

Article

SIGAN: A Multi-Scale Generative Adversarial Network for Underwater Sonar Image Super-Resolution

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

Department of Oceanography and Hydrography, Dalian Naval Academy, Dalian 116018, China; 18908414801@163.com (C.P.); trighosts@163.com (G.B.); 13998435151@163.com (Y.C.)

* Correspondence: jsh_1978@163.com

Abstract: Super-resolution (SR) is a technique that restores image details based on existing information, enhancing the resolution of images to prevent quality degradation. Despite significant achievements in deep-learning-based SR models, their application in underwater sonar scenarios is limited due to the lack of underwater sonar datasets and the difficulty in recovering texture details. To address these challenges, we propose a multi-scale generative adversarial network (SIGAN) for super-resolution reconstruction of underwater sonar images. The generator is built on a residual dense network (RDN), which extracts rich local features through densely connected convolutional layers. Additionally, a Convolutional Block Attention Module (CBAM) is incorporated to capture detailed texture information by focusing on different scales and channels. The discriminator employs a multi-scale discriminative structure, enhancing the detail perception of both generated and high-resolution (HR) images. Considering the increased noise in super-resolved sonar images, our loss function emphasizes the PSNR metric and incorporates the L2 loss function to improve the quality of the output images. Meanwhile, we constructed a dataset for side-scan sonar experiments (DNASI-I). We compared our method with the current state-of-the-art super-resolution image reconstruction methods on the public dataset KLSG-II and our self-built dataset DNASI-I. The experimental results show that at a scale factor of 4, the average PSNR value of our method was 3.5 higher than that of other methods, and the accuracy of target detection using the super-resolution reconstructed images can be improved to 91.4%. Through subjective qualitative comparison and objective quantitative analysis, we demonstrated the effectiveness and superiority of the proposed SIGAN in the super-resolution reconstruction of side-scan sonar images.

Keywords: image super-resolution; generative adversarial network; underwater sonar images; underwater target detection



Citation: Peng, C.; Jin, S.; Bian, G.; Cui, Y. SIGAN: A Multi-Scale Generative Adversarial Network for Underwater Sonar Image Super-Resolution. *J. Mar. Sci. Eng.* **2024**, *12*, 1057. <https://doi.org/10.3390/jmse12071057>

Academic Editor: Mohamed Benbouzid

Received: 30 May 2024
Revised: 19 June 2024
Accepted: 20 June 2024
Published: 24 June 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

The detection and identification of underwater targets such as shipwrecks, aircraft wreckages, pipelines, and reefs are crucial tasks in marine science research, resource exploration, and ocean mapping [1]. These tasks have significant implications for maritime traffic safety, marine fishery development, sonar detection, and military operations. Side-scan sonar, known for its high-resolution imaging capabilities, has long been the preferred technology for underwater target detection and identification [2–5]. However, due to limitations in range and the complexity of the measurement environment, a large number of low-resolution side-scan sonar images still exist, greatly hindering the development of underwater target identification. Super-resolution (SR), a technique that enhances image resolution through reconstruction, can improve the quality of underwater sonar images and play a crucial role in enhancing target detection accuracy [2–5], target segmentation [6–8], and other image scenarios [9–12].

Super-resolution (SR) is an image processing technology aimed at reconstructing high-resolution (HR) images from low-resolution (LR) images or videos [13]. Depending on the

number of input images, SR can be classified into single-image SR (SISR) or multi-image SR (MISR). SISR generates high-resolution images from a single low-resolution image and is widely used in environmental monitoring, medical image processing, video surveillance, and security. The key to SISR lies in exploring the mapping relationship between HR and LR images. Previous studies have proposed various methods to learn this mapping. Traditional filtering and interpolation techniques such as linear interpolation [14,15] generate HR images based on neighborhood information. Although these methods are computationally efficient, they simplify the mapping relationship in SISR, leading to overly smooth images that lack important details, especially texture and target edges.

With the development of deep learning, convolutional neural networks (CNNs) were introduced to the SR field by Dong et al. [16]. As deep learning neural networks evolved, many researchers designed powerful neural network models such as ResNet [17], DenseNet [18], and Residual Dense Blocks (RDBs) to further enhance SR models. These include multi-layer networks like VDSR, the enhanced deep SR (EDSR) network [19], the deep recursive convolutional network (DRCN) [20], and very deep residual channel attention networks (RCANs) [21].

Despite significant breakthroughs achieved by complex CNNs in image super-resolution reconstruction, issues such as blurred image edges, high noise levels, and poor perceived image quality persist. The advent of generative adversarial networks (GANs) led to their introduction into the SR domain by Ledig et al. [22]. Since the emergence of SRGAN [22], numerous GAN-based models have been applied to super-resolution image generation. Enhanced SRGAN (ESRGAN) reduced image artifacts by extending Residual Blocks (RRDBs), while fine-grained-attention GANs (FASRGANs) [23] improved the capability to generate high-quality images through image scoring.

Despite the significant achievements of super-resolution (SR) in optical images [24–27], the effectiveness of SR generation for underwater samples remains unsatisfactory. The complexity and diversity of interference in underwater environments result in substantial differences in texture and detail between sonar images and optical images. Existing algorithms such as RCAN [21], EDSR [19], SRResNe [28], and SRGAN have demonstrated suboptimal performance in the super-resolution reconstruction of side-scan sonar images. Specifically, EDSR and SRGANs exhibit limited capabilities in image restoration and texture reproduction, leading to unclear images. Furthermore, RCANs and SRResNe suffer from noise amplification issues while enhancing image resolution.

To address these issues and improve the image quality of super-resolution reconstruction for side-scan sonar images, this paper proposes a multi-scale adversarial generative network. The network details are as follows: The generator utilizes Residual Dense Blocks (RDB) to design a five-layer Residual Dense Network (RDN), extracting rich local features through densely connected convolutional layers [28–31]. The final layer of the generator incorporates a Convolutional Block Attention Module (CBAM) to capture detailed texture information by focusing on different scales and channels [32,33]. The discriminator adopted a multi-scale discriminative structure to enhance the detail perception of both generated and high-resolution (HR) images through the comprehensive scale-aware judgment of sonar images [34,35]. Additionally, considering the increased noise in super-resolved sonar images, our loss function emphasized the peak signal-to-noise ratio (PSNR) to reduce noise and improve output image quality. We also constructed a dataset (DNASI-I) for side-scan sonar experiments. Subjective qualitative comparisons and objective quantitative analyses with the current state-of-the-art super-resolution image reconstruction methods were conducted on the public dataset KLSG-II and our dataset DNASI-I. The experimental results demonstrated the significant effectiveness and superiority of SIGAN in the super-resolution reconstruction of side-scan sonar images.

2. Methods

2.1. Overall Structure of SIGAN

The SIGAN is a multi-scale generative adversarial network model specifically designed for underwater side-scan sonar images. Similar to traditional super-resolution generative adversarial network structures, as shown in Figure 1, the SIGAN mainly consists of two parts: the generator and the discriminator.

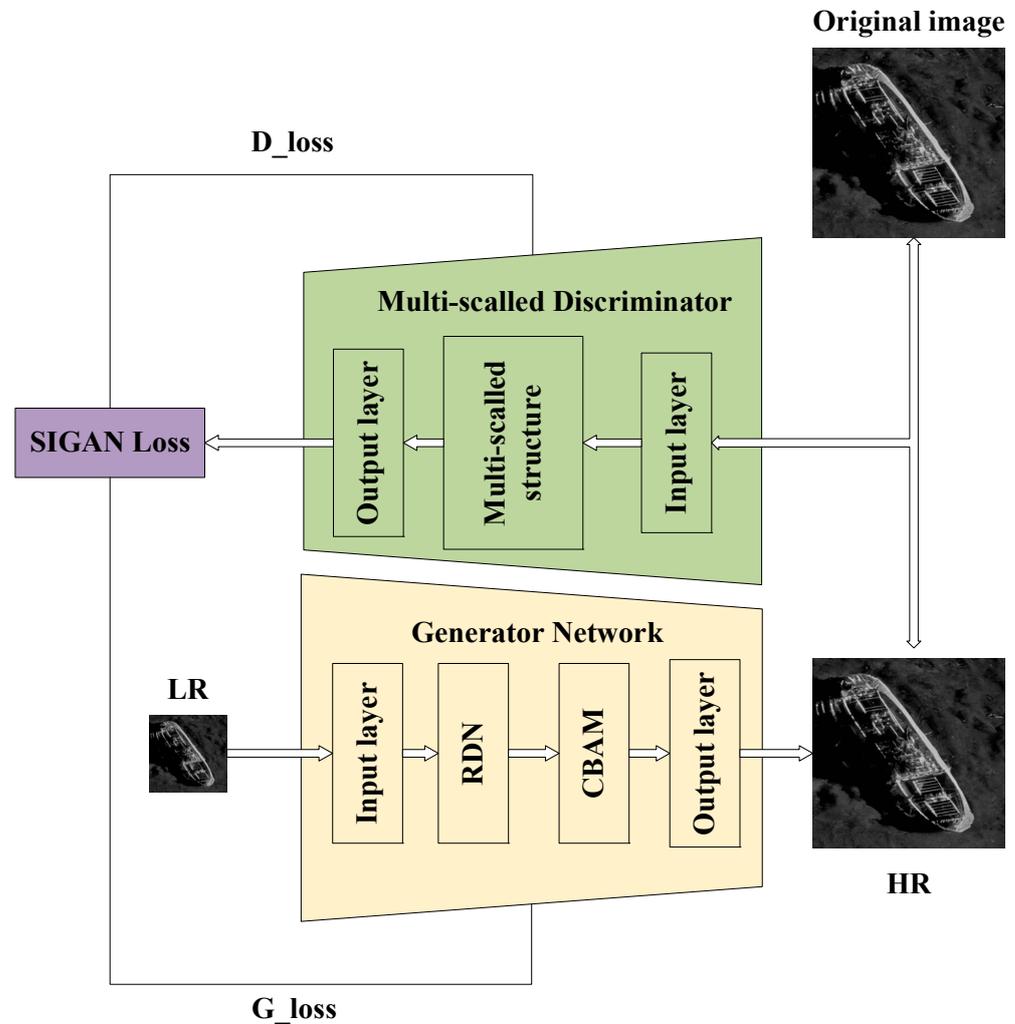


Figure 1. Overall structure of SIGAN.

As shown in Figure 1, the generator takes the low-resolution (LR) image as an input, which sequentially passes through the input layer, Residual Dense Network (RDN) structure, CBAM attention mechanism module, and finally through the output layer to produce the high-resolution (HR) reconstructed image. The HR image and the high-resolution original image (original image) are then simultaneously inputted into the multi-scale discriminator for discrimination. The VGG16 network is used to extract image features for comparison and to calculate the loss function, which is then backpropagated to both the generator and the discriminator.

During the entire training process, the goal of the generator is to reconstruct high-resolution images (HR) from low-resolution side-scan sonar images (LR) that the discriminator cannot distinguish. The goal of the discriminator is to differentiate between real side-scan sonar images (original image) and the reconstructed images (HR) generated by the generator.

2.2. CBAM-Based Generator Structure

Our proposed SIGAN generator consists of four parts, as shown in Figure 2. The input layer is a single convolutional layer activated by the ReLU function. The output layer is an upsampling layer to generate super-resolution reconstructed images. The main part of the generator consists of Residual Dense Blocks (RDB) and the CBAM attention mechanism module.

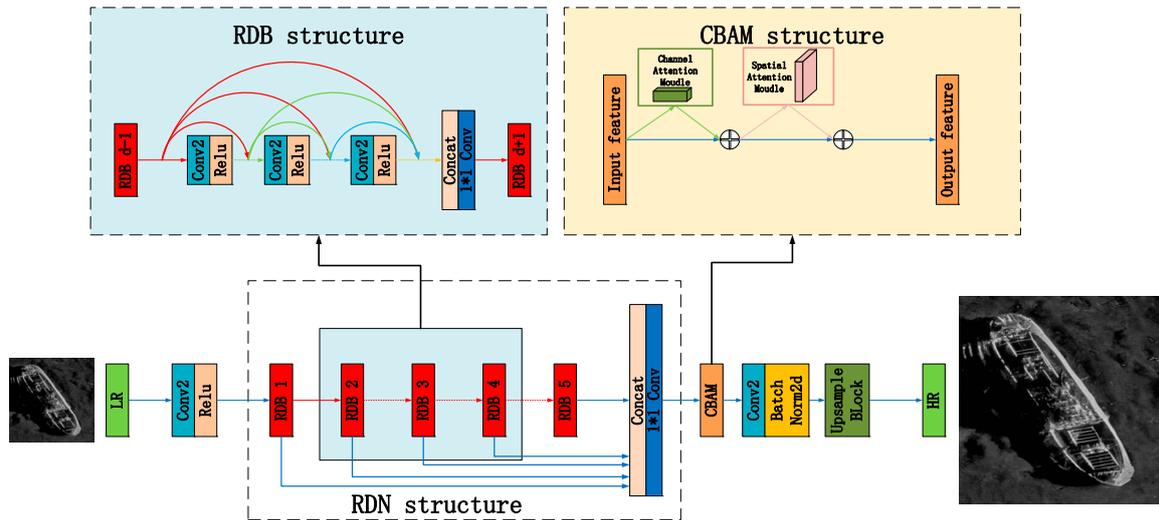


Figure 2. SIGAN generator structure.

2.2.1. RDN Structure

The Residual Dense Network (RDN) was proposed by Yulun Zhang in 2018 and effectively utilizes the hierarchical features of convolutional layers. The overall structure of the network has three main characteristics: First, the RDB module connects the state of the previous RDB module to each layer of the current RDB module through skip-connections, forming a continuous memory mechanism. Second, through local feature fusion within the RDB module, the local features of the previous RDB module are adaptively fused with the current module’s local features, resulting in more extensive and stable training. After adequately capturing the internal features of the module, global feature fusion (GFF) is used to jointly learn the global hierarchical features.

2.2.2. CBAM Structure

The adequate learning of target detail features and background features in side-scan sonar images is crucial for the generator 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) in the generator, placed after the RDN layer. Figure 2 shows the overall architecture after adding the CBAM module. The CBAM consists of two independent sub-modules: the Channel Attention Module (CAM) and the Spatial Attention Module (SAM). Compared to attention mechanisms that focus solely on spatial features, this design achieves better results while saving parameters and computational power and can be conveniently inserted into various layers of the network. This module aims to reduce information diffusion while amplifying cross-dimensional interactions in the channel and spatial dimensions of the image, thereby improving network performance. By focusing on relevant features and minimizing interference, the CBAM module enhances the detail and fineness of side-scan sonar images, which is crucial for generating realistic and high-quality output images.

2.3. Multi-Scale Discriminator Structure

In the field of super-resolution, traditional methods typically input the image directly into the discriminator to obtain a single score or a score map [36,37]. However, for side-scan sonar images rich in detail and texture features, using a single score for decision-making is too absolute as it ignores the local and multi-scale features of the image. Additionally, the complex background features of underwater images make it difficult to use the entire image as a reference for the quality of super-resolution image reconstruction. To address these issues, SIGAN, based on the SRGAN discriminator structure with the VGG16 style, designs a multi-scale discriminator structure with five scales. It mainly consists of four parts: the encoder, multi-scale module, decoder, and output layer, as shown in Figure 3.

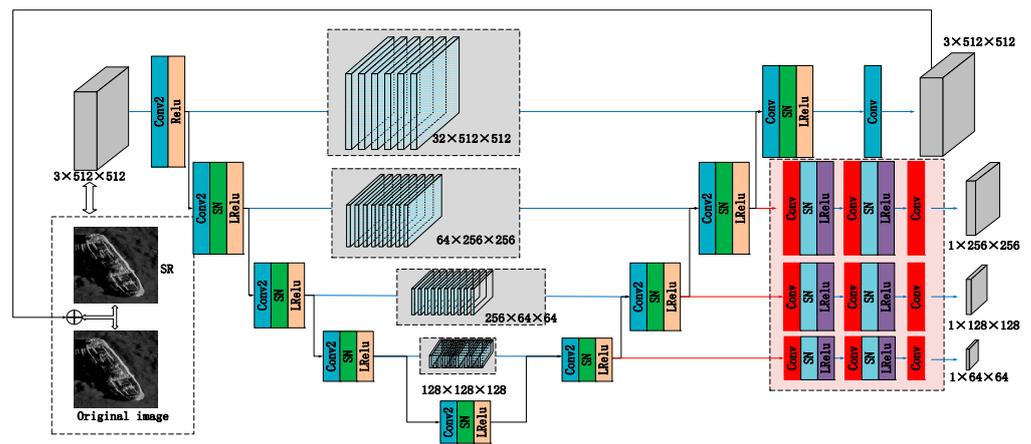


Figure 3. SIGAN discriminator structure.

The encoder extracts multi-scale image features of side-scan sonar through the VGG16 network, using convolutional layers to reduce the size of feature maps while increasing the number of feature maps. The encoder is designed with five layers, corresponding to image sizes of 512, 256, 128, 64, and 32; the decoder and output layer utilize bilinear feature interpolation for upsampling and use convolutional layers to reduce the number of feature maps.

Multi-scale module: In the SIGAN discriminator, we aimed to improve the discriminator’s ability to discriminate through receptive fields of different sizes at different scales, rather than relying solely on a global judgment from a single image. For example, for the 2nd, 3rd, and 4th scales, the image sizes are $64 \times 256 \times 256$, $128 \times 128 \times 128$, and $256 \times 64 \times 64$, respectively, with receptive field sizes of $3 \times 16 \times 16$, $3 \times 8 \times 8$, and $3 \times 4 \times 4$. The generated patch score maps, map_2, map_3, and map_4, provide more accurate feedback for the image’s local regions, thereby enhancing the network’s ability to recover detailed textures at different scales. This method is particularly effective in processing side-scan sonar images [38].

2.4. Loss Function Composition

The loss function proposed for SIGAN consists of four parts: adversarial loss, perceptual loss, TV loss, and l_2 loss introduced with a focus on the PSNR metric. The calculation method of the loss function is as follows:

$$l_{\text{SIGAN}} = \alpha \cdot l_2 + \beta \cdot l_{\text{adv}} + \chi \cdot l_{\text{pre}} + \delta \cdot l_{\text{TV}} \quad (1)$$

Among these, $\alpha/\beta/\chi/\delta$ represents the respective weights. Currently, mainstream super-resolution algorithms reduce the emphasis on the PSNR metric to pursue the visual quality of the reconstructed image. However, for underwater side-scan sonar images, the level of noise is a crucial factor affecting image discrimination. Therefore, in our loss

function definition, the weight of the l_2 loss is set relatively high. This approach aims to reduce the noise in the generated images. The calculation method for MSE is as follows:

$$l_2 = MSE = \frac{1}{mn} \sum_{i=1}^n \sum_{j=1}^m (x_{ij} - y_{ij})^2 \quad (2)$$

Among these, x_{ij} and y_{ij} represent the predicted and actual values of the ij th pixel of the image, respectively. The calculation method for $PSNR$ is as follows:

$$PSNR = 20 \bullet \log_{10} \left(\frac{MAX_I}{\sqrt{MSE}} \right) \quad (3)$$

where MAX_I represents the maximum pixel value in the image. The adversarial loss is defined as follows:

$$l_{adv} = - \sum_{i=1}^n \log D(I_i) - \sum_{i=1}^n \log(1 - D(\hat{I}_i)) \quad (4)$$

where I is the real side-scan sonar image, \hat{I} is the super-resolution reconstructed image, and $D(I_i)$ is the probability that the discriminator classifies the i th side-scan sonar image as real. To avoid reliance on pixel-level loss, we used the VGG loss based on the ReLU activation layers of the VGG network as defined in SRGAN. $\varphi_{i,j}$ denotes the feature maps obtained from the i th max-pooling layer before the j th convolutional layer (after activation) in the network. The perceptual loss in this paper is defined as follows:

$$l_{pre} = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\varphi_{i,j}(I^{HR})_{x,y} - \varphi_{i,j}(G_{\theta_G}(I^{LR})_{x,y}))^2 \quad (5)$$

In this formula, $W_{i,j}$ and $H_{i,j}$ describe the dimensions of the feature maps in the VGG network.

Total Variation Loss, derived from total variation denoising in image processing, was calculated by summing the square roots of the squared differences between each pixel and its right and bottom neighbors:

$$l_{TV} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n ((x_{i,j-1} - x_{i,j})^2 + (x_{i+1,j} - x_{i,j})^2)^{\frac{\beta}{2}} \quad (6)$$

3. Experimental Validation

To demonstrate the practicality and efficacy of the SIGAN in the super-resolution reconstruction of side-scan sonar images, we conducted comparative analyses with several state-of-the-art super-resolution (SR) networks, including EDSR, the SRGAN, the RCAN, and MSRRResNet. To ensure a fair comparison, these models were evaluated under the same dataset conditions. The hardware setup for model training comprised two Intel Xeon Silver 4410T processors and four NVIDIA GeForce RTX 4090 graphics cards. The software environment was configured with PyTorch 1.6.0, CUDA 11.8, and Python 3.10 on Windows 10.

3.1. Datasets

Given the scarcity of publicly available side-scan sonar datasets, we collected various underwater sonar images from diverse scenarios and locations across the nation. The data sources included publicly available datasets (e.g., KLSG-II), datasets obtained through collaborations with other universities, and data from in situ ocean exploration experiments. The dataset encompassed a wide range of content, including seabed shipwrecks, aircraft wreckage, underwater rocks, schools of fish, divers, and seabed sand dunes. We compiled

this data to create our dataset, named DNASI-I. To validate our methodology, we selected two datasets for experimentation: the publicly available side-scan sonar dataset KLSG-II and our proprietary dataset DNASI-I. The KLSG-II dataset is available for download on GitHub (URL: <https://github.com/HHUCzCz/-SeabedObjects-KLSG--II>, accessed on 3 June 2024). Both datasets underwent identical processing techniques. Specifically, we randomly cropped all images to 128×128 pixels, covering three categories: seabed shipwrecks, aircraft wreckages, and underwater rocks. Finally, we selected 100 images with distinct target texture structures and easy comparability from our dataset DNASI-I to serve as the test set (t100). Sample images from the datasets are shown in Figure 4.

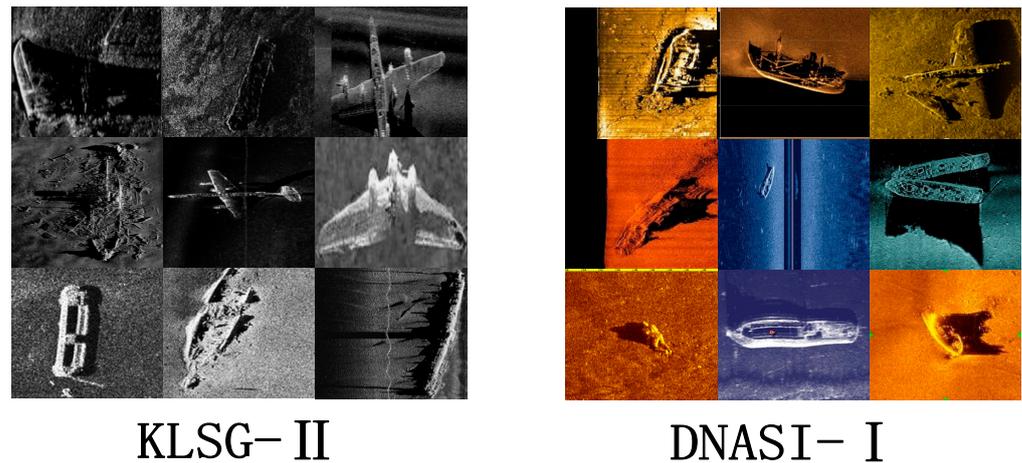


Figure 4. Sample images from the datasets. The left side shows selected images from KLSG-II, and the right side shows selected images from DNASI-I.

3.2. Objective Evaluation Metrics

The evaluation metrics adopted in this study included peak signal-to-noise ratio (PSNR), the Structural Similarity Index (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS). The PSNR was used to calculate the mean squared error between two images, deriving a peak signal-to-noise ratio to assess the similarity between the training and generated images. The SSIM quantifies structural information from the perspective of image composition, independent of brightness and contrast, reflecting the structural integrity of objects within the scene. LPIPS measures the perceptual differences between two images by learning the inverse mapping from the generated to real images, prioritizing perceptual similarity. These metrics collectively assist in comprehensively assessing the quality of generated images.

3.3. Visual Comparison with Other SR Methods

We trained the models using the RCAN, EDSR, MSRResNet, the SRGAN, and our method (the SIGAN). Subjective visual comparisons were conducted for magnification factors of $r = 2$, $r = 4$, and $r = 8$. Figure 5 shows the results generated by training our SIGAN model on the KLSG-II dataset. Figure 6 presents the results generated by different networks trained on the KLSG-II dataset.

A comparative analysis was performed using images of underwater rocks, aircraft, and shipwrecks. At a magnification of $r = 2$, images generated by the RCAN, EDSR, and the SRGAN exhibited pronounced sharpening effects, resulting in overly harsh edges and image distortions. Conversely, images from MSRResNet appeared overly smooth, leading to a loss in high-frequency information. At a magnification of $r = 4$, while EDSR-based methods managed to restore some textures, they also introduced artifacts not present in the original images, causing distortions. The RCAN and SRGAN methods produced images with unclear edges and blurring. The images from MSRResNet continued to suffer from excessive smoothness, resulting in distortions. At a magnification of $r = 8$, a detailed image

of a sunken ship was analyzed. Comparisons of the ship’s surface structure revealed that MSRResNet and EDSR restored images with uneven edges, though some high-frequency details were preserved. The RCAN and SRGAN handled edge details well but resulted in overall blurry images.

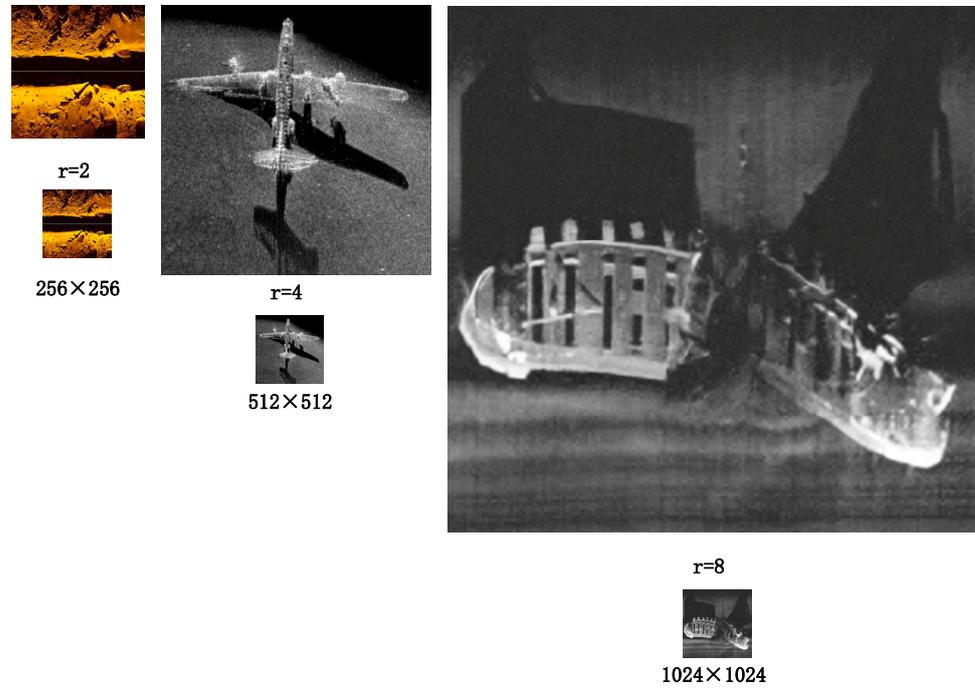


Figure 5. Results generated by training the SIGAN.

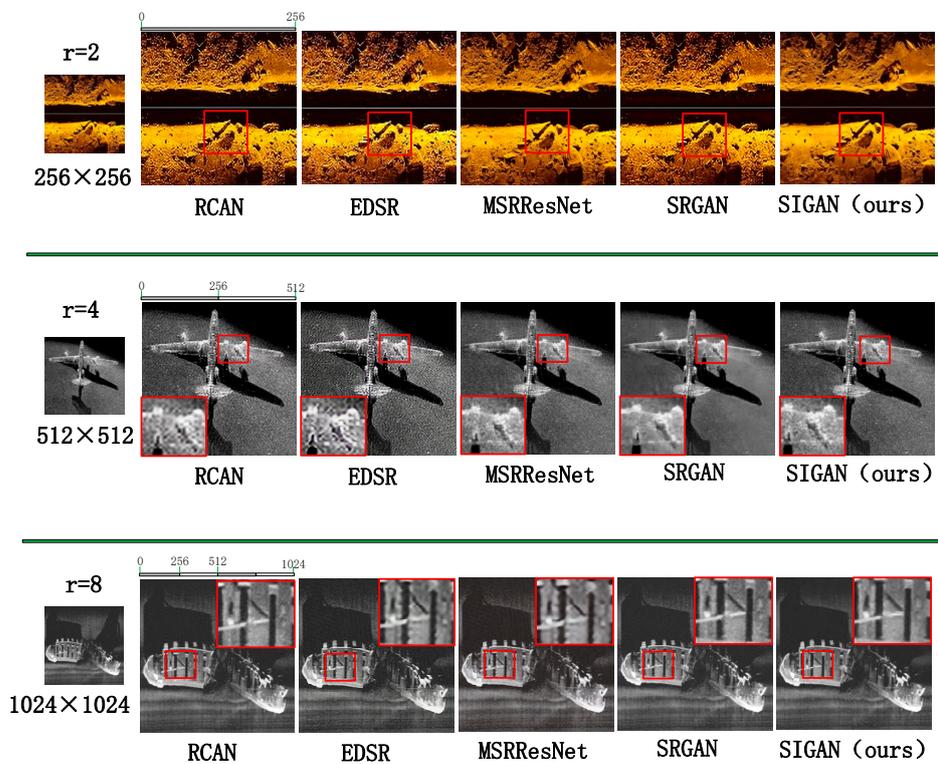


Figure 6. Results generated by different networks trained on the KLSG-II dataset. Note: to make the comparative analysis more intuitive, we scaled the 512×512 and 1024×1024 images down to 256×256 for uniformity. The actual image comparison is reflected in Figure 5.

In contrast, our SIGAN method, utilizing attention mechanisms, focused more adeptly on the super-resolution generation of targets, thus offering significant advantages in handling edge details. For instance, at a magnification of $r = 4$ in Figure 6, the SIGAN effectively managed the edge details of the aircraft wreckage, enhancing edge resolution while preserving high-frequency image details. Due to the loss function’s emphasis on the PSNR metric, our method produced noticeably fewer image artifacts compared to other methods, successfully avoiding the introduction of noise while enhancing resolution.

Training was also conducted on the DNASI-I dataset, with the results shown in Figure 7:

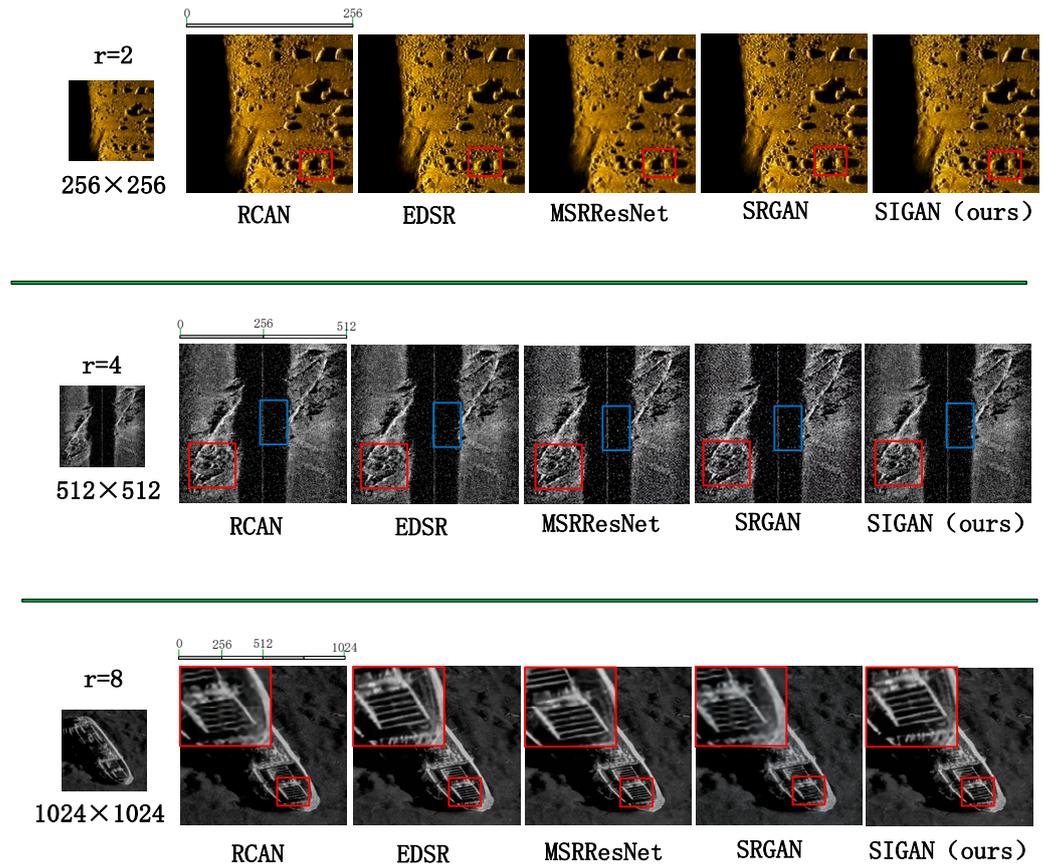


Figure 7. Results generated on the DNASI-I dataset using different networks.

Representative images were selected for comparison. At a magnification of $r = 2$, images generated by the RCAN, EDSR, and the SRGAN still exhibited pronounced sharpening effects. At a magnification of $r = 4$, a noisier image was selected for comparison, revealing significant noise increase in MSRResNet and the SRGAN, while our method managed noise effectively, not exacerbating it with increased resolution. At a magnification of $r = 8$, an image with detailed horizontal features of a sunken ship showed that the deck railing structures restored by the RCAN and SRGAN were blurry and discontinuous, yielding suboptimal results.

In contrast, our SIGAN approach provided clear advantages, especially noticeable at higher magnifications. The attention mechanism in the SIGAN allowed for superior detail preservation and clarity, demonstrating enhanced capability in handling complex image textures without compromising on noise, which is crucial for practical applications in underwater imaging.

3.4. Objective Evaluation with Other Methods

Table 1 presents the quantitative analysis of our SIGAN method on the KISG-II dataset, using magnification factors of $r = 2$, $r = 4$, and $r = 8$ for different categories such as

shipwrecks, aircraft, and underwater rocks. We calculated the average results for PSNR, SSIM, and LPIPS. The most outstanding performance indicators are highlighted in bold.

Table 1. Performance metrics on the KISG-II dataset for various methods. The underlined parts represent the data with better performance.

KLSG-II		RCAN			EDSR			MSRResNet			SRGAN			SIGAN		
r	Class	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
2	Shipwreck	28.209	0.7966	0.196	29.145	0.8276	0.153	29.641	0.8356	0.152	28.209	0.7966	0.16	30.1563	0.8524	0.199
	Aircraft	28.555	0.78	0.1458	29.244	0.8033	0.1303	31.054	0.9364	0.1046	28.555	0.78	0.1458	31.3621	0.8836	0.0828
	Rocks	22.707	0.5652	0.251	23.023	0.5625	0.2433	24.514	0.6765	0.2256	22.707	0.5652	0.251	26.5623	0.7246	0.228
	Set100	26.433	0.7167	0.188	28.809	0.7028	0.1826	29.134	0.7994	0.1589	26.433	0.7167	0.188	29.1156	0.8324	0.1285
4	Shipwreck	27.866	0.5667	0.4035	24.516	0.5843	0.314	25.595	0.6321	0.3141	23.866	0.5667	0.4035	27.2544	0.7315	0.1998
	Aircraft	24.812	0.652	0.2613	25.065	0.5825	0.2932	30.386	0.8362	0.1026	24.812	0.652	0.2613	29.3691	0.8210	0.1212
	Rocks	19.962	0.2297	0.3691	20.082	0.2254	0.3664	21.261	0.3122	0.378	19.962	0.2297	0.3691	24.2145	0.6615	0.321
	Set100	22.722	0.3971	0.3114	21.478	0.3911	0.3288	23.993	0.5635	0.2547	22.722	0.3971	0.3114	25.9541	0.761	0.25
8	Shipwreck	21.267	0.431	0.4035	22.06	0.4466	0.4171	22.958	0.4844	0.4327	21.267	0.431	0.4035	23.2155	0.631	0.35
	Aircraft	22.804	0.4571	0.2613	22.278	0.4247	0.4142	27.617	0.732	0.1482	22.804	0.4571	0.2613	25.4612	0.6315	0.2218
	Rocks	18.92	0.1439	0.3691	19.261	0.152	0.4214	19.552	0.1685	0.4028	18.92	0.1439	0.3691	21.1586	0.5365	0.499
	Set100	20.741	0.3396	0.3114	21.163	0.3448	0.4178	21.874	0.3997	0.3514	20.741	0.3396	0.3114	22.1498	0.636	0.3599

An analysis of Table 1 revealed that our SIGAN method consistently outperformed other methods across all scales, demonstrating significantly better average values for each evaluation metric, specifically:

- **PSNR:** the SIGAN and EDSR generally showed higher PSNR values, particularly at a magnification factor of 2, where both algorithms demonstrated strong performance across multiple categories. Notably, at $r = 4$, the SIGAN achieved an average PSNR of 25.95 on the Set100, nearly indistinguishable from high-resolution images.
- **SSIM:** the SIGAN consistently performed well across different magnification factors and categories, particularly in the “Shipwreck” and “Aircraft” categories, where the SSIM values were high, indicating good structural fidelity of the images.
- **LPIPS:** Low LPIPS values indicated that the perceived quality of the images was closer to the original. EDSR stood out in the “Set100” category, showcasing its advantages in perceived image quality. Similarly, the SIGAN showed lower LPIPS values at lower magnification factors (such as $2\times$) in the “Shipwreck” and “Aircraft” categories, indicating superior perceptual quality.

We also conducted a quantitative analysis of various methods on the DNASI-I dataset, with the results presented in Table 2. The best-performing data is highlighted in bold:

From the analysis of Table 2, it is evident that the SIGAN continued to excel on the DNASI-I dataset. Although the performance generally declines with increasing magnification factors, it is noteworthy that our method maintained a certain level of stability without significant drops. In Table 2, the SIGAN’s performance on the Set100 test set significantly surpasses other methods on every evaluation metric. The quantitative comparisons clearly demonstrate that our method is superior. At $r = 4$, the average PSNR on Set100 is higher by 3.5, illustrating the clarity and detail preserved by the SIGAN.

Comparing Tables 1 and 2, it is apparent that models trained on the DNASI-I dataset perform better than those trained on the KLSG-II dataset at the same scale. Whether it is our SIGAN approach or other networks like EDSR, this indirectly attests to the superior performance and higher quality of our dataset in super-resolution tasks.

Table 2. Performance metrics on the DNASI-I dataset for various methods. The underlined parts represent the data with better performance.

DNASI-I		RCAN			EDSR			MSRResNet			SRGAN			SIGAN		
r	Class	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
2	Shipwreck	30.429	0.8763	0.1764	32.059	0.9103	0.1377	32.605	<u>0.9392</u>	0.1368	30.429	0.8763	0.144	<u>33.1719</u>	<u>0.9377</u>	<u>0.1791</u>
	Aircraft	31.41	0.858	0.1312	32.169	0.8836	0.1173	<u>34.16</u>	1.03	0.0942	31.41	0.858	0.1312	<u>34.4983</u>	<u>0.9719</u>	<u>0.0745</u>
	Rocks	24.978	0.6217	0.2259	25.325	0.6188	0.2189	26.966	0.7442	0.2031	24.978	0.6217	0.2259	<u>29.2185</u>	<u>0.797</u>	<u>0.2052</u>
	Set100	29.076	0.7884	0.1692	31.69	0.7731	0.1643	32.047	0.8793	0.1431	29.076	0.7884	0.1692	<u>32.027</u>	<u>0.9157</u>	<u>0.1156</u>
4	Shipwreck	<u>30.652</u>	0.6234	0.3632	26.967	0.6428	0.2826	28.155	0.6953	0.2826	26.252	0.6234	0.3632	<u>29.9803</u>	<u>0.8046</u>	<u>0.1798</u>
	Aircraft	27.293	0.7172	0.2352	27.572	0.6408	0.2639	<u>33.425</u>	0.9198	0.0923	27.293	0.7172	0.2352	<u>32.306</u>	<u>0.9031</u>	<u>0.1091</u>
	Rocks	21.958	0.2527	0.3322	22.09	0.2479	0.3297	23.387	0.3434	0.3402	21.958	0.2527	0.3322	<u>26.6359</u>	<u>0.7277</u>	<u>0.2889</u>
	Set100	25.094	0.4368	0.2803	23.626	0.4302	0.296	26.392	0.6199	0.2293	25.094	0.4368	0.2803	<u>28.5495</u>	<u>0.8371</u>	<u>0.225</u>
8	Shipwreck	23.394	0.4741	0.3632	24.266	0.4911	0.3757	25.254	0.5328	0.3896	23.394	0.4741	0.3632	<u>25.5371</u>	<u>0.6941</u>	<u>0.315</u>
	Aircraft	25.085	0.5028	0.2352	24.506	0.4672	0.3725	<u>30.378</u>	0.8052	0.1334	25.085	0.5028	0.2352	<u>28.0073</u>	<u>0.6947</u>	<u>0.1996</u>
	Rocks	20.812	0.1583	0.3322	21.187	0.1672	0.3791	21.508	0.1852	0.3625	20.812	0.1583	0.3322	<u>23.2744</u>	<u>0.5902</u>	<u>0.4491</u>
	Set100	22.815	0.3736	0.2803	23.279	0.3794	0.376	24.061	0.4397	0.3163	22.815	0.3736	0.2803	<u>24.3644</u>	<u>0.5781</u>	<u>0.2232</u>

4. Discussion

The essence of deep-learning-based side-scan sonar image super-resolution is to extract edge detail features from targets, such as the bows, masts, and shadows of shipwrecks, from complex models. These features, learned from high-resolution images, are then applied to enhance the detail in low-resolution images. Considering that the ultimate goal of image super-resolution reconstruction is to improve the performance of downstream target recognition tasks, this discussion focuses on analyzing the performance of SIGAN-generated super-resolution reconstruction images in target recognition tasks. Additionally, we conducted ablation studies to explore the contributions of various components of the network, aiming to understand how super-resolution enhances the accuracy of target recognition.

4.1. Performance in Target Recognition

To assess the effectiveness of super-resolution reconstructed side-scan sonar images in target recognition tasks, we selected a dataset of 200 high-resolution side-scan shipwreck images, each measuring 600×600 pixels, for testing. As a baseline, we used 2181 low-resolution side-scan sonar images of 128×128 pixels, reconstructed using EDSR, the SRGAN, the RCAN, MSRResNet, and our proposed SIGAN, followed by annotation processing. The detection model employed was Yolov5, known for its light weight, accuracy, and ease of deployment. The experimental groups are detailed as follows Table 3:

Table 3. Target recognition testing groups.

Group	LR Images	HR Images
G1	2181	-
G2	2081	100 (EDSR)
G3	2081	100 (SRGAN)
G4	2081	100 (RCAN)
G5	2081	100 (MSRResNet)
G6	2081	100 (SIGAN)
Test images	-	200

Testing was conducted using 200 real high-resolution side-scan sonar shipwreck target images, and the models were evaluated using precision, recall, and average precision (AP), widely used metrics in the field of object detection. The results are as follows:

From the results in Table 4, it was observed that the introduction of high-resolution reconstructed images does not always directly improve the accuracy of target detection. For example, compared to Group G1 (using only low-resolution images), Group G3 (using

high-resolution images reconstructed by the SRGAN) saw a 2% decrease in precision. This highlights that super-resolution algorithms may negatively impact detection performance under certain conditions, potentially due to the loss of complex textures or excessive smoothing during the reconstruction process. Notably, the super-resolution approach based on EDSR (G2) achieved a precision of 91.1%, indicating a relatively high accuracy rate, yet its recall was lower than the control group (G1). This suggests that although this method can increase the probability of correctly identifying targets, it does not detect as many targets as desired. This may reflect some limitations of the EDSR method in reconstructing details, such as possibly losing some critical information while preserving local image features. In contrast, our method (G6, using high-resolution images reconstructed by SIGAN) performed excellently in both precision and recall, with a 2.6% increase in precision over the original low-resolution images (G1), significantly demonstrating the effectiveness of our approach. The high precision and recall indicated that images reconstructed by the SIGAN not only retained crucial textures and structural details but also enhanced the authenticity and diversity of the images, which are vital for applications in side-scan sonar imaging.

Table 4. Effectiveness of different training sets on real measurement side-scan sonar shipwreck target image detection. The bolded parts represent the data with better performance.

	Precision	Recall	AP0.5	AP0.5:0.95
G1	88.8%	89.9%	0.938	0.558
G2	91.1%	85.7%	0.935	0.576
G3	86.8%	92.6%	0.932	0.576
G4	89.7%	90.1%	0.945	0.581
G5	90.9%	94.1%	0.962	0.566
G6	91.4%	94.6%	0.959	0.579

To prevent dataset bias on a single detection model, multiple detection models (YOLOv5n, YOLOv5s, and YOLOv5m) were used for a precision comparison experiment (Table 5). The analysis of the results showed that there are differences in precision among the detection models due to varying complexities and training durations. However, a horizontal comparison across the three experimental groups consistently showed the best performance in Group G6, further validating the effectiveness of the experimental data.

Table 5. Precision comparison among different models. The bolded parts represent the data with better performance.

Model/Group	G1	G2	G3	G4	G5	G6
YOLOv5n	88.8%	91.1%	86.8%	89.7%	90.9%	91.4%
YOLOv5s	89.6%	90.2%	87.5%	89.5%	91.4%	92.3%
YOLOv5m	90.1%	91.5%	89.6%	90.3%	91.9%	92.9%

4.2. Ablation Study

To verify the role of each component in image super-resolution, we conducted ablation studies on the CBAM, the multi-scale structure of the discriminator, and the L2 loss function using the method of controlled variables. The evaluation metrics employed were PSNR, SSIM, and LPIPS. Four groups were designed for comparative experiments, with the experimental setup, training dataset, and evaluation data consistent with Chapter 3. The results are shown in Table 6.

Table 6. Ablation study.

Group	CBAM Model	Multi-Scale	L2 Loss	PSNR	SSIM	LPIPS
T1	-	-	-	30.68	0.828	0.232
T2	√	-	-	32.288	0.781	0.054
T3	-	√	-	33.721	0.92	0.155
T4	-	-	√	34.925	0.938	0.147
T5	√	√	√	35.92	0.969	0.035

From the table, it is evident that compared to the control group (T1), Group T2, which integrated the CBAM, focused on important channels and spatial regions, aiding the model in concentrating on crucial information in the image. This significantly improved the similarity of the reconstructed image on the target, as evidenced by the substantial improvement in the LPIPS metric. Comparing T1 and T3, the multi-scale structure helped the discriminator in distinguishing the authenticity of images, thereby enhancing the quality of the generated images, which was reflected in the improved PSNR and SSIM metrics. Comparing T1 with T4 shows the importance of the L2 loss function in enhancing the PSNR metric, which increased by 4.3, significantly reducing the noise in the generated images.

5. Conclusions

Addressing the challenges of limited underwater target samples, small dataset sizes, high acquisition difficulty, and low resolution in side-scan sonar imagery, we proposed a super-resolution reconstruction method named the SIGAN, targeting shipwrecks, airplane debris, and underwater reefs. We conducted comparative experiments with current state-of-the-art super-resolution reconstruction methods. The experimental results demonstrated that our proposed method exhibited significant superiority in both subjective performance and objective data analysis. The generated images have clear edges and complete structures, achieving a PSNR value of 32 on the DNASI-I dataset, which is significantly higher than other similar super-resolution algorithms.

Additionally, we discussed the results of target detection and ablation experiments. The results indicated that using the SIGAN for target detection improved accuracy by 2% compared to general methods and by 3.6% compared to the original images. These findings preliminarily demonstrated the applicability and superiority of the SIGAN method in handling the super-resolution reconstruction of underwater side-scan sonar images, providing an effective solution to the issues of underwater target sample scarcity and low image resolution. However, due to the current small scale of side-scan sonar datasets, our method's applicability to all side-scan sonar data remains unverified. Therefore, the experimental results of this study have certain limitations. More publicly available data and extensive experiments are needed in the future to further refine and validate our method.

Author Contributions: Methodology, C.P., S.J. and Y.C.; Software, S.J.; Validation, C.P. and G.B.; Resources, 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: Data are contained within the article.

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

References

1. Henriksen, L. Real-Time Underwater Object Detection Based on an Electrically Scanned High-Resolution Sonar. In Proceedings of the IEEE Symposium on Autonomous Underwater Vehicle Technology (AUV'94), Cambridge, MA, USA, 19–20 July 1994; pp. 99–104.
2. Huo, G.; Wu, Z.; Li, J. Underwater object classification in sidescan sonar images using deep transfer learning and semisynthetic training data. *IEEE Access* **2020**, *8*, 47407–47418. [[CrossRef](#)]
3. Wang, Z.; Guo, J.; Zeng, L.; Zhang, C.; Wang, B. MLFFNet: Multilevel feature fusion network for object detection in sonar images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5119119. [[CrossRef](#)]
4. Zhou, T.; Si, J.; Wang, L.; Xu, C.; Yu, X. Automatic detection of underwater small targets using forward-looking sonar images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4207912. [[CrossRef](#)]
5. Zhang, P.; Tang, J.; Zhong, H.; Ning, M.; Liu, D.; Wu, K. Self-trained target detection of radar and sonar images using automatic deep learning. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 4701914. [[CrossRef](#)]
6. Xu, H.; Zhang, L.; Er, M.J.; Yang, Q. Underwater Sonar Image Segmentation Based on Deep Learning of Receptive Field Block and Search Attention Mechanism. In Proceedings of the 2021 4th International Conference on Intelligent Autonomous Systems (ICoIAS), Wuhan, China, 14–16 May 2021; pp. 44–48.
7. Wang, Z.; Guo, J.; Huang, W.; Zhang, S. Side-scan sonar image segmentation based on multi-channel fusion convolution neural networks. *IEEE Sens. J.* **2022**, *22*, 5911–5928. [[CrossRef](#)]
8. Li, J.; Jiang, P.; Zhu, H. A local region-based level set method with Markov random field for side-scan sonar image multi-level segmentation. *IEEE Sens. J.* **2020**, *21*, 510–519. [[CrossRef](#)]
9. Dos Santos, M.M.; De Giacomo, G.G.; Drews, P.L.J.; Botelho, S.S. Matching color aerial images and underwater sonar images using deep learning for underwater localization. *IEEE Robot. Autom. Lett.* **2020**, *5*, 6365–6370. [[CrossRef](#)]
10. Zhou, T.; Wang, Y.; Chen, B.; Zhu, J.; Yu, X. Underwater multitarget tracking with sonar images using thresholded sequential Monte Carlo probability hypothesis density algorithm. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [[CrossRef](#)]
11. Nambiar, A.M.; Mittal, A. A Gan-Based Super Resolution Model for Efficient Image Enhancement in Underwater Sonar Images. In Proceedings of the OCEANS 2022-Chennai, Chennai, India, 21–24 February 2022; pp. 1–8.
12. Chen, W.; Gu, K.; Lin, W.; Yuan, F.; Cheng, E. Statistical and structural information backed full-reference quality measure of compressed sonar images. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *30*, 334–348. [[CrossRef](#)]
13. Freeman, W.T.; Pasztor, E.C.; Carmichael, O.T. Learning low-level vision. *Int. J. Comput. Vis.* **2000**, *40*, 25–47. [[CrossRef](#)]
14. Duchon, C.E. Lanczos filtering in one and two dimensions. *J. Appl. Meteorol. Climatol.* **1979**, *18*, 1016–1022. [[CrossRef](#)]
15. Keys, R. Cubic convolution interpolation for digital image processing. *IEEE Trans. Acoust. Speech Signal Process.* **1981**, *29*, 1153–1160. [[CrossRef](#)]
16. Yoon, Y.; Jeon, H.G.; Yoo, D.; Lee, J.Y.; So Kweon, I. Learning a deep convolutional network for light-field image super-resolution. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Santiago, Chile, 7–13 December 2015; pp. 24–32.
17. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
18. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
19. Lim, B.; Son, S.; Kim, H.; Nah, S.; Mu Lee, K. Enhanced Deep Residual Networks for Single Image Super-Resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 136–144.
20. Kim, J.; Lee, J.K.; Lee, K.M. Deeply-Recursive Convolutional Network for Image Super-Resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1637–1645.
21. Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. Image Super-Resolution Using Very Deep Residual Channel Attention Networks. In Proceedings of the 15th European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 286–301.
22. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Loy, C.C. Esrgan: Enhanced Super-Resolution Generative Adversarial Networks. In Proceedings of the 15th European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, 8–14 September 2018.
23. Yan, Y.; Liu, C.; Chen, C.; Sun, X.; Jin, L.; Peng, X.; Zhou, X. Fine-grained attention and feature-sharing generative adversarial networks for single image super-resolution. *IEEE Trans. Multimed.* **2021**, *24*, 1473–1487. [[CrossRef](#)]
24. Wang, Z.; Chen, J.; Hoi, S.C.H. Deep learning for image super-resolution: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 3365–3387. [[CrossRef](#)] [[PubMed](#)]
25. Liu, A.; Liu, Y.; Gu, J.; Qiao, Y.; Dong, C. Blind image super-resolution: A survey and beyond. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 5461–5480. [[CrossRef](#)] [[PubMed](#)]
26. Arefin, M.R.; Michalski, V.; St-Charles, P.L.; Kalaitzis, A.; Kim, S.; Kahou, S.E.; Bengio, Y. Multi-Image Super-Resolution for Remote Sensing Using Deep Recurrent Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 206–207.

27. Liu, F.; Yang, X.; De Baets, B. A deep recursive multi-scale feature fusion network for image super-resolution. *J. Vis. Commun. Image Represent.* **2023**, *90*, 103730. [[CrossRef](#)]
28. Zhou, D.; Duan, R.; Zhao, L.; Chai, X. Single image super-resolution reconstruction based on multi-scale feature mapping adversarial network. *Signal Process.* **2020**, *166*, 107251. [[CrossRef](#)]
29. Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; Fu, Y. Residual Dense Network for Image Super-Resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2472–2481.
30. Zhang, Y.; Zhang, K.; Chen, Z.; Li, Y.; Timofte, R.; Zhang, J.; Zhang, K.; Peng, R.; Ma, Y.; Jia, L.; et al. NTIRE 2023 Challenge on Image Super-Resolution (x4): Methods and Results. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 1864–1883.
31. Kong, D.; Gu, L.; Li, X.; Gao, F. Multi-Scale Residual Dense Network for the Super-Resolution of Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5612612. [[CrossRef](#)]
32. Qin, J.; Sun, X.; Yan, Y.; Jin, L.; Peng, X. Multi-resolution space-attended residual dense network for single image super-resolution. *IEEE Access* **2020**, *8*, 40499–40511. [[CrossRef](#)]
33. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional Block Attention Module. In Proceedings of the 15th European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
34. Wang, L.; Shen, J.; Tang, E.; Zheng, S.; Xu, L. Multi-scale attention network for image super-resolution. *J. Vis. Commun. Image Represent.* **2021**, *80*, 103300. [[CrossRef](#)]
35. Qin, X.; Gao, X.; Yue, K. Remote Sensing Image Super-Resolution Using Multi-Scale Convolutional Neural Network. In Proceedings of the 2018 11th UK-Europe-China Workshop on Millimeter Waves and Terahertz Technologies (UCMMT), HangZhou, China, 5–7 September 2018; Volume 1, pp. 1–3.
36. Li, Y.; Mavromatis, S.; Zhang, F.; Du, Z.; Sequeira, J.; Wang, Z.; Zhao, X.; Liu, R. Single-image super-resolution for remote sensing images using a deep generative adversarial network with local and global attention mechanisms. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 3000224. [[CrossRef](#)]
37. Wang, X.; Xie, L.; Dong, C.; Shan, Y. Real-Esrgan: Training Real-World Blind Super-Resolution with Pure Synthetic Data. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 1905–1914.
38. 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. [[CrossRef](#)]

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.