Journal of Physics:
Conference Series

PAPER • OPEN ACCESS

Small-target and diversity oriented underwater sonar image augmentation

To cite this article: Fengdan Wen *et al* 2025 *J. Phys.: Conf. Ser.* **3007** 012056

Journal of Physics: Conference Series **3007** (2025) 012056 doi:10.1088/1742-6596/3007/1/012056

Small-target and diversity oriented underwater sonar image augmentation

Fengdan WEN*, Shaohua JIN, Gang BIAN, Chengyang PENG

Department of Oceanography and Hydrography, Dalian Naval Academy, Dalian, Liaoning 116018, China

Email: wenfengdan@163.com

Abstract. Underwater sonar images are crucial in areas like oceanographic research for mapping the seabed and detecting resources, and in marine biology for understanding habitats. They are also important for naval and military uses such as navigation and surveillance. However, due to equipment and environmental limitations, the number of image samples is restricted, impeding further data-driven AI research. Although some works have explored data augmentation of underwater sonar images, they still face the following two problems: 1) inability to generate small-target images; 2) limited diversity of generated images. Toward this end, in this paper we propose a small-target and diversity oriented underwater sonar image augmentation method. Specifically, for small-target images, we propose to first detect and extract the target objects in the seabed sonar images, then perform scale scaling, and fuse them onto the background image using the Poisson fusion algorithm; for diverse images, we ingeniously combine mainstream image generation methods, including GAN, VAE, and Diffusion Models, using the diversity of the generative models to ensure the diversity of the generated images. Meanwhile, we design a Mixture-of-Experts (MoE) enhanced discriminator in GAN to screen the images generated by the three generative models to ensure the quality of the final augmented images. Experimental results prove that our method can effectively increase the proportion of small-target images and ensure the diversity of the augmented images, which further boost related researches based on underwater sonar images.

1. Introduction

Ocean exploration is pivotal for deepening our understanding of Earth, harnessing marine resources, and safeguarding the marine environment. With technological advancements, identifying seabed targets like sunken ships and coral reefs has become essential in oceanography, archaeology, resource exploration, and environmental monitoring[1-2]. These efforts not only unveil the ocean's mysteries but also significantly contribute to the sustainable development of the marine economy.

Marine mapping technology primarily depends on single-beam and multi-beam echosounders, as well as side-scan sonar systems. Underwater sonar images are crucial for seabed mapping and resource detection in oceanographic research[3-4], and indispensable for naval and military applications such as navigation and surveillance[5-6]. However, the scarcity of available image samples, due to equipment and environmental constraints, impedes further AI-driven research in object detection[7], semantic segmentation[8], and image classification[9-10]. Therefore, developing efficient underwater sonar image augmentation methods with minimal human intervention is of paramount importance.

Previous methods have utilized either rule-based approaches[11-12] or deep learning-based techniques[16-25] for data augmentation. Despite their promising success, these methods have overlooked two critical factors: *1) Lack of small-target samples*. In reality, due to the remote distance or limitations in equipment and environment, a significant proportion of objects in sonar images appear small. Additionally, certain categories, such as mines or debris from aircraft or ships, are inherently small. However, existing methods have neglected the augmentation of small-target samples. This may be attributed to the scarcity of available small-target samples, which increases the difficulty for current methods that require a substantial number of images for effective augmentation. *2) Neglecting the diversity of generated samples*. Mainstream underwater sonar image augmentation methods rely solely on one specific method. Although these methods can generate more samples, they inevitably result in low diversity. This means that the generated images often share specific patterns, lacking the informative content necessary for downstream tasks.

To address the above-mentioned limitations, we proposed a small-target and diversity oriented underwater sonar image augmentation method. Our method begins with a Detect-Scale-Fuse procedure designed to generate a sufficient number of samples featuring small targets. This procedure initially detects objects of normal size within the existing images. It then segments these objects and scales them down to a smaller size. Subsequently, we meticulously employ a Poisson fusion algorithm [28] to integrate the scaled-down objects with the background images. This process allows us to efficiently create a dataset with a significant proportion of small-target samples, which can serve as a robust foundation for subsequent deep learning-based augmentation. To tackle the challenge of diversity in image generation, we have integrated several leading image augmentation methods: Generative Adversarial Networks (GANs)[13], Variational Auto-Encoders (VAEs)[14], and Diffusion Models[15]. By combining these approaches, we harnessed their unique strengths to produce a wide variety of underwater sonar images. Furthermore, to ensure the quality of the final augmented images, we resorted to the discriminator in GANs to filter out low-quality images across all three augmentation methods. Essentially, The discriminator serves as a binary classifier that assesses the likelihood of an image being part of the real-world image distribution. To enhance the discriminator's capability to effectively filter out low-quality images, we meticulously designed a Mixture-of-Experts (MoE)[26-27] enhanced discriminator. The MoE enhanced discriminator incorporates multiple experts, each specialized in recognizing different types of inputs, thus improving the overall accuracy of the quality assessment. Finally, we can obtain augmented high-quality underwater sonar images that boast both a high percentage of small targets and a high degree of diversity.

Our main contribution can be summarized as follows:

- To the best of our knowledge, we are the first to address the small-target issue in underwater sonar image augmentation. We propose a Detect-Scaling-Fusion procedure to generate images containing small targets with Poisson fusion algorithm, which can serve as seed samples for subsequent generation processes.
- We pioneer the combination of three mainstream generative methods—GANs, VAEs, and Diffusion models—to ensure the diversity of synthesized samples.
- We innovatively designed a Mixture-of-Experts enhanced discriminator to filter out low-quality samples from the three generative methods, thereby ensuring the quality of the synthesized samples.

2. Related work

Generally, the sonar image augmentation methods can be classified into two groups: rule-based methods and deep learning-based methods.

2.1. Rule-based methods

In the early stages, researchers focused on designing rules for generating sonar images through techniques such as rotation, flipping, scaling, translation, blurring, noise addition, and elastic transformation on existing sample images. For instance, Nguyen et al.[11] attempted to expand the dataset by incorporating scattering, polarization, and geometric transformations. Subsequently, Huo et al.[12] enhanced the approach by combining image segmentation with the simulation of intensity distributions across different regions to generate images. While these methods successfully increased the dataset size, they were limited in their ability to produce more discriminative or informative images, as they heavily relied on transformations of existing images.

2.2. Deep learning-based methods
In recent years, the rapid evolution of deep learning has made generative methods the go-to approach for sonar image augmentation. Techniques such as GANs[13], VAEs[14], and Diffusion Models [15] have been at the forefront of this trend. Among these, GAN-based methods[16-20] have demonstrated significant success in sonar image augmentation, leveraging the adversarial dynamic between the generator and the discriminator. However, the training process for GANs is often unstable, which can hinder the reliability of even well-trained models. VAE-based methods[21-22], on the other hand, first encode input data into a latent space, typically a Gaussian distribution, and then decode samples from this latent distribution to produce synthetic images. By minimizing reconstruction error and regularizing the latent space, these methods offer a more stable training process. However, they are often criticized for producing images of lower quality. Recently, Diffusion Models[23-25] have garnered attention for their ability to generate high-quality images, offering a promising alternative in the field of sonar image augmentation.

2.3. Analysis
Although these methods have achieved success in augmenting sonar images, they often overlook the challenge of small-target detection. The objects in sonar images can be small due to factors such as equipment limitations or environmental conditions, which poses a significant challenge for subsequent detection or classification tasks. Moreover, relying solely on a single augmentation method can limit the diversity of the generated samples. In real-world application scenarios, however, diversity is essential for ensuring the robustness of downstream models. Different form the previous methods, our approach involves an initial step of generating small-target images as seed samples using rule-based techniques. We then use these seed samples to produce a variety of underwater sonar images through mainstream deep learning-based methods. Ultimately, this process yields a diverse set of high-quality underwater sonar images.

3. Methodology
In this section, we detail the proposed underwater sonar image augmentation method. The pipeline of our method is illustrated in Figure 1. This method begins with the small-target oriented seed samples acquisition, followed by diversity oriented high-quality samples augmentation.

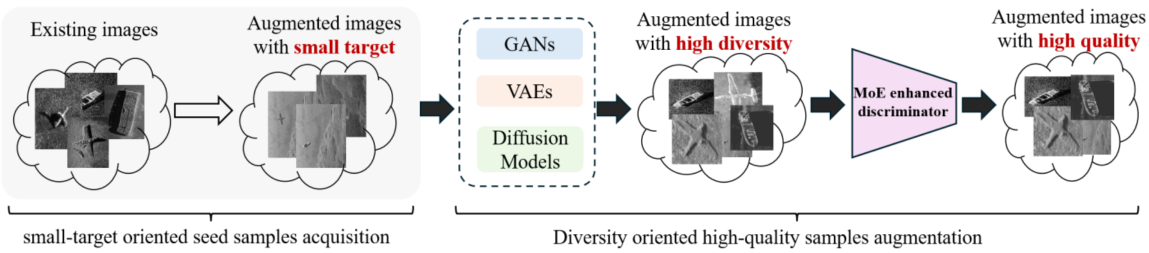

Figure 1. Pipeline of our proposed method.

3.1. Small-target oriented seed samples acquisition
To obtain images with small targets, we design a Detect-Scaling-Fusion procedure, which is illustrated in Figure 2. Specifically, we first employ the pre-trained object detection model, *e.g.*, YOLO-v8 [29-30], to detect normal-sized objects and segment them to obtain a collection. Thereafter, we perform object scaling to convert them into small size. Additionally, some rotation or transformation operations are carried out to ensure the diversity of small targets. Finally, we employ the Poisson fusion algorithm to fuse the small targets with the background to obtain the augmented images.

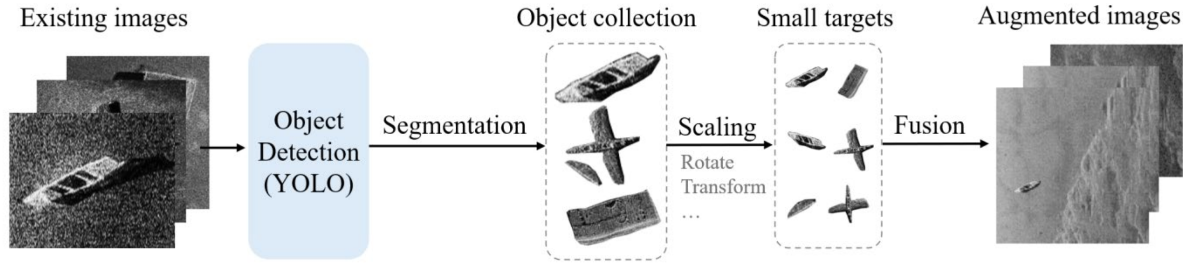

Figure 2. The proposed Detect-Scaling-Fusion procedure.

Essentially, the Poisson fusion algorithm is based on the solution of the Poisson equation, which allows for the reconstruction of a new image with smooth transitions between the small object image and the background image. The core idea of the Poisson fusion algorithm lies in minimizing the following energy function:

$$E(v) = \int_{\Omega} |\nabla v - \nabla g|^2 d\Omega \tag{1}$$

where v is the final augmented image that we want to obtain, g is the background image, and ∇ represents the gradient operator. Minimizing this energy function ensures that the gradient of the augmented image v is as close as possible to the gradient of the background image g. The Poisson equation is defined as follows:

$$\Delta v = \mathrm{div}(\nabla g) \tag{2}$$

where Δ is the Laplace operator and div is the divergence operator.
Here we introduce the detailed steps of this algorithm:

- Step1. Define a mask that indicates the region where the object will be inserted into the background. Let M be the mask, where $M(x,y) = 1$ inside the object region and $M(x,y) = 0$ outside.
- Stpe2. Calculate the gradient of the background image ∇g.
- Step3. Solve the Poisson equation $\Delta v = \mathrm{div}(\nabla g)$ subject to the boundary conditions determined by the mask M.
- Step4. Obtain the augmented image I by fusing the small target image O and the background image using the mask M. This is given as follows:

$$I(x,y) = \begin{cases} O(x,y) & \text{if } M(x,y) = 1 \\ v(x,y) & \text{if } M(x,y) = 0 \end{cases} \tag{3}$$

In this way, we can obtain natural-looking augmented images since this method takes into account the gradient information of the images. The derived images can act as a seed for the subsequent deep learning-based methods, which ensures the generation a sufficient percentage of small target images.

3.2. Diversity oriented high-quality samples augmentation
In this part, we first introduce the union of generative methods, including GANs, VAEs, and Diffusion Models, for ensuring the diversity of the augmented images. Then we elaborate on our designed MoE enhanced discriminator for maintaining high-quality images.

3.2.1. Union of generative methods.
The paradigm of the selected three generative methods is illustrated in Figure 3. Here we introduce the core of the three methods.
(1) Generative Adversarial Networks (GANs)
As shown in Figure 3(a), GANs consist of a generator and a discriminator. The core principle is to improve the generator's ability to generate realistic data through adversarial training between the

generator and the discriminator. The generator aims to produce data that is as realistic as possible. It takes a random noise vector z as input and transforms it into generated data $G(z)$ through a series of neural network layers. The discriminator's task is to distinguish between real data x and the data generated by the generator $G(z)$ and outputs a probability value indicating the likelihood that the input data is real. During training, the generator and discriminator are trained alternately. First, the discriminator is trained with the generator fixed to accurately distinguish between real and generated data. Then, with the discriminator fixed, the generator is trained to make the data it generates deceive the discriminator as much as possible. The objective function of GANs is as follows:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}\left[\log\left(1 - D\big(G(z)\big)\right)\right] \tag{4}$$

where $p_{\text{data}}(x)$ is the distribution of real data, $p_z(z)$ is the distribution of the noise vector z, $D(x)$ is the discriminator's result for real data x, and $D\big(G(z)\big)$ is the discriminator's result for generated data $G(z)$.

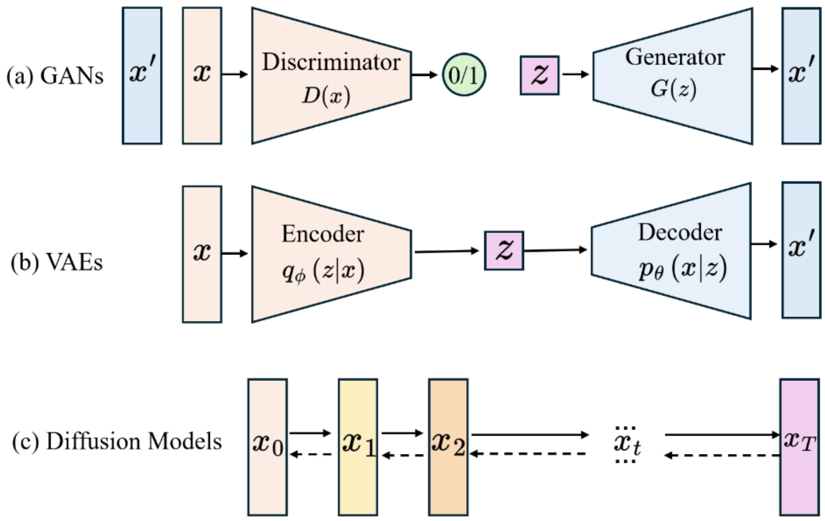

Figure 3. Paradigm of the utilized generative models, including (a) GANs, (b) VAEs, and (c) Diffusion Models.

(2) Variational Auto-Encoders (VAEs)
As shown in Figure 3(b), VAEs are generative models based on probabilistic models which encode data into a vector in the latent space and then decode the latent vector to reconstruct the data. The encoder maps the input data x to a vector z in the latent space, which follows a prior distribution, usually assumed to be a Gaussian distribution. The decoder decodes the latent vector z into reconstructed data x', attempting to recover the original input data from the latent space. The objective function of VAEs is as follows:

$$\mathcal{L}(\theta,\phi;x) = \mathbb{E}_{q_\phi(z|x)}[\log\, p_\theta(x \mid z)] - D_{KL}\big(q_\phi(z \mid x)\|p(z)\big) \tag{5}$$

where θ and ϕ are the parameters of the decoder and encoder respectively, $q_\phi(z \mid x)$ is the posterior distribution of the latent vector z given data x, $p_\theta(x \mid z)$ is the conditional distribution of data x given the latent vector z, $p(z)$ is the prior distribution of the latent vector z, and D_{KL} is the KL divergence used to measure the difference between two distributions.

(3) Diffusion Models
The core idea of diffusion models is to gradually add noise to the data and then learn how to recover the original data from the noisy data. As shown in Figure 3(c), in the forward diffusion process, starting from the original data x_0, Gaussian noise is gradually added to the data through a series of time steps t to obtain a series of noisy data x_t. The forward diffusion process can be expressed as:

$$q(x_t \mid x_{t-1}) = \mathcal{N}\left(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I\right) \tag{6}$$

where β_t is a noise coefficient that changes with time step t, and $\mathcal{N}$ represents a Gaussian distribution. In the reverse diffusion process, the goal is to learn a reverse diffusion process to gradually remove the noise from the noisy data x_T and finally recover the original data x_0. The reverse diffusion process can be modeled by a neural network, usually called a denoiser or generator $p_\theta(x_{t-1} \mid x_t)$, where θ is the parameter of the neural network. The objective function of diffusion models is as follows:

$$\begin{aligned} \mathcal{L}_{VLB} &= -\mathbb{E}_{q(x_0)} \log p_\theta(x_0) \\ &\le -\mathbb{E}_{q(x_0)} \mathbb{E}_{q(x_{1:T} \mid x_0)} \log \frac{p_\theta(x_{0:T})}{q(x_{1:T} \mid x_0)} \end{aligned} \tag{7}$$

where $q(x_{1:T} \mid x_0)$ is the joint distribution from x_0 to x_T in the forward diffusion process, and $p_\theta(x_{0:T})$ is the joint distribution from x_T to x_0 in the reverse diffusion process.
Overall, combining these methods can leverage their respective strengths and result in a more diverse and rich set of generated images.

3.2.2. MoE enhanced discriminator.
Although the three generative models are capable of producing a variety of underwater sonar images, the quality of the resulting augmented samples remains largely unoptimized. Manual selection by humans is not only time-consuming but also inefficient. Drawing inspiration from the discriminator in GANs, we meticulously design a MoE enhanced discriminator to filter out low-quality samples from the outputs of these three models. Specifically, within each hidden layer of the discriminator, we design an MoE CONV Layer. Each of these layers contains three convolutional experts, each corresponding to one of the three generative models. This can be formulated as follows:

$$\mathrm{CONV}^{MOE}(\mathbf{x}) = \sum_{i=1}^{3} \mathrm{CONV}(\mathbf{x})_i \cdot G(\mathbf{x})_i \tag{8}$$

where $\mathrm{CONV}(\mathbf{x})$ represents the convolutional operation, $\mathrm{G}(\mathbf{x})$ denotes the gating function.

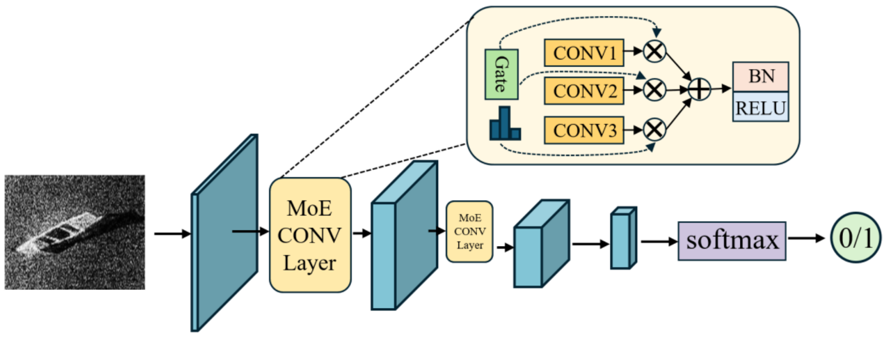

Figure 4. MoE enhanced discriminator.

The discriminator outputs a probability value p within the range $[0,1]$, which indicates the likelihood that the input data is real. This probability can also be interpreted as a measure of the sample's quality. Specifically, we predefined a threshold λ. Only if the probability value p is greater than or equal to λ will the generated samples be retained; otherwise, they will be discarded. By applying this threshold, we ensure that only high-quality augmented samples are selected for use.

4. Experiment

4.1. Experimental settings.
We employed SeabedObjects-KLSG[31] dataset as our initial dataset for augmenting underwater sonar images. This dataset contains 385 wreck, 36 drowning victim, 66 airplane, 129 mine and 578 seafloor images. Additionally, we incorporated the images obtained from our practical detection work, which include 62 airplane images, 115 fish swarm images, and 110 seafloor images. The threshold λ is set to 0.8. All experiments are conducted using PyTorch on a server equipped with a single A100-40G GPU.

4.2. Quantitative results.
We employed the commonly utilized image quality assessment metrics, namely the Fréchet Inception Distance (FID), Kernel Maximum Mean Discrepancy (MMD), and Inception Score (IS), to evaluate the augmented samples. Specifically, the FID is designed to measure the similarity between the distribution of the generated images and that of the real images. A lower FID score implies better quality and can also lead to good diversity. The MMD is a statistical test employed to ascertain whether two samples originate from different distributions. A lower MMD value indicates a high generation performance. The IS can be used to assess the diversity and clarity of the generated images. A higher IS value suggests a greater diversity of the generated samples. We compared our generated samples with those of three basic generative methods, namely GANs, VAEs, and Diffusion models, and presented the results in Table 1. It can be observed that our proposed method can ensure both the quality and the diversity of the generated images.

Table 1. Comparison of our proposed method towards FID, MMD, and IS.

Method	FID ($\downarrow$)	MMD ($\downarrow$)	IS ($\uparrow$)
GANs	115.3	0.24	3.22
VAEs	122.4	0.29	4.65
Diffusion model	92.8	0.18	4.02
Ours	85.2	0.15	5.14

4.3. Qualitative results.
Here we provided the qualitative results of our data augmentation method. The small-target oriented seed samples acquisition samples are illustrated in Figure 5, including the small target images and the fusion images.

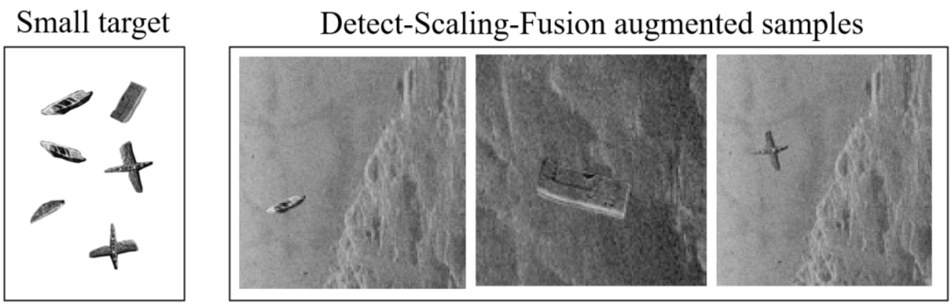

Figure 5. Small-target oriented seed samples acquisition samples.

The qualitative results of our method are presented in Figure 6, which demonstrates that the augmented images contain a relatively high percentage of small-target images characterized by high diversity and superior quality.

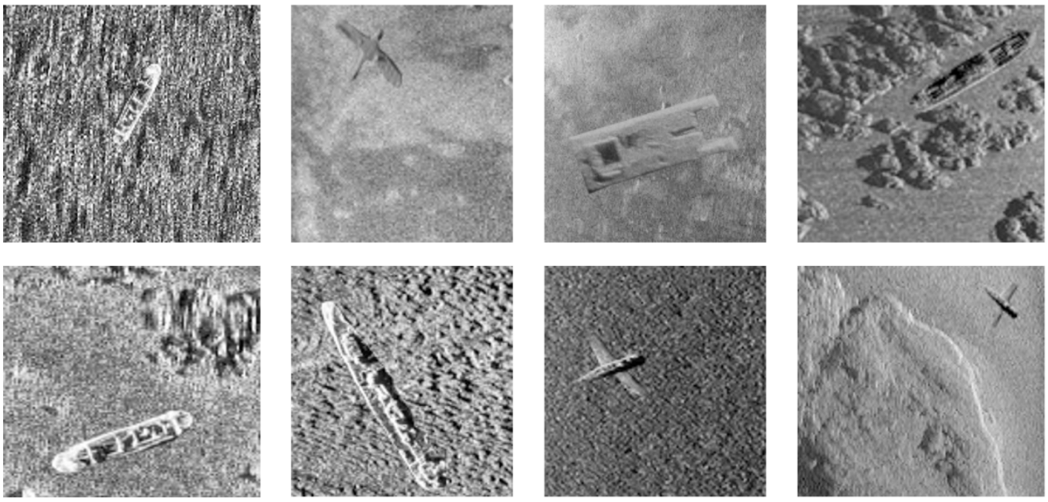

Figure 6. Final generated underwater sonar images.

5. Conclusion
In this work, we introduce an innovative method for augmenting underwater sonar images, designed to overcome the challenges associated with generating small targets and ensuring sample diversity. Our approach integrates a Detect-Scale-Fuse strategy with a combination of generative models, *i.e.*, GANs, VAEs, and Diffusion Models, resulting in a solution that significantly enhances the detection of small targets and maintains sample diversity. Furthermore, we have incorporated a MoE enhanced discriminator to ensure the quality of the generated images. Hopefully this method can provide a robust foundation for future research and practical applications in the fields of ocean exploration and surveillance.

References
[1] Fu, L., & Arko, R. A. (2014). Rapid Deployment of a RESTful Service for Data Collected by Oceanographic Research Cruises. In The Semantic Web: ESWC 2014 Satellite Events: ESWC 2014 Satellite Events, Anissaras, Crete, Greece, May 25-29, 2014, Revised Selected Papers 11 (pp. 209-212). Springer International Publishing.
[2] Pozorski, T., & Pozorski, S. (2018). Early complex society on the north and central peruvian coast: new archaeological discoveries and new insights. Journal of Archaeological Research, 26, 353-386.
[3] Zhu, P., Isaacs, J., Fu, B., & Ferrari, S. (2017, December). Deep learning feature extraction for target recognition and classification in underwater sonar images. In 2017 IEEE 56th annual conference on decision and control (CDC) (pp. 2724-2731). IEEE.
[4] Zhou, T., Si, J., Wang, L., Xu, C., & Yu, X. (2022). Automatic detection of underwater small targets using forward-looking sonar images. IEEE Transactions on Geoscience and Remote Sensing, 60, 1-12.
[5] Dos Santos, M. M., De Giacomo, G. G., Drews, P. L., & Botelho, S. S. (2020). Matching color aerial images and underwater sonar images using deep learning for underwater localization. IEEE Robotics and Automation Letters, 5(4), 6365-6370.
[6] Yu, Y., Zhao, J., Gong, Q., Huang, C., Zheng, G., & Ma, J. (2021). Real-time underwater maritime object detection in side-scan sonar images based on transformer-YOLOv5. Remote Sensing, 13(18), 3555.
[7] Karimanzira, D., Renkewitz, H., Shea, D., & Albiez, J. (2020). Object detection in sonar images. Electronics, 9(7), 1180.
[8] Rahnemoonfar, M., & Dobbs, D. (2019, July). Semantic segmentation of underwater sonar imagery with deep learning. In IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium (pp. 9455-9458). IEEE.

[9] Kumar, N., Mitra, U., & Narayanan, S. S. (2014). Robust object classification in underwater sidescan sonar images by using reliability-aware fusion of shadow features. IEEE Journal of oceanic engineering, 40(3), 592-606.
[10] Du, X., Sun, Y., Song, Y., Sun, H., & Yang, L. (2023). A comparative study of different CNN models and transfer learning effect for underwater object classification in side-scan sonar images. Remote Sensing, 15(3), 593.
[11] Nguyen, H., Lee, E., & Lee, S. (2019). Study on the Classification Performance of Underwater Sonar Image Classification Based on Convolutional Neural Networks for Detecting a Submerged Human Body †. Sensors (Basel, Switzerland), 20.
[12] Huo, G., Wu, Z., & Li, J. (2020). Underwater object classification in sidescan sonar images using deep transfer learning and semisynthetic training data. IEEE access, 8, 47407-47418.
[13] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2020). Generative adversarial networks. Communications of the ACM, 63(11), 139-144.
[14] Kingma, D. P. (2013). Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.
[15] Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. Advances in neural information processing systems, 33, 6840-6851.
[16] Jiang, Y., Ku, B., Kim, W., & Ko, H. (2020). Side-scan sonar image synthesis based on generative adversarial network for images in multiple frequencies. IEEE Geoscience and Remote Sensing Letters, 18(9), 1505-1509.
[17] Steiniger, Y., Kraus, D., & Meisen, T. (2021). Generating synthetic sidescan sonar snippets using transfer-learning in generative adversarial networks. Journal of Marine Science and Engineering, 9(3), 239.
[18] Yu, T., Chen, A., Chang, X., Xie, Y., & Liu, J. (2024, June). Research on sonar image generation technique based on improved CycleGAN. In International Conference on Remote Sensing, Mapping, and Image Processing (RSMIP 2024) (Vol. 13167, pp. 115-121). SPIE.
[19] Zhou, J., Li, Y., Qin, H., Dai, P., Zhao, Z., & Hu, M. (2024). Sonar image generation by MFA-CycleGAN for boosting underwater object detection of AUVs. IEEE Journal of Oceanic Engineering.
[20] Yu, M., Ye, X., Chen, A., & Wang, Z. (2023, October). Simulated Sonar Image Generation Method Based on Improved CycleGAN. In 2023 IEEE 11th International Conference on Computer Science and Network Technology (ICCSNT) (pp. 281-285). IEEE.
[21] Panisilvam, J., Castillón, M., Lawrance, N., & Siegwart, R. (2022, October). Conditioned deep feature consistent variational autoencoder for simulating realistic sonar images. In OCEANS 2022, Hampton Roads (pp. 1-5). IEEE.
[22] Xenaki, A., Gips, B., & Pailhas, Y. (2022). Unsupervised learning of platform motion in synthetic aperture sonar. The Journal of the Acoustical Society of America, 151(2), 1104-1114.
[23] Yang, Z., Zhao, J., Zhang, H., Yu, Y., & Huang, C. (2023). A side-scan sonar image synthesis method based on a diffusion model. Journal of Marine Science and Engineering, 11(6), 1103.
[24] Zhang, F., Hou, X., Wang, Z., Cheng, C., & Tan, T. (2024). Side-Scan Sonar Image Generator Based on Diffusion Models for Autonomous Underwater Vehicles. Journal of Marine Science and Engineering, 12(8), 1457.
[25] Natarajan, P., Basha, K., & Nambiar, A. (2024). Synth-SONAR: Sonar Image Synthesis with Enhanced Diversity and Realism via Dual Diffusion Models and GPT Prompting. arXiv preprint arXiv:2410.08612.
[26] Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). ªAdaptive Mixtures of Local Experts, º Neural Computation, vol. 3.
[27] Chen, Z., Deng, Y., Wu, Y., Gu, Q., & Li, Y. (2022). Towards understanding the mixture-of-experts layer in deep learning. Advances in neural information processing systems, 35, 23049-23062.
[28] Sun, J., Zhu, H., Xu, Z., & Han, C. (2013). Poisson image fusion based on Markov random field fusion model. Information fusion, 14(3), 241-254.
[29] Ma, Q., Jin, S., Bian, G., & Cui, Y. (2024). Multi-Scale Marine Object Detection in Side-Scan Sonar Images Based on BES-YOLO. Sensors, 24(14), 4428.
[30] Hussain, M. (2023). YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection. Machines, 11(7), 677.

[31] Huo, G., Wu, Z., & Li, J. (2020). Underwater object classification in sidescan sonar images using deep transfer learning and semisynthetic training data. IEEE access, 8, 47407-47418.