

Article

Zero-Shot Learning-Based Recognition of Highlight Images of Echoes of Active Sonar

Xiaochun Liu ¹, Yunchuan Yang ^{1,*}, Xiangfeng Yang ¹, Liwen Liu ¹, Lei Shi ¹, Yongsheng Li ¹ and Jianguo Liu ²

¹ Xi'an Precision Machinery Research Institute, Xi'an 710077, China; xiaochunliu@mail.nwpu.edu.cn (X.L.); ggdate@163.com (X.Y.); seasonsleo@163.com (L.L.); leishshi@163.com (L.S.); 210yezi@163.com (Y.L.)

² School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China; liujianguo@nwpu.edu.cn

* Correspondence: yunchuan yang@163.com

Abstract: Reducing the impact of underwater disturbance targets and improving the ability to recognize real moving targets underwater are important directions of active sonar research. In this paper, the highlight model of underwater targets was improved and a method was proposed to acquire highlight images of the echoes of these targets. A classification convolutional neural network called HasNet-5 was designed to extract the global features and local highlight features of the echo highlight images of underwater targets, which achieved the true/false recognition of targets via multi-classification. Five types of target highlight models were used to generate simulation data to complete the training, validation and testing of the network. Tests were performed using experimental data. The results indicate that the proposed method achieves 92% accuracy in real target recognition and 94% accuracy in two-dimensional disturbance target recognition. This study provides a new approach for underwater target recognition using active sonar.

Keywords: underwater target recognition; highlight images of echoes; zero-shot learning; target highlight models



Citation: Liu, X.; Yang, Y.; Yang, X.; Liu, L.; Shi, L.; Li, Y.; Liu, J. Zero-Shot Learning-Based Recognition of Highlight Images of Echoes of Active Sonar. *Electronics* **2024**, *13*, 457. <https://doi.org/10.3390/electronics13020457>

Academic Editors: George A. Tsihrintzis and Hirokazu Kobayashi

Received: 9 October 2023

Revised: 26 December 2023

Accepted: 20 January 2024

Published: 22 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Underwater target recognition is an important research direction in the field of hydroacoustic engineering. Eliminating the effects of disturbance targets and recognizing the target of concern are among the important research topics. Geometric features are crucial for underwater target recognition. However, they are not exclusive to real targets; numerous interfering targets also have geometric features [1,2]. Disturbance targets (such as towed acoustic decoy) have echo-scattering features, geometric features, and motion features similar to those of real targets, which makes it difficult to recognize underwater moving targets such as large underwater unmanned vehicles or submarines.

Active sonar, which radiates pulsed acoustic waves directionally through transducer arrays, plays an important role in the field of target recognition and is capable of obtaining feature information and classifying targets from target echoes at a long distance [3]. At present, active sonar recognition methods mainly include traditional target feature extraction methods, methods involving echo information extraction combined with deep learning, etc.

Active sonar is used to obtain the target's features, such as the distance, bearing, intensity, echo extension, echo energy and target size. Feature transform, fusion and logical judgment for target recognition belong to the traditional target feature extraction methods. The authors of [4–7] investigated a recognition method based on the undulation of the bearing of the target echoes, which can obtain the one-dimensional (1D) geometric scale of an underwater target, and solved the problem of recognizing point-source disturbance targets and real targets. The authors of [8–11] obtained the two-dimensional (2D) geometric scale of the target, implying that their methods had the ability to recognize point-source

disturbance targets, 1D disturbance targets and real targets. However, it is difficult for the above methods to discriminate between 2D disturbance targets and real targets.

Preprocessing, data conversion and low-level feature extraction are performed on active sonar echo data. Subsequently, through deep learning, the fine features of underwater targets can be extracted, improving the target recognition capability. Therefore, echo information extraction combined with deep learning is becoming a research hotspot.

In [12], an echo feature extraction method based on human auditory experience was proposed. In [13], the Wigner-Ville distribution (WVD) time-frequency features of the echoes were extracted and a Gustafson-Kessel (GK) clustering classification algorithm was proposed, which was validated using data from the scaled model of the pool. In [14], the power spectral statistics, linear predictive coding (LPC) coefficients and auto regressive (AR) coefficients of the echoes were extracted and a classification method based on the feed-forward neural network was proposed, in which the pool data of six types of targets were collected. The classification accuracy was higher than 90%. In [15], the target echo was acquired via multibeam sonar and the classification of the seabed substrate was achieved using the k-medoids algorithm. The feature extraction and classification of multiple materials and different volumes of spherical targets were conducted in [16] using convolutional neural networks (CNNs) based on short-time Fourier-transform time-frequency spectra and wavelet scale spectra. The simulation data demonstrated a geometric size classification accuracy of 97%. The above methods can extract fine features such as the geometry and material of underwater targets, but the influence of the hydroacoustic channel when the target is far away has not been verified.

In [17–22], imaging sonar (including forward sonar, side-scan sonar and synthetic aperture sonar) was used to acquire sonar images of underwater targets and deep learning networks, such as CNNs, were utilized to classify the underwater targets with a high target discrimination capability. However, the datasets used in the above studies consisted of near-range underwater static target data.

In [23], the multiple signal classification (MUSIC) algorithm was used to estimate the geometric scattering of underwater targets and was validated experimentally in an anechoic tank. In [24], the shape and size features of underwater targets were extracted using Wigner-Ville distribution time-frequency features and classified via a support vector machine, and the experiments indicated that the weak highlight echo signal of elastic scattering plays an important role in target recognition. In [25], chirplet atomic decomposition was used to improve the extraction of the geometric highlight features of underwater targets, and the simulation analysis indicated that the recognition capability can be improved. In [26], a combination of high-resolution direction of arrival (DOA) estimation, time-delay compensation and data fusion was proposed for estimating the geometric structure of underwater targets and was validated through simulation analysis. These studies indicate that the fine geometric structural features of underwater targets have attracted attention from industry. However, the main research focus is target geometric feature extraction.

The aforementioned algorithms can be classified into four distinct categories, and their respective merits and demerits are presented in Table 1. Target scale recognition algorithms face challenges in distinguishing between 2D disturbance targets and real targets. Acoustic imaging recognition algorithms encounter difficulties in recognizing moving targets at long distances (beyond 350 m). While time-frequency feature recognition algorithms and geometric structure recognition algorithms can extract the fine features of underwater targets, there has been no significant breakthrough in the recognition of long-distance moving targets.

Table 1. Summary of various algorithms.

Algorithm Type	Advantages	Unsolved Problems
Target scale recognition algorithms	The algorithm is straightforward and computationally efficient, rendering it suitable for stationary as well as low- and high-speed moving targets.	It is challenging to discriminate between 2D disturbance targets and real targets.
Time-frequency feature recognition algorithms	These algorithms are capable of extracting intricate features that accurately capture the geometric shape, size and material properties of targets.	The validation of these algorithms is based on the data obtained from pool tests, and the recognition capability for long-distance moving targets has not been verified.
Acoustic imaging recognition algorithms	The acquisition of images that capture the geometric shape and echo intensity of targets can be achieved.	Currently, long-distance moving target recognition remains unfeasible for application.
Geometric structure recognition algorithms	Relevant information regarding the target, such as its geometric shape, size and structure, can be effectively extracted.	The validation relies on pool test data, while the investigation of long-distance moving targets recognition remains unexplored.

To address the challenge of long-distance recognition of underwater moving targets, we propose the EHITRA (echo highlight image target recognition algorithm), an underwater target recognition method based on target echo highlight image extraction. Specifically, our contributions are as follows:

1. The underwater target highlight model has been enhanced to depict more accurately the distance, echo intensity, horizontal angle, pitch angle and frequency response of each highlight scattering region of targets. This model serves as the theoretical foundation for extracting highlight information and can be utilized to generate the simulation data of underwater targets in order to address the issue of data scarcity.
2. The paper proposes a methodology for acquiring the highlight image of underwater targets by utilizing cross-spectral directional or high-resolution DOA algorithms, thereby enabling the retrieval of multi-highlight information from moving targets at long distances. Furthermore, it employs the principle of orthogonal projection to derive the distribution map of the highlight scattering region on the target.
3. The HasNet-5 convolutional classification network is established, which leverages the concept of zero-shot learning. The network is trained using simulation data from four typical disturbance targets and an underwater vehicle, enabling it to effectively extract both global features and local highlight features of underwater targets. The effectiveness of the recognition method is validated through experimental data, demonstrating a recognition probability of 92% for actual targets and 94% for 2D disturbance targets.

The method proposed in this paper demonstrates the characteristics of a straightforward classification network and effortless applicability, rendering it suitable for active sonar systems on mobile platforms such as UUVs or torpedoes. However, it is recommended that the sonar system consists of at least 10 array elements to enhance the directional ability and signal-to-noise ratio (SNR). The proposed technology exhibits promising application potential in counter-jamming and acoustic decoy scenarios, as well as for the recognition of large underwater vehicles or submarines.

The remainder of this paper is organized as follows. Section 2 introduces the theoretical methods, including the improvement of the underwater target highlight model, the acquisition method for underwater target highlight images and the design of the HasNet-5 convolutional neural network. Section 3 describes the training, validation and testing of the network model using the generated simulation data. Finally, the target classification method is tested using experimental data and the results are analyzed and discussed. Section 4 summarizes the paper.

2. Theory and Methodology

2.1. Improvement of Underwater Target Highlight Model

According to the target highlight model, the spatial geometric structural characteristics of underwater targets (including disturbances, which can be considered as pseudo-targets) can be obtained, including structural characteristics such as the scale, shape and distribution of the strong scattering zones of the target [27]. From the perspective of a linear time-invariant system, considering only the three parameters of amplitude, time delay and phase jump of the target highlight or equivalent highlight echo, the highlight echo model of the underwater target can be obtained using a single-frequency signal ω_0 [27,28], as follows:

$$H(\mathbf{r}, \omega) = \sum_{k=1}^N A(\mathbf{r}, \omega) e^{j\omega\tau_k} e^{j\varphi_k}, \quad (1)$$

where $\omega = \omega_0 + \Delta\omega$, with $\Delta\omega$ being the Doppler shift. k represents the k th target echo highlight. N represents the number of target highlights. $A(\mathbf{r}, \omega)$ denotes the amplitude reflection factor, which is a function of the vector \mathbf{r} and the signal frequency ω . $\tau_k = 2d_k/c$ denotes the time-delay factor, which is determined by the difference in acoustic range of the k th highlight with respect to the reference point d_k and c represents the speed of sound. φ_k denotes the phase jump factor of the k th highlight.

The target highlight echo model established in the literature [29] divides the amplitude factor into two parts—the hydroacoustic channel loss and the target reflection factor—and introduces the horizontal and pitch angles of the sonar with respect to the target, as follows:

$$H(\mathbf{r}, \omega) = \sum_{k=1}^N A_k^t(\theta, \psi, \omega) A_k^c(\mathbf{r}, \omega) e^{j\omega\tau_k} e^{j\varphi_k}, \quad (2)$$

where $A_k^t(\theta, \psi, \omega)$ denotes the local plane wave reflection factor of the k th highlight of the target, which is a function of the horizontal angle θ , the pitch angle ψ and the signal frequency ω . $A_k^c(\mathbf{r}, \omega)$ denotes the hydroacoustic channel propagation loss factor.

In this study, the above underwater target highlight echo model is modified according to the requirements of extracting underwater target echo highlight images. Taking the acoustic reference center of the active sonar as the observation point and the target reference center as the origin, the horizontal and pitch angles of the observation point relative to the target highlight or equivalent highlight are refined. Furthermore, the target highlight distance is used instead of the target distance to obtain a more accurate model of the underwater target highlight, as given by Equation (3).

$$H(\mathbf{R}, \omega) = \sum_{k=1}^{N(\theta, \psi)} A(\mathbf{r}_k, \omega) V_k(\theta_k, \psi_k, \omega) e^{j\omega\tau_k} e^{j\varphi_k} \quad (3)$$

Here, $N(\theta, \psi)$ denotes the number of highlights when the horizontal angle of the sonar relative to the target reference center is θ and the pitch angle is ψ . $A(\mathbf{r}_k, \omega)$ denotes the hydroacoustic channel propagation loss factor and \mathbf{r}_k denotes the vector distance of the k th highlight of the target relative to the observation point. $V_k(\theta_k, \psi_k, \omega)$ denotes the local plane wave reflection factor of the k th highlight of the target, which indicates the spatial characteristics of the echo intensity in the scattering zone of each highlight of the target and is a function of the horizontal angle θ_k , the pitch angle ψ_k and the signal frequency ω . $\mathbf{R} = \{\mathbf{r}_k\}$ is a second-order tensor. Let the vector distance \mathbf{r}_k be (x_k, y_k, z_k) ; then, the relationship among \mathbf{r}_k , θ_k and ψ_k can be expressed as $x_k = |\mathbf{r}_k| \cdot \cos(\psi_k) \cos(\theta_k)$, $y_k = |\mathbf{r}_k| \cdot \cos(\psi_k) \sin(\theta_k)$ and $z_k = |\mathbf{r}_k| \cdot \sin(\psi_k)$.

Examples of highlight models for underwater vehicles and four types of typical disturbances are given below to provide simulation data generation models for studying the proposed method.

Underwater vehicles (such as submarines) are generally streamlined. Under active sonar excitation, they exhibit several types of scattering, such as geometric specular reflection, angular scattering, multilayer structural scattering and elastic wave scattering [30]. Previously developed underwater vehicle target highlight models [28,31–35] treat the target's echo as a superposition of 3–12 strong highlights. In [23,25,26,36], theoretical analysis and pool tests confirmed that weak highlights exist in underwater targets and that the features of the weak highlights of underwater targets can be extracted via high-resolution time-delay estimation, time-frequency analysis and chirplet atomic decomposition. It can be seen that 3–12 strong highlights are insufficient for the desired effect of simulating the echo highlights of underwater vehicles and that its weak highlights should also be considered. In this study, to more accurately simulate the echo-scattering signals of an underwater vehicle, a highlight model that can characterize its fine features is established, including seven strong highlights, six weak highlights and six dim highlights. The improved highlight model of the underwater vehicle is illustrated in Figure 1. In Figure 1, the term “Bow” denotes the bow section of the underwater vehicle, “Foreship” refers to its frontal part, while “Hull1” and “Hull2”, respectively, represent the two middle sections. Lastly, “Stern” designates the rear section of the underwater vehicle. Black circles represent strong highlights, gray circles represent weak highlights, and light gray circles represent dim highlights.

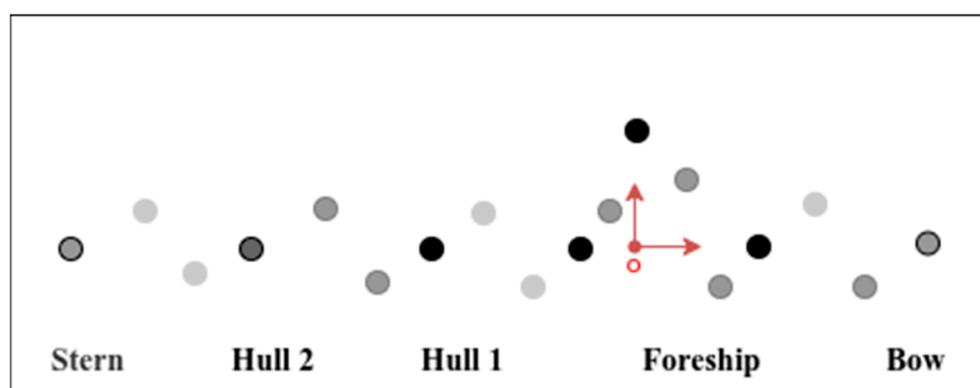


Figure 1. Improved highlight model of the underwater vehicle.

Some active sonar disturbances have receiving and transmitting sensors. A spatial distribution of multiple acoustic signal transmitting sensors is used to simulate the distribution of strong highlights of underwater vehicles to achieve the purpose of confusing the active sonar. Typical 1D disturbance (such as towed anti-torpedo acoustic decoy) can only model the horizontal strong highlight distribution characteristics of underwater vehicles [37]. The ongoing research focuses on the development of an enhanced active sonar acoustics decoy, aiming to accurately simulate both the horizontal and vertical dimensions of underwater vehicles. According to the developmental trajectory of active sonar jammers, active sonar may encounter four types of disturbances, as illustrated in Figure 2, and their highlight models are established with six to seven distinct focal points.

Type I is a 1D disturbance target with 1D geometric features and highlight distribution characteristics. Types II–IV are 2D disturbance targets with 2D geometric features and highlight distribution characteristics. The Type IV distribution closely resembles the highlight distribution of strong highlights on underwater vehicles, posing significant challenges for active sonar in terms of recognition. However, it remains challenging to accurately simulate the characteristics of localized weak highlights exhibited by real targets. In this study, the highlight images of underwater targets based on the above target highlight models are obtained. The data from these five types of targets (four types of disturbance targets and underwater vehicles) are used to study and validate the proposed method.

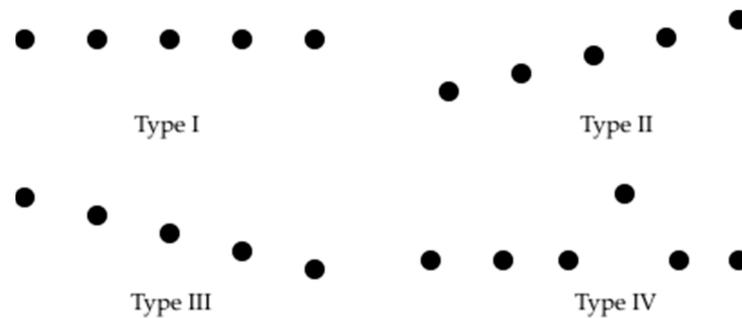


Figure 2. Models of four types of typical disturbance targets.

2.2. Method for Underwater Target Highlight Image Acquisition

During dynamic target recognition, the active sonar carrier and the target are generally in motion at relatively high speeds. It is difficult to achieve hydroacoustic imaging of the target in this case; however, using the highlight model described in the previous section, the highlight distribution image of the target can be obtained by acquiring the highlight information of the target. Cross-spectral orientation techniques [38] or high-resolution DOA estimation algorithms [23,24] can be used to extract the horizontal bearing, vertical bearing, distance and energy amplitude of the target's multiple highlights. The acquired multi-highlight bearings and distances of the target are generally based on a spherical coordinate system, and they are transformed into coordinates in the three-dimensional (3D) Cartesian coordinate system. The distribution of some of the close-range echo highlights of common underwater vehicles in the 3D Cartesian coordinate system is shown in Figure 3.

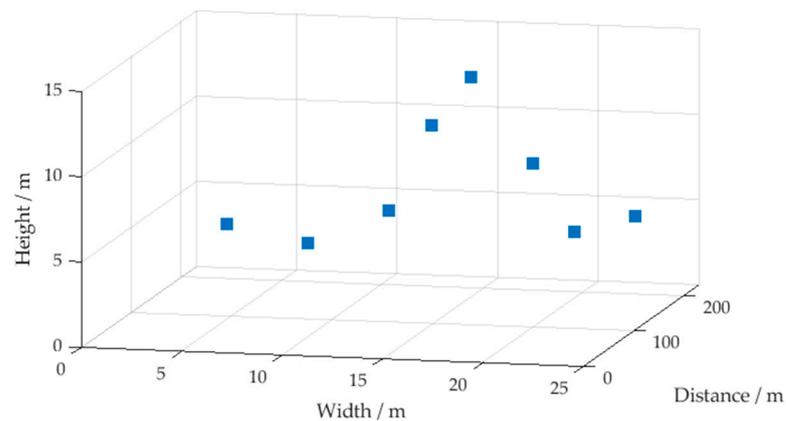


Figure 3. Distribution of some of the echo highlights of the underwater vehicle.

According to the principle of orthogonal projection, the distribution of highlights is transformed from the 3D space to the 2D space. The projection is performed along the target distance axis onto the surface with target horizontal and vertical scales. To use the convolutional classification network, the highlight distribution image needs to be further processed into a canonical grayscale image. The brief transformation process is as follows: (1) determine the size of the field of view according to the size characteristics of the target; (2) if the horizontal or vertical size is larger than the field of view, shrink all of the highlights isometrically into the field of view and calculate their pixel positions in the field of view; (3) if the horizontal or vertical size is smaller than the field of view, enlarge all of the highlights isometrically in the field of view appropriately and calculate their pixel positions in the field of view; (4) normalize the energy values of all of the highlights according to the highlight with the largest echo energy value. The normalized values are multiplied

by 255 and rounded to the nearest whole number, and the gray value $Gr[x_i, y_i]$ of the i th highlight of a certain detection cycle is calculated as

$$Gr[x_i, y_i] = f(p_w) = \frac{p_w[i]}{\arg \max_{1 \leq k \leq N} p_w[k]} + \frac{1}{2}, \tag{4}$$

where N represents the number of target highlights acquired in a given detection cycle, $p_w[i]$ denotes the energy value of the i th highlight and $i \in [1, N]$. According to the establishment conditions of the linear time-invariant system, the target channel is a time-varying space-varying channel considering different detection cycles of the active sonar. The number of echo highlights N of the underwater target should be a time-varying function. Therefore, N can be simply expressed as $N(t_n)$, i.e., the number of target highlights in the n th detection cycle of the active sonar. Let the width of the field of view of the highlight image be W and the height be H . The positions and gray values of the $N(t_n)$ highlights are represented into the field of view, and the gray values of the coordinates of the highlights that do not exist in the grayscale image are set to 0. As such, a highlight image of the target is obtained. Describing the data from the image point of view, the highlight image data of the target t_n cycle $Im(t_n, x_i, y_i)$ can be expressed as follows:

$$Im(t_n, x_i, y_i) = \frac{p_w[t_n, x_i, y_i]}{\arg \max_{(x_i, y_i)} p_w[t_n, x_i, y_i]} + \frac{1}{2}, \tag{5}$$

where $0 \leq x_i \leq W - 1, 0 \leq y_i \leq H - 1$. Some of the echo highlights of the underwater vehicle shown in Figure 3 are graphically represented in Figure 4. In Figure 4, the pixel gray value increases proportionally with the darkness of the color.

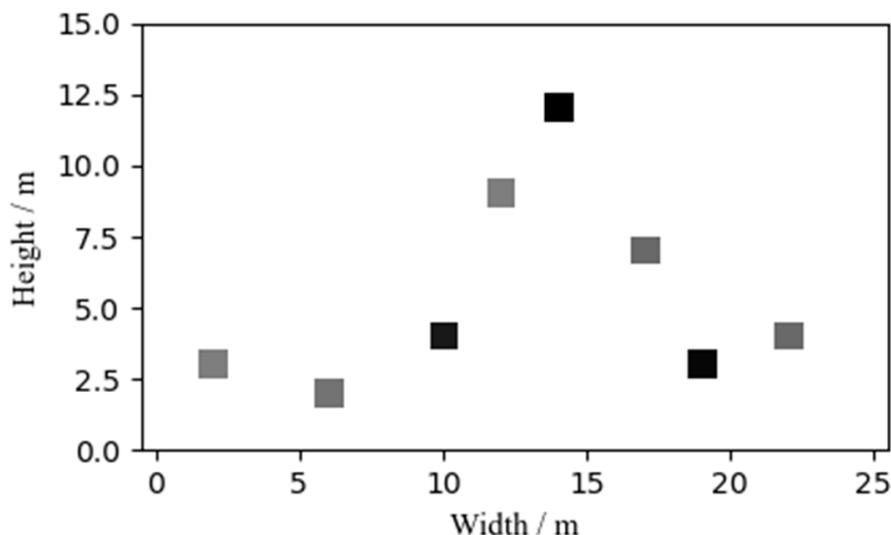


Figure 4. Local description of underwater vehicle’s echo highlights.

2.3. Design of the HasNet-5 Convolutional Neural Network

Underwater target recognition is generally a binary classification problem, i.e., recognizing the true target among many disturbances. In this study, considering the diversity of the geometric features of underwater disturbances, the target recognition problem is transformed into a multi-classification problem and a true-false judgment is performed on the target. For the classification of underwater target highlight images, a streamlined active sonar highlight model CNN network called HasNet-5 based on the LeNet-5 network structure was designed. The network structure of HasNet-5 is shown in Figure 5.

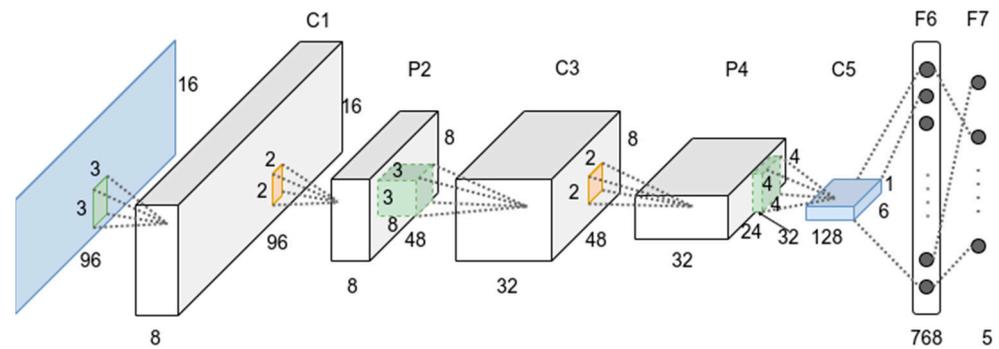


Figure 5. Structure of the HasNet-5 CNN.

The network consists of seven layers: three convolutional layers, two max pooling layers and two fully connected layers. It receives a 96×16 grayscale image as input, and the output corresponds to the probabilities of the five types. The first fully connected layer (F6) directly compresses the feature maps of the stereo output of the third convolutional layer (C5) into a 1D vector containing 768 neurons. The second fully connected layer (F7) contains five neurons, i.e., the probabilities of the five types, obtained using the softmax classifier. Table 2 presents the network parameters for each layer of HasNet-5.

Table 2. HasNet-5 model summary.

	Layer	Kernel Size	Stride	Number of Filters	Activation	Output Shape
Input	Image	-	-	-	-	96×16
1	Conv1	3×3	1	8	ReLu	96×16
2	Pool2	2×2	1	-	-	48×8
3	Conv3	3×3	1	32	ReLu	48×8
4	Pool4	2×2	1	-	-	24×4
5	Conv5	4×4	4	128	ReLu	6×1
6	FC6	-	-	-	-	768
Output	FC7	-	-	-	Softmax	5

The main reasons for HasNet-5 to be based on the LeNet-5 network structure are as follows: underwater target highlight images are similar to images of handwritten digits and LeNet-5 can efficiently classify images of handwritten digits and is a small model, which is favorable for engineering implementation. The differences between HasNet-5 and LeNet-5 include the following: the activation function of the convolutional layer is modified to the rectified linear unit (ReLu), which reduces the number of calculations; the convolutional layer uses a complementary method; the thickness of the first two convolutional layers is increased; and one fully connected layer is eliminated. In this study, the HasNet-5 network is implemented using the deep learning framework PyTorch.

3. Validation and Analysis

3.1. Dataset

The training, validation and test samples of the HasNet-5 classification network were generated by employing the established highlight models of underwater vehicles and four types of disturbance targets based on Equation (3) of the underwater target highlight model. Considering the general application of underwater target recognition, samples were generated at 10° intervals from 30° to 60° and from 120° to 150° for the horizontal bow angle of the active sonar relative to the center of the target reference and at -4° , -2° , 0° , 2° , and 4° for the pitch angle. The distance between the sonar reference center and the target centroid was assumed to be 400 m. The target speed was 4 m/s, and the sonar platform speed was 10 m/s. The sonar operates using a linear frequency modulation (LFM) signal, with a center frequency of 30 kHz and a frequency bandwidth of 1.5 kHz. The signal pulse

width was set to 15 ms, while the noise followed a Gaussian distribution. Additionally, the SNR was 6 dB. Based on these sonar parameters, the array sensor data for active sonar were generated using the target highlight model examples and Equation (3) in Section 2.1. Furthermore, the method described in Section 2.2 was employed to obtain the highlight images from the generated data. The overall process is illustrated in Figure 6. A total of 40 combinations of bow angles and pitch angles were generated, with each combination producing 250 samples for five types of targets. A total of 50,000 samples were obtained as the generated training sample set. The sample parameter configurations are presented in Table 3.

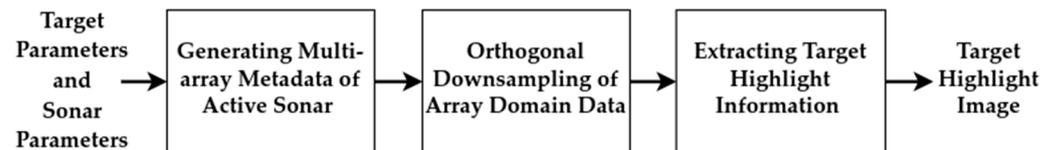


Figure 6. Process of sample generation.

Table 3. Parameter configuration for the generated samples.

Target Type	Horizontal Bow Angle (°)	Tilt Angle (°)	Number of Samples
Underwater vehicle	30, 40, 50, 60, 120, 130, 140, 150	−4, −2, 0, 2, 4	10,000
Type I disturbance	30, 40, 50, 60, 120, 130, 140, 150	−4, −2, 0, 2, 4	10,000
Type II disturbance	30, 40, 50, 60, 120, 130, 140, 150	−4, −2, 0, 2, 4	10,000
Type III disturbance	30, 40, 50, 60, 120, 130, 140, 150	−4, −2, 0, 2, 4	10,000
Type IV disturbance	30, 40, 50, 60, 120, 130, 140, 150	−4, −2, 0, 2, 4	10,000
Total			50,000

Following the aforementioned 40 combinations, a total of 10,000 samples were re-generated for the five types of targets. Each combination generated 50 samples for each target, with 5000 of them used as the validation sample set and 5000 of them used as the test sample set. The validation and test samples were invisible to the network during the training process. The generated validation samples were used to evaluate the model training effectiveness after completing a round of training. The generated test samples were used to evaluate the performance of the trained HasNet-5 network.

Echo data from Type I–IV simulated disturbances and underwater vehicles were collected via active sonar in a lake/sea to form a test sample set for the experimental data. The experimental data included 200 samples of Type I disturbance A (three array elements), 100 samples of Type I disturbance B (four array elements), 50 samples of Type II disturbance (four array elements), 50 samples of Type III disturbance (four array elements), 50 samples of Type IV disturbance (five array elements) and 100 samples of underwater vehicle targets.

The dataset comprises a generated training sample set, a generated validation sample set, a generated test sample set and an experimental data test sample set. The specific number of samples in each set is presented in Table 4.

Table 4. Quantification of each dataset.

Type of Sample Set	Generated Training Sample Set	Generated Validation Sample Set	Generated Test Sample Set	Test Sample Set of Experimental Data
Number of samples	50,000	5000	5000	550

3.2. Evaluation Metrics

According to the principle of minimizing the risk of missed detection, the HasNet-5 network was evaluated using the recall rate and its confusion matrix. The network was

evaluated using the classification error rate in the training and validation process, as given by Equation (6). Lower error rates indicate better network training.

$$ErrorRate = 1 - Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (6)$$

Here, TP , TN , FP and FN represent the numbers of true positive, true negative, false positive and false negative cases, respectively.

The confusion matrix was used to evaluate the classification effectiveness of the trained network model for the five types of target test data. The recall rate was used to evaluate the generalization ability of the trained network for target classification, and it was calculated using Equation (7). A higher recall indicates a higher rate of correct classification for a particular type of data.

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

Finally, experimental data samples were used to verify that the algorithm can discriminate among the four types of typical 2D disturbance targets to recognize the real targets.

3.3. Validation

As depicted in Table 1, the progress in identifying moving targets at long distances remains limited for time-frequency feature recognition algorithms, sonar imaging recognition algorithms and geometric structure feature recognition algorithms. Consequently, this study initially verifies the target scale recognition algorithm (TSRA) using experimental data while analyzing its existing issues. Subsequently, the proposed echo highlight image recognition method is validated and compared against the target scale recognition algorithm. The methods mentioned in the literature [8–11] essentially employ the target scale recognition algorithm, which involves determining both the horizontal and/or vertical scales of targets. Utilizing the experimental data from Section 3.1, we conduct a target classification test using the TSRA and present the classification results (confusion matrix) in Table 5.

Table 5. Test results of the TSRA.

Data Type	Classification Results					Classification Correctness Rate	Recognition Correctness Rate
	Type I Disturbance A	Type II Disturbance B	Type III Disturbance	Type IV Disturbance	Underwater Vehicle		
Type I disturbance A	197	0	0	0	3	98.5%	98.5%
Type I disturbance B	96	0	0	0	4	96%	96%
Type II disturbance	1	0	0	0	49	2%	2%
Type III disturbance	2	0	0	0	48	4%	4%
Type IV disturbance	0	0	0	0	50	0%	0%
Underwater vehicle	1	0	0	0	99	99%	99%

The classification correctness rates of the samples for Type I disturbance A and disturbance B were 98.5% and 96%, respectively, and the misclassified samples were all recognized as underwater vehicles. Type II–IV disturbances were incorrectly recognized as underwater vehicles because most of the samples exceeded the threshold values for the horizontal and vertical sizes. The samples of underwater vehicle targets were correctly recognized 99% of the time, with one sample recognized as a Type I disturbance. In Table 5, the “classification correctness rate” is the percentage of test results that were identical to the label (recall rate) and the “recognition correctness rate” is the percentage of times that

the disturbance was not recognized as an underwater vehicle or the underwater vehicle was not recognized as disturbance during the test.

The classification results of the TSRA indicate that it only had a high recognition rate for Type I disturbance; it could not discriminate among 2D disturbance targets and the underwater vehicle. It can be seen that simply extracting the horizontal and vertical sizes of the target is insufficient for classifying the 2D disturbance targets and the underwater vehicle. In this paper, a method for obtaining the underwater target's active sonar echo highlight image is proposed. The global and local features of the target's echo highlight image are extracted, and the classification and recognition of the target are achieved using the convolutional network.

Active sonar echo data for underwater targets are scarce owing to the high cost of acquiring experimental data for various underwater targets at different distances, bow angles and navigational states. In this study, to solve the problem of the lack and imbalance of measured experimental data from underwater targets, a generation model for highlight image data based on hydroacoustic physics is established as the theoretical basis and combined with examples of highlight models of underwater vehicles and the four types of disturbance targets given in Section 2.1. The generation model has similar ideas to the approach of establishing generative models based on visual features and semantic vectors in zero-shot learning [39,40] but with explicit physical implications. Drawing on the idea of data invisibility in the target-domain tasks for zero-shot learning, generated data classification is considered as a source classification task in which visual features and lexical vectors are used to perform one-to-one matching in the target-domain classification task. Therefore, the validation process was as follows: training, validation and testing of the network using generated data to obtain a classification network that performs well for the source classification task, followed by testing of the network using experimental data to verify the performance of the recognition algorithm in the target-domain task. The process of training and testing is illustrated in Figure 7.

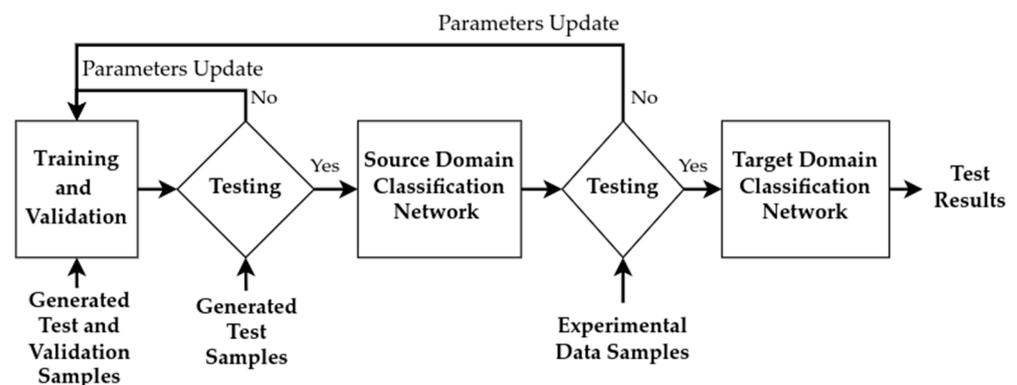


Figure 7. Training and testing process.

First, the training and validation of the network were performed using the generated 50,000 samples and 5000 validation samples, with the learning rate set to 0.0001 for 10 epochs of training and validation. The training and validation error curves obtained after training are shown in Figure 8.

The error curves of the training process indicate that the HasNet-5 network completed training. Classification tests were performed using 5000 test samples of the generated five types of targets. In testing, HasNet-5 classified more than 99% of the simulated data correctly. Partial feature maps of the second convolutional layer obtained using the learning network from Type I–IV disturbances and underwater vehicle simulation data are shown in Figures 9–11. In Figures 9–11, the pixel gray value increases proportionally with the brightness of the color. The feature maps reveal distinct global and local highlight features between the underwater vehicle and Type I–III disturbances. While the underwater vehicle shares similar global highlight features with Type IV disturbances, they exhibit noticeable

differences in terms of local highlight features. The target recognition method proposed in this paper aims to utilize the highlight features for achieving target classification. Therefore, the extracted highlight feature maps and the classification accuracy of the simulation data demonstrate the robust capability of HasNet-5 in both feature extraction and classification.

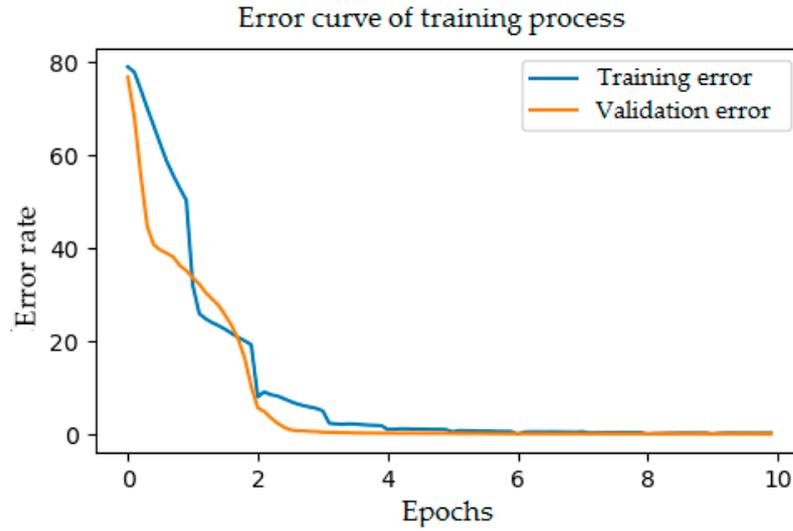


Figure 8. Error curves of the training process.

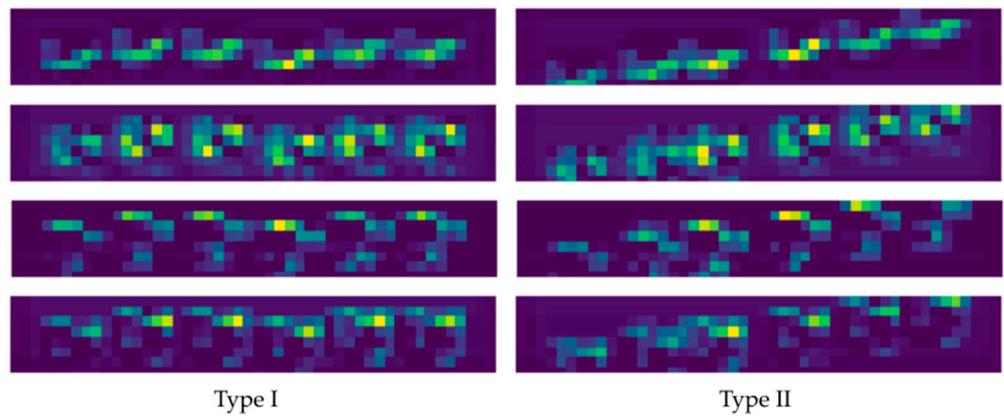


Figure 9. Feature maps of the second convolutional layer for Type I and II disturbances.

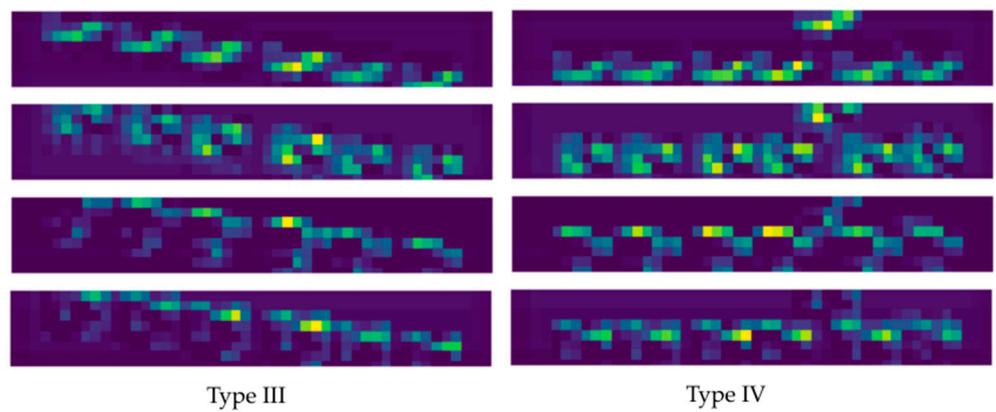


Figure 10. Feature maps of the second convolutional layer for Type III and IV disturbances.

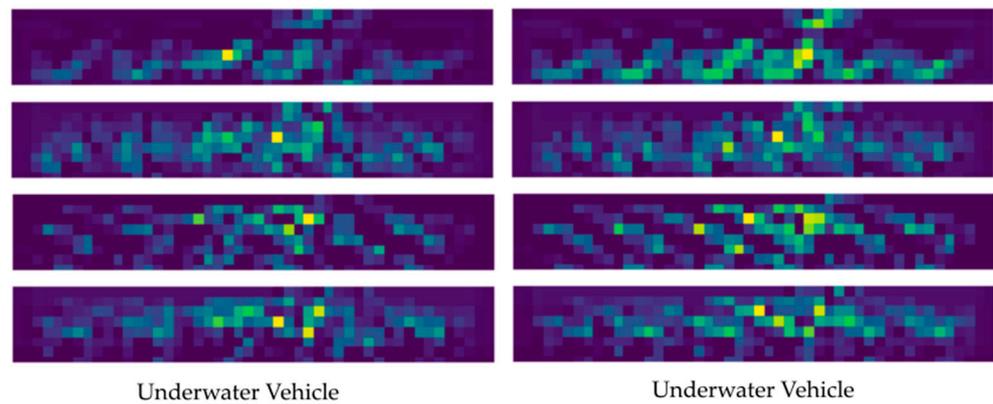


Figure 11. Feature maps of the second convolutional layer for the underwater vehicle.

Finally, the classification network was tested using the experimental data samples to evaluate its generalization performance and ability to classify target echo highlight images. The test results (confusion matrix) are presented in Table 6. Some of the feature maps of the second convolutional layer of the highlight image of the underwater vehicle in Figure 3 are shown in Figure 12, which have a high degree of similarity to the middle portion of Figure 11.

Table 6. Test results for HasNet-5 experimental data.

Data Type	Classification Results				Underwater Vehicle	Classification Correctness Rate	Recognition Correctness Rate
	Type I Disturbance A	Type II Disturbance B	Type III Disturbance	Type IV Disturbance			
Type I disturbance A	188	9	0	0	3	94%	98.5%
Type I disturbance B	96	3	0	0	1	96%	99%
Type II disturbance	1	49	0	0	0	98%	100%
Type III disturbance	2	0	48	0	0	96%	100%
Type IV disturbance	1	0	0	46	3	92%	94%
Underwater vehicle	1	1	2	4	92	92%	92%

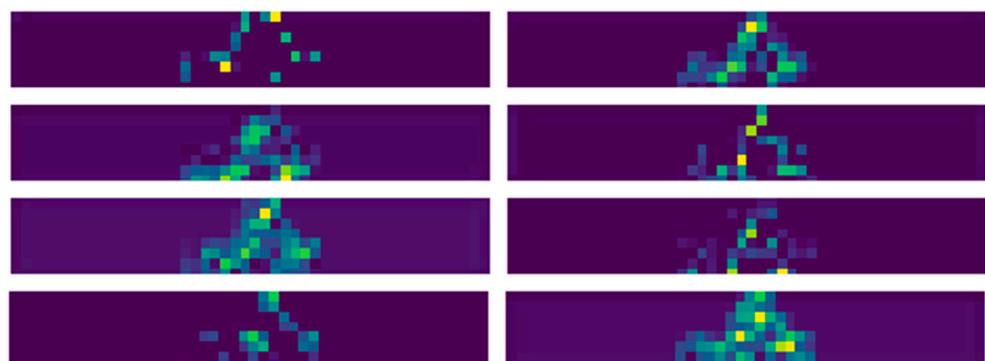


Figure 12. Feature maps of the second convolutional layer for the underwater vehicle.

3.4. Analysis

In the training of HasNet-5, convergence was generally achieved after 4–5 epochs. A classification correctness of higher than 99% was achieved using the generated test data, indicating that the algorithm extracted and classified the local and global features

of the echo highlights of the underwater target with excellent performance in the source classification task. Tests were conducted using 550 samples of experimental data to verify the generalization ability of the model and the effectiveness of the algorithm for the target-domain classification task. As shown in Table 6, during the testing process, it was found that the classification correctness rates of the experimental data samples for Type I disturbance A and Type I disturbance B were 94% and 96%, respectively. Meanwhile, there was a possibility of the disturbance being recognized as an underwater vehicle in the case of misclassification. This was found to be due to the fact that the simulated data of the underwater vehicle used in the training had a small size in the vertical direction and the echo image was highly similar to that of the Type I disturbance features. The classification correctness rates of the experimental data samples for Type II and Type III disturbances were 98% and 96%, respectively. Meanwhile, there was a possibility of the disturbance being recognized as a Type I disturbance in the case of misclassification. The classification correctness rate of the experimental data samples for Type IV disturbances was 92%, and there was a possibility of the disturbance being recognized as an underwater vehicle or a Type I disturbance in the case of misclassification. The experimental data samples from the underwater vehicle were classified correctly 92% of the time, and among the misclassifications, there was a high probability of the vehicle being recognized as Type IV disturbance. This phenomenon can be attributed to the similarity of global highlight features between the two types of targets. The classification network primarily distinguishes Type IV and underwater vehicles based on the local highlight features characterized by weak highlights. In addition, in the network training process, appropriately increasing the thickness of the convolutional layer and adjusting the length of the fully connected layer significantly improved the feature extraction and generalization ability of the classification network. Compared with LeNet-5, removing one fully connected layer enhanced the ability to extract the weak highlight features and significantly increased the classification correctness rate for the underwater vehicle and Type IV disturbance. Finally, the target recognition results of the proposed algorithm were compared with those of the TSRA, as shown in Figure 13. The validation results indicated that the proposed underwater target highlight image recognition method is capable of effectively recognizing 2D disturbances and underwater vehicles.

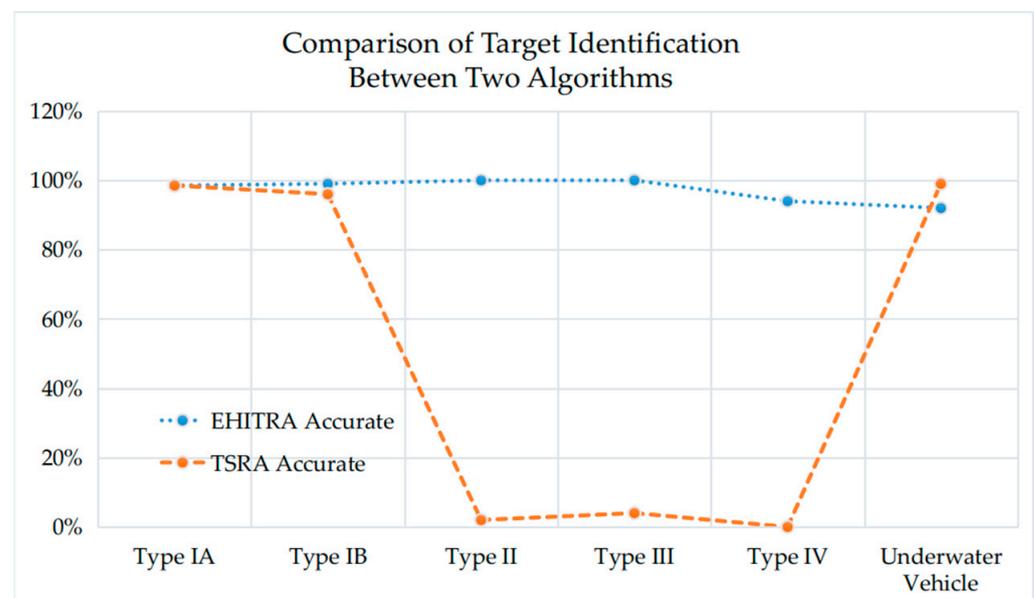


Figure 13. Comparison of classification ability between the EHITRA and TSRA.

4. Conclusions

The underwater target highlight model is enhanced in this study, wherein a novel method for the feature extraction from underwater moving target highlight images is proposed. Highlight model concrete examples are established for four typical disturbances and an underwater vehicle, and the simulation data for these five target types under multiple scenarios are generated. The HasNet-5 classification network is constructed, trained, validated and tested using the generated data from the five types of targets. This enables the network to effectively extract and classify the global and local highlight features of underwater targets. Finally, experimental data from the five types of targets are used to obtain test samples for evaluating the performance of the classification network. The results demonstrate that the recognition probability for underwater vehicles reaches 92%, while disturbances have a recognition probability higher than 94%, indicating excellent performance in classifying targets within the designated domain. Consequently, the proposed active sonar echo highlight image classification method exhibits robust capability in mitigating 2D disturbances and accurately identifying real underwater targets. The method is proved advantageous in long-distance identification tasks for active sonars from moving platforms (e.g., UUVs or torpedoes), surpassing traditional target scale recognition algorithms. It exhibits promising application potential in countering disturbances and acoustic decoy scenarios as well as identifying large underwater vehicles or submarines. However, further investigation is warranted for the classification approach of underwater target echo highlight imagery due to the limited and unbalanced test data, with the potential exploration of additional types of underwater target recognition problems.

Author Contributions: Conceptualization, X.L. and Y.Y.; methodology, X.L.; software, L.L. and Y.L.; validation, X.L., Y.Y., J.L. and X.Y.; formal analysis, L.L.; investigation, X.L.; resources, X.Y.; data curation, L.S.; writing—original draft preparation, X.L.; writing—review and editing, X.L., Y.Y., X.Y. and J.L.; visualization, X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are contained within the article.

Acknowledgments: The authors appreciate the linguistic assistance from Xiaojia Jiao during the revision of this paper.

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

References

1. Liang, K.; Wang, K. Using Simulation and Evolutionary Algorithms to Evaluate the Design of Mix Strategies of Decoy and Jammers in Anti-Torpedo Tactics. In Proceedings of the 2006 Winter Simulation Conference, Monterey, CA, USA, 3–6 December 2006; pp. 1299–1306.
2. Chen, Y.C.; Guo, Y.H. Optimal Combination Strategy for Two Swim-Out Acoustic Decoys to Countermeasure Acoustic Homing Torpedo. In Proceedings of the 2017 4th International Conference on Information Science and Control Engineering (ICISCE), Changsha, China, 21–23 July 2017; pp. 1061–1065.
3. Sun, T.; Jin, J.; Liu, T.; Zhang, J. Active sonar target classification method based on Fisher's dictionary learning. *Appl. Sci.* **2021**, *11*, 10635. [[CrossRef](#)]
4. Yu, L.; Cheng, Y.; Li, S.; Liang, Y.; Wang, X. Tracking and length estimation of underwater acoustic target. *Electron. Lett.* **2017**, *53*, 1224–1226. [[CrossRef](#)]
5. Yong, J.; Chen, Y.; Jia, B.; Zhang, Y. Simulation of Phase Characteristics of Underwater Target Acoustic Scattering. In Proceedings of the 2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP), Chongqing, China, 11–13 December 2019; pp. 1–5.
6. Liu, Z.; Li, Z.; Ma, G.; Wang, M. Submarine Target Identification based on Short-time Cross-spectrum. *Fire Control. Command. Control.* **2005**, *1*, 103–106.
7. Hao, B.; Wang, M.; Xiao, L. A Research of Echo-Bearing Fluctuation Feature of Underwater Target. *Torpedo Technol.* **2004**, *12*, 20–23.
8. Wang, Z.; Wu, J.; Wang, H.; Hao, Y.; Wang, H. A torpedo target recognition method based on the correlation between echo broadening and apparent angle. *Appl. Sci.* **2022**, *12*, 12345. [[CrossRef](#)]
9. Xu, Y.; Yuan, B.; Zhang, H. An improved algorithm of underwater target feature abstracting based on target azimuth tendency. *Appl. Mech. Mater.* **2012**, *155–156*, 1164–1169.

10. Wu, H.; Pan, M. Direction Identification Method of the Two-dimensional Scale Sound Target. *Ship Electron. Eng.* **2016**, *36*, 139–142.
11. Liu, X.; Dong, C. A Method of Distinguishing Submarine and Acoustic Decoy Based on Features of Target Space Dimension. *Torpedo Technol.* **2008**, *5*, 46–50.
12. Philips, S.; Pitton, J.; Atlas, L. Perceptual Feature Identification for Active Sonar Echoes. In Proceedings of the OCEANS, Boston, MA, USA, 18–21 September 2006; pp. 1–6.
13. Ou, H.H.; Au, W.W.L.; Syrmos, V.L. Underwater Ordnance Classification Using Time-Frequency Signatures of Backscattering Signals. In Proceedings of the OCEANS, Seattle, WA, USA, 20–23 September 2010; pp. 1–8.
14. Malarkodi, A.; Manamalli, D.; Kavitha, G.; Latha, G. Acoustic Scattering of Underwater Targets. In Proceedings of the Ocean Electronics (SYMPOL), Kochi, India, 23–25 October 2013; pp. 127–132.
15. Yu, X.; Zhai, J.; Zou, B.; Shao, Q.; Hou, G. A novel acoustic sediment classification method based on the k-medoids algorithm using multibeam echosounder backscatter intensity. *J. Mar. Sci. Eng.* **2021**, *9*, 508. [[CrossRef](#)]
16. Kubicek, B.; Sen Gupta, A.; Kirsteins, I. Feature extraction and classification of simulated monostatic acoustic echoes from spherical targets of various materials using convolutional neural networks. *J. Mar. Sci. Eng.* **2023**, *11*, 571. [[CrossRef](#)]
17. Zhang, T.; Liu, S.; He, X.; Huang, H.; Hao, K. Underwater target tracking using forward-looking sonar for autonomous underwater vehicles. *Sensors* **2020**, *20*, 102. [[CrossRef](#)] [[PubMed](#)]
18. Palomeras, N.; Furfaro, T.; Williams, D.P.; Carreras, M.; Dugelay, S. Automatic target recognition for mine countermeasure missions using forward-looking sonar data. *IEEE J. Ocean. Eng.* **2022**, *47*, 141–161. [[CrossRef](#)]
19. Zhang, B.; Zhou, T.; Shi, Z.; Xu, C.; Yang, K.; Yu, X. An underwater small target boundary segmentation method in forward-looking sonar images. *Appl. Acoust.* **2023**, *207*, 109341. [[CrossRef](#)]
20. Chungath, T.T.; Nambiar, A.M.; Mittal, A. Transfer learning and few-shot learning based deep neural network models for underwater sonar image classification with a few samples. *IEEE J. Ocean. Eng.* **2023**, *99*, 1–17. [[CrossRef](#)]
21. Kriminger, E.; Tory Cobb, J.; Príncipe, J.C. Online active learning for automatic target recognition. *IEEE J. Ocean. Eng.* **2015**, *40*, 583–591. [[CrossRef](#)]
22. Gerg, I.D.; Monga, V. Structural prior driven regularized deep learning for sonar image classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–16. [[CrossRef](#)]
23. Xia, Z.; Li, X.; Meng, X. High resolution time-delay estimation of underwater target geometric scattering. *Appl. Acoust.* **2016**, *114*, 111–117. [[CrossRef](#)]
24. Wu, Y.; Li, X.; Wang, Y. Extraction and classification of acoustic scattering from underwater target based on Wigner-Ville distribution. *Appl. Acoust.* **2018**, *138*, 52–59. [[CrossRef](#)]
25. Li, X.; Xu, T.; Chen, B. Atomic decomposition of geometric acoustic scattering from underwater target. *Appl. Acoust.* **2018**, *140*, 205–213. [[CrossRef](#)]
26. Rui, L.; Junying, A.; Gang, C. Target Geometric Configuration Estimation Based on Acoustic Scattering Spatiotemporal Characteristics. In Proceedings of the IEEE International Conference on Signal, Information and Data Processing (ICSIDP), Chongqing, China, 11–13 December 2019; pp. 1–4.
27. Tang, W. Highlight model of echoes from sonar targets. *J. Acoust.* **1994**, *19*, 92–100.
28. Song, Z.; Lanrui, L.; Xinhua, Z.; Dawei, Z.; Mingyuan, L. Simulation of Backscatter Signal of Submarine Target Based on Spatial Distribution Characteristics of Target Intensity. In Proceedings of the OES China Ocean Acoustics (COA), Harbin, China, 14–17 July 2021; pp. 234–239.
29. Zhao, A.; He, C.; Hui, J.; Niu, F. Research of Sonar Echo Highlights Measurement. In Proceedings of the OCEANS 2014, Taipei, Taiwan, 7–10 April 2014; pp. 1–9.
30. Jiang, Y.; Hao, X.; Feng, H.; Hui, J. A Study on 2-Dimensional Highlight Distribution of Underwater Target. *J. Acoust.* **1997**, *22*, 79–86.
31. Chen, H.; Fengzhen, Z.; Zhaohui, Z.; Yuan, P.; Yi, J. Echo Highlight Model of Underwater Target and Design of FPGA Signal Simulation Module. In Proceedings of the 2020 5th International Conference on Communication, Image and Signal Processing (CCISP), Chengdu, China, 13–15 November 2020; pp. 53–57.
32. Sun, R.; Ma, X.; Shu, X. Simulation of Echoes from Submarine in Shallow Waters. In Proceedings of the 2013 2nd International Conference on Measurement, Information and Control, Harbin, China, 16–18 August 2013; pp. 851–854.
33. Kim, B.I.; Lee, H.U.; Park, M.H. A Study on Highlight Distribution for Underwater Simulated Targets. In Proceedings of the IEEE International Symposium on Industrial Electronics Proceedings (Cat. No. 1), Busan, Republic of Korea, 12–16 June 2001; pp. 1988–1992.
34. Liu, W.; Zhao, J.; Song, Y.; Zhang, J. Underwater Target Modeling Technology Based on Modified Highlight Model. *Torpedo Technol.* **2010**, *18*, 352–356.
35. Deng, K.; Xiang, X.; Gu, J. Multi-highlight model of scaling acoustic decoy. *Acoust. Technol.* **2011**, *30*, 201–205.
36. Wang, B.; Wang, W.; Fan, J.; Zhao, K.; Zhou, F.; Tan, L. Modeling of bistatic scattering from an underwater non-penetrable target using a Kirchhoff approximation method. *Def. Technol.* **2022**, *18*, 1097–1106. [[CrossRef](#)]
37. Zhao, J. Based on scale constraint K-means underwater target highlights clustering algorithm. In Proceedings of the 2023 4th International Conference on Computer Engineering and Application (ICCEA), Hangzhou, China, 7–9 April 2023; pp. 456–459.

38. Luo, T.; Xing, G.; Ge, C.; Niu, X. DOA Estimation based on Cross-Spectrum through a Co-Prime Array. In Proceedings of the 2021 4th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE), Changsha, China, 26–28 March 2022; pp. 817–822.
39. Long, Y.; Liu, L.; Shen, F.; Shao, L.; Li, X. Zero-shot learning using synthesized unseen visual data with diffusion regularisation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 2498–2512. [[CrossRef](#)]
40. Larochelle, H.; Erhan, D.; Bengio, Y. Zero-data Learning of New Tasks. In Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence, Chicago, IL, USA, 13–17 July 2008; pp. 646–651.

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.