

Article

Method for Augmenting Side-Scan Sonar Seafloor Sediment Image Dataset Based on BCEL1-CBAM-INGAN

Haixing Xia, Yang Cui *, Shaohua Jin, Gang Bian, Wei Zhang and Chengyang Peng

Department of Oceanography and Hydrography, Dalian Naval Academy, Dalian 116018, China; a1466448779@163.com (H.X.); jsh_1978@163.com (S.J.); trighosts@163.com (G.B.); 18590361852@163.com (W.Z.); 18908414801@163.com (C.P.)

* Correspondence: 13998435151@163.com; Tel.: +86-13998435151

Abstract: In this paper, a method for augmenting samples of side-scan sonar seafloor sediment images based on CBAM-BCEL1-INGAN is proposed, aiming to address the difficulties in acquiring and labeling datasets, as well as the insufficient diversity and quantity of data samples. Firstly, a Convolutional Block Attention Module (CBAM) is integrated into the residual blocks of the INGAN generator to enhance the learning of specific attributes and improve the quality of the generated images. Secondly, a BCEL1 loss function (combining binary cross-entropy and L1 loss functions) is introduced into the discriminator, enabling it to focus on both global image consistency and finer distinctions for better generation results. Finally, augmented samples are input into an AlexNet classifier to verify their authenticity. Experimental results demonstrate the excellent performance of the method in generating images of coarse sand, gravel, and bedrock, as evidenced by significant improvements in the Fréchet Inception Distance (FID) and Inception Score (IS). The introduction of the CBAM and BCEL1 loss function notably enhances the quality and details of the generated images. Moreover, classification experiments using the AlexNet classifier show an increase in the recognition rate from 90.5% using only INGAN-generated images of bedrock to 97.3% using images augmented using our method, marking a 6.8% improvement. Additionally, the classification accuracy of bedrock-type matrices is improved by 5.2% when images enhanced using the method presented in this paper are added to the training set, which is 2.7% higher than that of the simple method amplification. This validates the effectiveness of our method in the task of generating seafloor sediment images, partially alleviating the scarcity of side-scan sonar seafloor sediment image data.

Keywords: sample amplification; side-scan sonar; background image; convolutional attention mechanism; BCEL1loss function



Citation: Xia, H.; Cui, Y.; Jin, S.; Bian, G.; Zhang, W.; Peng, C. Method for Augmenting Side-Scan Sonar Seafloor Sediment Image Dataset Based on BCEL1-CBAM-INGAN. *J. Imaging* **2024**, *10*, 233. <https://doi.org/10.3390/jimaging10090233>

Academic Editor: Toon Goedemé

Received: 7 August 2024

Revised: 18 September 2024

Accepted: 19 September 2024

Published: 20 September 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

With the continuous development and utilization of marine resources, research on seafloor sediment environments has gained increasing attention. Side-scan sonar technology, as a crucial underwater detection method, plays a significant role in marine surveys, underwater archeology, and seafloor geological studies [1–4]. Side-scan sonar systems transmit sound waves and record their reflections to generate seafloor sediment images, which are essential for understanding seabed topography, detecting underwater targets, and assessing changes in marine environments [5,6].

However, traditional side-scan sonar seafloor sediment images are often constrained by factors such as underwater environmental complexity, imaging resolution limitations, and insufficient lighting conditions. These issues result in noise, blurriness, and occlusions in the images, limiting their usability and application scope. Moreover, acquiring datasets of seafloor sediment images is challenging due to high labeling costs and a lack of diversity and quantity of data samples [7–9]. Therefore, effectively enhancing and augmenting side-scan sonar seafloor sediment images to improve their quality and informativeness

have become current research hotspots. With the advancement and widespread adoption of deep learning technologies, as well as the increasing demand in marine research, deep learning-based object detection methods far surpass traditional machine learning approaches, garnering extensive attention in underwater detection fields [10–13]. The automatic identification and classification of seafloor sediment images also hold significant importance in marine geological surveys and seabed resource exploration. There are studies on methods for augmenting images of underwater targets obtained using side-scan sonar, such as the method designed by Tang Yulin et al. [14] based on CSLS-CycleGAN for augmenting high-quality samples of zero- and small-sample underwater representative targets, methods for augmenting samples of side-scan sonar sediment images are lacking.

Currently, the mainstream method for image augmentation involves using generative adversarial networks (GANs) for image generation, with most networks requiring large datasets for training. However, there is a significant scarcity of side-scan sonar images corresponding to certain types of sediment in existing datasets. For instance, Quanyin Zhang et al. [15] encountered the problem of having only one sample point for coarse sand in their sediment classification study, merging coarse sand with fine sand into the sand category. Therefore, there is a need to design GANs capable of augmenting small samples to address the scarcity of side-scan sonar seafloor sediment image data. In 2019, Assaf Shocher et al. [16] proposed the INGAN network, which is suitable for training on a single input image and learning block distributions within it to synthesize numerous new natural images of different sizes. However, using the INGAN network to train side-scan sonar seafloor sediment images tends to generate unrealistic images. Thus, to enhance the quality of the generated images, this paper proposes an improved method based on CBAM-BCEL1-INGAN for augmenting side-scan sonar seafloor sediment image datasets. This method incorporates an attention mechanism into the residual blocks of the generator to enhance the learning of specific attributes and improve the quality of the generated images. Additionally, the BCEL1 loss function is introduced into the discriminator, allowing it to focus on both global image consistency and finer distinctions. The aim is to effectively utilize existing data to expand the dataset and enhance the performance and generalization ability of deep learning models. Experimental results demonstrate that the proposed method for augmenting side-scan sonar seafloor sediment image datasets can learn features of various sediment sonar images and generate a large number of augmented samples, providing an effective solution to the problem of the lack of diversity and quantity of side-scan sonar seafloor sediment image data samples.

2. Materials and Methods

2.1. InGAN Model

The purpose of InGAN is to learn the internal blocks of images rather than making structural and stylistic transformations to the images. However, using the INGAN network to train side-scan sonar seafloor sediment images tends to generate unrealistic images. Given the distinct grayscale and texture features of side-scan sonar images, this paper proposes a method for augmenting side-scan sonar seafloor sediment image datasets based on CBAM-BCEL1-INGAN. The specific network process is illustrated in Figure 1. Firstly, an attention mechanism is incorporated into the residual blocks of the INGAN network generator to enhance the learning of specific attributes and improve the quality of the generated images. Secondly, the discriminator introduces the BCEL1 loss function (a combination of binary cross-entropy loss and L1 loss). This allows the discriminator to focus on both global image consistency (L1 loss) and finer distinctions (binary cross-entropy loss), thereby achieving better generation results. Finally, the augmented samples are added to the test set for sediment classification in order to confirm their authenticity.

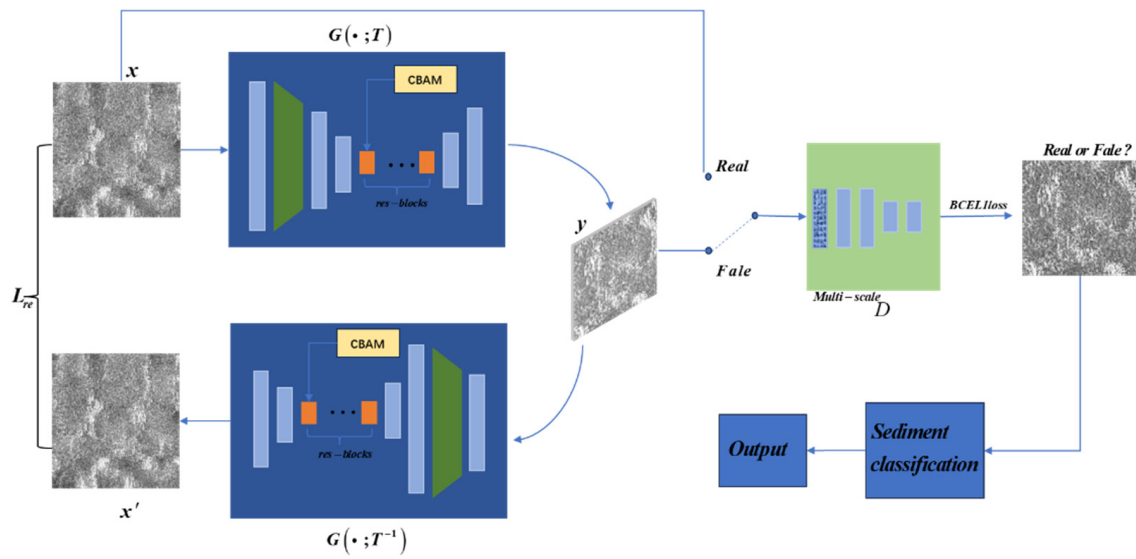


Figure 1. CBAM-BCE1-INGAN network flowchart.

The InGAN model consists of a generator G and a multiscale discriminator D . The generator redirects input x to output y , whose size or shape is determined by a geometric transformation T . The multiscale discriminator D learns to distinguish the block statistics of the fake output y from the real block statistics of the input images. Additionally, by leveraging the self-homomorphism of G and by using G and its inverse transformation T^{-1} , we reconstruct input x from y . This introduces a concept similar to “cycle consistency”, where y generated from x is fed back into the generator to reconstruct the original scale image x' , ideally making x and x' identical.

However, this process differs fundamentally from approaches such as CycleGAN [17,18], which involve dual generators and discriminators: one for $A \rightarrow B$ and another for $B \rightarrow A$ transformations. In contrast, InGAN operates with a single pathway, focusing on learning internal blocks of images without structural or stylistic transformations.

The generator in InGAN comprises three parts: convolutional layers for up or down-sampling and image feature extraction, geometric transformation layers for image scale transformation, and residual layers for deepening image feature extraction. The structure of the generator is illustrated in Figure 2.

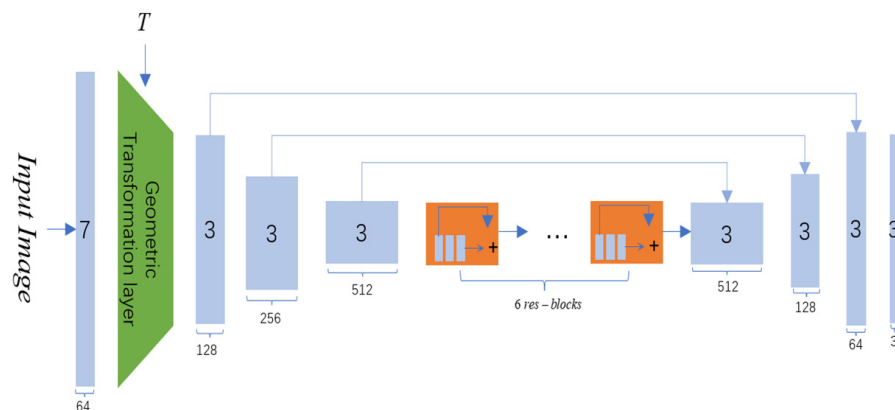


Figure 2. Generator structure diagram.

In the optimization phase, as shown in Figure 3, discriminator D is a multiscale discriminator. It evaluates the authenticity of generated images at different scales by comparing them with real images, weighting the scores obtained from the different scales

to optimize the adversarial loss. This approach enhances the control and stability of the GAN in generating high-quality images.

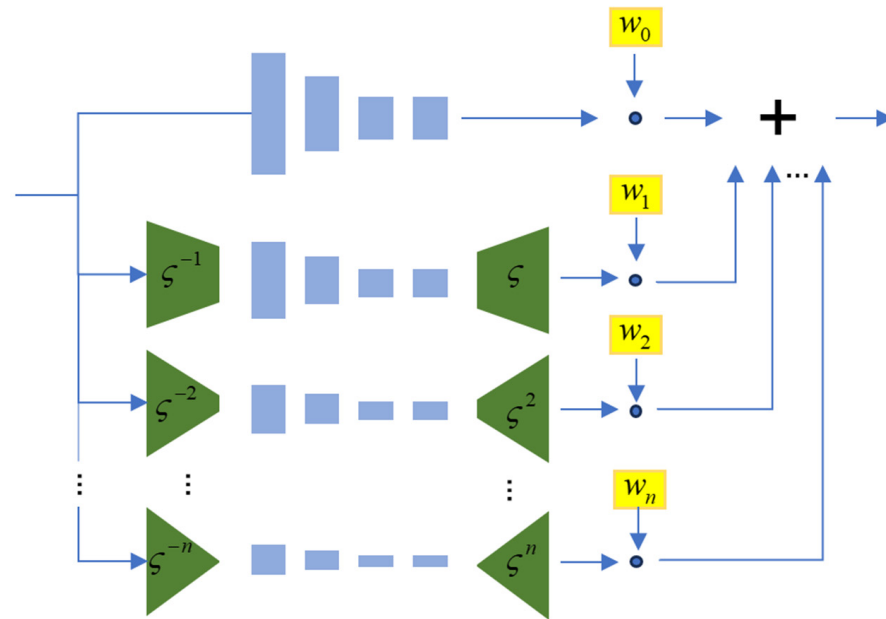


Figure 3. Multiscale discriminator.

The discriminator employs a fully convolutional structure, as depicted in the figure. At a single scale, it typically consists of four convolutional layers: a convolutional extraction layer, followed by a downsampling layer, a conventional convolutional layer, and, finally, a sigmoid activation layer producing scores in the range [0, 1]. This setup analyzes the structure of the image at that scale to determine authenticity.

For a multiscale evaluation, the discriminator computes weighted scores for each scale and aggregates them to produce the final output. Scale weights are designed to enable multiscale discrimination, enhancing the model’s ability to analyze images across different resolutions effectively.

$$n = \left\lceil \log_{\zeta} \left(\frac{InputSize}{ReceptiveField} \right) \right\rceil \tag{1}$$

$$\sum_{i=0}^n w_i = 1 \tag{2}$$

where n represents the number of scales or layers in the multiscale discriminator; ζ indicates the downsampling factor; $InputSize$ denotes the spatial resolution (number of pixels) of the input image at scale n ; $ReceptiveField$ represents the effective field of the discriminator, indicating the range over which each position in the network receives input; and w_i signifies the weights allocated to each scale, used to weight the outputs of the discriminator at different scales to obtain the final discriminative result.

The generator uses the geometric transformation T^{-1} and the forged y to obtain the reconstructed version of x' . The optimization method similar to LSGAN [19] is as follows:

$$L_{GAN}(G, D) = E_{y \sim P_{data}(x)} \left[(D(x) - 1)^2 \right] + E_{x \sim P_{data}(x)} \left[D(G(x))^2 \right] \tag{3}$$

where x represents a real sample, $D(x)$ indicates the score given to the real sample by the discriminator. A score closer to 1 means that the discriminator considers the sample more realistic. $G(x)$ represents a sample generated by the generator based on x , which follows the same distribution. $D(G(x))$ is the score given by the discriminator to the generated sample. If D considers the generated sample more fake, the score $D(G(x))$ will be closer to 0.

The reconstruction process is as follows: first, $y = G(x; T)$, and then $x' = G(y; T^{-1})$. The reconstruction loss is as follows:

$$L_{re} = \left\| G(G(x; T); T^{-1}) - x \right\|_1 \tag{4}$$

Through the loss of the confrontation process and the reconstruction process, the final loss function of InGAN is obtained as follows:

$$L_{INGAN} = L_{GAN} + \lambda \cdot L_{re} \tag{5}$$

2.2. Residual Block Based on CBAM

In the research using GANs to augment side-scan sonar image data, it is crucial to fully learn the background and texture features of the images. In the generator of the INGAN network, six residual layers are added to deepen the network. In this study, to enhance the generator’s learning of specific attributes and improve the quality of the generated images, a Convolutional Block Attention Module (CBAM) is incorporated into the residual blocks, as illustrated in Figure 4.

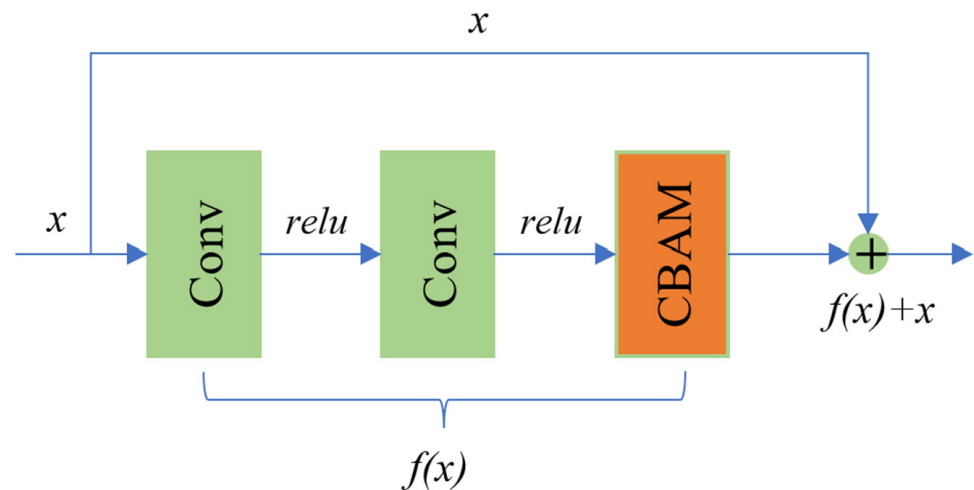


Figure 4. Residual block structure based on CBAM.

The CBAM [20] consists of two submodules: channel attention and spatial attention. Channel attention is used to adjust the importance between different channels of the feature map, while spatial attention is used to adjust the importance between the spatial positions of the feature map. The combination of these two submodules can make the model more focused on important features and help to improve its generalization ability and performance.

In CBAM-based residual blocks, features are first extracted through convolution operations. Then, the CBAM is used to adjust the channel attention and spatial attention to these features. Finally, the adjusted features are residually connected to the input in order to retain the information of the original features and enhance the representation of important features.

In the residual block, two convolution layers are used to extract the features of the input tensor, the attention mechanism is used to adjust the features to make the features more prominent, and then the attention-adjusted features are added to the original input tensor to form the final output. The aim is to prevent the problems of increasing training difficulty and information loss when training the deep network.

2.3. Discriminator Based on BCE L1 Loss Function

In our study, we choose binary cross-entropy loss and L1 loss as loss functions in the generative adversarial network (GAN). These two loss functions each provide effective

optimization targets for different aspects of the GAN. Binary cross-entropy loss is mainly used in discriminator training. It measures the difference between the real sample and the generated sample based on the log-likelihood principle, and it enables the discriminator to distinguish the two more accurately. Specifically, the binary cross-entropy loss function (BCE) is as follows:

$$BCE(D(x), y) = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\sigma(D(x_i))) + (1 - y_i) \log(1 - \sigma(D(x_i)))] \quad (6)$$

where y is a binary label (0 or 1) representing the true class of the sample and $D(x)$ denotes the output of the discriminator, which represents the probability that the sample is real.

$L1$ loss, in contrast, focuses on pixel-level differences in the generated image, helping to maintain structural and semantic consistency between the generated image and the target image. The $L1$ loss function is as follows:

$$L1(G(z), t) = \frac{1}{N} \sum_{i=1}^N |G(z_i) - t_i| \quad (7)$$

where $G(z)$ is the generated image, t is the target image, and N is the number of pixels.

The reason for choosing these two loss functions is that they emphasize different aspects of the adversarial generation network, allowing the model to optimize the quality and consistency of the generated images more comprehensively. However, due to the different measurement scales of these two loss functions, directly adding them together may result in one loss function contributing excessively to the total loss. To avoid this issue, the total loss function is defined as the comparison and sum of the two loss functions, with normalization applied to both loss functions.

$$BCE_n = \frac{BCE}{BCE_{init}} \quad (8)$$

$$L1_n = \frac{L1}{L1_{init}} \quad (9)$$

$$TotalLoss = \frac{1}{2}(BCE_n + L1_n) \quad (10)$$

where BCE_{init} and $L1_{init}$ are the initial values of the binary cross-entropy loss function and the $L1$ loss function, respectively. By using the above method, both loss functions can maintain a similar scale during the training process, thus avoiding bias towards one particular loss function.

3. Experimental and Results

Based on the CBAM-BCEL1-INGAN framework, augmenting the seabed sediment images obtained from side-scan sonar data is a crucial component of this study. To evaluate the feasibility and effectiveness of our approach, we assess the performance of the proposed GAN method. Firstly, qualitative and quantitative analyses are conducted on the quality of the generated images. Subsequently, through ablation experiments, we qualitatively and quantitatively analyze the effectiveness of the strategies employed within the GAN. Finally, a trained classification model is utilized to predict the classes of the generated images, validating their effectiveness.

The dataset used in our experiments consists of side-scan sonar data collected in 2019 from the Jiaozhou Bay area, Qingdao, using a Klein4000 side-scan sonar system manufactured by Klein Corporation of United States. We selected representative images depicting mud-sand, sandy mud, fine sand, coarse sand, gravel, and bedrock sediment types for experimentation. Examples of some sample images are shown below, see Figure 5.

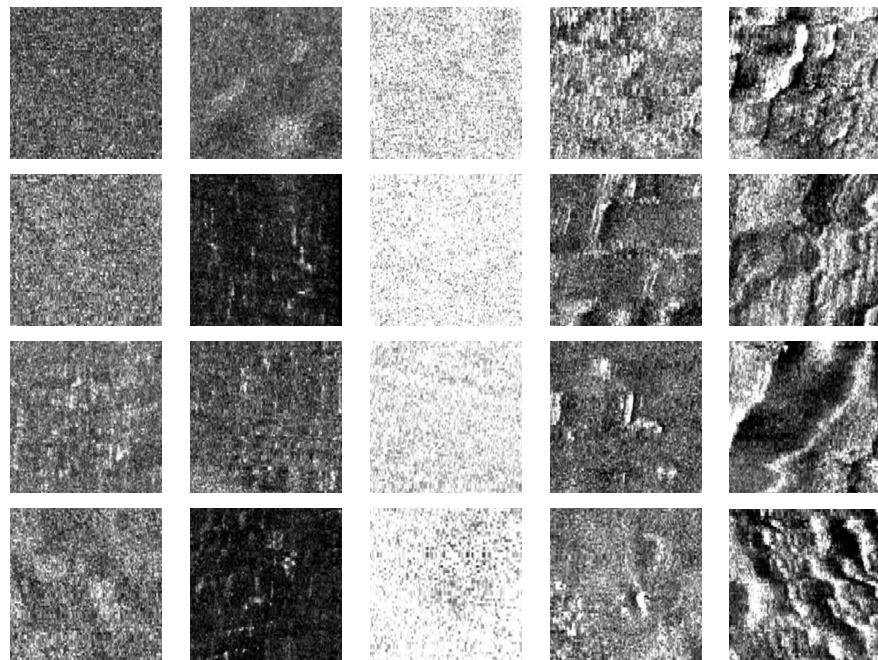


Figure 5. Parts of the samples in the dataset.

3.1. Evaluation Index

For the generated image, we mainly consider two factors, namely, the sharpness and the diversity, while also considering the similarity between the generated image and the real image; therefore, in this study, the Frechet Inception Distance (*FID*), Kernel Maximum Mean Discrepancy (*MMD*), Inception Score (*IS*), Peak Signal-to-Noise Ratio (*PSNR*), and Structural Similarity (*SSIM*) are selected as evaluation indicators to analyze the quality of the generated images.

FID is an indicator used to assess the quality and variety of images generated by a generative model. It works by comparing the distribution of the generated image and the real image in a specific space, and it represents the distance between the feature vector of the generated image and the feature vector of the real image. The calculation formula is

$$FID = \|\mu_g - \mu_r\|_2^2 + Tr(\Sigma_g + \Sigma_r - 2(\Sigma_g \Sigma_r)^{1/2}) \tag{11}$$

where $\|\mu_g - \mu_r\|_2^2$ is the *L2* norm of the square of the mean vector difference; *Tr* represents the trace of the matrix (i.e., the sum of the diagonal elements of the matrix); and $(\Sigma_g \Sigma_r)^{1/2}$ is the square root of the product of the covariance matrix Σ_g and Σ_r , representing the matrix obtained by taking the square root of the eigenvalues of the product of two matrices.

According to the formula, it can be observed that the greater the difference in the mean vector between the generated and sample images, the larger the *FID* value. The main objective of this study is to ensure that single-image augmentation methods generate diverse images. This diversity allows for the generated images to possess characteristics similar to those of the sample images while maximizing the difference in the mean vector from the sample images. Therefore, a larger *FID* value indicates greater diversity in the generated images.

The *MMD* is the maximum average difference between two distributed samples that are small enough to consider the two distributions the same; otherwise, they are considered different. In image generation, the lower the *MMD* value, the more realistic the generated image. The calculation formula is

$$MMD^2(P_r, P_g) = E_{x_r, x_r' \sim P_r, x_g, x_g' \sim P_g} [k(x_r, x_r') - 2k(x_r, x_g) + k(x_g, x_g')] \tag{12}$$

where x_r is the source domain data, and x_g is the target domain data, which measures the difference between the real distribution P_r and the generated distribution P_g given a fixed kernel function k .

The Inception Score is a metric used to measure the sharpness and variety of the generated images.

$$IS(G) = \exp(E_x D_{KL}(p(y|x) \parallel p(y))) \tag{13}$$

where $p(y|x)$ represents the probability distribution of the generated images belonging to each category, and $p(y)$ represents the probability distribution of the label vectors obtained from the generated samples.

The PSNR is a vital measure in image quality assessment and is widely used in image processing. In image denoising, the noise power is determined by computing the Mean Squared Error (MSE) between the denoised image and the original image. For an original image (I) and its denoised version (R), the MSE is computed as follows:

$$MSE = \frac{1}{MN} \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} [I(i, j) - R(i, j)]^2 \tag{14}$$

The PSNR is defined as

$$PSNR = 10 \log_{10} \left(\frac{Max_I^2}{MSE} \right) \tag{15}$$

where Max_I^2 represents the maximum pixel value of the original image. In image denoising research, a higher PSNR indicates that the denoised image R contains less noise, thus implying better denoising effectiveness.

Structural Similarity (SSIM) is a metric used to gauge image similarity, and it is also applicable for assessing compressed image quality. This study utilizes SSIM to evaluate denoised image quality by computing the Structural Similarity between the denoised and original images. SSIM assesses image similarity based on three factors: luminance, contrast, and structure.

Given two input images, x and y , the definition of SSIM is as follows:

$$SSIM(x, y) = [l(x, y)]^\alpha [c(x, y)]^\beta [s(x, y)]^\gamma (\alpha, \beta, \gamma > 0) \tag{16}$$

With

$$l(x, y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \tag{17}$$

$$c(x, y) = \frac{2\sigma_{xy} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \tag{18}$$

$$s(x, y) = \frac{\sigma_{xy} + c_3}{\sigma_x + \sigma_y + c_3} \tag{19}$$

where $l(x, y)$ is the brightness comparison, $c(x, y)$ is the contrast comparison, and $s(x, y)$ represents the structure comparison. μ_x and μ_y denote the mean values of x and y , respectively, while σ_x and σ_y represent the standard deviations of x and y , respectively. σ_{xy} denotes the covariance between x and y , and c_1 , c_2 , and c_3 are constants used to avoid division by zero errors.

In practical calculations, it is common to set $\alpha = \beta = \gamma = 1$ and $c_3 = c_2/2$. This simplifies the definition of SSIM to

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{20}$$

SSIM ranges from 0 to 1, with a higher value indicating less disparity between the output image and the undistorted image, indicating superior image quality. When two images are identical, SSIM equals 1.

3.2. Experimental Design

To verify the effectiveness of the CBAM-BCEL1-INGAN side-scan sonar image dataset amplification method proposed in this paper, the side-scan sonar images of six types of sediment—muddy sand, sandy mud, fine sand, coarse sand, gravel, and bedrock—were amplified, and images with representative characteristics were selected for training. New natural images of different sizes were generated. The model training was implemented using Python language based on the Pytorch framework, and the hardware environment was as follows: the operating system was Windows 11; the CPU was 12th Gen Intel (R) Core (TM) i9-12900H 2.50GHz; the GPU was one NVIDIA GeForce RTX3050; and the memory was 4GB.

3.2.1. Validity Verification of Amplified Images

To verify the effectiveness of substrate image amplification using this method, representative images of different substrates were selected for training. Some amplification samples are shown in Table 1 below.

The FID, MMD, IS, PSNR, and SSIM indices were calculated to evaluate the image quality of nine amplified samples of the above six types of substrates.

When the batch image is amplified, the smaller the FID value, the higher the quality and diversity of the generated image, but, for single-image amplification, the smaller the FID value, the closer the generated image to the original image; thus, in the process of single-image amplification, a larger FID value indicates that the generated image has higher quality and diversity. An analysis of Table 2 shows that coarse sand, gravel, and bedrock have high FID values, indicating that these three substrates have better effects in the single-image substrate amplification experiment. The MMD values of all categories are around 1.02, which indicates that there is a certain similarity between the generated image and the real image in terms of feature statistics. Regarding value size, there is no significant difference between several substrates. Regarding the IS index, the larger the value, the better the clarity and diversity of the generated image, and an analysis of the value shows that the three substrates of coarse sand, gravel, and bedrock have better effects. By analyzing the PSNR and SSIM metrics, it is found that the PSNR values for all categories are relatively low, and the SSIM values are close to 0. This indicates that the generated images have a low similarity to the original images, which is sufficient to demonstrate the diversity of the generated images.

For the above analysis, the entropy of the gray co-occurrence matrix is used for a quantitative analysis. When all values in the co-occurrence matrix are equal, or the pixel values show the greatest randomness, the entropy is the highest. Therefore, the entropy value indicates the complexity of the gray distribution of the image. The higher the entropy value, the more complex the image. It can be seen from the entropy value in Table 2 that the entropy of the three substrates, coarse sand, gravel, and bedrock, is larger.

3.2.2. Ablation Experiment and Evaluation

The role of each module in the performance of this model was verified. Ablation experiments were conducted on the CBAM and BCEL1 loss function by using the control variable method, and the evaluation indices were the FID, MMD, IS, PSNR, and SSIM. Four groups of control experiments were designed with the bedrock 1 image as the experimental object, and the experimental results are shown in Table 3. By comparing Groups 1 and 2, it can be seen that the quality of the images generated by the model after incorporating the attention mechanism is higher, which proves the effectiveness of the residual block based on the CBAM proposed in this paper for the model. By comparing Groups 3 and 1, we can see the superiority of the BCEL1 loss function proposed in this paper. By comparing Group

4 with Groups 2 and 3, it can be seen that the model with the combination of the CBAM and BCEL1 loss function has better performance than that with only a single strategy, indicating that the combination plays a crucial role in improving the overall performance of the model, thus reflecting the effectiveness of the method proposed in this paper.

Table 1. Partial amplification examples.

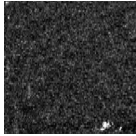
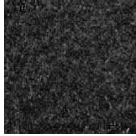
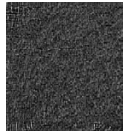
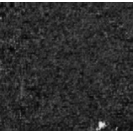
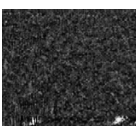

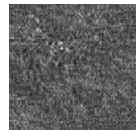
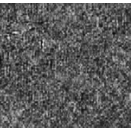
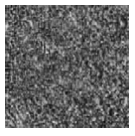
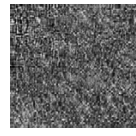

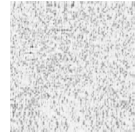
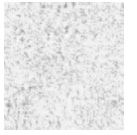
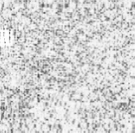


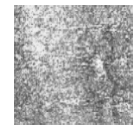
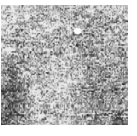








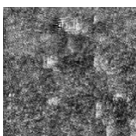
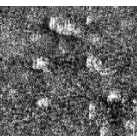
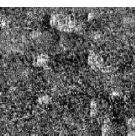
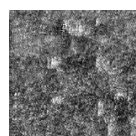
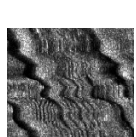

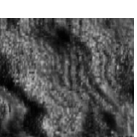


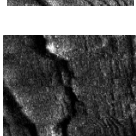

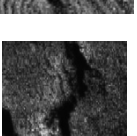
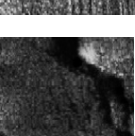
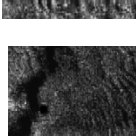

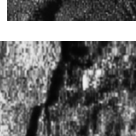



Substrate	Training Sample		Generated Sample		
Muddy sand					
Sandy mud					
Fine sand					
Coarse sand					
Gravel 1					
Gravel 2					
Bedrock 1					
Bedrock 2					
Bedrock 3					

Table 2. Image quality evaluation.

Group	Entropy of Gray Co-Occurrence Matrix	FID	MMD	IS	PSNR	SSIM
Muddy sand	11.9461	448.58	1.023	1.2474 ± 0.1040	14.32	0.19
Sandy mud	13.2604	403.44	1.022	1.2379 ± 0.0602	9.52	0.06
Fine sand	11.6167	284.60	1.017	1.1847 ± 0.1110	9.44	0.11
Coarse sand	14.2414	902.46	1.017	1.6031 ± 0.2075	6.67	0.02
Gravel 1	13.8789	1001.79	1.020	1.4307 ± 0.1345	8.12	0.10
Gravel 2	13.5145	686.24	1.020	1.4044 ± 0.1315	10.19	0.09
Bedrock 1	13.4833	1024.62	1.033	1.4070 ± 0.1720	6.49	0.03
Bedrock 2	12.3735	987.43	1.020	1.4499 ± 0.1658	10.13	0.09
Bedrock 3	13.8416	576.86	1.015	1.3268 ± 0.0914	9.00	0.18

Table 3. Network performance of different methods.

Model	CBAM Model	BCEL1 Loss	FID	MMD	IS	PSNR	SSIM
1	—	—	877.12	1.035	1.2711 ± 0.1176	7.10	0.069
2	—	✓	895.29	1.033	1.3861 ± 0.1671	7.03	0.065
3	✓	—	944.60	1.033	1.3667 ± 0.0935	7.25	0.066
4	✓	✓	1024.62	1.033	1.4070 ± 0.1720	6.44	0.058

The partial amplification of bedrock 1 images by the four groups of models trained with different strategies is shown in Figure 6. By comparing Models 2 and 1, it can be seen that the model with the CBAM can improve the quality and diversity of the generated images in terms of the evaluation indicators, but some images show unnecessary details. By comparing Models 3 and 1, it can be seen that the model using the BCEL1 loss function can achieve better image generation, but it also produces unnecessary defects while improving the index. By comparing Models 4 and 1, it can be seen that the model combining the CBAM and BCEL1 loss function performs well in terms of the evaluation indicators and the image texture, edge, and other details, which proves the effectiveness of the proposed method.

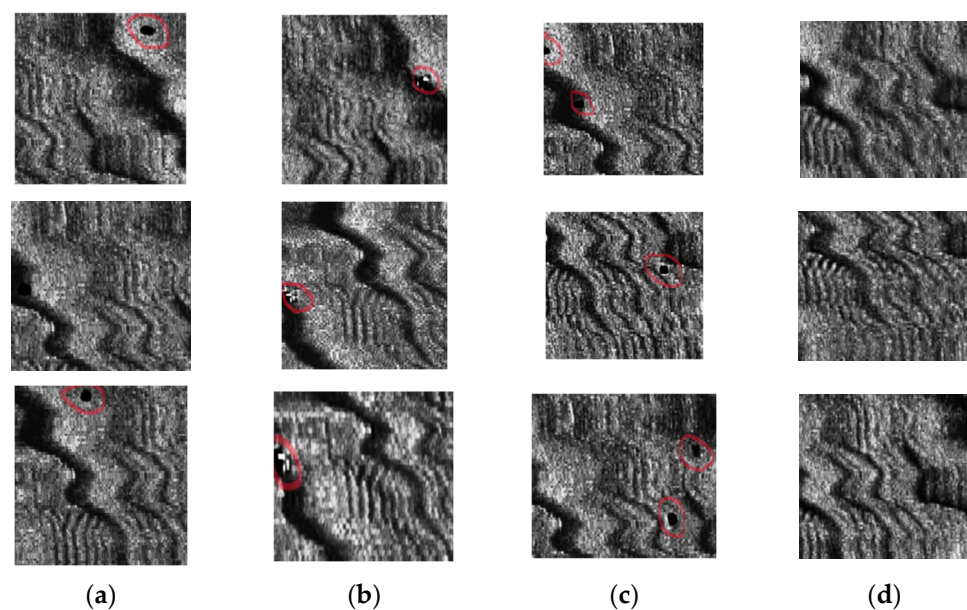


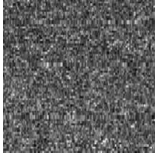
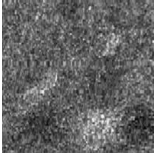




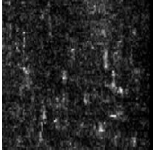

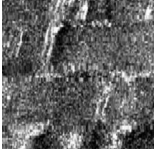







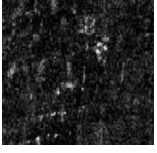



Figure 6. Amplification of 4 groups of models, Figures (a–d) are partial amplifications generated by models 1–4 (The red circles are features of unnecessary detail).

3.2.3. Classification Model Verification

Considering that the purposes of this study are to expand side-scan sonar substrate images so as to improve the performance of deep learning-based substrate classification models and to expand the training set because it contains a small number of samples, a deep learning-based object detection model is selected for comparative experiments. At present, there are many detection models, and the AlexNet network, as a lightweight, fast, and mature detection model, is suitable for this experiment. Therefore, the AlexNet network is selected to train real images, obtain a classification model, and classify the generated images.

The training and testing datasets consist of side-scan sonar data collected in 2019 from the Jiaozhou Bay area in Qingdao, utilizing a Klein4000 side-scan sonar system. There are a total of 27,202 sonar images across five different sediment types. These include 5123 images of sandy mud sediment, 5851 images of muddy sand sediment, 7126 images of fine sand sediment, 4406 images of gravel sediment, and 4696 images of bedrock sediment. For detailed information about the image library, please refer to Table 4.

Table 4. A side-scan sonar image library was used in the experiment.

	Sandy Mud	Muddy Sand	Sand	Gravel	Bedrock
Example diagram					
					
					
					
Quantity	5123	5851	7126	4406	4696
Total	27,202				
Size	256 × 256				

Currently, there are many detection models. This article uses five common models—AlexNet, GoogleNet, VggNet, ResNet, and DenseNet—to train the dataset, and the hyperparameter selection is shown in Table 5.

Table 5. Basic parameter settings of the network.

Training Parameters	Parameter Settings
Training Epochs	100
batch_size	32
Learning Rate	0.0001

The training accuracy curve and training time are shown in Table 6.

Table 6. Training results of five models.

Model	Training Duration (min)	Validation Accuracy Curve
AlexNet	121.8	
GoogleNet	129.1	
VggNet	161.9	
ResNet	124.3	
DenseNet	135.6	

It is evident from the above training results that the AlexNet network has the shortest training time, and the convergence value of the validation accuracy curve is the highest, at approximately 92%, making it more suitable for this experiment. The AlexNet network consists of approximately 630 million connections, and it includes five convolutional layers and three pooling layers. It utilizes fully connected layers and a softmax layer for image

classification, as shown in Figure 7. Each convolutional layer consists of convolutional kernels, bias terms, rectified linear unit activation functions, and local response normalization modules. The first, second, and fifth convolutional layers are followed by a max pooling layer, and the last three layers are fully connected layers. The final output layer is a softmax layer, which converts the network output into the probability values used for predicting the image’s class [21].

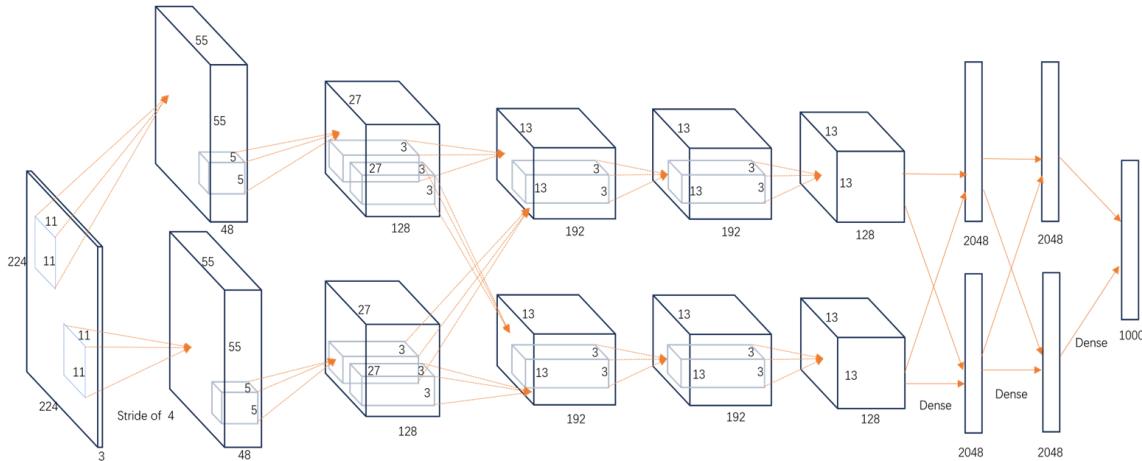


Figure 7. Structure of the AlexNet network model (The direction the arrow points in indicates the path of data from one layer to the next).

By using the AlexNet network model trained on real side-scan sonar substrate images, we randomly selected 100 images from the augmented images of bedrock image 1 and the real image test set in Section 3.2.2 for validation. The validation was conducted 10 times, and the average value was taken. The specific details are shown in Table 7.

Table 7. Detection results of the generated images using the classification model.

Group	CBAM Model	BCEL1 Loss	Generated Image Bedrock Recognition Rate
1	—	—	90.50%
2	—	✓	93.60%
3	✓	—	92.20%
4	✓	✓	97.30%
5	Original dataset		92.60%

An analysis of Table 7 shows that both the model with only the CBAM (Group 2) and the model using only the BCEL1 loss function (Group 3) improve the accuracy of the generated images compared with Group 1. However, the model that integrates the CBAM and the BCEL1 loss function achieves the greatest accuracy improvement, with an increase of 6.8% compared with Group 1. Additionally, it still maintains a high recognition rate when tested against the real image dataset. The above experiments demonstrate that the images generated using the method proposed in this paper are closer in realism to the actual side-scan sonar substrate images.

Due to the limited sampling points of coarse sand in the training set, six types of substrate sonar images were selected from the original dataset, and several were chosen from the generated images. Three groups of datasets were designed to train the AlexNet network: one dataset containing only the original images, one dataset containing both the original images and the images generated using the proposed augmentation method, and one dataset containing both the original images and images augmented using simple methods (such as flipping and rotating). The specific details can be found in Table 8.

Table 8. Composition of the three groups of datasets.

Group	Original Image	Images Augmented Using the Proposed Method	Images Augmented Using the Simple Method
1	260	—	—
2	200	60	—
3	200	—	60

The dataset is divided into training and validation sets in a ratio of 2:1, and 100 real images of bedrock substrate are selected from the original images to test the model's performance. The results are shown in Table 9.

Table 9. Detection results of real bedrock substrate images with different training sets.

	Accuracy
AlexNet-1	82.50%
AlexNet-2	87.70%
AlexNet-3	85.00%

Table 9, comparing the detection results of groups 1 and 2, shows that the classification accuracy of bedrock-type matrices is improved by 5.2% when images enhanced using the method presented in this paper are added to the training set. In contrast, a comparison of groups 1 and 3 showed that the accuracy was only 2.5% better when images enhanced with a simple method were added to the training set. The reason for this is that images enhanced with a simple method do not increase the diversity of the base image simply because the increase in number improves the accuracy. This indicates that the improvement of the model performance is mainly due to the use of the enhanced data generated using the method proposed in this paper, which further indicates that the enhanced image meets the requirements of the realism and diversity of the side-scan sonar image.

4. Discussion

In this paper, the CBAM-BCEL1-INGAN method based on side-scan sonar shows good results in image generation tasks of different bottom types. Through the experimental analysis and result evaluation in this paper, the following conclusions are drawn:

(1) The images generated using the method proposed in this paper perform well in terms of quality and diversity. In particular, for sediment types such as coarse sand, gravel, and bedrock, the images generated show high quality and diversity in evaluation indicators such as the FID value and IS index. This shows that the proposed method can effectively enlarge the images of these complex substrates, enrich the dataset, and improve the generalization ability of the model. (2) Through an entropy analysis of the gray co-occurrence matrix, the complexity and authenticity of the generated image are further verified. The larger the entropy, the more complex the gray distribution of the image. The entropy of the coarse sand, gravel, and bedrock images produced in this study is high, which proves that these images are close to the real images in terms of detail and texture and have high complexity and naturalness. (3) Ablation experiment results show that both the CBAM and BCEL1 loss functions contribute significantly to the quality improvement of the generated images. Specifically, the introduction of the CBAM improves the attention mechanism of the model and enriches the details of the generated images. Although the BCEL1 loss function restricts image generation, it improves the overall quality and consistency of the generated image. The model with the fusion of these two modules shows the best performance on all evaluation indices, which verifies the effectiveness and superiority of the proposed method in the image generation task. (4) In the verification experiment of the classification model, the AlexNet classification model is used to test the accuracy of the images generated using the proposed method, and it is concluded that the image generated using the model combining the CBAM and BCEL1 loss function has

the highest recognition rate. This further proves that the proposed method can generate high-quality amplified images.

In addition to the INGAN network used in this paper, the sinGAN [22] network, has achieved good results in the field of image augmentation. However, its performance in augmenting side-scan sonar substrate images is not ideal. As shown in Figure 8, some images generated using the sinGAN network during training have lost the characteristics of the original images, and some of them are no longer similar to the side-scan sonar images.

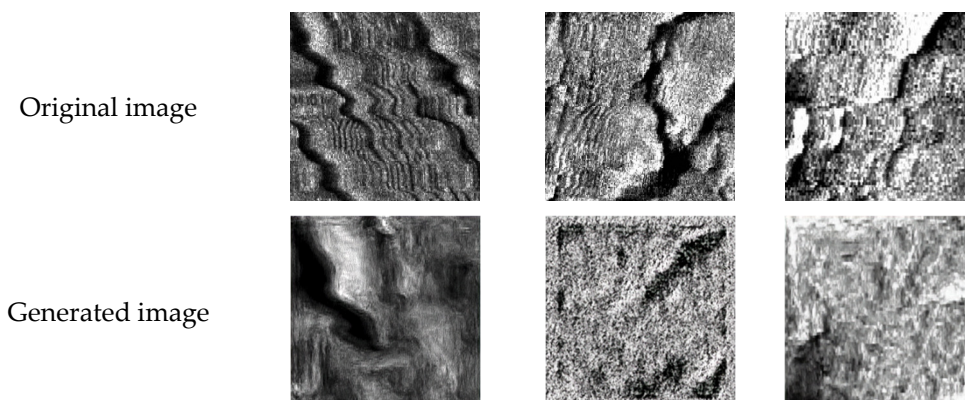


Figure 8. Some images were generated using the sinGAN network during training.

The method proposed in this paper has significant value in practical applications. The acquisition of side-scan sonar images is costly and challenging, while the method presented here can effectively augment the dataset, reducing the reliance on a large number of real images and lowering costs. Specifically, the generated high-quality images can be used for the following practical seabed exploration tasks: (1) Training more robust models: The generated high-quality images can be used to train more robust seabed substrate classification models, thereby improving the accuracy and reliability of the models in practical applications. This is of great significance for automated seabed exploration and resource assessment. (2) Applications in data-scarce environments: In situations where it is difficult to obtain large-scale real data, the generated images can serve as supplementary data, effectively addressing the issue of data scarcity. For example, in deep-sea exploration, obtaining a large number of high-quality sonar images is extremely difficult due to harsh environments and high costs. The method proposed in this paper can significantly reduce the reliance on actual collected data. (3) Reducing data annotation costs: By using the generated images, the demand for manual annotation can be reduced, thereby lowering data annotation costs. This provides a cost-effective solution, especially in machine learning tasks that require a large amount of annotated data.

In summary, the CBAM-BCEL1-INGAN image amplification method proposed in this paper has excellent performance in the amplification task of side-scan sonar submarine bottom images, and its effectiveness and superiority are verified by several experiments. Future studies can further optimize the model structure and training strategies, explore more amplification methods suitable for different application scenarios, and further improve the quality and diversity of the generated images.

5. Conclusions

Aiming to address the problems of difficulty in acquiring seafloor sediment image datasets, high labeling costs, and the insufficient diversity and quantity of data samples, this paper proposes a method for the sample amplification of seafloor sediment images obtained using side-scan sonar based on CBAM-BCEL1-INGAN. A residual block based on the CBAM is designed to retain information about the original features and enhance the representation of important features. The BCEL1 loss function is designed based on the original L1 loss function so that the discriminator can pay attention to both the

global image consistency and more subtle differences at the same time in order to obtain a better generation effect. Through experiments on existing sediment image datasets, it is confirmed that the proposed method performs well in the task of sediment image generation, and it solves the problem of the lack of side-scan sonar sediment image data to a certain extent. The images generated using the method proposed in this paper can be effectively used for practical underwater exploration tasks, reducing the reliance on a large number of real images, lowering costs, and improving the accuracy and reliability of seabed substrate classification and other related tasks. This not only enriches the dataset but also significantly reduces the costs of data acquisition and labeling. Future research can further optimize the model structure and training strategies to explore more augmentation methods suitable for different application scenarios, further enhancing the quality and diversity of the generated images.

Author Contributions: Conceptualization, S.J.; methodology, H.X.; formal analysis, G.B.; resources, C.P. and W.Z.; writing—original draft preparation, H.X.; writing—review and editing, S.J. and Y.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Sample data are included in the article.

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

References

1. Ku, A.; Zhou, X.; Peng, C. The Research on Development of Side-scan Sonar Detection Technology. *Hydrogr. Surv. Charting* **2018**, *38*, 50–54.
2. Vehicle, U. Marine Science and Engineering. *J. Mar. Sci. Eng.* **2020**, *8*, 557.
3. Chen, E.; Guo, J. Real Time Map Generation Using Sidescan Sonar Scanlines for Unmanned Underwater Vehicles. *Ocean Eng.* **2014**, *91*, 252–262. [[CrossRef](#)]
4. Buscombe, D. Shallow Water Benthic Imaging and Substrate Characterization Using Recreational-grade Sidescan-sonar. *Environ. Model. Softw.* **2017**, *89*, 1–18. [[CrossRef](#)]
5. Yang, F. Fusing and Classifying Multi-Beam Sonar and Side-Scan Sonar Data. Ph.D. Dissertation, Wuhan University, Wuhan, China, 2003.
6. Zhang, W.; Liang, S.; Yang, R. Application of Side-scanning Sonar and Shallow Stratum Profiler in Surface Silt Detection of Colombo Offshore Seabed. *Port Eng. Technol.* **2014**, *51*, 86–91.
7. Xu, Y.; Wang, Y.; Huang, J.; Fang, P. Influence of Seabed Sediment on Sonar Detection Performance. *Ship Electron. Eng.* **2017**, *37*, 123–125.
8. Lou, Z.; Zou, C.; Wu, C. Advances and challenges in seabed substrate classification. In Proceedings of the 2020 China Earth Science Joint Academic Conference (26th)—Special Session 76: Deep Mineral Resource Exploration Technology and Application, Special Session 77: Rock Physics and In-Well Detection Frontiers, Special Session 78: In-Well Geophysics and Deep Drilling, Wuhan, China, 18–21 October 2020; Beijing Botong Electronic Publishing House: Beijing, China, 2020; pp. 158–159.
9. Zhang, K.; Yuan, F.; Cheng, E. Analysis on Influence of Bottom Sediment Types on Side-scan Sonar Imaging Properties. *Int. J. Math. Models Methods Appl. Sci.* **2018**, *12*, 60–66.
10. Zhu, P.; Isaacs, J.; Fu, B.; Ferrari, S. Deep Learning Feature Extraction for Target Recognition and Classification in Underwater Sonar Images. In Proceedings of the IEEE 56th Annual Conference on Decision and Control, Melbourne, Australia, 12–15 December 2017; Volume 12, pp. 12–15.
11. Feldens, P.; Darr, A.; Feldens, A.; Tauber, F. Detection of Boulders in Side Scan Sonar Mosaics by a Neural Network. *Geosciences* **2019**, *9*, 159. [[CrossRef](#)]
12. Huo, G.; Yang, S.X.; Li, Q.; Zhou, Y. A Robust and Fast Method for Sidescan Sonar Image Segmentation Using Nonlocal Despeckling and Active Contour Model. *IEEE Trans. Cybern.* **2016**, *47*, 855–872. [[CrossRef](#)] [[PubMed](#)]
13. Coiras, E.; Mignotte, P.Y.; Petillot, Y.; Bell, J.; Lebart, K. Supervised target detection and classification by training on augmented reality data. *IET Radar Sonar Navig.* **2007**, *1*, 83–90. [[CrossRef](#)]
14. Tang, Y.; Wang, L.; Yu, D.; Li, H.; Liu, M.; Zhang, W. CSLS-CycleGAN based side-scan sonar sample augmentation method for underwater target image. *Syst. Eng. Electron.* **2024**, *46*, 1514–1524.
15. Zhang, Q.; Zhao, J.; Li, S.; Zhang, H. Seabed Sediment Classification Using Spatial Statistical Characteristics. *J. Mar. Sci. Eng.* **2022**, *10*, 691. [[CrossRef](#)]

16. Shocher, A.; Bagon, S.; Isola, P.; Irani, M. InGAN: Capturing and Remapping the “DNA” of a Natural Image. *Computer Vision and Pattern Recognition. arXiv* **2019**, arXiv:1812.00231.
17. Li, B.Q.; Huang, H.N.; Liu, J.Y.; Li, Y. Optical Image-to-Underwater Small Target Synthetic Aperture Sonar Image Translation Algorithm Based on Improved CycleGAN. *Acta Electron. Sin.* **2021**, *49*, 1746–1753.
18. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In Proceedings of the IEEE International Conference on Conference and Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2242–2251.
19. Mao, X.; Li, Q.; Xie, H.; Lau, R.Y.; Wang, Z.; Paul Smolley, S. Least Squares Generative Adversarial Networks. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.
20. Ferrari, V.; Hebert, M.; Sminchisescu, C.; Weiss, Y. *Computer Vision—ECCV 2018*; Springer International Publishing: Berlin/Heidelberg, Germany, 2018.
21. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet Classification with Deep Convolutional Neural Networks. *Commun. Acm* **2017**, *60*, 84–90. [[CrossRef](#)]
22. Shaham, T.R.; Dekel, T.; Michaeli, T. Singan: Learning a Generative Model from a Single Natural Image. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27–28 October 2019.

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.