Journal of
Marine Science and Engineering

MDPI

Article

Sample Augmentation Method for Side-Scan Sonar Underwater Target Images Based on CBL-sinGAN

Chengyang Peng, Shaohua Jin *, Gang Bian, Yang Cui and Meina Wang

Department of Oceanography and Hydrography, Dalian Naval Academy, Dalian 116018, China; 18908414801@163.com (C.P.); 13998435151@163.com (Y.C.)
* Correspondence: jsh_1978@163.com

Abstract: The scarcity and difficulty in acquiring Side-scan sonar target images limit the application of deep learning algorithms in Side-scan sonar target detection. At present, there are few amplification methods for Side-scan sonar images, and the amplification image quality is not ideal, which is not suitable for the characteristics of Side-scan sonar images. Addressing the current shortage of sample augmentation methods for Side-scan sonar, this paper proposes a method for augmenting single underwater target images using the CBL-sinGAN network. Firstly, considering the low resolution and monochromatic nature of Side-scan sonar images while balancing training efficiency and image diversity, a sinGAN network is introduced and designed as an eight-layer pyramid structure. Secondly, the Convolutional Block Attention Module (CBAM) is integrated into the network generator to enhance target learning in images while reducing information diffusion. Finally, an L1 loss function is introduced in the network discriminator to ensure training stability and improve the realism of generated images. Experimental results show that the accuracy of shipwreck target detection increased by 4.9% after training with the Side-scan sonar sample dataset augmented by the proposed network. This method effectively retains the style of the images while achieving diversity augmentation of small-sample underwater target images, providing a new approach to improving the construction of underwater target detection models.

Keywords: sample amplification; side-scan sonar; imaging mechanism; style transfer; sinGAN

Citation: Peng, C.; Jin, S.; Bian, G.; Cui, Y.; Wang, M. Sample Augmentation Method for Side-Scan Sonar Underwater Target Images Based on CBL-sinGAN. *J. Mar. Sci. Eng.* **2024**, *12*, 467. https://doi.org/10.3390/jmse12030467

Academic Editor: Dmitry A. Ruban

Received: 14 January 2024
Revised: 6 March 2024
Accepted: 6 March 2024
Published: 8 March 2024

1. Introduction

Deeper into the ocean, the demand for seabed topography exploration has been increasing, especially in the areas of seabed target identification and detection. This plays a vital role in fields such as navigational safety, marine surveying, maritime search and rescue, and military missions. Currently, marine mapping primarily utilizes single-beam multi-beam echo sounding systems and Side-scan sonar systems. Among these, Side-scan sonar, with its high-resolution acoustic imaging capability of the seabed, has a distinct advantage in seabed target identification [1–5]. Seabed target identification largely relies on manual detection and recognition, a method fraught with issues such as low efficiency, time consumption, and strong subjectivity. Therefore, research into automatic detection methods for underwater targets is of great significance.

Some scholars have adopted machine learning techniques, combined with manual features and classification technologies, to achieve automated underwater target detection [6,7]. However, these methods are limited when dealing with complex seabed environments, as Side-scan sonar images often suffer from low resolution, insufficient features, high noise, and deformation. The advancements in deep learning technologies in the field of computer vision have significantly improved the performance of target detection and are thus widely used in the field of underwater intelligent detection [8–10]. Models based on Deep Convolutional Neural Networks (DCNN) are effective but require high-quality training data, which are often scarce and limited in representativeness for Side-scan sonar

images [11–14]. Therefore, there is an urgent need for sample augmentation research for small-sample underwater targets in Side-scan sonar images.

With the development of sample augmentation techniques in the optical imaging field, data augmentation techniques in the underwater acoustics field have also emerged [15–20]. Currently, the main methods for Side-scan sonar image augmentation are of two types: one is the image style transfer method represented by GAN (Generative Adversarial Networks) [21–33], and the other is based on the diffusion model for image generation [34]. For instance, Ye Xiufen [23] used the AdaIN network for style transfer and achieved good results in target detection; Yang Zhiwei [24] adopted an improved DDIM model for data augmentation, successfully enhancing the model's accuracy; Huang Chao [21] utilized the metal style network for data augmentation from geometric and physical perspectives, obtaining a rich set of Side-scan sonar images. However, both of these models require a large dataset of Side-scan sonar images and preprocessing, which increases the workload and limits the threshold for generation.

Therefore, there are few GAN networks that satisfy the need for data enhancement with a small number of samples [35,36]. In 2019, a network named SinGAN was proposed in a collaborative research project between the Technion—Israel Institute of Technology and Google [37]. This model, through in-depth learning of a single natural image, can grasp the distribution characteristics of internal patches in the image. Through such learning, SinGAN is capable of producing a series of both high-quality and diverse image samples. Therefore, this paper selects the SinGAN network as the main method for research on Side-scan sonar data augmentation. However, since the original network's dataset mostly consists of colorful artistic style images and natural landscape images like lakes and birds, when using the SinGAN network to augment black and white Side-scan sonar underwater target images, it was found that the targets appeared unrealistic and illogical.

Therefore, to enhance the network's learning of targets and consider the characteristics of black and white waterfall images in Side-scan sonar, this paper proposes a single-image sample augmentation method for Side-scan sonar underwater targets based on CBL-sinGAN. Firstly, an eight-layer pyramid network structure is designed according to the characteristics of Side-scan sonar images, which improves the diversity of generated images while fully learning the image textures. Secondly, the Convolutional Block Attention Module (CBAM Module) is integrated into the generator to enhance target learning while reducing information diffusion [38–41]. Then, an L1 loss-based loss function is introduced in the discriminator to strengthen its ability to discern the authenticity of targets, eliminate unrealistic, fake images, and improve the quality of generated images while enhancing training stability and avoiding training mode collapse. Finally, based on this transformation model, high-quality augmentation of existing small-sample Side-scan sonar images is performed, and the augmented images are used for recognition and detection with YOLOv5. Experiments in this paper prove that the single-image sample augmentation method for Side-scan sonar underwater targets based on CBL-sinGAN (CBL refers to the integration of two modules into the original SinGAN network: the CBAM attention mechanism and the L1 loss function) proposed in this paper can generate a large number of high-quality Side-scan sonar augmented samples according to the characteristics of Side-scan sonar images, such as limited samples, monochromatic color, and diverse target shapes, Introducing a novel approach to address the issue of small target sample augmentation in Side-scan sonar and high-performance underwater target detection models. The structure of this paper is as follows: Section 1 is the introduction, the network structure is elaborated in Section 2, the experimental design and corresponding parameter calculations are described in Section 3, the experimental conclusions are discussed in Section 4, and the conclusion and prospect are summarized in Section 5.

2. Methods

2.1. The Basic Structure of CBL-sinGAN

SinGAN was proposed in 2019 as part of a collaborative research project between the Technion—Israel Institute of Technology and Google. It is an unconditional generative model capable of learning from a single natural image. After training, it can capture the internal distribution of patches within the image, thereby generating high-quality, diverse samples with the same visual content as the image. SinGAN consists of a fully convolutional GAN pyramid, where each pyramid level is responsible for learning the distribution characteristics of the image at different scales. This allows for the generation of new samples of arbitrary size and deformation, maintaining significant variability while preserving the global structure and fine texture of the image. However, since the original network's dataset mostly consists of colorful artistic-style images and natural landscapes like lakes and birds, it was found that when using the SinGAN network to augment black and white Side-scan sonar underwater target images, the targets appeared unrealistic and illogical. Therefore, to enhance the network's learning of targets and consider the characteristics of black and white waterfall images in Side-scan sonar, this paper proposes a single-image sample augmentation method for Side-scan sonar underwater targets based on CBL-sinGAN. Firstly, we design an eight-layer pyramid network structure that is specifically tailored to the characteristics of Side-scan sonar images. This structure enhances the diversity of generated images while fully capturing the intricate textures present in the images. Secondly, we integrate the Convolutional Block Attention Module (CBAM Module) into the generator. This module enhances the learning of target features while reducing information diffusion, resulting in improved target detection performance. Furthermore, we introduce an L1 loss-based loss function in the discriminator. This loss function strengthens the discriminator's ability to distinguish authentic targets from fake ones. It effectively eliminates unrealistic, fake images, thereby improving the overall quality of the generated images. Moreover, the introduction of this loss function enhances training stability and prevents the occurrence of training mode collapse. The proposed network structure is illustrated in Figure 1.

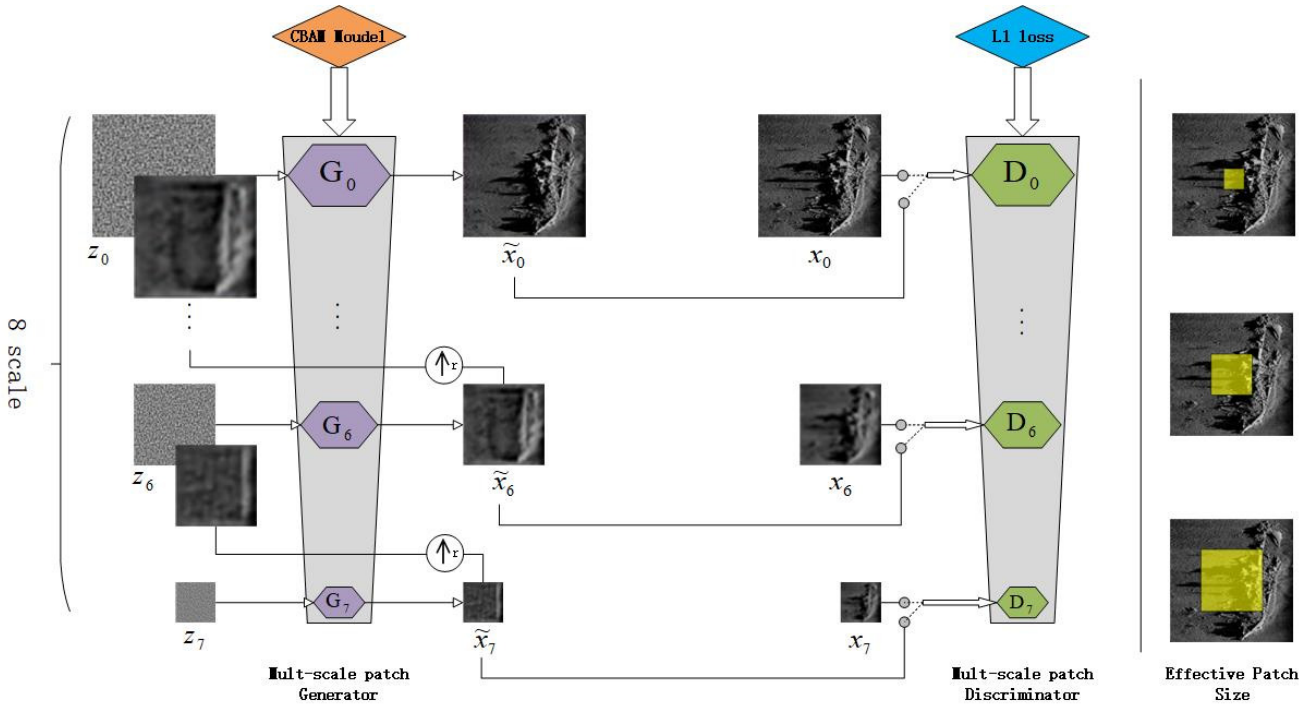

Figure 1. CBL-sinGAN's multi-scale structure. On the left is the training process of the generator, which is divided into eight scales, and on the right is the training process of the discriminator. After each scale, move on to the next scale. See Section 2.1 for a detailed description of the process.

2.2. The Multi-Scale Architecture of the SinGAN Module

As shown in Figure 1, our model consists of a GAN pyramid, where both the training and generation processes are completed in a coarse-to-fine manner. At each scale, the generator G_N learns and is responsible for generating images of different scales, adopting a coarse-to-fine approach, with its input being noise Z and the upsampled output from the previous level. The discriminator D_N's function is to discern whether its input samples are real or forged.

It is important to note that the training target of the generator G_N is patches of a single image, not the entire image sample. To handle targets in Side-scan sonar, such as shipwrecks, airplanes, etc., the generator needs to capture the layout and shape of large objects in the image, such as the hull structure of shipwrecks, the wing features of airplanes and other global attributes. To achieve this, the generator framework consists of a series of hierarchical patch-GANs (Markovian discriminators) [41,42], each responsible for capturing the patch distribution at different scales of the training image. These patch-GANs have small receptive fields and limited capacity to prevent memorization of a single image. Although traditional GANs have also explored similar architectures [43–47], and SinGAN has been applied in multiple fields for image augmentation generation, we are the first to apply it for single-image Side-scan sonar image augmentation.

Our model consists of a generator and a discriminator pyramid. The generator $\{G_1, G_2, G_3, G_4, G_5, G_6, G_7\}$ targets images at eight different scales for pyramid training: $\left\{\widetilde{x}_0, \widetilde{x}_1, \widetilde{x}_2, \widetilde{x}_3, \widetilde{x}_4, \widetilde{x}_5, \widetilde{x}_6, \widetilde{x}_7\right\}$, where $\widetilde{x}_N$ is the downsampled sample of X, with a downsampling factor of r_n, where is greater than 1. Each generator, through adversarial training with the discriminator, generates real image samples X_N that best match the patch distribution of the corresponding images. The goal of the generator is to deceive the discriminator, whose objective is to distinguish between the patches of generated samples ($\widetilde{x}_N$) and those of real samples (x_N).

Image generation starts from the coarsest scale and progresses upward through each level of training up to the finest scale, with noise added at each scale. The generators and discriminators at the same level have the same size of receptive field, which decreases progressively with each level of generation. At the coarsest scale, the image is purely generated, meaning G_7 maps the spatial Gaussian white noise z_7 to the image sample $\widetilde{x}_7$.

$$\widetilde{x}_7 = G_7(z_7) \tag{1}$$

The receptive field at this level is generally set to half the height of the image, so the generator focuses more on the overall global structure of the image. As the scale progresses, the receptive field is set smaller, and the generator pays more attention to the detailed texture of the image. In addition to spatial noise z_N, the input also includes the upsampled sample of the image $\widetilde{x}_{\mathrm{rec}}^{\mathrm{N+1}} \uparrow^r$ generated at the previous scale, that is:

$$\widetilde{x}_{\mathrm{N}} = G_N\left(z_N, \widetilde{x}_{\mathrm{rec}}^{\mathrm{N+1}} \uparrow^r\right), N < 7 \tag{2}$$

The training process of the generator progresses from the coarsest to the finest scale, training the multi-scale architecture layer by layer. Once the generator at each layer is trained, it remains fixed. The training loss for the generator at the N layer includes an adversarial term and a reconstruction term:

$$\min_{G_N} \max_{D_{\mathrm{N}}} L_{adv}(G_N, D_{\mathrm{N}}) + \alpha L_{rec}(G_N) \tag{3}$$

Adversarial Loss: See Section 2.4 for details

Each generator G_N is paired with a Markovian discriminator D_N, which discriminates D_N the overlapping patches of the input. Based on the original WGAN-GP loss function, we introduce an L1 loss for discrimination, which further increases the stability of the

training. The final discrimination score is averaged. The loss defined here is for the entire image, not just for individual patches, allowing the network to learn boundary conditions. The receptive field architecture of G_N and D_N is the same, both being 11×11 in size.

Reconstruction Loss: To ensure that the generator can reconstruct the original image from a specific set of input noise images. Let the image generated and reconstructed at the Nth layer be $\tilde{x}_{\mathrm{rec}}^{\mathrm{N+1}} \uparrow^r$, and the random Gaussian white noise generated at this layer be Z_N. Then, for $N < 7$, we have:

$$L_{rec} = \left\| G_N\left(z_N, \tilde{x}_{\mathrm{rec}}^{\mathrm{N+1}} \uparrow^r\right) - x_{\mathrm{N}} \right\|^2 \tag{4}$$

For $N = 7$, the calculation method for the reconstruction loss is as follows:

$$L_{rec} = \|G_7(z_7) - x_7\|^2 \tag{5}$$

wherein the standard deviation σ_N of the random noise Z_N at each layer is determined by the reconstructed image $\tilde{x}_{\mathrm{rec}}^{\mathrm{N+1}} \uparrow^r$ and the real image x_{N}, indicating the number of image details that need to be added under the training at that level. The specific calculation method is as follows:

$$\sigma_N = \sqrt{\left\| \tilde{x}_{\mathrm{rec}}^{\mathrm{N+1}} \uparrow^r - x_{\mathrm{N}} \right\|^2} \tag{6}$$

After the training at each level, the fake Side-scan sonar images $\tilde{x}_{\mathrm{N}}$ are generated at various scales, and the real Side-scan sonar images x_{N} are simultaneously sent to the discriminators D_N at each level for discrimination. After the discrimination, the results are fed back to the generator, and the generator is retrained after adjusting the loss.

2.3. Generator Based on the CBAM Model

The thorough learning of target detail features and background characteristics in Side-scan sonar images is key to the generator's ability to produce high-quality images. To enhance the learning of global information and local features in the input image and to strengthen the interaction between channel and spatial dimensions, this paper introduces the Convolutional Block Attention Module (CBAM Module) at the second layer of the generator. It is placed after the body layer of the generator, as shown in Figure 2:

Figure 2 above shows the overall architecture after adding the attention mechanism CBAM module. It can be seen that the CBAM module includes two independent sub-modules: the Channel Attention Module (CAM) and the Spatial Attention Module (SAM). Compared to attention mechanisms that only focus on spatial aspects, this approach achieves better results. It allows for easy insertion into various levels of the network while saving parameters and computational power.

This module aims to reduce information diffusion and amplify the cross-dimensional interaction of channels and space in images, thereby enhancing network performance. By focusing on relevant features and minimizing interference, the CBAM module makes the representation of Side-scan sonar images more delicate and detailed, which is crucial for producing realistic and high-quality output images.

2.4. Discriminator Based on L1 Loss Function

L1 loss is a mean squared error loss that focuses on targets, as these false targets contribute more after squared loss, thereby providing feedback to the generator, 'prompting' it to generate higher quality and more representative samples of Side-scan sonar target images. This paper incorporates L1 loss into the discriminator, calculating it alongside WGAN-GP, thereby strengthening the discriminator's ability to discern false targets. Adversarial Loss: Each generator G_N is paired with a Markovian discriminator D_N, which discriminates D_N the overlapping patches of the input. Based on the original WGAN-GP loss function, we introduce

an L1 loss for discrimination, which further increases the stability of the training. The final discrimination score is averaged. The loss defined here is for the entire image, not just for individual patches, allowing the network to learn boundary conditions. The receptive field architecture of G_N and D_N is the same, both being 11 × 11 in size.

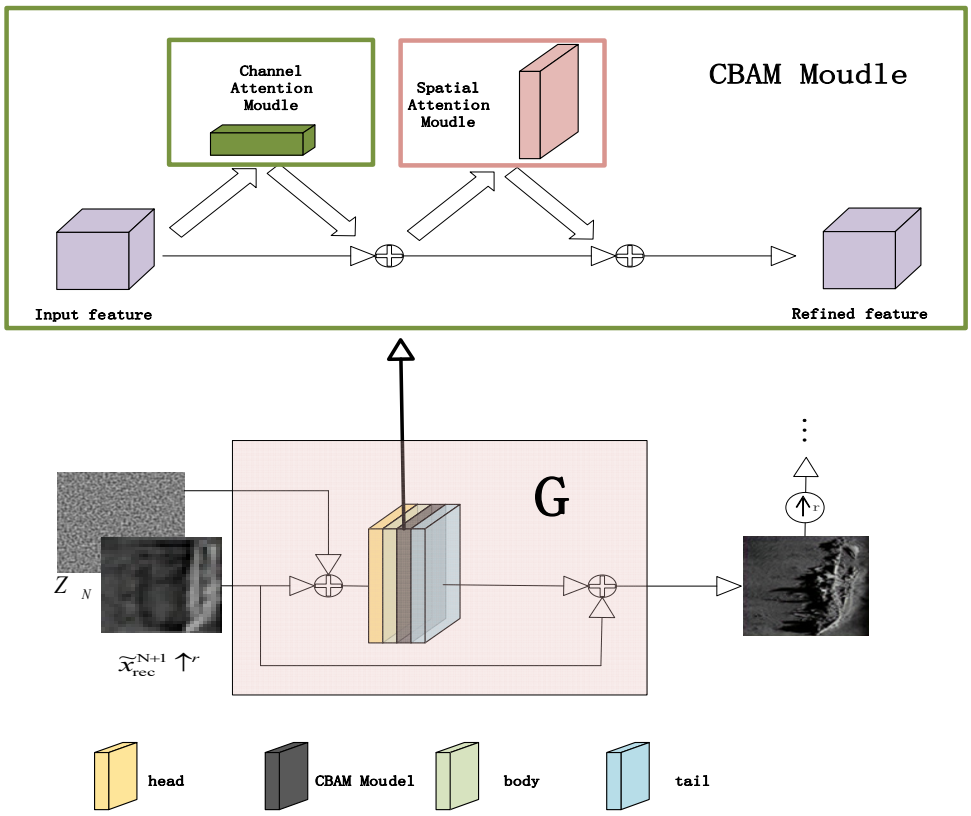

Figure 2. Overall single-scale generator architecture after adding the attention mechanism CBAM module.

The calculation principle of L1 loss is as follows:

$$loss(x, y) = \frac{1}{n}\sum_{i-1}^{n}|y_i - f(x_i)| \tag{7}$$

wherein $f(x_i)$ and y_i represent the predicted value of the ith fake Side-scan sonar image generated at the same scale and the corresponding real Side-scan sonar image, respectively, and n is the number of generated images.

3. Experimental Validation

3.1. Dataset Description and Experimental Equipment Parameters

To validate the performance of the single-image sample augmentation method for Side-scan sonar underwater targets based on CBL-sinGAN in target detection, this paper designed various comparative experiments. For the augmentation of shipwreck targets, 215 shipwreck images with different shapes and backgrounds, 62 airplane images, 115 fish swarm images, and 2 underwater diver images were selected. Using the network, 2650 augmented images were generated. These augmented images, along with other real Side-scan sonar shipwreck images, were used in several experiments. The hardware used for model training included an Intel® Core™ i7-13700KF CPU and an NVIDIA GeForce RTX 4070 GPU with 12 GB. The software compilation environment was PyTorch 1.6.0, CUDA 11.8, and Python 3.10 under Windows 10. Some of the datasets are shown as follows in Figure 3:

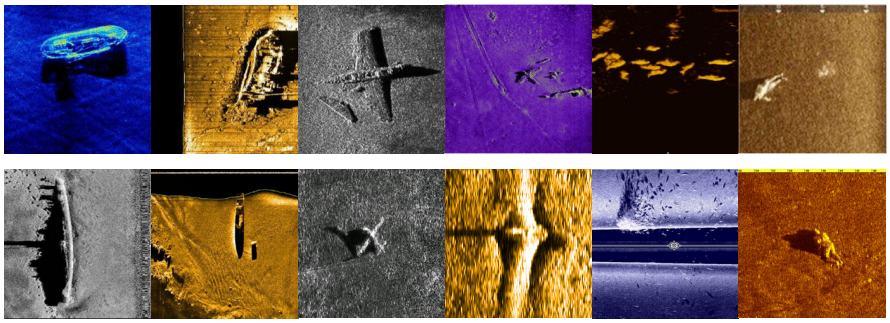

Figure 3. Partial dataset.

3.2. Evaluation Metrics

The evaluation of images is mainly based on feature diversity and structural similarity to assess the quality of style transfer images. According to the study in [37], in this paper, we select Fréchet Inception Distance (FID), Kernel Maximum Mean Discrepancy (MMD), and Inception Score (IS) as the metrics for image quality assessment. FID is used to measure the quality of images generated by Generative Adversarial Networks (GANs). It compares the similarity between the distribution of generated images and real images. A lower FID score indicates that the distribution of generated images is closer to that of real images, implying better quality. It is particularly adept at capturing the diversity and variation in generated images. MMD is a statistical test used to determine if two samples come from different distributions. A smaller MMD value indicates that the two distributions are more similar. In the context of generative models, a lower MMD means that the generated data are closer to the real data distribution. IS is another metric used to evaluate the quality of images generated by GANs. It measures the diversity of generated images and the clarity or distinctiveness of these images. A higher IS indicates that the model has generated a variety of unique images, each with clear and confident category predictions.

3.3. Analysis of Augmented Image Quality

In this paper, image augmentation was performed for shipwrecks, airplanes, fish swarms, and underwater divers. Representative images with different backgrounds were selected for training. Some examples of the augmented samples are shown in the following Figure 4:

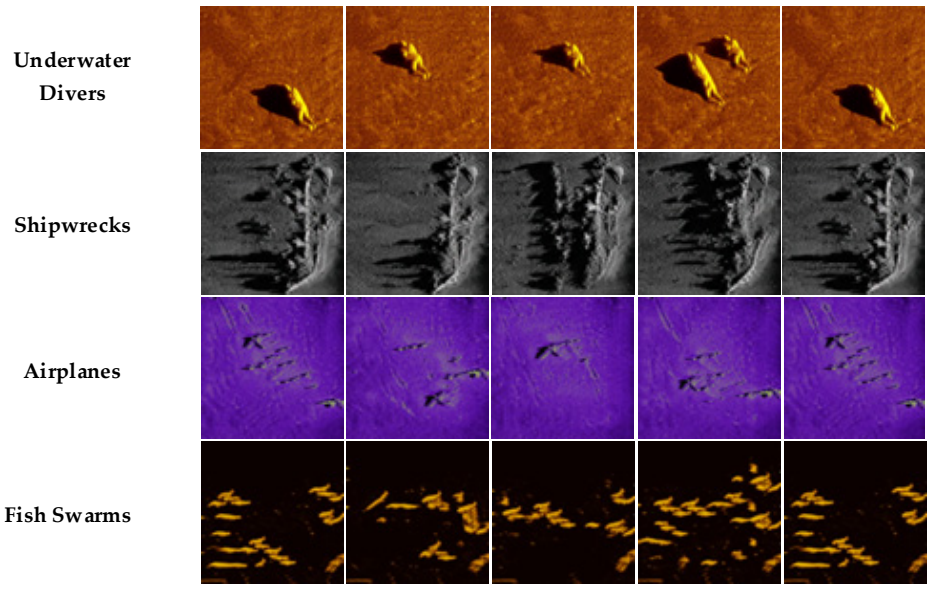

Figure 4. Augmented samples. Amplification tests were carried out on submarine frogmen, airplanes, shipwrecks, and fish.

For the four types of expansion targets above, calculations were performed for their respective categories' FID, MMD, and IS metrics to assess image quality. Among these metrics, a smaller FID is better, MMD closer to 0 is better, and a larger IS is better.

In an overarching analysis, it is evident that the FID values for all categories fall within a relatively high range, as illustrated in Table 1. It is worth considering that FID may have varying benchmark values for Side-scan sonar images. From the perspective gleaned from the work by Tang et al. [11], these numerical values still signify commendable generative quality in the context of side-scan sonar image augmentation. Turning our attention to the IS metric, it is noteworthy that all categories exhibit IS values surpassing the threshold of 3.9. This suggests that the generated images possess a discernible degree of clarity and diversity. Additionally, it is worth noting that the MMD values for all categories hover around 0.2, indicating a notable similarity in statistical characteristics between the generated images and their real counterparts.

Table 1. Test metrics. The corresponding FID, IS, and MMD indicators are calculated for the amplified images.

Group	FID	IS	MMD
plane	127.7	3.923 ± 0.226	0.221
People	123.0	4.346 ± 0.197	0.217
Fish	152.6	4.965 ± 0.367	0.237
Boat	118.9	4.456 ± 0.300	0.139

Engaging in a vertical comparative analysis of the metrics reveals the following trends:

FID: Among the categories, the "Boat" category displays the most favorable performance with the lowest FID score. An analysis of this phenomenon suggests that the abundance of shipwreck images may facilitate enhanced model learning and style replication, consequently yielding images that closely resemble real ones.

IS: The "Fish" category outperforms others in terms of the IS metric. This can potentially be attributed to the substantial presence of fish targets, often characterized by smaller dimensions. This observation implies that the Singan network excels when confronted with the challenge of generating diverse and complex images of small targets.

MMD: In the context of the MMD metric, once again, the "Boat" category exhibits superior performance with the lowest MMD score. This underscores the proximity of the generated shipwreck images to real images in terms of statistical characteristics.

Upon analysis of the evaluation metrics, it becomes apparent that the single-image sample augmentation method for underwater targets using CBL-sinGAN manifests varying effects across different Side-scan sonar targets (such as shipwrecks, airplanes, fish schools, and divers). Notably, in the domains of FID and MMD metrics, the "Boat" category distinguishes itself, possibly indicating the model's efficiency and precision when handling such images. Conversely, the elevated IS score observed in the "Fish" category implies the model's proficiency in generating highly diverse and intricate images of small targets. These findings underscore the notion that different image categories present divergent challenges to the generative model. Nonetheless, the model demonstrates distinct strengths in addressing these challenges, highlighting a degree of versatility in the augmentation model across various Side-scan sonar target categories.

3.4. Performance of the Model on Object Detection

Considering the aim of this paper is to augment underwater target sample images obtained from Side-scan sonar to enhance the performance of deep learning-based object detection models, the subsequent sections of this paper involve comparative experiments using deep learning-based object detection models. Currently, there is a plethora of object detection models available, and for this experiment, we have chosen the YOLOv5 detection model due to its lightweight nature, speed, and maturity, making it well-suited for this study.

The selected target images are shipwrecks, and three sets of datasets have been designed for training and deploying the YOLOv5 model. These sets consist of the following: datasets containing only real shipwreck images, datasets containing only augmented shipwreck images, and datasets containing a combination of both real and augmented shipwreck images. For the evaluation of the detection model's performance, 100 authentic Side-scan sonar shipwreck images were chosen. The specifics of this evaluation can be found in the table. It is important to note that the augmented shipwreck image data underwent a screening process to remove images of subpar generation quality. The specific grouping of the datasets is shown in Table 2.

Table 2. The composition of the YOLOv5 detection model's dataset and validation set.

Group	Real Shipwreck Images	Augmented Shipwreck Images
1	50	-
2	-	424
3	50	424
Detection Images	100	-

The model was tested using 100 real Side-scan sonar images of shipwrecks after training. The evaluation metrics adopted were precision, recall, and average precision (AP), which are widely used in the field of target detection. The detection results are as follows in Table 3:

Table 3. The effect of different training sets on the detection of real measured side-scan sonar shipwreck target images.

	Precision	Recall	AP0.5	AP0.5:0.95
YOLOv5-1	90.0%	91.2%	0.924	0.546
YOLOv5-2	94.8%	95.8%	0.958	0.593
YOLOv5-3	94.9%	96.0%	0.961	0.61

Observations from Table 3 show that, compared to YOLOv5-1 and YOLOv5-2, the model trained with the method described in this paper for augmenting Side-scan sonar images of shipwreck targets demonstrates higher precision, recall, and average precision. This proves the crucial role of augmented images in enhancing the model's performance. Comparing YOLOv5-2, which used only augmented images, with YOLOv5-3, which used both augmented and real Side-scan sonar images, shows little difference in evaluation metrics. This indicates that the improvement in model performance is mainly due to the Side-scan sonar augmented data generated by the method proposed in this paper, further illustrating that the augmented images meet the requirements of authenticity and diversity for Side-scan sonar images.

To prevent dataset bias in a single detection model, multiple detection models (YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5x) were employed for comparative experiments on precision. The analysis of the experimental results (Table 4) shows that, due to differences in the complexity and training time of the detection models, there are variations in precision among different models. However, a horizontal comparison across the three experimental groups indicates that precision is consistently higher in Group 3 than in Group 2 and Group 1, thereby further proving the validity of the experimental data.

Table 4. Comparison of precision rates between different models.

Detection Model/Group	1	2	3
YOLOv5n	82.7%	89.2%	91.5%
YOLOv5s	83.1%	92.3%	92.8%
YOLOv5m	86.1%	94.9%	95.3%
YOLOv5x	90.0%	94.8%	94.9%

4. Discussion

4.1. The Unique Advantages and Comparative Analysis of CBL-sinGAN

CBL-sinGAN, by retaining the style and texture of the original Side-scan sonar images, can generate more realistic Side-scan sonar images while simulating object distortions caused by underwater conditions, thereby improving the practicality of the images. This method is particularly effective for datasets with few samples, as it can generate images with a consistent style, addressing the limitations of traditional networks in this regard. Additionally, CBL-sinGAN considers the unique relationship between the target and its shadow during the image generation process, enhancing the realism of the images.

In the context of international research, style transfer networks such as cycleGAN and WGAN have achieved significant results in learning image textures and styles, especially with large datasets. However, these networks do not perform well under data-limited conditions, particularly in the field of Side-scan sonar image enhancement. The emergence of CBL-sinGAN provides a unique solution to this challenge, especially when the number of samples is limited, by generating high-quality images with a strong sense of realism, marking a significant advancement in existing technology. CBL-sinGAN fills a gap in existing technology, showcasing the new potential of deep learning in the field of Side-scan sonar image processing and offering a new direction for the development of future underwater target recognition and detection models. Figures 5 and 6 below display the style images generated from training with thirty Side-scan sonar images and optical domain images:

Figure 5. Style transfer results of the cycleGAN network trained with 30 Side-scan sonar images.

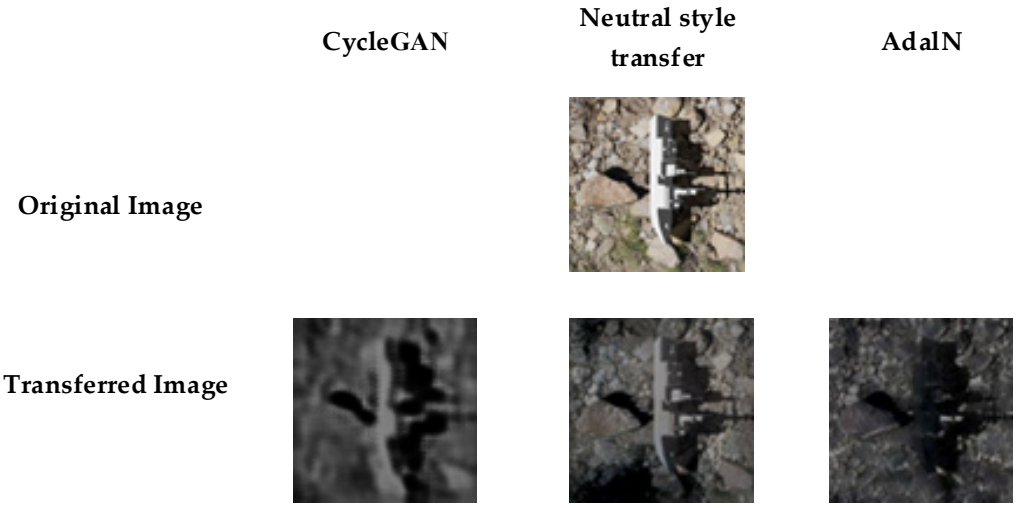

Figure 6. Single image style transfer effects of different networks.

The Side-scan sonar underwater target single image sample augmentation method based on CBL-sinGAN is essentially a process of rearranging and recombining real Side-scan sonar images. It augments images by reorganizing the image background, stretching

and scaling the target morphology, changing the number of targets, and altering target positions. As sinGAN is a new field for processing Side-scan sonar images, we encountered many issues during model training and made corresponding improvements.

4.2. Scale Selection in Image Generation

Unlike artistic style images, Side-scan sonar images need to maintain the stability and authenticity of the target during augmentation to prevent the generation of unrealistic targets. Experiments have found (Figure 7) that the choice of training scale greatly affects the quality of image generation. At lower scales, the network focuses on learning image details, leading to insufficient learning of the overall target and resulting in generated images with only basic textures and scattered targets. As the training scale increases, the learning of the target gradually takes shape. However, an increase in scale leads to longer training times, and setting too many scales can reduce the diversity of the generated images. The calculation of related metrics is as follows in Table 5:

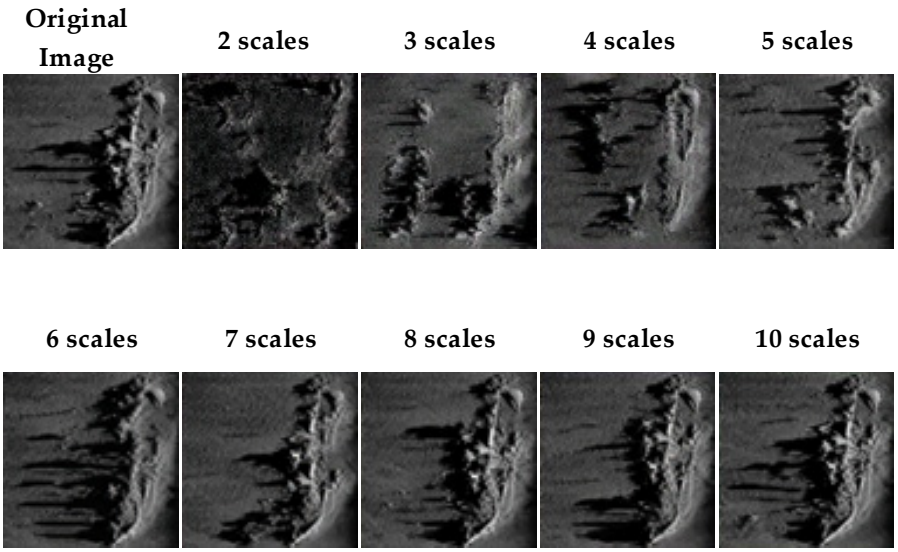

Figure 7. Image migration effects at different scales.

Table 5. Calculation of image metrics at different generative scales..

Scale	FID	IS	MMD
2	405.947	4.509	0.7412
3	368.951	4.378	0.4852
4	380.985	4.585	0.4600
5	274.182	4.563	0.3936
6	216.276	4.534	0.3349
7	204.740	4.522	0.3278
8	141.345	4.583	0.2496
9	183.675	4.512	0.2306
10	164.144	4.364	0.1901

Based on the analysis of the table, it can be observed that when the scale is set to 8, the generated images exhibit the best performance in terms of the FID index, indicating that the model has learned the style of Side-scan sonar images well. Simultaneously, the IS index is 5.583, which is also excellent, showing good diversity in the generated images. The higher diversity at scale 3 might be due to the model learning only the texture of the images, not considering the overall target, thus resulting in greater diversity. In terms of the MMD index, 0.24 is also within a good range, showing significant similarity between the generated and original images. Therefore, considering all factors, the structure of the

CBL-sinGAN pyramid network is set to 8 layers, where the model can improve the diversity of generated images while sufficiently learning image textures.

4.3. Ablation Experiment and Evaluation

To verify the role of each module in the performance of our model, ablation experiments were conducted on the CBAM module and the L1 loss function, using FID, MMD, and IS as evaluation metrics. Four groups were designed for comparative experiments, with the same experimental setup, training dataset, and evaluation data as mentioned in a previous Section 3.3. The results are shown in Table 6.

Table 6. Calculation of image metrics for ablation studies.

Group	CBAM Model	L1 Loss	FID	IS	MMD
1	-	-	131.61	4.299 ± 0.285	0.169
2	√	-	122.35	4.385 ± 0.210	0.154
3	-	√	157.92	4.306 ± 0.275	0.261
4	√	√	118.89	4.456 ± 0.300	0.139

From the table, it is evident that Groups 2 and 4, which incorporated the CBAM module, showed improved feature expression by focusing on important channels and spatial regions, as reflected in the improved FID and IS metrics compared to the control group (Group 1). Comparing Groups 1 and 3, the improvement in the IS index indicates that L1 Loss performs well in handling datasets with outliers or in feature selection. The combination of CBAM and L1 Loss in Group 4 not only improved image quality (lower FID and higher IS) but also increased the diversity of generated images (lower MMD). This is because CBAM enhanced feature expression, while L1 Loss increased the model's robustness to outliers, allowing for diverse augmentation of images while maintaining their authenticity.

The effects of different modules on the transformation of some Side-scan sonar images are shown in Figure 8a–e.

From the images, it can be seen that the images generated by Group 1 have the lowest realism, not conforming to the normal structural morphology of targets. For example, there are overlapping phenomena of two upper bodies in humans, overlapping ghost images in ships, blurred targets in airplanes, and excessive focus on background learning in fish swarms, neglecting target learning and augmentation, leading to large areas of black in the generated images. Additionally, observing shipwreck 'c', it can be noted that the generated image's shadow does not correspond with the target, resulting in situations where there is only a shadow without a target. After adding the CBAM in Group 2, compared to Group 1, there is a certain improvement in the model's ability to learn targets during generation, but the phenomenon of overlapping ghost images can still be observed in shipwreck targets. Comparing Groups 3 and 1, it is evident that the model's ability to maintain the morphology of targets during generation has improved, effectively preventing distortion in targets, but the diversity of the model has not increased, and the morphology of underwater divers remains unchanged. Comparing Group 4 with Group 1, it is observed that after integrating CBAM and L1 loss, the model performs very well in learning the texture details of targets, maintaining postures, and augmenting backgrounds. It generates samples with complete details fewer distortions and maintains a good one-to-one correspondence between targets and shadows.

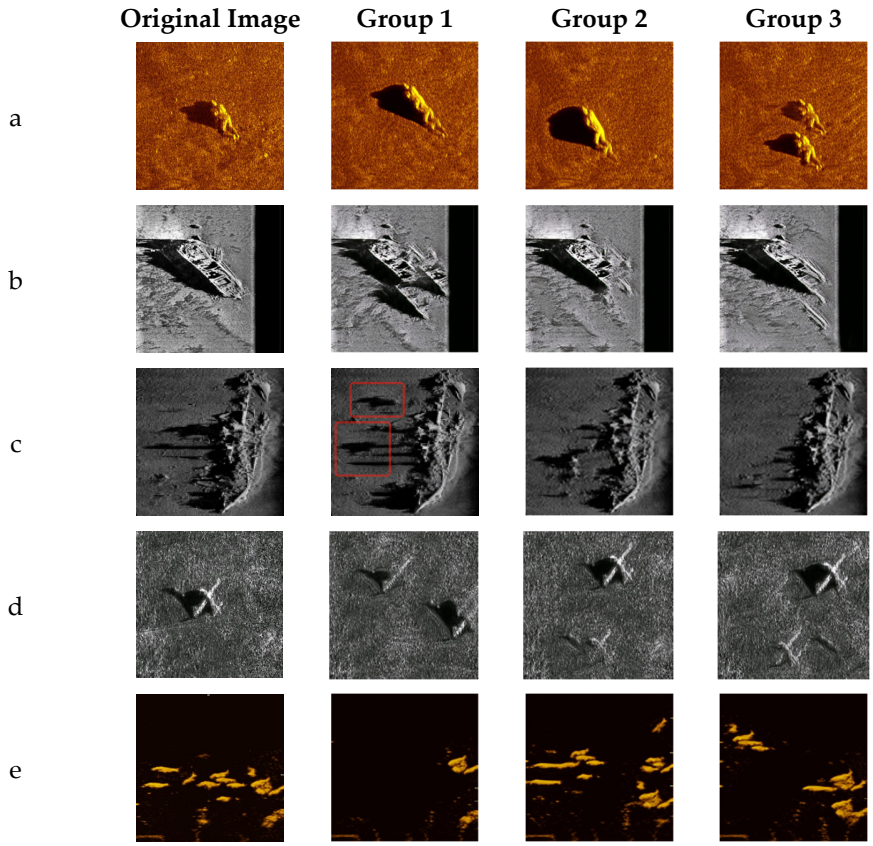

Figure 8. Different modules on the transformation of some Side-scan sonar images.

5. Conclusions

In response to the challenges of scarce Side-scan sonar underwater target images, the difficulty in forming a style with a limited number of samples, challenges in acquisition, and high costs, which all contribute to the poor performance of deep learning-based underwater obstacle detection models, we propose a single image sample augmentation method for Side-scan sonar underwater targets based on CBL-sinGAN. This method is tailored to the resolution and target size characteristics of Side-scan sonar images, utilizing an 8-layer scale GAN pyramid network structure. It enhances the diversity of generated images while thoroughly learning image textures. The CBAM module is integrated into the generator to enhance target learning while reducing information diffusion. Furthermore, a loss function based on L1 loss is introduced in the discriminator to strengthen its ability to discern the authenticity of targets, eliminating unrealistic, fake images. This improves the quality of generated images and training stability, preventing training mode collapse. The advantages of CBL-sinGAN include its ability to maximally retain the style, background texture, and integrity and authenticity of the original Side-scan sonar images, thereby generating more realistic Side-scan sonar images; it can simulate object distortions caused by underwater noise and water interference through deformations and blurring of objects, generating images that more closely match the real conditions of Side-scan sonar operations; for datasets with too few samples, which are insufficient to form a fixed style, the augmentation has significant advantages; when generating images, it takes into account the relationship between the target and its shadow, a unique feature of Side-scan sonar images. Finally, we conducted an analysis of evaluation metrics for the augmented 2650 images and comparative experiments for target recognition with real Side-scan sonar images. Ablation experiments were also used to assess the performance of the modules. Our method has been proven to generate a large number of high-quality augmented Side-scan sonar images from a small set of original images. It also improves the accuracy of automatic underwater obstacle detection and recognition models using these augmented images, achieving the goal of high-quality augmentation with few

samples. This method addresses, to a certain extent, the accuracy issues in deep learning-based target recognition and detection models caused by the scarcity of underwater samples.

Author Contributions: Conceptualization, C.P. and S.J.; methodology, C.P.; software, G.B.; validation, C.P., S.J. and Y.C.; formal analysis, G.B.; investigation, Y.C.; resources, M.W.; data curation, M.W.; writing—original draft preparation, C.P.; writing—review and editing, S.J.; visualization, Y.C.; supervision, S.J. 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.

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

References

1. Buscombe, D. Shallow water benthic imaging and substrate characterization using recreational-grade side scan-sonar. *Environ. Model. Softw.* **2017**, *89*, 1–18. [CrossRef]
2. Flowers, H.J.; Hightower, J.E. A novel approach to surveying sturgeon using side-scan sonar and occupancy modeling. *Mar. Coast. Fish.* **2013**, *5*, 211–223. [CrossRef]
3. Johnson, S.G.; Deaett, M.A. The application of automated recognition techniques to side-scan sonar imagery. *IEEE J. Ocean. Eng. J. Devoted Appl. Electr. Electron. Eng. Ocean. Environ.* **1994**, *19*, 138–144. [CrossRef]
4. Burguera, A.; Bonin-Font, F. On-line multi-class segmentation of side-scan sonar imagery using an autonomous underwater vehicle. *J. Mar. Sci. Eng.* **2020**, *8*, 557. [CrossRef]
5. Chen, E.; Guo, J. Real time map generation using Side-scan sonar scanlines for unmanned underwater vehicles. *Ocean Eng.* **2014**, *91*, 252–262. [CrossRef]
6. Shin, J.; Chang, S.; Bays, M.J.; Weaver, J.; Wettergren, T.A.; Ferrari, S. Synthetic Sonar Image Simulation with Various Seabed Conditions for Automatic Target Recognition. In Proceedings of the OCEANS 2022, Hampton Roads, VA, USA, 17–20 October 2022; pp. 1–8.
7. Neupane, D.; Seok, J. A review on deep learning-based approaches for automatic sonar target recognition. *Electronics* **2020**, *9*, 1972. [CrossRef]
8. Topple, J.M.; Fawcett, J.A. MiNet: Efficient deep learning automatic target recognition for small autonomous vehicles. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 1014–1018. [CrossRef]
9. Zhou, X.; Yu, C.; Yuan, X.; Luo, C. Deep Denoising Method for Side Scan Sonar Images without High-quality Reference Data. In Proceedings of the 2022 2nd International Conference on Computer, Control and Robotics (ICCCR), Shanghai, China, 18–20 March 2022; pp. 241–245.
10. 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]
11. Tang, Y.L.; Jin, S.H.; Bian, G.; Zhang, Y. Shipwreck target recognition in side-scan sonar images by improved YOLOv3 model based on transfer learning. *IEEE Access* **2020**, *8*, 173450–173460.
12. Tang, Y.; Li, H.; Zhang, W.; Bian, S.; Zhai, G.; Liu, M.; Zhang, X. Lightweight DETR-YOLO method for detecting shipwreck target in side-scan sonar. *Syst. Eng. Electron.* **2022**, *44*, 2427.
13. Nguyen, H.-T.; Lee, E.-H.; Lee, S. Study on the classification performance of underwater sonar image classification based on convolutional neural networks for detecting a submerged human body. *Sensors* **2019**, *20*, 94. [CrossRef]
14. Tang, Y.; Wang, L.; Yu, D.; Li, H.; Liu, M.; Zhang, W. Sample Augmentation Method for Side-scan sonar Underwater Target Images Based on CSLS-CycleGAN. *Syst. Eng. Electron.* **2023**, *45*, 1–16. Available online: http://kns.cnki.net/kcms/detail/11.2422.TN.20230807.1406.002.html (accessed on 7 January 2024).
15. Rajani, H.; Gracias, N.; Garcia, R. A Convolutional Vision Transformer for Semantic Segmentation of Side-Scan Sonar Data. *arXiv* **2023**, arXiv:2302.12416. [CrossRef]
16. Álvarez-Tuñón, O.; Marnet, L.R.; Antal, L.; Aubard, M.; Costa, M.; Brodskiy, Y. SubPipe: A Submarine Pipeline Inspection Dataset for Segmentation and Visual-inertial Localization. *arXiv* **2024**, arXiv:2401.17907.
17. Ge, Q.; Ruan, F.; Qiao, B.; Zhang, Q.; Zuo, X.; Dang, L. Side-scan sonar image classification based on style transfer and pre-trained convolutional neural networks. *Electronics* **2021**, *10*, 1823. [CrossRef]
18. Huo, G.; Wu, Z.; Li, J. Underwater object classification in Side-scan sonar images using deep transfer learning and semisynthetic training data. *IEEE Access* **2020**, *8*, 47407–47418. [CrossRef]
19. Li, C.; Ye, X.; Cao, D.; Hou, J.; Yang, H. Zero shot objects classification method of side scan sonar image based on synthesis of pseudo samples. *Appl. Acoust.* **2021**, *173*, 107691. [CrossRef]
20. Tang, Y.; Wang, L.; Jin, S.; Zhao, J.; Huang, C.; Yu, Y. AUV-Based Side-Scan Sonar Real-Time Method for Underwater-Target Detection. *J. Mar. Sci. Eng.* **2023**, *11*, 690. [CrossRef]
21. Huang, C.; Zhao, J.; Yu, Y.; Zhang, H. Comprehensive Sample Augmentation by Fully Considering SSS Imaging Mechanism and Environment for Shipwreck Detection Under Zero Real Samples. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5906814. [CrossRef]

22. Kim, S.; Kim, J.; Kim, T.; Heo, H.; Kim, S.; Lee, J.; Kim, J.H. Unpaired Panoramic Image-to-Image Translation Leveraging Pinhole Images. Available online: https://openreview.net/forum?id=bRm0rul3SZ (accessed on 7 January 2024).
23. Xi, J.; Ye, X.; Li, C. Sonar Image Target Detection Based on Style Transfer Learning and Random Shape of Noise under Zero Shot Target. *Remote Sens.* **2022**, *14*, 6260. [CrossRef]
24. 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 Electon. Sin.* **2021**, *49*, 1746.
25. Bird, J.J.; Barnes, C.M.; Manso, L.J.; Ekárt, A.; Faria, D.R. Fruit quality and defect image classification with conditional GAN data augmentation. *Sci. Hortic.* **2022**, *293*, 110684. [CrossRef]
26. Mikołajczyk, A.; Majchrowska, S.; Carrasco Limeros, S. The (de) biasing effect of gan-based augmentation methods on skin lesion images. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Singapore, 18–22 September 2022; Springer Nature: Cham, Switzerland, 2022; pp. 437–447.
27. Razavi, A.; Van den Oord, A.; Vinyals, O. Generating diverse high-fidelity images with vq-vae-2. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 1–11.
28. Karras, T.; Laine, S.; Aittala, M.; Hellsten, J.; Lehtinen, J.; Aila, T. Analyzing and improving the image quality of stylegan. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 8110–8119.
29. 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 Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
30. Xu, Y.; Ben, K.; Wang, T. Research on DCGAN Model Improvement and SAR Image Generation. *Comput. Sci.* **2020**, *47*, 93–99.
31. Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. *J. Big Data* **2019**, *6*, 1–48. [CrossRef]
32. Jiang, Y.; Ku, B.; Kim, W.; Ko, H. Side-scan sonar image synthesis based on generative adversarial network for images in multiple frequencies. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 1505–1509. [CrossRef]
33. Steiniger, Y.; Kraus, D.; Meisen, T. Generating synthetic Side-scan sonar snippets using transfer-learning in generative adversarial networks. *J. Mar. Sci. Eng.* **2021**, *9*, 239. [CrossRef]
34. Yang, Z.; Zhao, J.; Zhang, H.; Yu, Y.; Huang, C. A Side-Scan Sonar Image Synthesis Method Based on a Diffusion Model. *J. Mar. Sci. Eng.* **2023**, *11*, 1103. [CrossRef]
35. Liu, W.; Piao, Z.; Tu, Z.; Luo, W.; Ma, L.; Gao, S. Liquid warping gan with attention: A unified framework for human image synthesis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 5114–5132. [CrossRef]
36. Jiang, W.; Liu, S.; Gao, C.; Cao, J.; He, R.; Feng, J.; Yan, S. Psgan: Pose and expression robust spatial-aware gan for customizable makeup transfer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 5194–5202.
37. 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 October–2 November 2019; pp. 4570–4580.
38. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
39. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 1–11.
40. Li, X.; Hu, X.; Yang, J. Spatial group-wise enhance: Improving semantic feature learning in convolutional networks. *arXiv* **2019**, arXiv:1905.09646.
41. Li, C.; Wand, M. Precomputed real-time texture synthesis with markovian generative adversarial networks. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016*; Proceedings, Part III 14; Springer International Publishing: Cham, Switzerland, 2016; pp. 702–716.
42. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
43. Karras, T.; Aila, T.; Laine, S.; Lehtinen, J. Progressive growing of gans for improved quality, stability, and variation. *arXiv* **2017**, arXiv:1710.10196.
44. Wang, T.C.; Liu, M.Y.; Zhu, J.Y.; Tao, A.; Kautz, J.; Catanzaro, B. High-resolution image synthesis and semantic manipulation with conditional gans. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8798–8807.
45. Karras, T.; Laine, S.; Aila, T. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 4401–4410.
46. Denton, E.L.; Chintala, S.; Fergus, R. Deep generative image models using a laplacian pyramid of adversarial networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 1–9.
47. Huang, X.; Li, Y.; Poursaeed, O.; Hopcroft, J.; Belongie, S. Stacked generative adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 5077–5086.