Journal of
Marine Science and Engineering

Article

Multi-Scale Frequency-Adaptive-Network-Based Underwater Target Recognition

Lixu Zhuang [1], Afeng Yang [1,*], Yanxin Ma [2] and David Day-Uei Li [3]

[1] School of Communication Engineering, Hangzhou Dianzi University, Hangzhou 310005, China; 2780909232xx@gmail.com
[2] College of Meteorology and Oceanography, National University of Defense Technology, Changsha 410073, China; mayanxin@nudt.edu.cn
[3] Department of Biomedical Engineering, University of Strathclyde, Glasgow G1 1XQ, UK; david.li@strath.ac.uk
* Correspondence: yangafeng@hdu.edu.cn

Abstract: Due to the complexity of underwater environments, underwater target recognition based on radiated noise has always been challenging. This paper proposes a multi-scale frequency-adaptive network for underwater target recognition. Based on the different distribution densities of Mel filters in the low-frequency band, a three-channel improved Mel energy spectrum feature is designed first. Second, by combining a frequency-adaptive module, an attention mechanism, and a multi-scale fusion module, a multi-scale frequency-adaptive network is proposed to enhance the model's learning ability. Then, the model training is optimized by introducing a time–frequency mask, a data augmentation strategy involving data confounding, and a focal loss function. Finally, systematic experiments were conducted based on the ShipsEar dataset. The results showed that the recognition accuracy for five categories reached 98.4%, and the accuracy for nine categories in fine-grained recognition was 88.6%. Compared with existing methods, the proposed multi-scale frequency-adaptive network for underwater target recognition has achieved significant performance improvement.

Keywords: underwater target recognition; Mel energy spectrum; frequency adaptation; attention mechanism; multi-scale fusion

check for updates

Citation: Zhuang, L.; Yang, A.; Ma, Y.; Li, D.D.-U. Multi-Scale Frequency-Adaptive-Network-Based Underwater Target Recognition. *J. Mar. Sci. Eng.* **2024**, *12*, 1766. https://doi.org/10.3390/jmse12101766

Academic Editor: Mohamed Benbouzid

Received: 29 August 2024
Revised: 30 September 2024
Accepted: 30 September 2024
Published: 5 October 2024

1. Introduction

Underwater target recognition utilizes sonar-received radiated noise or echo signals to identify target objects [1]. These targets include various underwater entities, such as ships, schools of fish, and submarines, and information regarding their locations and motion states. Radiated noise features excellent concealment and long operational ranges [2], making it a crucial method for underwater target recognition. However, the complexity of underwater environments, the low signal-to-noise ratio (SNR) characteristics of radiated noise [3], and the adversarial nature of underwater targets significantly increase the difficulty of effective recognition. Consequently, underwater target recognition remains a highly challenging task.

Underwater target recognition based on radiated noise mainly involves two primary processes: feature extraction and classifier design [4]. Traditional feature extraction techniques encompass time-domain features, such as zero crossing rates [5], energy spectrum [6], and autocorrelation [7], and frequency-domain features like LOFAR [8] and DEMON [9]. With rapid advances in computing technologies, it has been discovered that features extracted using short-time Fourier transform (STFT) [10], wavelet transform [11], and other time–frequency-domain methods exhibit superior applicability and representational capability for target recognition. Consequently, researchers have developed auditory features based on auditory perception principles [12]. Mel frequency spectrum diagrams were extracted utilizing Mel filter banks, leveraging the human ear's high resolution for low-frequency sounds and lower resolution for high-frequency sounds [13]. Wang et al. reported high accuracy of 94.3% using gamma-frequency cepstral coefficients (GFCC) [14]. Since most useful information in underwater signals is concentrated in low frequencies, time–frequency features based on auditory characteristics have demonstrated more potent information representation capabilities for radiated noise.

In recent years, deep learning underwater target recognition has become a trend [15]. Liu et al. utilized a one-dimensional convolution neural network (1D CNN) model to identify the envelope modulation spectrum of underwater target radiated noise [9], demonstrating good generalization ability. Hong et al. combined features and an 18-layer residual network for underwater target classification [16]. Jin et al. introduced a sparse autoencoder (SAE) to recognize and classify underwater acoustic signals [17]. Han et al. used a 1D CNN and a long short-term memory (LSTM) network [18], improving recognition accuracy. Unlike traditional classifiers, deep learning requires only input rich in target characteristics and learns through nonlinear decision-making [19]. Multi-scale fusion [20] and attention mechanisms [21] are two crucial deep learning concepts. Multi-scale fusion involves merging features at different scales, including both semantic and spatial features, to mitigate information loss during convolution. Yan et al. proposed a multi-scale asymmetric CNN [22], which conducts multi-resolution analysis to extract deeper multi-scale features from the time and frequency domains, thus enhancing accuracy. Attention mechanisms can focus on crucial information in input data and ignore irrelevant information, improving the model performance. Liu et al. introduced a channel attention mechanism into the model [23], enhancing the target's intrinsic features. Fei et al. proposed a residual attention CNN [24], achieving high accuracy.

In summary, current underwater target recognition techniques still focus on two key steps: feature extraction and classifier design. In feature extraction, the research has evolved from traditional statistical features to bioinspired features mimicking the human auditory system. Regarding classifier design, the field has gradually transitioned from traditional expert systems towards deep learning methods. The approaches for underwater target recognition have evolved from conventional mathematical computations to artificial intelligence [25].

In this manuscript, we have innovated in feature extraction and classifier design by introducing an enhanced three-channel Mel energy spectrum feature for radiation noise and proposing a multi-scale frequency-adaptive network. To obtain more comprehensive frequency information, we have devised three Mel filters with varying distribution densities across the entire frequency range to obtain Mel energy spectrum features with diverse frequency resolutions. Drawing inspiration from multi-scale fusion and attention mechanisms, we have introduced a multi-scale frequency-adaptive network, which employs visual geometry group (VGG) as the primary network for feed-forward operations, with the network core containing a multi-scale fusion module. Within the multi-scale fusion module are three branches of frequency-adaptive residual modules and a spatial and channel squeeze-and-excitation module. The frequency-adaptive module applies two different-sized frequency pooling kernels to the frequency enhancement algorithm for adaptive computation and frequency attention learning. Integrating multi-scale fusion, attention mechanisms, and adaptive computation enhances the model's learning ability and improves recognition performance. Additionally, we have introduced a sample augmentation strategy based on time–frequency masking and data confusion and a focal loss function to optimize the network architecture during the model learning process and enhance performance on challenging samples. The main contributions of this study are summarized as follows:

(1) Based on the varying distribution densities of Mel filters across different frequency ranges, we designed a three-channel improved Mel energy spectrum feature. This fused feature exhibits superior inter-class separability.

(2) We propose a multi-scale frequency-adaptive network to enhance the model's learning capability. This model employs a frequency-adaptive module and an attention mechanism, combined with a multi-scale fusion module.

(3) To address imbalanced sample distribution problems, we introduce a sample augmentation strategy based on a time–frequency mask, data confusion, and a focal loss function. This improves the model's performance on complex samples.

(4) Systematic experiments were conducted using the ShipsEar dataset to evaluate the improved features and the proposed model. The results demonstrate that the recognition accuracy reached 98.4% for five classes and 88.6% for nine classes, outperforming existing models.

The organization of this paper is as follows: Section 2 details the methodology, including feature extraction and model design; Section 3 covers the experimental analysis, detailing the experimental data, settings, and a comprehensive analysis of the results.

2. Materials and Methods

2.1. Feature Extraction

The Mel energy spectrum is a spectrogram derived from the short-time spectrum via a Mel filter bank. These filters are designed based on the human auditory system, which is more sensitive to lower frequencies and less sensitive to higher frequencies. A Mel filter bank typically consists of multiple triangular band-pass filters with a dense distribution in the low-frequency range and a sparse distribution in the high-frequency range. Adjacent triangular band-pass filters overlap, each responding to a specific frequency range [26]. Therefore, the Mel energy spectrum retains more low-frequency information, benefiting target classification and recognition. The particular steps for extracting the Mel energy spectrum features are as follows.

(1) Conduct short-time Fourier transform to the signal $s(n)$ to obtain the amplitude spectrum $|S(t,f)|$:

$$|S(t,f)| = \left| \sum_{m=-\infty}^{+\infty} s(n)w(m-n)e^{-\frac{j2\pi kn}{N}} \right|, \ 0 \le k \le N-1, \tag{1}$$

where t and f are the time and the, and $w(n)$ is the Hamming window.

(2) Obtain the energy density function $P(t,f)$ of $s(n)$ based on $|S(t,f)|$:

$$P(t,f) = S(t,f) \times S^*(t,f), \tag{2}$$

where $S^*(t,f)$ is the complex conjugate of $S(t,f)$.

(3) Translate the energy density function $P(t,f)$ through a set of Mel filters; the energy density function $\widetilde{P}_{mel}(t,f)$ after filtering is expressed as

$$\widetilde{P}_{mel}(t,f) = \sum_{m=1}^{N} \sum\nolimits_{k=f_{m-1}}^{f_{m+1}} P(t,f)H_m[k], \tag{3}$$

where N represents the number of triangular bandpass filters in the Mel filter bank, and $H_m[k]$ is the frequency response function of a triangular bandpass filter with a center frequency f_m and a response frequency range of (f_{m-1}, f_{m+1}). $H_m[k]$ is defined as follows:

$$H_m[k] = \begin{cases} 0 & k < f_{m-1} \text{ or } k > f_{m+1} \\ \frac{2(k-f_{m-1})}{(f_{m+1}-f_{m-1})(f_m-f_{m-1})} & f_{m-1} \le k \le f_m \\ \frac{2(f_{m+1}-k)}{(f_{m+1}-f_{m-1})(f_m-f_{m-1})} & f_m \le k \le f_{m+1}, \end{cases} \tag{4}$$

where the center frequency f_m of the Mel filter can be obtained from the corresponding frequency f:

$$f_m = 2959 \times \log_{10}(1 + \frac{f}{700}). \tag{5}$$

(4) Transform the energy density function $\widetilde{P}_{mel}(t, f)$ of the Mel-filtered signal into a logarithmic scale to obtain the logarithmic Mel energy spectrum $P_{mel}(t, f)$:

$$P_{mel}(t, f) = \log_{10}(\widetilde{P}_{mel}(t, f)). \tag{6}$$

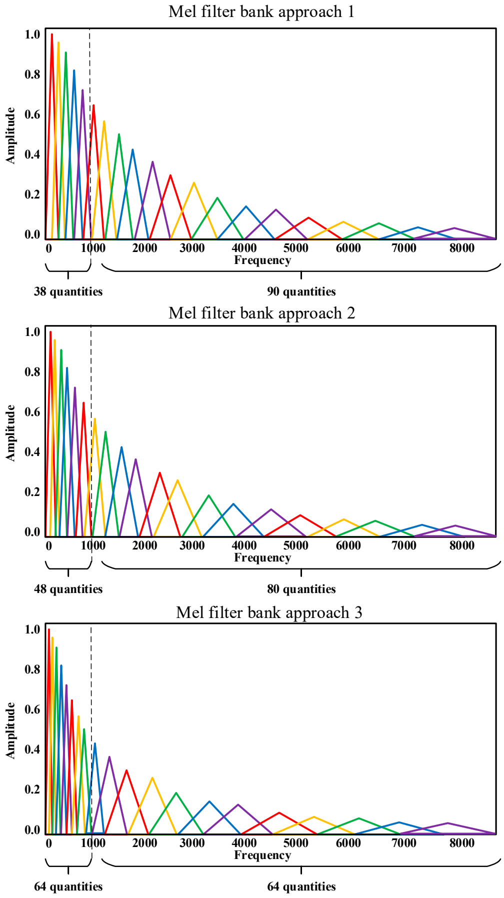
The characteristic information of ship-radiated noise is primarily concentrated in the low-frequency range. To obtain more comprehensive low-frequency information, we adjust the number of Mel filters in the high-frequency and low-frequency ranges while extracting Mel energy spectrum features. We designed different feature extraction schemes by increasing the number of Mel filters in the low-frequency range.

A frequency of 1000 Hz is set as the boundary between the low and high-frequency ranges. The total number of Mel filters is N, the number of Mel filters in the low-frequency range is denoted as N_l, and the center frequency of the Mel filters in the low-frequency range is denoted as $f_l = [f_1, f_2, \ldots, f_{N_l}]$. Similarly, the number of Mel filters in the high-frequency range is denoted as N_h, and the center frequency of the Mel filters in the high-frequency range is denoted as $P_{mel}(t, f) = [P_{mel}(t, f_l), P_{mel}(t, f_h)]$. The Mel energy spectrum feature is denoted as $P_{mel}(t, f) = [P_{mel}(t, f_l), P_{mel}(t, f_h)]$. Figure 1 presents different design schemes for the Mel filter bank The extraction schemes are as follows:

(1) The default scheme consists of 128 Mel filters, with 38 filters in the low-frequency range.

(2) The number of Mel filters is adjusted to 48 in the low-frequency range and 80 in the high-frequency range.

(3) The number of Mel filters in both the low- and high-frequency ranges is 64 each.


Mel filter bank approach 1

Mel filter bank approach 2

Mel filter bank approach 3

Figure 1. The different design schemes for the Mel filter bank.

Figure 2 illustrates the three-channel improved extraction process of the Mel energy spectrum. The Mel energy spectrum features $P_{mel}(t, f)$ is denoted as P, and the features generated by the three schemes are denoted as $P_{default}, P_{low48}, P_{low64}$, respectively. Combining the Mel energy spectrum features from the three schemes, the three-channel improved Mel energy spectrum graph feature $P_{fused} = [P_{default}, P_{low48}, P_{low64}]$ was constructed. Compared to a single Mel energy spectrum feature, the three-channel feature provides richer frequency distribution information, enhancing recognition accuracy.

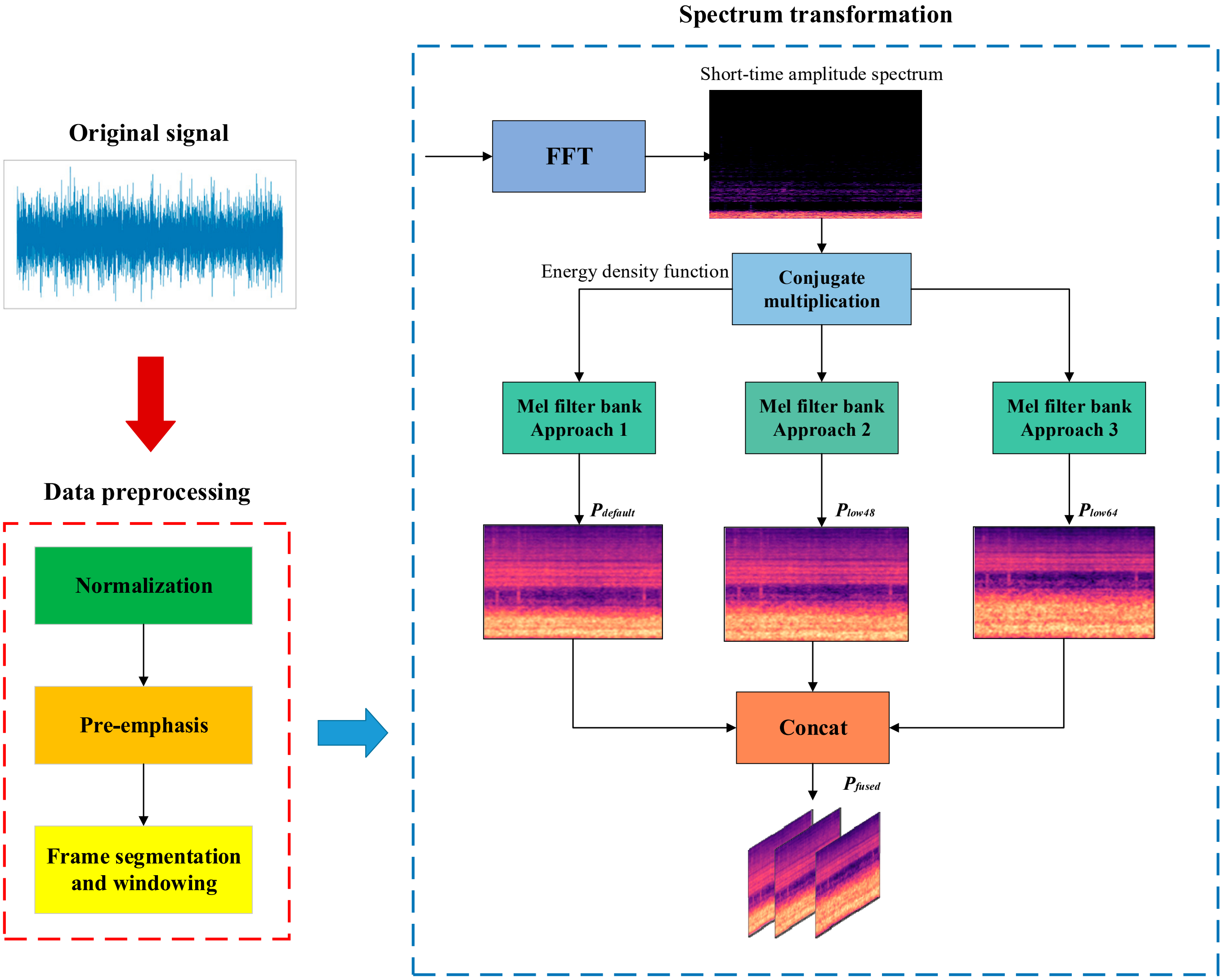


Figure 2. The extraction process of the three-channel improved Mel energy spectrum.

2.2. Data Augmentation

Data augmentation techniques have been proposed to improve recognition performance. The study employs SpecAugment [27] and Mixup [28] methods to simulate the signal loss caused by external factors during signal acquisition, thereby increasing data diversity. Figure 3 shows an example of the features of the enhanced Mel energy spectrum.

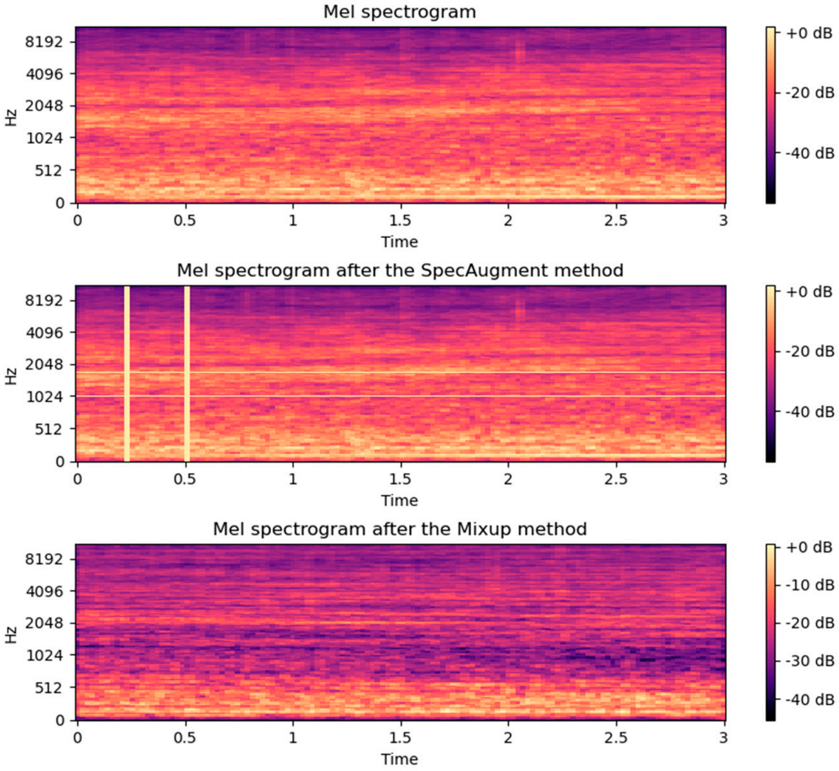


Figure 3. The Mel spectrogram after data augmentation.

The SpecAugment method enhances data by masking the spectrum's time–frequency domain. Masking in the frequency domain simulates the attenuation characteristics of underwater signals.

During the frequency domain masking, the band-stop filter in the frequency range $[f_1, f_2]$ is formulated as

$$X'(f) = \begin{cases} 0, \ f_1 \leq f \leq f_2 \\ X(f), \ \text{otherwise} \end{cases}. \tag{7}$$

The expression for random masking in the time domain range $[t_1, t_2]$ is

$$X'(t) = \begin{cases} 0, \ t_1 \leq t \leq t_2 \\ X(t), \ \text{otherwise} \end{cases}. \tag{8}$$

The Mixup method augments the dataset and reduces the influence of noisy samples on the model by mixing two different audio samples and their corresponding labels in arbitrary proportions. Specifically, a new sample and label are formed via weighted linear interpolation of the features and labels of two randomly selected samples (x_i, y_i) and (x_j, y_j):

$$\begin{aligned} \hat{x} &= \lambda x_i + (1-\lambda)x_j \\ \hat{y} &= \lambda y_i + (1-\lambda)y_j. \end{aligned} \tag{9}$$

where λ represents the interpolation factor, $\lambda \in (0,1)$.

2.3. Multi-Scale Frequency-Adaptive Network

This section introduces the principle and framework of the multi-scale frequency-adaptive network. Figure 4 shows the overall structure of the network, with VGG [29] serving as the backbone network for feedforward computation. The VGG structure comprises three convolutional modules, where each module has its internal convolutional modules replaced by multi-scale frequency-adaptive modules. The network comprises two convolutional layers, one residual layer, and a multi-scale fusion (MSF) module. The MSF module includes three frequency-adaptive residual (FAR) branches and a spatial and channel squeeze-and-excitation (scSE) module.

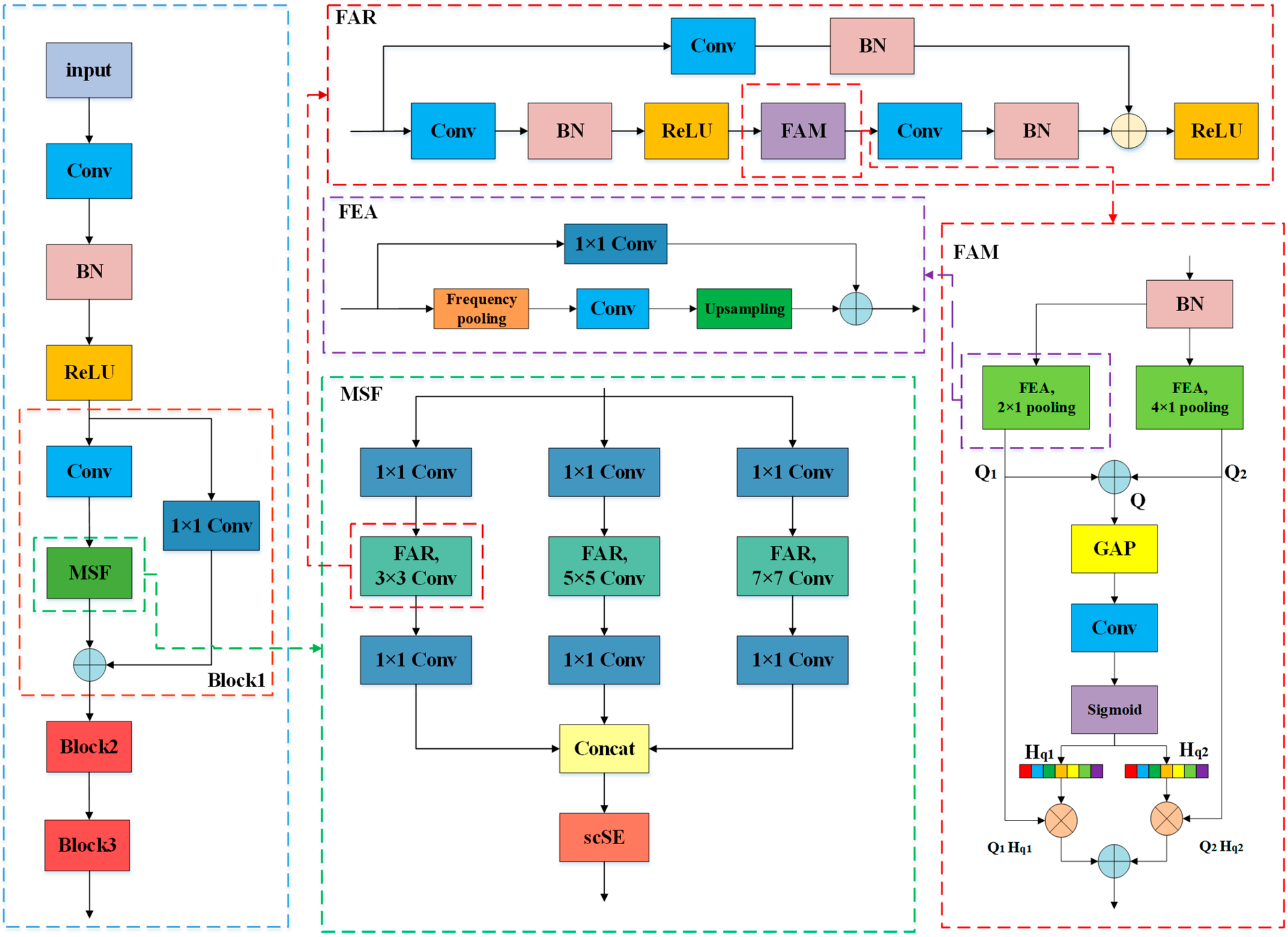


Figure 4. The structure of a multi-scale frequency-adaptive network.

2.3.1. Frequency-Adaptive Residual Module

In underwater target recognition, the time–frequency spectrum of radiated noise contains fewer critical details in the time dimension due to ships' relatively stable motion state and mechanical structure over short periods. However, it contains a wealth of critical information in the frequency dimension. Hence, this paper proposes a frequency enhancement algorithm (FEA) to enhance the frequency information in the features, thereby improving the model's ability to learn features. The principle of the FEA is illustrated in Figure 5. The input feature map x undergoes frequency pooling operations to compress the frequency dimension. This is followed by feature extraction through convolutional layers, which indirectly expands the receptive field in the frequency dimension. Subsequently, upsampling restores the dimensions to match the original input feature map size. Finally, the result is summed with the input feature map x to obtain x_1 to preserve the practical information from the original feature map. The enhanced feature map x_1 can be expressed as

$$x_1 = U[wD[x]_{r\times 1} + b]_{r\times 1} + x, \tag{10}$$

where $D[\cdot]_{r\times 1}$ represents a pooling operation with a pooling kernel size of $r \times 1$; $U[\cdot]_{r\times 1}$ represents an upsampling operation with an upsampling kernel size of $r \times 1$; and w and b denote the convolutional layer's weight matrix and the bias value, both of which are $r \times r$.

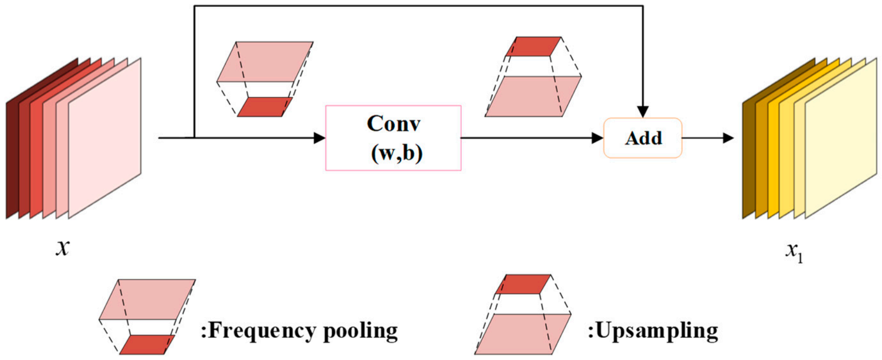


Figure 5. The structure of the frequency-adaptive algorithm.

The frequency-adaptive module (FAM) utilizes three pooling kernels of different sizes to the FEA and adaptively computes and allocates the frequency weights for each branch. The model structure is shown in the left part of Figure 6. Initially, the feature size is reduced through a convolutional layer, followed by two frequency enhancement blocks with pooling kernel sizes of 2 × 1 and 2n × 1, respectively, resulting in two feature maps Q_1 and Q_2 that possess different frequency receptive fields. The features Q_1 and Q_2 from the two branches are combined to obtain a merged feature Q. Subsequently, two rounds of global average pooling (GAP) and fully connected (FC) layers produce two frequency tensors: H_{q1} and H_{q2}. This process is represented by the following formula:

$$H_{qi} = \mathrm{FC}_i(\mathrm{GAP}(Q)), \tag{11}$$

where i represents the number of frequency channels, and H_{qi} denotes the frequency tensor adaptively selected. $i \in 1,2$, $H_{qi} \in R^{H\times1\times1}$, and H represent the frequency information.

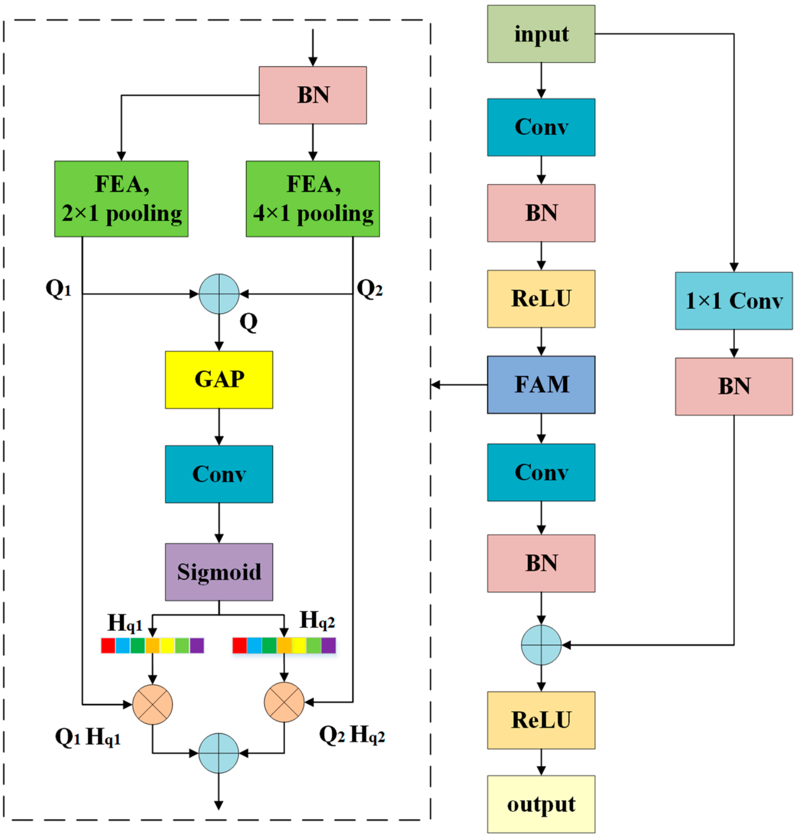


Figure 6. The frequency-adaptive residual module.

The final output of the module is represented as

$$\hat{Q} = Q_1 \times H_{q1} + Q_2 \times H_{q2}. \tag{12}$$

As shown in Figure 6, the residual network backbone consists of the following components: two convolutional layers, two batch normalization (BN) layers, and two rectified linear unit (ReLU) activation functions. FAM is positioned between the two convolutional layers of the residual network backbone to extract more frequency features and capture deeper semantic features.

2.3.2. Squeeze-and-Excitation Module

The squeeze-and-excitation (SE) module is an implementation of an attention mechanism with which to emphasize important features selectively [30], which measures the importance of features along specific dimensions to enhance meaningful features and suppress irrelevant ones. Assuming an input feature map X with the size $H \times W \times C$, the goal of the SE module is to recalibrate X to generate a corrected feature map $\hat{X}$ through operation $F_{SE}(\cdot)$. $F_{SE}(\cdot)$ is a function that maps X during the correction process to $\hat{X}$ and can be constructed based on different types of modules.

Spatial Squeeze and Channel Excitation Module (cSE)

The cSE module is illustrated in Figure 7. The input feature map X is passed through global average pooling to generate the vector Z, $Z \in \mathbb{R}^{1\times1\times C}$. The k-th element in vector Z is referred to as

$$Z_k = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_k(i,j),\ k = 1,2,\ldots,C, \tag{13}$$

where $X_k(i,j)$ represents the k-th channel feature map of X.

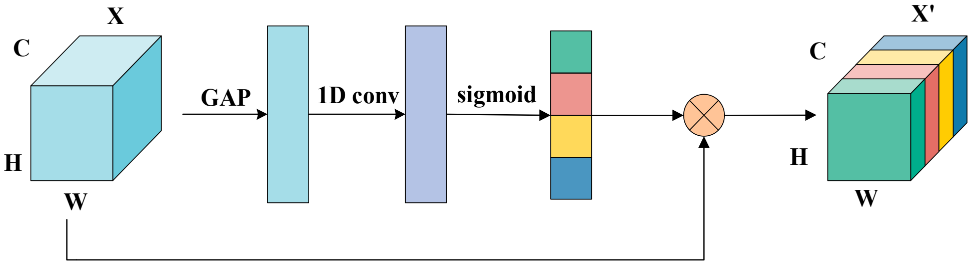


Figure 7. A spatial compression and channel excitation module.

During the activation process, the traditional approach involves first reducing dimensionality and then increasing it through two fully connected layers to extract dependencies among feature channels. Although this approach reduces the complexity of the model, the dimensionality reduction process disrupts the direct correspondence between channels and weights, adversely affecting the inter-channel dependencies. Wang et al. [31] introduced a local cross-channel interaction method to address this issue, which efficiently executes through one-dimensional convolution while adaptively determining the size of the convolutional kernel. Specifically, the compressed channel vector is used to compute the weights of each channel through one-dimensional convolution. Subsequently, a sigmoid activation function is applied to map the weight vector to the range of 0–1. Finally, the channel weight vector is multiplied by the input feature map X to generate the corrected feature map X'. X' can be represented as

$$X' = \sigma[(w_1 Z + b_1)] \cdot X, \tag{14}$$

where w_1 and b_1 represent the weight matrix and bias of the one-dimensional convolution, respectively; $\sigma[\cdot]$ denotes the sigmoid function; and the symbol "·" denotes the dot product of vectors or matrices.

The formula for calculating the adaptive convolutional kernel size is

$$k = \left| \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right|, \tag{15}$$

where C denotes the number of channels of vector Z, $b = 1$, and $\gamma = 2$.

Channel Squeeze and Spatial Excitation Module (sSE)

The sSE module is illustrated in Figure 8. The input feature map X undergoes a convolution with 1 output channel and a kernel size of 1×1, resulting in a feature map of size $H \times W \times 1$. The weight matrix is mapped using the sigmoid function. Finally, the weight matrix is multiplied with the original feature map in the spatial dimension to generate the corrected feature map X''. The corrected feature map X'' can be represented as

$$X'' = \sigma[(w_2 Z + b_2)] \cdot X, \tag{16}$$

where w_2 and b_2 represent the weight matrix and bias of the convolution with an output channel of 1 and a kernel size of 1×1; $\sigma[\cdot]$ denotes the sigmoid function.

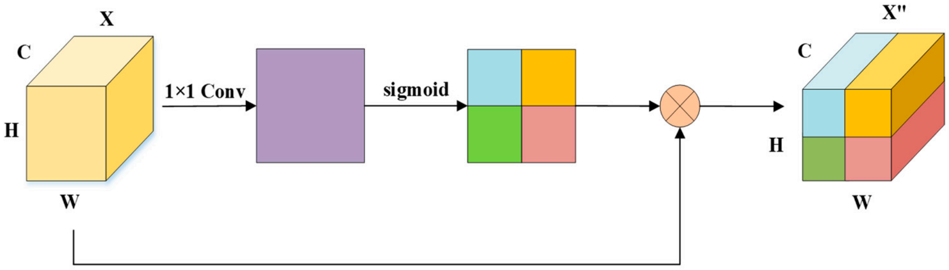


Figure 8. Channel squeeze and spatial excitation module.

Spatial and Channel Squeeze Excitation Module (scSE)

The scSE module is illustrated in Figure 9, which is achieved by adding the outputs of the parallel cSE and sSE modules. Thus, the corrected feature map X''' is as follows:

$$X''' = X' + X''. \tag{17}$$

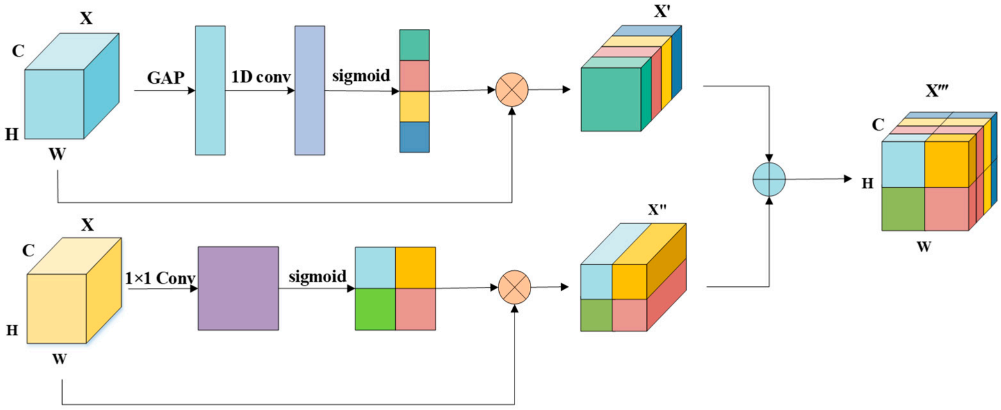


Figure 9. A spatial and channel squeeze excitation module.

2.3.3. Multi-Scale Fusion Module

The MSF module is the core of the entire multi-scale frequency-adaptive network, extracting rich and deep semantic features. The principle of the multi-scale fusion module in this paper is to use three convolution branches of different scales to capture and fuse information at varying receptive ranges. These branches share weights, differing only in their receptive field sizes. This approach helps reduce model parameters and mitigates the risk of overfitting commonly associated with complex models. The parallel-connected network architecture dramatically enhances the model's training efficiency. The structure of the multi-scale fusion module is depicted on the right of Figure 10, consisting of multi-scale feature extraction and multi-scale feature fusion parts. The feature extraction part incorporates frequency-adaptive residual modules in the three branches, using convolution kernels of 3 × 3, 5 × 5, and 7 × 7. The 1 × 1 convolutions before and after the module are used to reduce parameter volume and lower the computation complexity of the model. The feature fusion part involves merging features extracted from each branch in the channel dimension through concatenation, and feature enhancement using squeeze excitation modules. Residual connections are implemented to mitigate gradient explosion and vanishing issues during training.

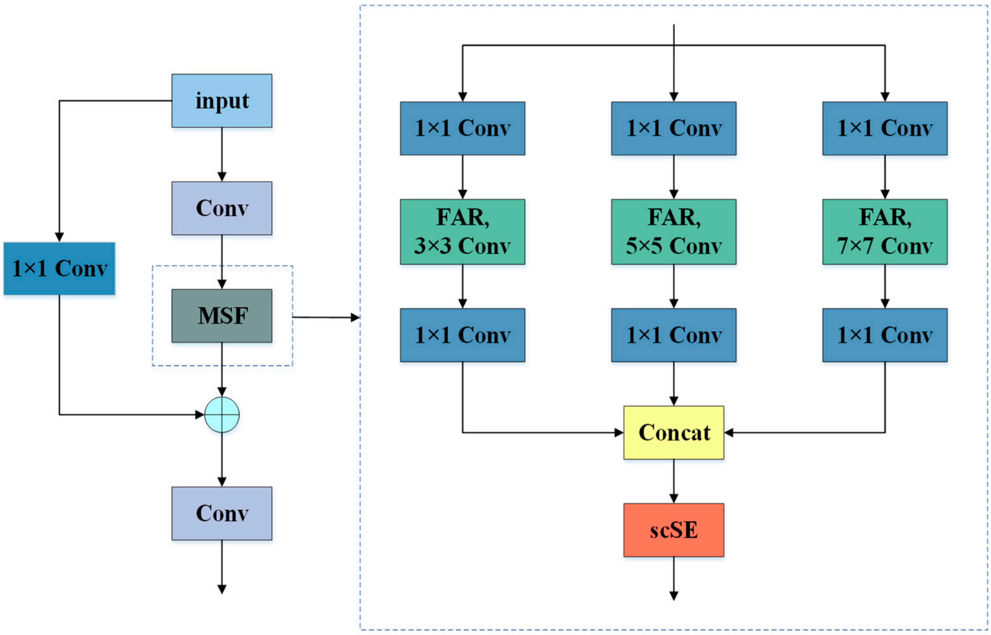


Figure 10. A multi-scale fusion module.

2.3.4. Focal Loss Function

Due to the challenges in acquiring underwater acoustic signals and the substantial annotation workload [32], the available datasets for underwater target recognition are relatively small. Additionally, the varying occurrence frequencies and durations of different vessels contribute to imbalanced data labels. To address these issues, this paper applies the focal loss function [33] to underwater target recognition. The focal loss function introduces a modulation factor to the standard cross-entropy loss which adjusts the weight of easily classified samples, causing the model to focus more on difficult-to-classify samples. The formula for focal loss is as follows:

$$L(p_i) = -\sum_{i=1}^{n} (1-p_i)^{\gamma} \log(p_i), \tag{18}$$

where p_i is the probability that the model predicts the i-th sample as belonging to the actual class, and γ is the tuning parameter that controls the attenuation degree of the loss for readily classified samples, and it is commonly set to 2. $(1-p_i)^{\gamma}$ is referred to as the decay factor, which reduces the contribution of the loss arising from readily classified samples.

To address the imbalance between positive and negative samples, a balance factor α is introduced into the focal loss function. The improved focal loss function is expressed as follows:

$$L(y, p_i) = -\alpha_i \sum_{i=1}^{n} (1-p_i)^{\gamma} \log(p_i), \alpha_i = \begin{cases} \alpha & y = 1 \\ 1-\alpha & y = 0 \end{cases}, \quad \alpha \in [0,1], \tag{19}$$

where y represents the true label of the sample.

3. Results

3.1. Experimental Data

The dataset used in this experiment is the ShipsEar dataset [34]. This dataset was recorded under various sea conditions at the Vigo port in Spain during autumn 2012 and summer 2013. The dataset consists of 90 audio samples with durations ranging from 15 s to 10 min, including 11 categories of ships and 1 category for background noise. According to the ship tonnage standard, the samples are categorized into four major groups: A, B, C, and D, with background noise classified as E. The specific classification of the samples is shown in Table 1. In the data-preprocessing stage, the samples, recorded at a sampling rate of 22,050 Hz, were segmented into 3 s intervals with a 33% overlap, totaling 5582 segments. Subsequently, the samples were divided into training, validation, and test sets in a ratio of 7:1.5:1.5.

Table 1. Sample classification in the ShipsEar dataset.

Category	Quantity	Target
A	925	Dredger/Fishing boat/Mussel/Trawler/Towboat
B	766	Motorboat/Pilot boat/Sailboat
C	2112	Passenger ship
D	1218	Ocean liner/Ro-Ro ship
E	561	Natural noise

3.2. Experimental Setup

The software platform for this experiment is the PyTorch 2.2.2 framework, running on a Windows 11 system. The hardware platform features an Intel Core i9-13900K CPU with 64 GB of RAM, and an NVIDIA RTX 4090 GPU with 16 GB of VRAM. The Adam optimizer was employed for training, utilizing the focal loss function mentioned in this paper. The initial learning rate was 0.001, with an adaptive update strategy utilized to converge the model, eventually reducing it to 0.00001. The batch size was set to 32, and each experiment was trained for 100 epochs, and repeated 30 times.

3.3. Experimental Results and Analysis

After preprocessing the ShipsEar dataset based on the established experimental parameters, experiments were conducted and the results were analyzed. Initially, the frequency pooling kernel in the frequency-adaptive algorithm was optimized by designing and evaluating different kernel sizes to determine the optimal parameter configuration. Precision, recall, and F1 scores were used as evaluation metrics, and a confusion matrix was utilized to analyze the ship attributes within the dataset. Subsequent ablation experiments analyzed the integration of various modules in the model and their contributions. Comparative experiments were conducted using various features, loss functions, and models. Finally, the dataset was carefully partitioned for detailed underwater target recognition.

3.3.1. Parameter Optimization

To investigate the impact of the frequency pooling kernel within the frequency-adaptive module on the model, we conducted experiments using various sizes of pooling kernels. In the experiments, a pooling kernel of size 2 was used as the baseline, and the ratio of the pooling kernel size in the other branch to the baseline was defined as the pooling kernel size ratio, serving as a reference indicator. The experimental results, presented in Figure 11, indicated that initially, as the pooling kernel size ratio increased, accuracy also gradually increased, suggesting that a larger pooling kernel could increase the receptive field, allowing the model to learn more frequency information. When the pooling kernel size ratio reached 4, the accuracy peaked at 98.4%. However, as the ratio increased further, the accuracy began to decline. This indicated that as the receptive field continued to expand, the frequency information within the features gradually diminished, leading to the decline in the effectiveness of the frequency-adaptive algorithm in the other branch and adversely affecting the model's learning, resulting in a sharp drop in recognition accuracy. Therefore, the pooling kernel's selection range must be carefully controlled. We selected a pooling kernel size ratio of 4 in the subsequent experiments.

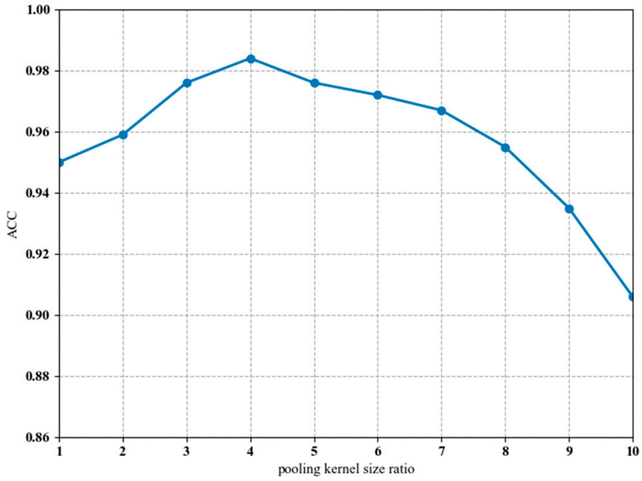


Figure 11. The impact of the frequency pooling kernel.

3.3.2. Performance Evaluation

In this experiment, precision, recall, and F1-score were used to evaluate the performance of the model. The formulas used to calculate each metric are as follows:

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{20}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{21}$$

$$\text{F1} - \text{score} = \frac{2TP}{2TP + FP + FN}, \tag{22}$$

where TP represents true positives, which signifies the number of samples correctly predicted as the positive class by the model; FP represents false positives, which indicate the number of samples incorrectly predicted as the positive class by the model; and FN represents false negatives, which denotes the number of samples incorrectly predicted as the negative class by the model.

Apart from the precision, recall measures the model's coverage rate of the positive class, and the F1-score considers both metrics, making it particularly suitable for scenarios with imbalanced classes. The recognition results for the five types of ships are shown in Table 2.

Table 2. The recognition results for five types of ships.

Category	Precision	Recall	F1-Score	Support
A	0.967	0.978	0.973	92
B	0.962	0.962	0.962	52
C	0.991	0.987	0.989	225
D	0.990	0.990	0.990	96
E	1.0	1.0	1.0	47
Average	0.984	0.984	0.984	512

Table 2 reveals that the precision, recall, and F1-score for each type of ship are above 96%, indicating a high recognition accuracy for the model. The metrics for category E are perfect, with a recognition accuracy of 100%, suggesting a solid capability to discriminate against background noise and other types of ships. The metrics for categories A, B, and C are relatively lower, potentially due to the uneven distribution of sample categories.

The confusion matrix in Figure 12 further confirms the points mentioned above, where the intensity of color in the confusion matrix represents the number of samples. In the matrix, category C has many samples, making it more prone to prediction errors. The diagonal elements correspond to the number of samples correctly predicted for each category, while the off-diagonal elements correspond to the number of samples incorrectly predicted. Categories B and C each had one or two samples incorrectly predicted as category A, possibly because category A includes shallow-water vessels like dredgers that might produce more considerable ocean background noise in shallow waters, potentially affecting prediction results.

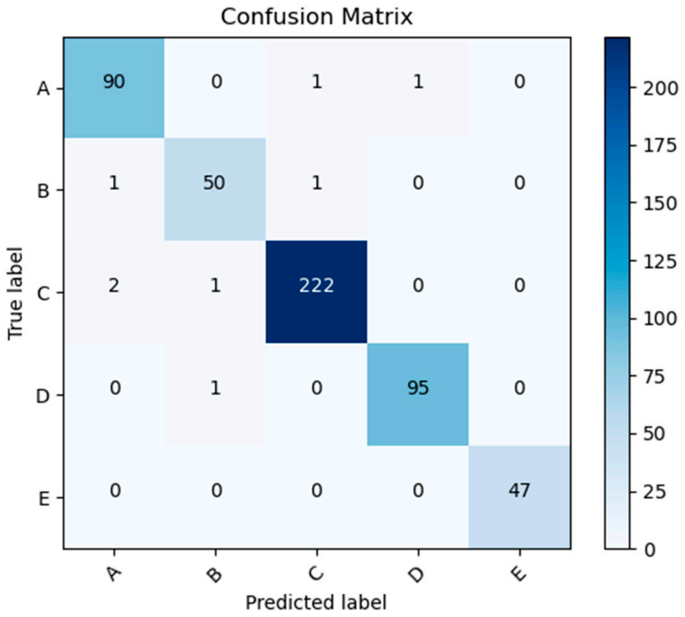


Figure 12. Confusion matrix for five types of vessels.

3.3.3. Ablation Study

To validate the effectiveness of the multi-scale frequency-adaptive network modules proposed in this paper, ablative experiments were conducted on the ShipsEar dataset. To facilitate result analysis, subsequent experiments employed accuracy as the evaluation metric. The three-channel improved Mel spectrogram features, in conjunction with data augmentation, were employed as model input. The residual network replaced the previously described VGG module, serving as the baseline (BS) for the ablative experiments. The FAM, scSE, and MSF of the MSFAN were individually embedded into the baseline model, ensuring consistency in training parameters and environmental conditions across experiments. Table 3 displays the recognition accuracy achieved after embedding different modules into the model. The data in the table show that incorporating scSE and FAM significantly increased recognition accuracy by 1.6% and 2.7%, respectively. This indicates that the frequency-adaptive and squeeze-and-excitation modules enhance the neural network's extraction of deep features. Adding the multi-scale fusion module resulted in a 4.2% improvement in recognition accuracy compared to the baseline, indicating a substantial enhancement. Further improvements in accuracy by integrating the other two modules with the multi-scale fusion module demonstrate the module's ability to expand the network's receptive field, enhance feature representation, and, when integrated with other modules, enable more precise target identification. Therefore, incorporating the described modules can significantly enhance model recognition accuracy.

Table 3. Results of ablation experiments on different modules.

Module	Recognition Accuracy
BS	91.8%
BS-scSE	93.4%
BS-FAM	94.5%
BS-MSF	96.0%
BS-MSF-scSE	96.1%
BS-MSF-FAM	97.2%
BS-MSF-FAM-scSE	98.4%

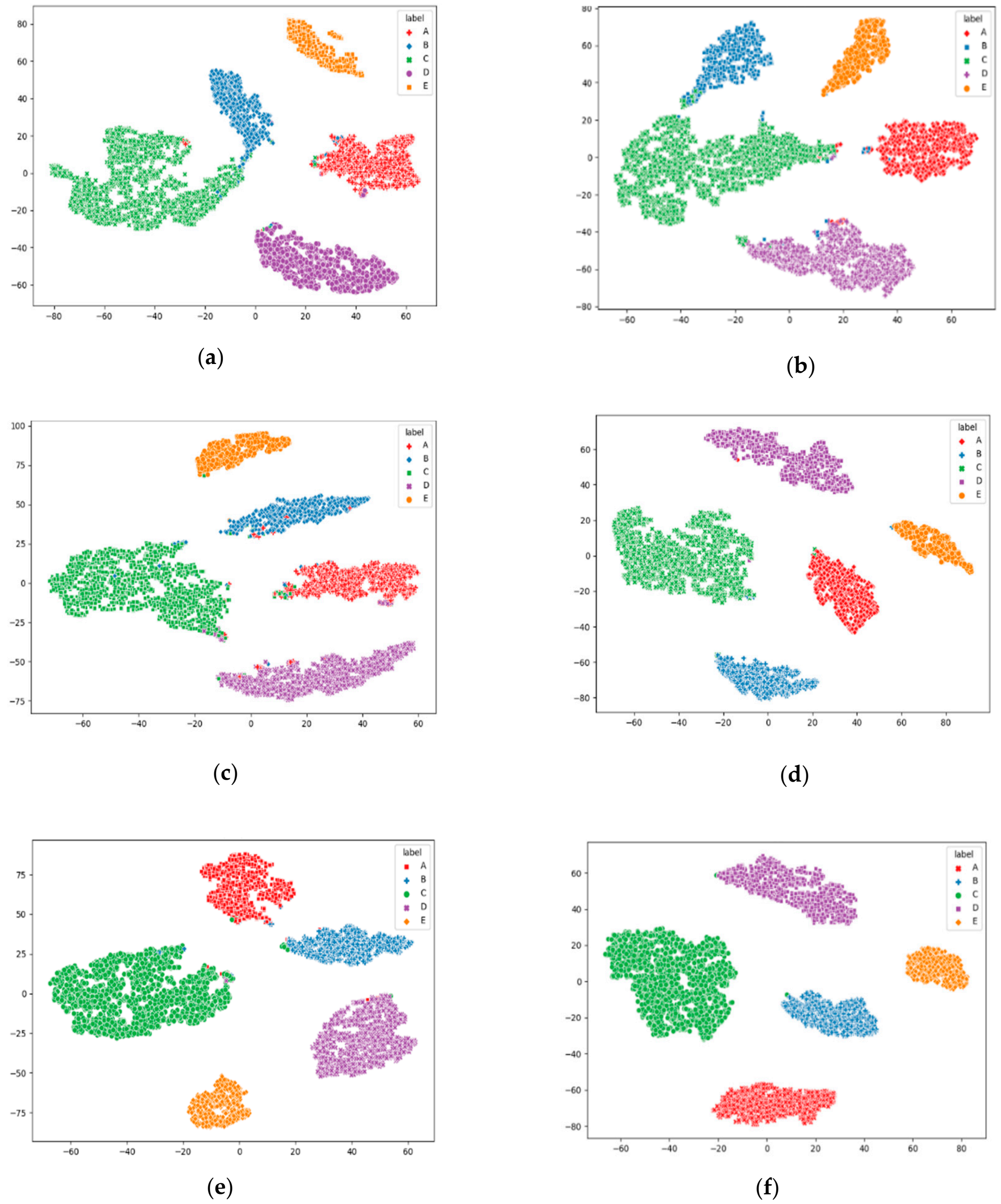
3.3.4. Comparison of Different Features

Table 4 displays the recognition accuracy of different features processed through the model. It is observed fusion features significantly enhances recognition performance compared to traditional single features. Among single features, the Mel frequency cepstrum coefficient (MFCC) achieves a recognition accuracy of 93.9%, 1.5% higher than the Mel spectrogram and substantially higher than the 83.6% accuracy of STFT. This indicates that both MFCC and the Mel spectrograms effectively represent the spectral features of underwater acoustic signals, with MFCC yielding better results. In contrast, STFT struggles to accurately describe these characteristics. Feature fusion further improved accuracy due to comprehensive feature representation. Incorporating differential features (DF) as part of the fusion, DF-Mel achieves a 2.3% higher accuracy over the single Mel spectrogram. DF-MFCC shows a 2.1% improvement over a single MFCC, highlighting that differential features can enhance overall feature quality. The improved Mel filter bank designed in this paper provides corresponding MFCC fusion features (3C-IMFCC). However, its recognition performance is lower than that of the improved Mel spectrogram. This might be due to the potential loss of high-order information during the conversion from the Mel spectrogram to MFCC via discrete cosine transform (DCT), which removes spectral correlations. Therefore, the three-channel improved Mel (3C-IMel) spectrogram features exhibit superior performance.

Table 4. Recognition accuracy following model-based feature processing.

Features	Recognition Accuracy
STFT	83.6%
Mel	92.4%
MFCC	94.0%
DF-Mel	94.7%
DF-MFCC	96.1%
3C-IMel	98.4%
3C-IMFCC	97.1%

To assess the discriminability between different features visually, this paper employs the t-SNE algorithm to visualize the feature separation post-training. Figure 13 displays scatter plots of Mel, MFCC, DF-Mel, DF-MFCC, 3C-IMFCC, and 3C-IMel features, mapped from high- to two-dimensional vectors. Different colors are used to represent various categories. The three-channel improved Mel spectrogram features show an apparent separation effect with compact intra-class clusters. This demonstrates their strong inter-class separability and intra-class cohesion, which are crucial characteristics of effective features.



(**a**)

(**b**)

(**c**)

(**d**)

(**e**)

(**f**)

Figure 13. t-SNE results of different features: (**a**) t-SNE results of the Mel; (**b**) t-SNE results of the MFCC; (**c**) t-SNE results of the DF-Mel; (**d**) t-SNE results of the DF-MFCC; (**e**) t-SNE results of the 3C-IMFCC; (**f**) t-SNE results of the 3C-Imel.

3.3.5. Comparison of Loss Functions

To illustrate the effectiveness of the focal loss function used in this paper during model training, we compared the performance of various loss functions on the model's recognition accuracy. As shown in Table 5, compared to the cross-entropy loss function, the focal loss function, which performs excellently on imbalanced datasets, shows an improvement in recognition accuracy. This indicates that the focal loss function is more prominent when dealing with imbalanced data.

Table 5. Comparative analysis of recognition accuracy across loss functions.

Loss Function	Cross-Entropy Loss Function	Focal Loss Function
Accuracy	97.5%	98.4%

To fully demonstrate the focal loss function's ability to enhance model performance on imbalanced data, we balanced the quantities of each class in the ShipsEar dataset. Table 6 shows the adjusted distribution of samples in the ShipsEar dataset after balancing.

Table 6. Sample classification in the balanced ShipsEar dataset.

Category	Quantity	Target
A	500	Dredger/Fishing boat/Mussel/Trawler/Towboat
B	500	Motorboat/Pilot boat/Sailboat
C	500	Passenger ship
D	500	Ocean liner/Ro-Ro ship
E	500	Natural noise

The performance of each loss function on the model trained with the adjusted dataset provides a contrast, with the results shown in Table 7. The recognition accuracy with focal loss is similar to cross-entropy loss on the balanced dataset. This suggests that the adjustment parameters in the focal loss function are not significant in a balanced dataset but are crucial for optimizing performance on imbalanced datasets.

Table 7. Recognition accuracy across loss functions in the balanced dataset.

Loss Function	Cross-Entropy Loss Function	Focal Loss Function
Accuracy	98.9%	99.1%

3.3.6. Comparison of Different Models

To validate the performance of the multi-scale frequency-adaptive network presented in this paper, we compared its recognition accuracy from 30 repeated experiments with that of other popular models. The results are illustrated in Figure 14.

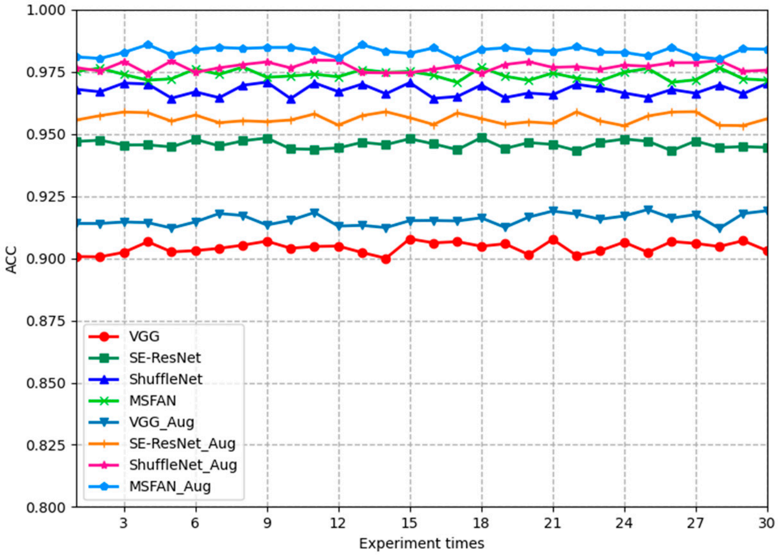


Figure 14. Experiments times with different models.

The results show that recognition accuracy improved for each model after data augmentation. This indicates that data augmentation enhances model performance and generalization, reduces overfitting risk, and improves robustness. The proposed model demonstrated consistently higher recognition performance across the 30 repeated experiments than all other models, suggesting that it is both effective and robust.

3.3.7. Fine-Grained Recognition

Ref. [34] partitions the dataset by categorizing ships based on tonnage attributes. This method of classification, using similar attributes across categories, can introduce errors in recognition tasks. We adopted the partitioning method from ref. [35] to enable a more comprehensive analysis of the ShipsEar dataset. A subset with nine categories, termed ShipsEar2, is selected for recognition. These nine categories include dredger, fishing boat, motorboat, sailboat, container ship, passenger ship, roll-on/roll-off ship, sailing ship, and natural noise. Table 8 shows the training-testing split, with an additional 15% of samples randomly selected from the training set for validation.

Table 8. Training-testing dataset splitting situation.

Category	ID in Training Set	ID in Test Set
Dredger	80, 93, 94, 96	95
Fish boat	73, 74, 76	75
Motorboat	21, 26, 33, 39, 45, 51, 52, 70, 77, 79	27, 50, 72
Mussel boat	46, 47, 49, 66	48
Ocean liner	16, 22, 23, 25, 69	24, 71
Passenger ship	06, 07, 08, 10, 11, 12, 14, 17, 32, 34, 36, 38, 40, 41, 43, 54, 59, 60, 61, 63, 64, 67	9, 13, 35, 42, 55, 62, 65
Ro-Ro ship	18, 19, 58	20, 78
Sailboat	37, 56, 68	57
Natural noise	81, 82, 84, 85, 86, 88, 90, 91	83, 87, 92

Table 9 shows the recognition results of ShipsEar2 using MSFAN. The average metric for each ShipsEar2 category is 0.88, indicating excellent overall model performance and feature processing in the nine-class recognition task. Lower results in the five-class recognition task are attributed to increased classification difficulty and dataset partitioning changes. Metrics below 0.5 were observed in the Dredger and Sailboat categories due to insufficient samples, with only six Sailboat samples in the test set. Thus, insufficient samples can introduce significant randomness, affecting the results.

Table 9. Recognition results of ShipsEar2 using the MSFAN.

Category	Precision	Recall	F1-Score	Support
Dredger	1.0	0.45	0.621	20
Fish boat	0.75	0.75	0.75	16
Motor boat	0.893	0.704	0.787	71
Mussel boat	0.703	0.963	0.813	27
Ocean liner	0.852	0.821	0.836	56
Passenger ship	0.934	0.929	0.931	351
Ro-Ro ship	0.868	0.894	0.881	66
Sailboat	0.208	0.833	0.333	6
Natural noise	1.0	1.0	1.0	91
Average	0.908	0.886	0.891	704

Figure 15 illustrates the use of the t-SNE algorithm to visualize model-processed features across different categories in ShipsEar2.

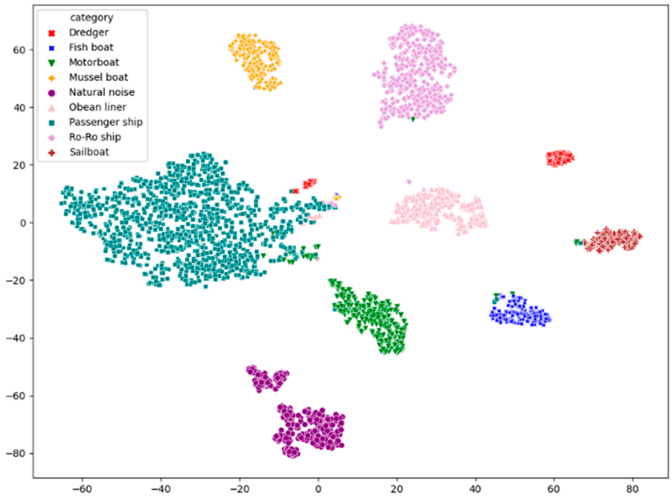


Figure 15. t-SNE results in ShipsEar2.

To demonstrate the effectiveness of ShipsEar2, we compare the proposed MSFAN model with SE-ResNet [36]. The three-channel improved Mel energy spectrum is compared against the reference contrast features used in previous experiments. The recognition results are shown in Figure 16.

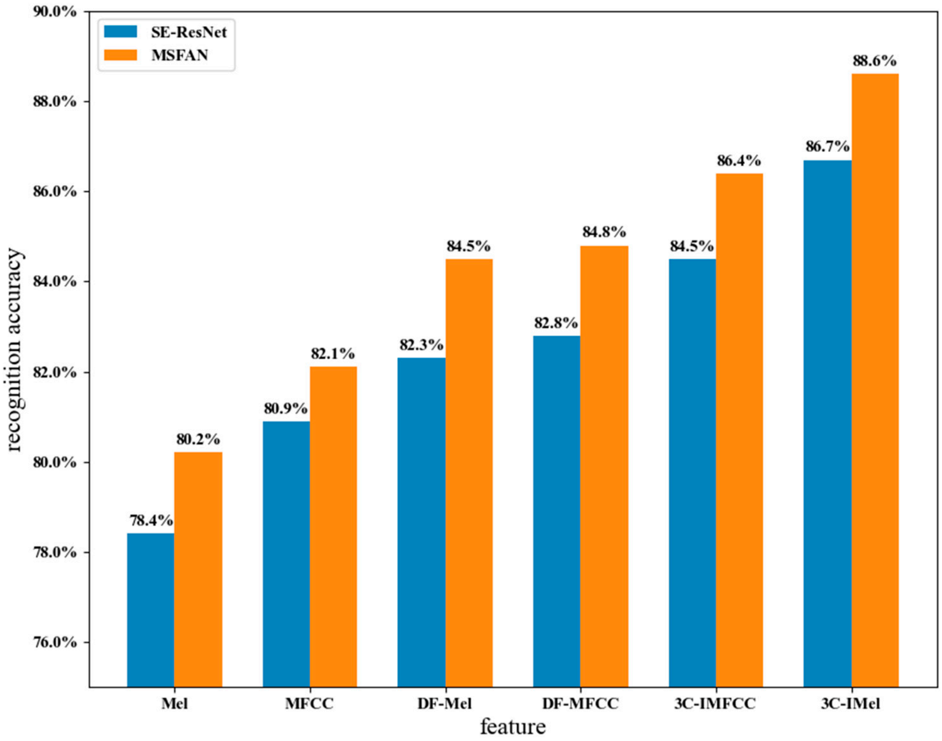


Figure 16. Comparative analysis of recognition results of ShipsEar2.

The results indicate that in the complex task of fine-grained recognition, MSFAN significantly outperforms the SE-ResNet model regarding recognition accuracy, with improvements reaching up to 2%. The enhancement in recognition performance is more pronounced with the 3C-Imel feature than other features.

4. Conclusions

This paper introduces a feature extraction method using a three-channel improved Mel energy spectrum. Experiments show that this feature provides stronger inter-class separability and intra-class cohesion than traditional MFCC and Mel energy spectrum features. Additionally, we propose a novel MSFAN model for underwater acoustic target recognition. This network combines frequency-adaptive residual modules, squeeze-and-excitation modules, and multi-scale fusion mechanisms, enhancing feature representation capacity. Rigorous evaluation of the ShipsEar dataset using a focal loss function shows that the proposed MSFAN model enhances recognition accuracy and generalization, achieving 98.4% accuracy, surpassing existing models.

Moreover, fine-grained classification on a nine-category subset achieves 88.6% accuracy, marking considerable advancements over existing techniques. The MSFAN-based methodology marks a significant leap forward in underwater target recognition, especially in feature extraction, model design, and training strategies. This work provides novel insights and a robust framework, offering new perspectives in underwater acoustic target recognition.

Author Contributions: Methodology, L.Z. and A.Y.; software, L.Z., A.Y., Y.M. and D.D.-U.L.; validation, A.Y. and Y.M.; investigation, Y.M.; writing—original draft preparation, L.Z. and A.Y.; writing—review and editing, L.Z. and D.D.-U.L.; supervision, A.Y. and Y.M.; project administration, A.Y. All authors have read and agreed to the published version of the manuscript.

Funding: Supported by the Fundamental Research Funds for the Provincial Universities of Zhejiang (GK249909299001-006).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding author.

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

References

1. Fang, S.; Du, S.; Luo, X.; Han, N.; Xu, X. Development of Underwater Acoustic Target Feature Analysis and Recognition Technology. *Bull. Chin. Acad. Sci.* **2019**, *34*, 297–305. [CrossRef]
2. Zhao, S.; Sun, C.-Y.; Chen, X.-H.; Tian, T. The analysis of tow ship radiated noise characteristics and the cancellation of the noise. *Tech. Acoust.* **2013**, *32*, 92–95.
3. Jiang, X.Y.; Du, X.M. Detection of torpedo radiated noise in strong interference background. *Tech. Acoust.* **2010**, *29*, 18–22.
4. Hao, Y.X. Ship Radiated Noise Classification Method Based on Deep Neural Network. Master's Thesis, Harbin Engineering University, Harbin, China, 2019.
5. Li, J.; Wang, B.; Cui, X.; Li, S.; Liu, J. Underwater Acoustic Target Recognition Based on Attention Residual Network. *Entropy* **2022**, *24*, 1657. [CrossRef] [PubMed]
6. Luo, X.; Feng, Y. An Underwater Acoustic Target Recognition Method Based on Restricted Boltzmann Machine. *Sensors* **2020**, *20*, 5399. [CrossRef]
7. Ke, X.; Yuan, F.; Cheng, E. Underwater Acoustic Target Recognition Based on Supervised Feature-Separation Algorithm. *Sensors* **2018**, *18*, 4318. [CrossRef] [PubMed]
8. Wang, B.; Wu, C.; Zhu, Y.; Zhang, M.; Li, H.; Zhang, W. Ship Radiated Noise Recognition Technology Based on ML-DS Decision Fusion. *Comput. Intell. Neurosci.* **2021**, *2021*, 8901565. [CrossRef]
9. Liu, D.; Zhao, X.; Cao, W.; Wang, W.; Lu, Y. Design and Performance Evaluation of a Deep Neural Network for Spectrum Recognition of Underwater Targets. *Comput. Intell. Neurosci.* **2020**, *2020*, 8848507. [CrossRef]
10. Zhang, Q.; Da, L.; Zhang, Y.; Hu, Y. Integrated Neural Networks Based on Feature Fusion for Underwater Target Recognition. *Appl. Acoust.* **2021**, *182*, 108261. [CrossRef]
11. Wu, C.; Wang, B.; Xu, Q.; Zhu, Y. Ship radiated noise recognition technology based on wavelet packet decomposition and PCA-Attention-LSTM. *Tech. Acoust.* **2022**, *41*, 264–273.
12. Liu, C.; Hong, F.; Feng, H.; Hu, M. Underwater Acoustic Target Recognition Based on Dual Attention Networks and Multiresolution Convolutional Neural Networks. In Proceedings of the OCEANS 2021: San Diego—Porto, San Diego, CA, USA, 20–23 September 2021; IEEE: New York, NY, USA, 2021; pp. 1–5.
13. Liu, F.; Shen, T.; Luo, Z.; Zhao, D.; Guo, S. Underwater Target Recognition Using Convolutional Recurrent Neural Networks with 3-D Mel-Spectrogram and Data Augmentation. *Appl. Acoust.* **2021**, *178*, 107989. [CrossRef]
14. Wang, X.; Liu, A.; Zhang, Y.; Xue, F. Underwater Acoustic Target Recognition: A Combination of Multi-Dimensional Fusion Features and Modified Deep Neural Network. *Remote Sens.* **2019**, *11*, 1888. [CrossRef]

15. Hinton, G.E.; Salakhutdinov, R.R. Reducing the Dimensionality of Data with Neural Networks. *Science* **2006**, *313*, 504–507. [CrossRef]
16. Hong, F.; Liu, C.; Guo, L.; Chen, F.; Feng, H. Underwater Acoustic Target Recognition with ResNet18 on ShipsEar Dataset. In Proceedings of the 2021 IEEE 4th International Conference on Electronics Technology (ICET), Chengdu, China, 7–10 May 2021; IEEE: New York, NY, USA, 2021; pp. 1240–1244.
17. Jin, G.; Liu, F.; Wu, H.; Song, Q. Deep Learning-Based Framework for Expansion, Recognition and Classification of Underwater Acoustic Signal. *J. Exp. Theor. Artif. Intell.* **2020**, *32*, 205–218. [CrossRef]
18. Han, X.C.; Ren, C.; Wang, L.; Bai, Y. Underwater Acoustic Target Recognition Method Based on a Joint Neural Network. *PLoS ONE* **2022**, *17*, e0266425. [CrossRef] [PubMed]
19. Huang, Q.; Zeng, X. An underwater acoustic target recognition method combining wavelet decomposition and an improved convolutional neural network. *J. Harbin Eng. Univ.* **2022**, *43*, 159–165. [CrossRef]
20. Huang, H. Multi-Scale Fusion Acoustic Scene Classification Based on Attention Mechanism. Master's Thesis, Fuzhou University, Fuzhou, China, 2021.
21. Xue, L.; Zeng, X.; Jin, A. A Novel Deep-Learning Method with Channel Attention Mechanism for Underwater Target Recognition. *Sensors* **2022**, *22*, 5492. [CrossRef]
22. Yan, C.; Yan, S.; Yao, T.; Yu, Y.; Pan, G.; Liu, L.; Wang, M.; Bai, J. A Lightweight Network Based on Multi-Scale Asymmetric Convolutional Neural Networks with Attention Mechanism for Ship-Radiated Noise Classification. *J. Mar. Sci. Eng.* **2024**, *12*, 130. [CrossRef]
23. Liu, D.; Yang, H.; Hou, W.; Wang, B. A Novel Underwater Acoustic Target Recognition Method Based on MFCC and RACNN. *Sensors* **2024**, *24*, 273. [CrossRef]
24. Fei, H.; Wu, W.; Li, P.; Cao, Y. Acoustic Scene Classification Method Based on Mel Spectrogram Separation and LSCNet. *J. Harbin Inst. Technol.* **2022**, *54*, 124–130, 123. [CrossRef]
25. Zhang, Q.; Da, L.; Wang, C.; Zhang, Y.; Zhuo, J. An Overview on Underwater Acoustic Passive Target Recognition Based on Deep Learning. *J. Electron. Inf. Technol.* **2023**, *45*, 4190–4202. [CrossRef]
26. Abdul, Z.K.; Al-Talabani, A.K. Mel Frequency Cepstral Coefficient and Its Applications: A Review. *IEEE Access* **2022**, *10*, 122136–122158. [CrossRef]
27. Park, D.S.; Chan, W.; Zhang, Y.; Chiu, C.-C.; Zoph, B.; Cubuk, E.D.; Le, Q.V. SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition. In Proceedings of the Interspeech 2019, Graz, Austria, 15–19 September 2019; ISCA: Singapore, 2019; pp. 2613–2617.
28. Zhang, H.; Cisse, M.; Dauphin, Y.N.; Lopez-Paz, D. Mixup: Beyond Empirical Risk Minimization. *arXiv* **2018**, arXiv:1710.09412.
29. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, 7–9 May 2015.
30. Naranjo-Alcazar, J.; Perez-Castanos, S.; Zuccarello, P.; Cobos, M. Acoustic Scene Classification with Squeeze-Excitation Residual Networks. *IEEE Access* **2020**, *8*, 112287–112296. [CrossRef]
31. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; IEEE: New York, NY, USA, 2020; pp. 11531–11539.
32. Zhang, Y.H.; Wang, C.; Zhang, Q.; Li, Q.; Da, L.L. A Review of Underwater Acoustic Target Detection and Recognition Technology Based on Information Fusion. *J. Signal Process.* **2023**, *39*, 1711–1727. [CrossRef]
33. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal Loss for Dense Object Detection. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; IEEE: New York, NY, USA, 2017; pp. 2999–3007.
34. Santos-Domínguez, D.; Torres-Guijarro, S.; Cardenal-López, A.; Pena-Gimenez, A. ShipsEar: An Underwater Vessel Noise Database. *Appl. Acoust.* **2016**, *113*, 64–69. [CrossRef]
35. Xie, Y.; Ren, J.; Xu, J. Unraveling Complex Data Diversity in Underwater Acoustic Target Recognition through Convolution-Based Mixture of Experts. *Expert Syst. Appl.* **2024**, *249*, 123431. [CrossRef]
36. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; IEEE: New York, NY, USA, 2018; pp. 7132–7141.

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