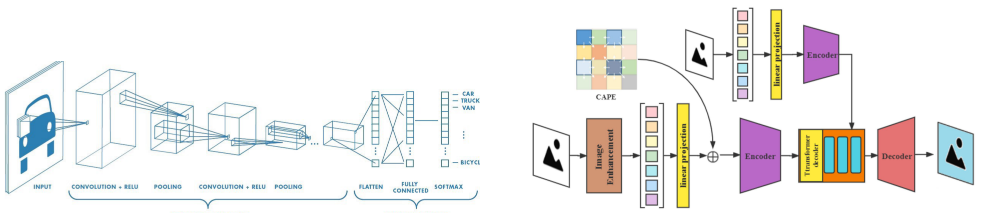
As shown in Figure 4, deep learning and neural networks are pivotal in underwater target recognition, enhancing audio classification and spectral analysis systems. The first two examples depict audio classification systems utilizing test templates, text encoders, and audio decoders to categorize underwater sounds. The third example highlights a deep learning framework for spectral analysis, emphasizing cross-temporal and cross-spectral analysis with a Variational Autoencoder (VAE). These architectures demonstrate the transformative potential of deep learning in advancing underwater target recognition [16, 15, 10].

4.2 Innovative Techniques and Frameworks

Innovative techniques and frameworks in neural networks have significantly advanced underwater acoustic target recognition, addressing marine environment challenges. Integrating CNNs with Recurrent Neural Networks (RNNs) captures both spatial and temporal features, enhancing recognition accuracy [9, 8]. The Convolution-based Mixture of Experts (CMoE) framework dynamically assigns expert layers to inputs, managing high intra-class diversity and inter-class similarity [6].

Transfer learning mitigates data scarcity, allowing models pre-trained on extensive datasets to adapt to specific tasks [23]. The UART framework employs tri-modal contrastive learning, integrating audio, text, and spectrogram inputs to enrich data perspectives [15]. Self-supervised learning techniques enable models to learn representations from unlabeled data, reducing reliance on labeled datasets [22].

These innovative techniques and frameworks highlight the transformative potential of neural networks in underwater acoustic target recognition, driving advancements in marine signal analysis [9].



(a) Convolutional Neural Network (CNN) Architecture[24]

(b) Deep Learning Framework for Image Enhancement and Recognition[33]

Figure 5: Examples of Innovative Techniques and Frameworks

As depicted in Figure 5, underwater target recognition has advanced through deep learning and neural networks, exemplified by innovative techniques and frameworks. The CNN architecture for image recognition features layers that process and extract features from input images, enhancing recognition accuracy. Complementing this is a deep learning framework for image enhancement, utilizing components like Contrast-Aware Pixel Enhancement (CAPE) to improve image quality and recognition capabilities, addressing challenges such as poor visibility and contrast in underwater environments [24, 33].

4.3 Data Augmentation and Preprocessing Strategies

Data augmentation and preprocessing strategies are vital for enhancing neural network performance in underwater acoustic target recognition. These strategies address data scarcity and diversity by expanding training datasets and improving model robustness across acoustic environments. Deep convolutional generative adversarial networks (DCGAN) generate realistic samples mimicking underwater signals, enhancing recognition accuracy [31].

Data augmentation involves manipulating acoustic signals to create variations through time-stretching, pitch-shifting, or noise addition, introducing variability that enhances model generalization. Multimodal data fusion and few-shot learning strategies enable effective recognition and localization despite noise and data scarcity, leveraging audio-visual-textual information and self-supervised learning frameworks for improved performance [2, 18, 22].

Preprocessing strategies prepare acoustic data for analysis, facilitating tasks like audio source separation across sampling frequencies, raw waveform modeling with SincNet, and data representation choices like spectrograms and MFCCs tailored for specific applications. Techniques like spectrogram generation transform time-domain signals into frequency-domain representations, capturing intricate underwater sound features, while advanced preprocessing methods like multi-window spectral analysis (MWSA) enhance spectral representation by offering multiple perspectives on the same data [25, 23, 26, 17].

Implementing data augmentation and preprocessing strategies is crucial for overcoming limited and imbalanced datasets in underwater acoustic target recognition. By enhancing data diversity through convolution-based mixtures of experts and advanced feature extraction methods like local-global feature fusion, these strategies develop robust neural network models designed to address challenges in underwater acoustic target recognition, achieving high accuracy and reliability in complex marine environments [4, 6, 18].

AI-generated, for reference only.

4.4 Challenges and Solutions in Deep Learning Applications

The application of deep learning in underwater acoustic target recognition faces challenges that hinder its effectiveness. The instability of underwater acoustic signals, subject to environmental fluctuations, complicates robust recognition system development, especially with limited annotated data [8]. High intra-class diversity and inter-class similarity can lead to misrecognition when data is scarce or targets exhibit similar characteristics [4].

CNNs face challenges with reliance on computationally intensive statistical tools, limiting real-time applicability. The double-descent phenomenon and overfitting arise from high similarity among training samples, degrading performance [8]. The interpretability of deep learning models remains a critical challenge, as they often function as "black boxes," obscuring prediction rationale. This lack of transparency is exacerbated by limited generalization ability in real-world environments, where models may struggle with new data. Traditional reliance on time-frequency representations can result in feature loss, impacting recognition accuracy [4].

Innovative solutions include advanced algorithms like CNNs for improved recognition and localization, multimodal approaches integrating audio-visual-textual information, and transfer learning leveraging pre-trained models for enhanced accuracy [9, 24, 2, 36, 12]. The Mobile_ViT method addresses feature homogenization and long-range dependencies, improving performance in acoustic scenarios. The Dourbal algorithm enhances computational efficiency, enabling quicker processing and minimizing resource demands, facilitating practical application of advanced neural networks in resource-limited environments [9, 37, 24, 38]. Innovative approaches proposed by Xu et al. enhance model robustness against noise and improve generalization, reducing reliance on high-quality augmented data.

While challenges persist, ongoing research yields innovative methodologies and frameworks addressing these obstacles. By confronting challenges like data limitations and complex environmental factors, advanced deep learning technologies can revolutionize marine signal analysis, enhancing accuracy and efficiency of recognition systems. Innovative frameworks like tri-modal contrastive learning and hybrid networks integrating local and global feature fusion improve recognition capabilities, facilitating better generalization across diverse underwater environments, leading to reliable and precise identification of marine targets [18, 33, 15, 16, 29].

5 Convolutional Neural Networks for Marine Signal Analysis

Convolutional Neural Networks (CNNs) are integral to marine signal analysis, offering robust frameworks for processing and interpreting complex acoustic data. This section examines the core components of CNNs, which are crucial for feature extraction and signal classification, facilitating their application in marine environments.

5.1 Fundamental Components of CNNs

CNNs, designed for grid-like data structures such as images and spectrograms, utilize convolutional operations to extract features [38]. The architecture of CNNs comprises convolutional, pooling, and fully connected layers, each playing a vital role in feature extraction and classification [24]. Convolutional layers detect patterns using filters that slide over the grid, generating feature maps through element-wise multiplications and summations. This hierarchical learning captures spatial hierarchies of features, enhancing performance in tasks like image recognition [36, 24, 38].

Pooling layers follow, reducing the spatial dimensions of feature maps to decrease computational load and mitigate overfitting. Techniques like max and average pooling distill salient features, improving model performance [25, 24]. Fully connected layers aggregate features from previous layers to produce final predictions, enabling effective classification of images and other data types [9, 24]. Advanced techniques, such as adversarial learning, have been integrated into CNN frameworks to enhance robustness in challenging environments like underwater acoustic target recognition. AMTNet, for example, uses multi-task learning with adversarial strategies to bolster recognition systems against acoustic variations [39].

Innovative methods, including linear pruning layers, optimize CNN performance by reducing model complexity while maintaining accuracy [4]. These advancements highlight CNNs' versatility in addressing marine signal analysis challenges.

5.2 Application of CNNs in Marine Signal Classification

CNNs have revolutionized marine signal classification by leveraging their powerful feature extraction capabilities and adaptability to complex data structures. CNNs process input data through layers applying convolutional filters, pooling operations, and activation functions to extract features and make predictions [38]. This layered approach enables CNNs to learn hierarchical representations, advantageous for processing intricate marine acoustic signals.

CNNs automate feature extraction, reducing reliance on manual methods that require domain expertise and computational resources [24]. By efficiently classifying marine signals, CNNs identify subtle patterns and variations often missed by manual methods, crucial in underwater environments with signal variability and noise challenges.

The effectiveness of CNNs in image recognition informs their adaptation for marine signal classification, incorporating advanced techniques such as data augmentation and residual architectures. Models like ResNet18, combined with data augmentation, significantly enhance recognition performance in underwater acoustic target recognition, addressing limited training samples. Hybrid approaches, such as Mobile_ViT, combine CNNs with Transformer architectures to improve local and global feature representation, achieving accuracy rates of 98.50

Furthermore, integrating CNNs with joint neural network methods has shown superior performance compared to traditional architectures [14]. These methods combine CNNs with other neural networks, such as RNNs, to capture spatial and temporal dependencies, providing comprehensive analysis and improving classification robustness in dynamic underwater environments.

The implementation of CNNs in marine signal classification underscores their significant impact on underwater acoustic target recognition, as evidenced by studies demonstrating improved accuracy through advanced techniques like data augmentation, hybrid networks, and innovative feature extraction methods. For instance, the ResNet18 model with data augmentation enhances recognition performance, while the Mobile_ViT framework effectively integrates feature fusion, achieving accuracy rates up to 98.50

5.3 Optimization and Performance Enhancement Techniques

Optimization and performance enhancement techniques are crucial for maximizing CNN effectiveness in marine signal analysis. These techniques refine network architecture and training processes to improve accuracy and computational efficiency. CNNs' ability to learn hierarchical feature representations directly from data reduces the need for manual extraction and enhances model performance [38].

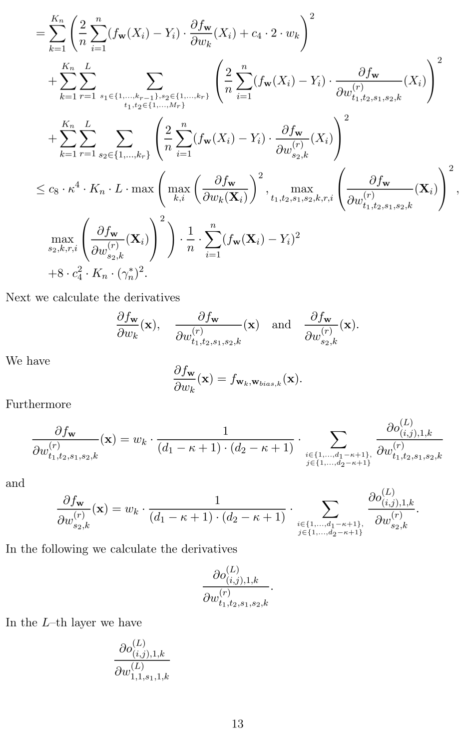
Strategies to optimize CNN performance include advanced learning rate schedules, such as cyclical learning rates, which adjust the learning rate dynamically during training to accelerate convergence and avoid local minima. This method enhances the training process by exploring a broader parameter space, improving generalization from limited data and increasing accuracy in tasks like underwater acoustic target recognition [27, 28, 24].

Regularization methods, such as dropout and batch normalization, enhance deep learning model performance. Dropout randomly deactivates neurons during training to prevent reliance on single features, mitigating overfitting. Batch normalization stabilizes learning by normalizing layer inputs, maintaining consistent activation distributions, and accelerating convergence [25, 27, 40, 24]. These methods are particularly effective in underwater environments, where data variability can lead to overfitting if not managed properly.

Transfer learning enhances CNN performance, especially with scarce labeled training data. It allows models to leverage knowledge from previously learned tasks to improve accuracy and generalization on new, related tasks [9, 37, 24]. By fine-tuning pre-trained models on large datasets for specific marine signal classification tasks, CNNs can enhance accuracy and reduce training time, beneficial in underwater applications with limited labeled datasets.

Network pruning and quantization significantly improve CNN efficiency by reducing memory footprint and computational demands while maintaining high accuracy. These methods streamline networks by removing redundant connections and reducing weight precision, enabling faster inference and deployment on resource-constrained devices, crucial for applications in computer vision and real-time processing [9, 24, 37, 36, 38]. These techniques make CNNs more suitable for deployment in resource-constrained environments, such as underwater monitoring systems.

The continuous development of optimization and performance enhancement techniques is essential for advancing CNN capabilities in marine signal analysis. By refining deep learning models' design and optimization, such as implementing Residual Networks and innovative data augmentation techniques, researchers can significantly improve underwater acoustic target recognition systems' accuracy and efficiency. Experimental results indicate accuracy rates exceeding 94.3



$$
\begin{aligned}
&= \sum_{k=1}^{K_n} \left(\frac{2}{n} \sum_{i=1}^{n} (f_{\mathbf{w}}(X_i) - Y_i) \cdot \frac{\partial f_{\mathbf{w}}}{\partial w_k}(X_i) + c_4 \cdot 2 \cdot w_k \right)^2 \\
&\quad + \sum_{k=1}^{K_n} \sum_{r=1}^{L} \sum_{\substack{s_1 \in \{1,\ldots,k_{r-1}\}, s_2 \in \{1,\ldots,k_r\} \\ t_1,t_2 \in \{1,\ldots,M_r\}}} \left(\frac{2}{n} \sum_{i=1}^{n} (f_{\mathbf{w}}(X_i) - Y_i) \cdot \frac{\partial f_{\mathbf{w}}}{\partial w^{(r)}_{t_1,t_2,s_1,s_2,k}}(X_i) \right)^2 \\
&\quad + \sum_{k=1}^{K_n} \sum_{r=1}^{L} \sum_{s_2 \in \{1,\ldots,k_r\}} \left(\frac{2}{n} \sum_{i=1}^{n} (f_{\mathbf{w}}(X_i) - Y_i) \cdot \frac{\partial f_{\mathbf{w}}}{\partial w^{(r)}_{s_2,k}}(X_i) \right)^2 \\
&\leq c_8 \cdot \kappa^4 \cdot K_n \cdot L \cdot \max \left(\max_{k,i} \left(\frac{\partial f_{\mathbf{w}}}{\partial w_k(\mathbf{X}_i)} \right)^2, \max_{t_1,t_2,s_1,s_2,k,r,i} \left(\frac{\partial f_{\mathbf{w}}}{\partial w^{(r)}_{t_1,t_2,s_1,s_2,k}}(\mathbf{X}_i) \right)^2, \right. \\
&\quad \left. \max_{s_2,k,r,i} \left(\frac{\partial f_{\mathbf{w}}}{\partial w^{(r)}_{s_2,k}}(\mathbf{X}_i) \right)^2 \right) \cdot \frac{1}{n} \cdot \sum_{i=1}^{n} (f_{\mathbf{w}}(\mathbf{X}_i) - Y_i)^2 \\
&\quad + 8 \cdot c_4^2 \cdot K_n \cdot (\gamma_n^*)^2.
\end{aligned}
$$

Next we calculate the derivatives

$$
\frac{\partial f_{\mathbf{w}}}{\partial w_k}(\mathbf{x}), \quad \frac{\partial f_{\mathbf{w}}}{\partial w^{(r)}_{t_1,t_2,s_1,s_2,k}}(\mathbf{x}) \quad \text{and} \quad \frac{\partial f_{\mathbf{w}}}{\partial w^{(r)}_{s_2,k}}(\mathbf{x}).
$$

We have

$$
\frac{\partial f_{\mathbf{w}}}{\partial w_k}(\mathbf{x}) = f_{\mathbf{w}_k, \mathbf{w}_{bias,k}}(\mathbf{x}).
$$

Furthermore

$$
\frac{\partial f_{\mathbf{w}}}{\partial w^{(r)}_{t_1,t_2,s_1,s_2,k}}(\mathbf{x}) = w_k \cdot \frac{1}{(d_1 - \kappa + 1) \cdot (d_2 - \kappa + 1)} \cdot \sum_{\substack{i \in \{1,\ldots,d_1-\kappa+1\}, \\ j \in \{1,\ldots,d_2-\kappa+1\}}} \frac{\partial o^{(L)}_{(i,j),1,k}}{\partial w^{(r)}_{t_1,t_2,s_1,s_2,k}}
$$

and

$$
\frac{\partial f_{\mathbf{w}}}{\partial w^{(r)}_{s_2,k}}(\mathbf{x}) = w_k \cdot \frac{1}{(d_1 - \kappa + 1) \cdot (d_2 - \kappa + 1)} \cdot \sum_{\substack{i \in \{1,\ldots,d_1-\kappa+1\}, \\ j \in \{1,\ldots,d_2-\kappa+1\}}} \frac{\partial o^{(L)}_{(i,j),1,k}}{\partial w^{(r)}_{s_2,k}}.
$$

In the following we calculate the derivatives

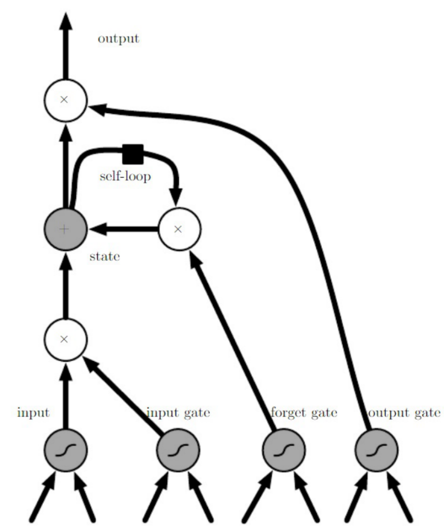
$$
\frac{\partial o^{(L)}_{(i,j),1,k}}{\partial w^{(r)}_{t_1,t_2,s_1,s_2,k}}.
$$

In the L–th layer we have

$$
\frac{\partial o^{(L)}_{(i,j),1,k}}{\partial w^{(L)}_{1,1,s_1,1,k}}
$$

13

(a) The image contains a mathematical equation and a table.[40]



(b) A neural network architecture with multiple gates and self-loop[20]

Figure 6: Examples of Optimization and Performance Enhancement Techniques

As shown in Figure 6, CNNs have emerged as powerful tools for optimizing performance in marine signal analysis. The first component illustrates a complex mathematical expression with a corresponding table elucidating numerical results, highlighting the analytical depth required for effective signal processing. The second component presents a sophisticated neural network architecture, featuring multiple gates and a self-loop, facilitating dynamic state transitions within the network. This architecture exemplifies advanced techniques in CNNs pivotal for optimizing performance in marine signal analysis, showcasing the interplay between theoretical foundations and practical innovations [40, 20].

5.4 Specialized CNN Architectures for Underwater Applications

Specialized CNN architectures address the unique challenges of underwater applications, where complex acoustic environments necessitate innovative solutions. The Multiscale Residual Deep Neural Network (MSRDN) employs multiscale residual units to classify underwater acoustic targets effectively by capturing features across different scales [5]. This approach enhances adaptability to underwater variability and noise, improving recognition accuracy.

AI-generated, for reference only.

The Mobile_ViT architecture combines local feature extraction capabilities of CNNs with global feature analysis from Transformers. This hybrid approach enhances recognition capabilities and efficiency, making it particularly suitable for underwater applications with limited computational resources [18]. Additionally, integrating CNNs with complex-valued signal processing allows for direct processing of modulation patterns from complex-valued inputs without explicit feature extraction [34].

SincNet, utilizing rectangular band-pass filters for raw waveform acoustic modeling, demonstrates potential for improving recognition accuracy in contexts like speech recognition [26]. Its parameter efficiency makes it attractive for underwater applications, where model size and computational demands are critical.

Advanced CNN architectures also incorporate domain-specific knowledge, as seen in the DEMONet method, which integrates physical characteristics with discriminative time-frequency features to enhance underwater acoustic recognition [10]. This underscores the importance of combining advanced signal processing techniques with a deep understanding of underlying physical phenomena.

Additionally, the development of a sampling-frequency-independent (SFI) convolution layer facilitates audio source separation across various sampling frequencies through the impulse invariant method. This innovation allows models to adapt to diverse sampling conditions, enhancing applicability in different acoustic scenarios [25].

These specialized CNN architectures showcase the potential of innovative designs to address the complex challenges of underwater acoustic environments, advancing the capabilities of signal processing systems in marine applications. The continuous evolution and optimization of CNN architectures, as explored in various domains, further emphasize the transformative impact of these technologies on acoustic signal processing [9].

6 Challenges and Future Directions

In underwater acoustic target recognition, overcoming challenges such as data scarcity and class imbalance is crucial for enhancing model efficacy and reliability. These issues significantly impact the performance and applicability of recognition systems in real-world scenarios. Figure 7 illustrates the hierarchical structure of these challenges and their corresponding mitigation strategies, focusing on data scarcity and class imbalance, environmental variability and signal complexity, computational limitations and model efficiency, as well as generalization and overfitting. This section delves into these challenges, their implications, and potential mitigation strategies.

6.1 Data Scarcity and Class Imbalance

Effective model development for underwater acoustic target recognition is hindered by data scarcity and class imbalance. The limited availability of labeled data restricts robust model training, essential for capturing the diverse variability of acoustic signals [4]. This scarcity is further complicated by the challenging nature of underwater environments, making data collection and annotation difficult [5]. Class imbalance skews model training towards more frequent classes, resulting in poor generalization to less represented classes, which is particularly problematic for rare marine species or specific underwater structures [7, 2]. Performance may decline when descriptive templates lack sufficient auxiliary information to compensate for data shortages [6].

To address these challenges, leveraging large amounts of unlabeled data through semi-supervised learning can improve model training and recognition performance [2]. Integrated Contrastive Learning (ICL) enhances recognition accuracy in data-scarce conditions by utilizing complementary features and providing regularization during training [6]. Adversarial multi-task learning frameworks also tackle data scarcity and class imbalance, boosting model performance [4]. The Dourbal algorithm, which reduces elementary operation numbers for correlation calculations, offers computational efficiency under these constraints [6]. By adopting innovative strategies and leveraging unlabeled data, researchers can develop more effective models for accurate underwater target recognition, enhancing marine signal analysis [5].

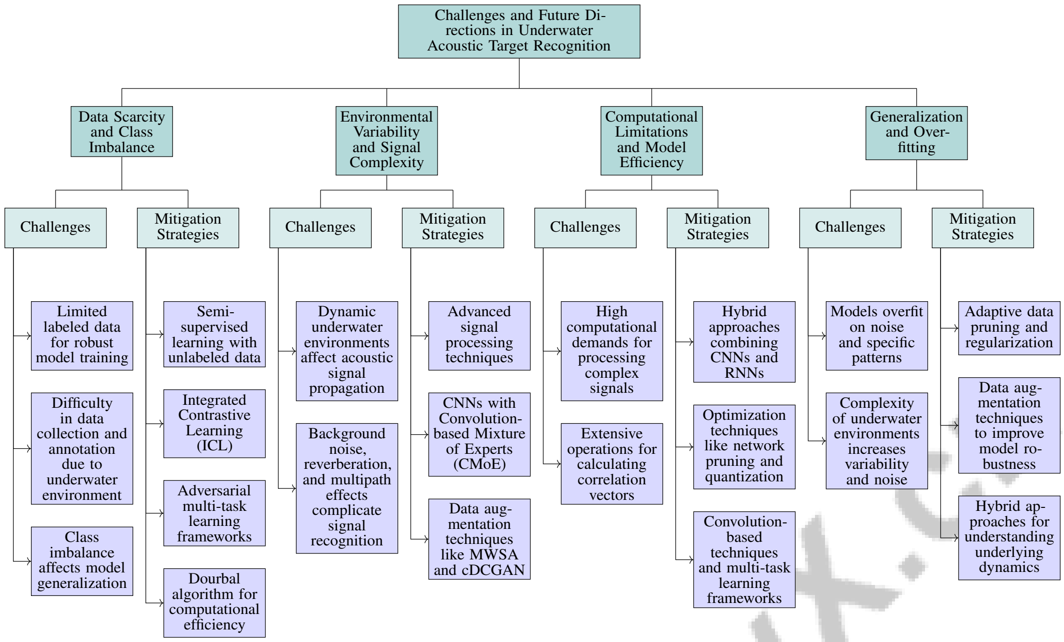


Figure 7: This figure illustrates the hierarchical structure of challenges and mitigation strategies in underwater acoustic target recognition, focusing on data scarcity and class imbalance, environmental variability and signal complexity, computational limitations and model efficiency, and generalization and overfitting.

6.2 Environmental Variability and Signal Complexity

Environmental variability and signal complexity significantly challenge underwater acoustic target recognition, affecting system accuracy and reliability. Underwater environments' dynamic nature, influenced by temperature, salinity, pressure, and currents, causes acoustic signal propagation fluctuations that complicate consistent recognition [4]. These environmental changes alter acoustic properties, hindering classification and reducing model effectiveness [5].

Background noise, reverberation, and multipath effects further exacerbate signal complexity, masking or distorting underwater targets' acoustic signatures. This complexity challenges traditional methods, which struggle to differentiate between target signals and noise, especially in environments with high intra-class diversity and inter-class similarity [6]. Robust models capable of adapting to varying conditions are necessary for accurate target identification.

Advanced signal processing techniques and machine learning models enhance recognition accuracy. Integrating Convolutional Neural Networks (CNNs) with frameworks like the Convolution-based Mixture of Experts (CMoE) improves recognition reliability by dynamically allocating inputs to the most appropriate expert based on characteristics [4]. This approach mitigates environmental variability impact by tailoring classification to specific input features.

Innovative data augmentation techniques, such as multi-window spectral analysis (MWSA) combined with Conditional Deep Convolutional Generative Adversarial Networks (cDCGAN), enhance data diversity and improve classification accuracy [7]. These techniques simulate a wide range of acoustic scenarios, allowing models to learn from comprehensive inputs, increasing resilience to environmental variability and signal complexity. Employing advanced methodologies and innovative frameworks enables researchers to develop robust models for accurate underwater target recognition, enhancing marine signal analysis [5].

6.3 Computational Limitations and Model Efficiency

Deploying deep learning models for underwater acoustic target recognition faces significant computational limitations due to the high demands of processing complex acoustic signals. Calculating

AI-generated, for reference only.

correlation vectors directly requires extensive operations for each new observation, demanding substantial computational resources [35]. This requirement can impede real-time processing and limit recognition systems' scalability in resource-constrained environments.

Advanced techniques, such as the Dimension-wise Class Activation Map (dCAM), increase computational complexity due to processing multiple input dimension permutations, affecting execution time in rapid processing applications [19]. Efficient algorithms and architectures are needed to deliver high performance without compromising speed or resource requirements.

Hybrid approaches combining different neural network architectures, such as CNNs with Recurrent Neural Networks (RNNs), demonstrate high accuracy in classifying dynamic systems with limited data points, highlighting their computational efficiency [8]. These approaches optimize feature extraction and classification processes, reducing computational overhead.

Optimization techniques, including network pruning and quantization, enhance deep learning model efficiency, particularly CNNs. By reducing parameters and simplifying data representation, these methods significantly decrease model size and computational requirements while maintaining accuracy. This optimization is beneficial in applications like image classification and computer vision, where efficient processing is essential for real-time performance [40, 36, 24, 4]. Removing redundant parameters and compressing networks make them suitable for deployment in resource-limited environments.

Addressing computational limitations and improving model efficiency are crucial for advancing underwater acoustic target recognition. Implementing convolution-based techniques, multi-task learning frameworks, and tri-modal contrastive learning approaches enhances neural network models' efficiency and accuracy for marine signal analysis. These strategies address underwater acoustic signals' complexities, characterized by high intra-class diversity and challenging environmental factors, improving recognition systems' robustness and scalability across various underwater scenarios. Optimized models deliver superior target recognition performance and expand machine learning applicability in diverse marine engineering contexts [15, 6, 20, 32].

6.4 Generalization and Overfitting

Generalization and overfitting present critical challenges in developing neural network models for underwater acoustic target recognition. These issues arise when models perform well on training data but fail with unseen data, often due to overfitting on noise and specific patterns rather than capturing the true data distribution [8]. The complexity of underwater environments, characterized by high variability and noise, exacerbates these challenges, necessitating robust solutions to improve model adaptability across diverse conditions.

Adaptive data pruning and regularization techniques effectively address overfitting, enhancing model generalization by selectively retaining informative data and reducing complexity [4]. These techniques prevent models from memorizing noise and irrelevant details, improving their ability to generalize to new data.

Moreover, the hybrid approach by Han et al. emphasizes understanding the underlying dynamics being classified, as misuse by non-specialists can lead to misinterpretation and reduced performance [8]. This highlights the need for models that are not only accurate but also interpretable, allowing users to grasp their predictions' basis.

Data augmentation remains crucial for enhancing model robustness, particularly in low signal-to-noise ratio (SNR) scenarios. By artificially expanding the training dataset through techniques like noise addition, time-stretching, and pitch-shifting, models can improve their ability to recognize patterns across diverse conditions. This approach enhances generalization and addresses challenges posed by limited data availability in complex environments, such as underwater acoustic recognition. Techniques like smoothness-inducing regularization and spectrogram-based data augmentation further support model robustness by ensuring simulated signals remain relevant to real-world scenarios, mitigating training process bias [23, 26, 28, 24].

Effectively tackling generalization and overfitting challenges in machine learning models requires a multifaceted strategy integrating innovative architectural designs, sophisticated regularization methods, and targeted data augmentation techniques, as evidenced by advancements in fields like computer vision and underwater acoustic target recognition [27, 24]. Leveraging these methodologies enables researchers to develop robust and reliable recognition systems capable of accurately identifying underwater targets across diverse environmental conditions.

7 Conclusion

The exploration of deep learning and neural networks in underwater acoustic target recognition reveals their substantial influence on enhancing classification accuracy and robustness. The deployment of Convolutional Neural Networks (CNNs) and the Convolution-based Mixture of Experts (CMoE) framework has proven effective in navigating the complexities of underwater acoustic environments, particularly by mitigating the challenges posed by intra-class diversity and inter-class similarity. Achieving a remarkable classification accuracy highlights the potential of these methodologies in surpassing traditional recognition approaches.

Moreover, the implementation of data augmentation strategies, notably with the ResNet18 model, has significantly contributed to improving recognition accuracy by diversifying the data inputs. The integration of pre-trained models and frameworks, such as dCAM, has further refined the identification of discriminative features, demonstrating their efficacy in underwater signal processing. The notable performance on datasets like ShipsEar underscores the advanced capabilities of these methods in acoustic target recognition.

This survey underscores the critical need for continued innovation and research in this interdisciplinary area. The development of novel strategies and methodologies not only enhances recognition accuracy on datasets with limited resources but also opens avenues for further advancements in the field. The demonstrated success of neural network-based approaches, including novel training methodologies, affirms their transformative role in advancing underwater acoustic target recognition.

References

[1] Xinwei Luo, Minghong Zhang, Ting Liu, Ming Huang, and Xiaogang Xu. An underwater acoustic target recognition method based on spectrograms with different resolutions. *Journal of Marine Science and Engineering*, 9(11):1246, 2021.

[2] Shipei Liu, Xiaoya Fan, and Guowei Wu. More perspectives mean better: Underwater target recognition and localization with multimodal data via symbiotic transformer and multiview regression, 2023.

[3] Mohamed Elawady. Sparse coral classification using deep convolutional neural networks, 2015.

[4] Yuan Xie, Tianyu Chen, and Ji Xu. Advancing underwater acoustic target recognition via adaptive data pruning and smoothness-inducing regularization, 2023.

[5] Feng Hong, Chengwei Liu, Lijuan Guo, Feng Chen, and Haihong Feng. Underwater acoustic target recognition with a residual network and the optimized feature extraction method. *Applied Sciences*, 11(4):1442, 2021.

[6] Yuan Xie, Jiawei Ren, and Ji Xu. Unraveling complex data diversity in underwater acoustic target recognition through convolution-based mixture of experts, 2024.

[7] Xiaoyang Du and Feng Hong. Introducing the brand new qiandaoear22 dataset for specific ship identification using ship-radiated noise, 2024.

[8] Daniel Han, Giuseppe Orlando, and Sergei Fedotov. Identification of the nature of dynamical systems with recurrence plots and convolution neural networks: A preliminary test, 2021.

[9] Zewen Li, Fan Liu, Wenjie Yang, Shouheng Peng, and Jun Zhou. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*, 33(12):6999–7019, 2021.

[10] Yuan Xie, Xiaowei Zhang, Jiawei Ren, and Ji Xu. Demonet: Underwater acoustic target recognition based on multi-expert network and cross-temporal variational autoencoder, 2024.

[11] Shengzhao Tian, Duanbing Chen, Hang Wang, and Jingfa Liu. Deep convolution stack for waveform in underwater acoustic target recognition. *Scientific reports*, 11(1):9614, 2021.

[12] Present status and challenges of.

[13] Xinwei Luo, Lu Chen, Hanlu Zhou, and Hongli Cao. A survey of underwater acoustic target recognition methods based on machine learning. *Journal of Marine Science and Engineering*, 11(2):384, 2023.

[14] Xing Cheng Han, Chenxi Ren, Liming Wang, and Yunjiao Bai. Underwater acoustic target recognition method based on a joint neural network. *Plos one*, 17(4):e0266425, 2022.

[15] Yuan Xie, Jiawei Ren, and Ji Xu. Underwater-art: Expanding information perspectives with text templates for underwater acoustic target recognition, 2024.

[16] Yuan Xie, Jiawei Ren, and Ji Xu. Underwater-art: Expanding information perspectives with text templates for underwater acoustic target recognition. *The Journal of the Acoustical Society of America*, 152(5):2641–2651, 2022.

[17] L. Wyse. Audio spectrogram representations for processing with convolutional neural networks, 2017.

[18] Haiyang Yao, Tian Gao, Yong Wang, Haiyan Wang, and Xiao Chen. Mobile_vit: underwater acoustic target recognition method based on local–global feature fusion. *Journal of Marine Science and Engineering*, 12(4):589, 2024.

[19] Paul Boniol, Mohammed Meftah, Emmanuel Remy, and Themis Palpanas. dcam: Dimension-wise class activation map for explaining multivariate data series classification, 2022.

[20] J P Panda. Machine learning for naval architecture, ocean and marine engineering, 2021.

[21] Daniel Tandler, Sebastian Dörner, Sebastian Cammerer, and Stephan ten Brink. On recurrent neural networks for sequence-based processing in communications, 2019.

[22] Shengzhao Tian, Di Bai, Junlin Zhou, Yan Fu, and Duanbing Chen. Few-shot learning for joint model in underwater acoustic target recognition. *Scientific Reports*, 13(1):17502, 2023.

[23] Amirmohammad Mohammadi, Tejashri Kelhe, Davelle Carreiro, Alexandra Van Dine, and Joshua Peeples. Transfer learning for passive sonar classification using pre-trained audio and imagenet models, 2024.

[24] Sumit Saha. A comprehensive guide to convolutional neural networks—the eli5 way. *Towards data science*, 15:15, 2018.

[25] Koichi Saito, Tomohiko Nakamura, Kohei Yatabe, Yuma Koizumi, and Hiroshi Saruwatari. Sampling-frequency-independent audio source separation using convolution layer based on impulse invariant method, 2021.

[26] Joachim Fainberg, Ondřej Klejch, Erfan Loweimi, Peter Bell, and Steve Renals. Acoustic model adaptation from raw waveforms with sincnet, 2019.

[27] Ji Xu, Yuan Xie, and Wenchao Wang. Underwater acoustic target recognition based on smoothness-inducing regularization and spectrogram-based data augmentation. *Ocean Engineering*, 281:114926, 2023.

[28] Ji Xu, Yuan Xie, and Wenchao Wang. Underwater acoustic target recognition based on smoothness-inducing regularization and spectrogram-based data augmentation, 2024.

[29] Yuan Xie, Jiawei Ren, and Ji Xu. Guiding the underwater acoustic target recognition with interpretable contrastive learning, 2024.

[30] Andre G. C. Pacheco. Classificação de espécies de peixe utilizando redes neurais convolucional, 2019.

[31] Qihai Yao, Yong Wang, and Yixin Yang. Underwater acoustic target recognition based on data augmentation and residual cnn. *Electronics*, 12(5):1206, 2023.

[32] Yuan Xie, Jiawei Ren, Junfeng Li, and Ji Xu. Advancing robust underwater acoustic target recognition through multi-task learning and multi-gate mixture-of-experts, 2024.

[33] Xiaoteng Zhou, Changli Yu, Shihao Yuan, Xin Yuan, Hangchi Yu, and Citong Luo. Learning visual representation of underwater acoustic imagery using transformer-based style transfer method, 2022.

[34] Timothy J O'Shea, Johnathan Corgan, and T. Charles Clancy. Convolutional radio modulation recognition networks, 2016.

[35] Pavel Dourbal and Mikhail Pekker. Fast calculation of correlations in recognition systems, 2016.

[36] Sascha Saralajew, Lars Holdijk, Maike Rees, and Thomas Villmann. Prototype-based neural network layers: Incorporating vector quantization, 2019.

[37] Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, Jianfei Cai, et al. Recent advances in convolutional neural networks. *Pattern recognition*, 77:354–377, 2018.

[38] Jianxin Wu. Introduction to convolutional neural networks. *National Key Lab for Novel Software Technology. Nanjing University. China*, 5(23):495, 2017.

[39] Yuan Xie, Ji Xu, Jiawei Ren, and Junfeng Li. Adversarial multi-task underwater acoustic target recognition: towards robustness against various influential factors, 2024.

[40] Michael Kohler, Adam Krzyzak, and Benjamin Walter. Analysis of the rate of convergence of an over-parametrized convolutional neural network image classifier learned by gradient descent, 2024.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.

AI-generated, for reference only.