MetaSegNet: An Edge Computing-Oriented Meta-Learning Method for Few-shot Sonar Image Segmentation

First Author
Institution1
Institution1 address
firstauthor@i1.org

Second Author
Institution2
First line of institution2 address
http://www.author.org/~second

Abstract

The complexity of the underwater environment leads to difficulties in acquiring sonar image samples. Due to the limited availability of data, traditional neural networks often struggle to achieve accurate image segmentation. Therefore, few-shot learning algorithms need to be developed to address the scarcity of annotated sonar image samples and enhance the accuracy of underwater acoustic image segmentation. In this paper, we propose MetaSegNet, a few-shot segmentation model for sonar images based on meta-learning. Our method integrates an Adaptive Scale and Shift (A2S) module and a Channel and Spatial Skip Attention Gate (CSSAG) mechanism to enhance segmentation performance and support generalization across diverse few-shot scenarios. Experimental results show that our algorithm consistently outperforms mainstream networks on public sonar datasets. In addition, we have designed a customized forward-looking sonar system as an embedded edge computing platform, with the Atlas 200 chip serving as its core component. By deploying the algorithm, collecting data, and conducting segmentation tests in Qiandao Lake, Hangzhou, Zhejiang, China, our system achieved a favorable balance between energy efficiency and segmentation performance. The forward-looking sonar system demonstrated satisfactory segmentation accuracy and real-time capability in practical engineering applications.

***Keywords:** Sonar images, Image segmentation, Few-shot learning, Edge computing, Embedded systems.*

1. Introduction

Image segmentation is a crucial step in sonar image analysis [21, 4, 31]. Through segmentation, target objects in a sonar image can be separated from the background, allowing for accurate localization of their positions and shapes. This process provides the basis for subsequent tasks such as target identification, classification, shape analysis, and motion analysis [43, 29].

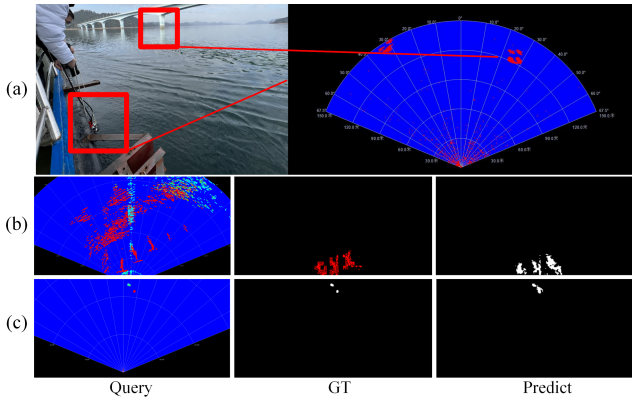


Figure 1. Forward-looking sonar equipment testing and imaging results. (a) Testing environment. (b) Fishing boat moored in the harbor. (c) Float ball.

The increasing demand for ocean-related activities such as bathymetric mapping and environmental modelling [36], underwater salvage and rescue [9], shipwreck detection [25], and the identification of underwater threats like torpedoes and submarines [26], underscores the strategic importance of sonar equipment in both civilian and military domains. As maritime territories play a vital role in national security and global stability, sonar-based technologies have become essential for resource exploration and defense operations. Light waves are significantly attenuated in seawater, making them unsuitable for long-distance transmission [35, 38]. In contrast, sound waves propagate with minimal attenuation, enabling efficient underwater transmission. Consequently, sonar has become the primary technique for underwater target imaging. Due to the changes in angle and velocity caused by the influence of seawater on the sonar, the collected data often does not represent the true echo strength of the seabed landform, and parsing the data often produces various aberrations [1, 42, 18], and it is easy to miss small targets [14], which significantly impact the subsequent processing of sonar images [33, 39, 45].

To address these challenges, deep learning methods have

been widely applied in sonar image segmentation [32, 41]. The U-Net model proposed by Ronneberger *et al.* [22] makes it a suitable network architecture for sonar image segmentation by virtue of its ability to capture contextual information, its skip-connection mechanism, its ability to fuse multi-scale features, and its robustness to labeled data [27]. However, the disadvantage of deep learning is that it requires a large amount of labeled training data to obtain good results [8]. Acquisition of sonar image data is difficult. Meanwhile, data acquisition and sharing are limited in cases where sonar image data involves sensitive information. The scarcity of sonar image data raises a broader issue of few-shot learning [30]. Due to the limited nature of sonar image data, models must be trained and adapted using only small amounts of data to generalize to new samples [12, 37]. Given limited computing resources, energy efficiency, and real-time requirements in marine edge environments [34], deploying suitable segmentation models on embedded edge devices is essential. In response to the challenges posed by limited training data and constrained deployment resources, meta-learning has emerged as an effective paradigm for enabling rapid adaptation to new tasks under few-shot settings. For this purpose, our approach designed a lightweight meta-learning framework for real-time sonar image segmentation on embedded edge devices. In addition, the paper's application of meta-learning to few-shot sonar image segmentation has the following two advantages: through meta-learning algorithms, adaptive optimization algorithms can be designed to automatically learn the optimal parameter configurations and model selection strategies for a given sonar task and dataset, improving the segmentation of sonar image data. Moreover, by pre-training on large-scale datasets, the model convergence can be accelerated and overall performance is improved.

In summary, our main contributions are listed as follows.

- To the best of our knowledge, we first introduced a meta-learning-based network model into the application of few-shot sonar image segmentation, and developed MetaSegNet, which effectively addresses the challenges that traditional neural networks face in limited-sample scenarios.

- We designed a multi-scale attention module (Channel and Spatial Skip Attention Gate, CSSAG) to effectively integrate sonar features across scales, and designed an adaptive factor-based parameter modulation method (Adaptive Scale and Shift, A2S) to enhance model generalization and adaptability under few-shot learning scenarios.

- We developed a custom-designed forward-looking sonar as an embedded edge computing platform based on the Atlas 200 chip. The model was deployed for real-time segmentation in real few-shot sonar image scenarios, as shown in Figure 1. Experimental results demonstrate strong adaptability and superior performance compared to other platforms.

2. Related Work

The goal of sonar few-shot image segmentation is to accurately segment sonar images using only a small number of samples. The challenge lies in the fact that traditional neural networks struggle to generalize under such limited data conditions. Few-shot segmentation methods can generally be categorized into three types: sample expansion-based methods, memory augmented-based methods, and parameter update-based methods.

2.1. Sample Expansion-Based Methods

The sample expansion model increases effective support samples by training a generation module to create similar, synthetic data, improving query recognition. For example, Antoniou *et al*. [2] used GANs to synthesize additional samples, while Sandfort *et al.* [23] applied CycleGAN to generate synthetic medical images, enhancing segmentation across domains. Liu *et al.* [16] expanded datasets via basic transformations. Recently, Zhang *et al.* [40] proposed a generative framework using segmentation-guided multi-level optimization to synthesize image-mask pairs, effectively augmenting data in ultra low-data scenarios.

2.2. Memory Augmented-Based Methods

Traditional neural networks require iterative training with large amounts of data and must relearn internal parameters when encountering new data, making them inefficient for few-shot scenarios. Memory augmented-based methods enable rapid reasoning by retrieving information from external storage. Google first introduced the memory-augmented neural network model, which integrates read and write modules into an external memory component and stores information based on the least-recently-used principle [24]. However, this approach suffers from information forgetting. To alleviate memory forgetting, Cai *et al.* proposed MM-Net [3], which compresses memory access into compact features for reasoning. Building on this, Cheng *et al.* [5] introduced a disentangled representation framework with implicit memory structures, boosting generalization and consistency across tasks.

2.3. Parameter Update-Based Methods

The parameter update-based model primarily utilizes meta-learning to refine model parameters using support sets. Meta-learning is a machine learning approach designed to enable algorithms to "learn how to learn", thereby improving their efficiency and flexibility in adapting to new

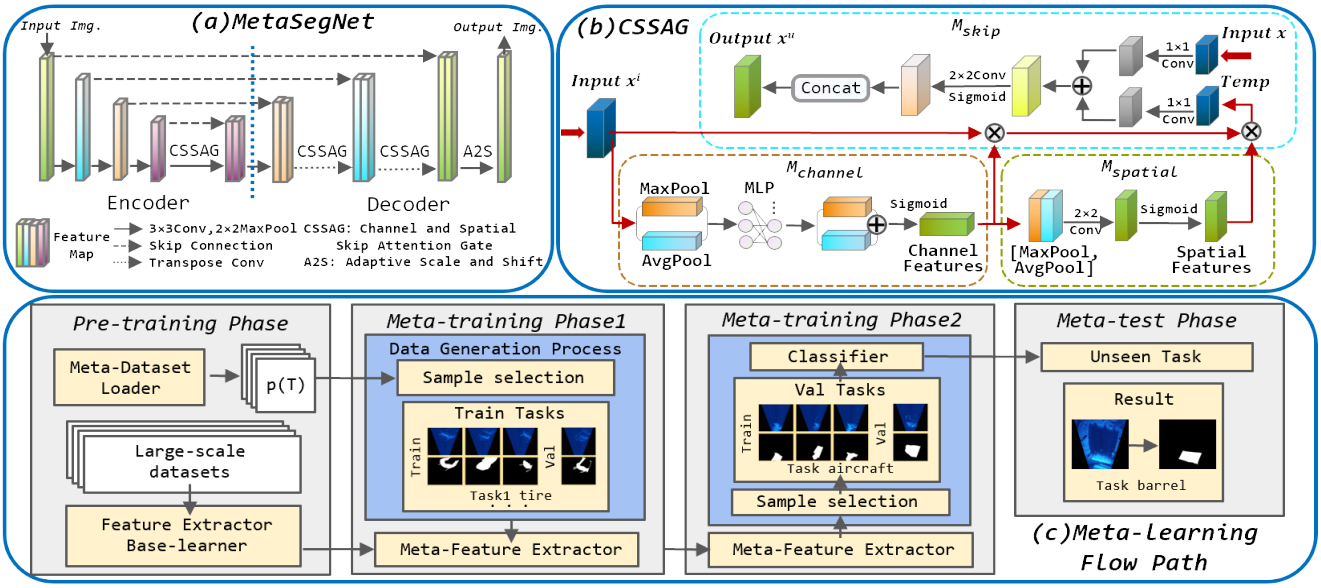


Figure 2. The overview of our method and process. (a) MetaSegNet framework including the improved U-Net architecture, CSSAG attention module and A2S module. (b) CSSAG attention module. (c) Meta-learning flow path.

tasks by leveraging experience and knowledge gained from multiple tasks [11, 13]. Model-Agnostic Meta-Learning (MAML) [7] is a model-agnostic algorithm that trains models to quickly adapt to new task data through iterative training on related tasks, resulting in improved performance during testing; however, it is prone to overfitting due to inconsistencies between training and testing data. To address this, Sun *et al.* introduced Meta-Transfer Learning (MTL) [28], which uses transfer learning features to improve classification accuracy while effectively mitigating overfitting in few-shot scenarios. Li *et al.* proposed Few-Shot Segmentation (FSS) [15], which achieves comparable accuracy with small datasets by training the network from scratch without pre-trained weights, unlike approaches relying on large-scale pre-training. Hong *et al.* introduced Volumetric Aggregation with Transformers (VAT) [10], which employs a pyramid encoder structure to allow coarse-level aggregation to guide finer-level feature learning and complementary matching scores, achieving excellent performance on public datasets.

Recently, parameter adaptation strategies have been explored in sonar imagery, where severe acoustic noise and limited labeled data pose significant challenges. For example, Zhu *et al.* [46] proposed a coarse-to-fine framework for few-shot sonar object segmentation, highlighting the difficulty of accurate contour delineation in cluttered underwater environments. In parallel, lightweight vision models for edge deployment, such as MobileViT [44], have been investigated in few-shot or self-supervised segmentation. However, these transformer-based approaches typically rely on sufficient data diversity and stable feature representations, which remain challenging in forward-looking sonar imagery. Therefore, our work focuses on parameter-level adaptation within a convolutional backbone to balance noise robustness and computational efficiency for edge-oriented few-shot sonar segmentation.

3. Method

This paper adopts the improved U-Net architecture as the backbone of MetaSegNet, which consists of a feature extractor enhanced by a Channel and Spatial Skip Attention Gate (CSSAG) module and an adaptive meta-training mechanism based on the Adaptive Scale and Shift (A2S) method. The model's encoder uses standard convolutional and pooling layers to extract hierarchical features, while the decoder integrates CSSAG to improve multi-scale feature fusion through attention-based skip connections, enabling more precise segmentation results. The overall meta-learning training flow path is illustrated in Figure 2(c). The entire network will be deployed on the Atlas 200 edge computer, integrated into the self-developed forward-looking sonar device to achieve real-time edge computing in marine environments.

3.1. Meta-Learning Feature Extractor Design

Based on the MetaSegNet architecture, we improved its mechanism of skip connections, as shown in Figure 2(a). The model's encoder employs a 3x3 convolutional layer with ReLU activation, followed by 2x2 max-pooling to gradually downscale feature maps. The decoder inputs the vector output from the previous layer of the model and its corresponding coding part output into the CSSAG attention

mechanism module, concatenating the CSSAG output with the corresponding encoder output, followed by convolution and deconvolution for upsampling. Different from applying attention on backbone feature maps, CSSAG is integrated into the skip connections of the encoder–decoder architecture. The attention gate selectively filters encoder features before fusion with decoder representations, which is particularly important for suppressing noise-dominated high-resolution features in forward-looking sonar images.

The CSSAG attention mechanism is introduced in the decoder, as shown in Figure 2(b), where x and x_i denote the skip-connected output of the i-th layer of the network structure and output of the previous layer of the model, respectively. x and x_i are first weighted and summed by a 1×1 convolutional kernel and then processed with the ReLU activation function, a 1×1 convolution, and the sigmoid activation function, finally the Hadamard product with x again to obtain the output x^u, as shown in Eq. (1).

$$x^u = x \odot \sigma\left(\text{Conv}_{1\times1}\left(\text{ReLU}\left(\text{Conv}_{1\times1}(x + x_i)\right)\right)\right) \quad (1)$$

Unlike CBAM [6], which applies channel and spatial attention sequentially to feature maps, CSSAG integrates attention into skip connections, allowing direct feature fusion between encoder and decoder layers. This design preserves fine-grained details while suppressing noise in sonar images. We did not adopt lightweight ViT variants such as MobileViT or EdgeViT [17, 19] although recent lightweight vision transformers are promising for edge deployment, because the limited sonar dataset and the small image resolution (224×224) do not provide enough data diversity for transformer-based patch modeling. Moreover, the convolution-based CSSAG yields better stability under high-noise marine imagery.

3.2. Adaptive Factor-Based Network Parameter Training Method

This section discusses instance segmentation using meta-transfer learning. To further enhance the model's adaptability under few-shot settings, we proposed A2S, a method for dynamic weight adjustment in a meta-learner via tunable convolutional kernels. It enhances model generalization and adaptability, tackling inaccuracies in sonar image segmentation by pre-training and refining model parameters for better robustness. Unlike traditional parameter optimization that seeks a single optimal set of parameters, A2S learns task-adaptive scaling and offset parameters through a meta-learning process, enabling rapid adaptation with minimal updates.

Different from feature-wise modulation methods such as FiLM and weight scaling strategies in MTL, A2S performs task-conditioned adaptation directly on convolutional weights and biases. This parameter-level modulation embeds task-specific information into the model parameters,

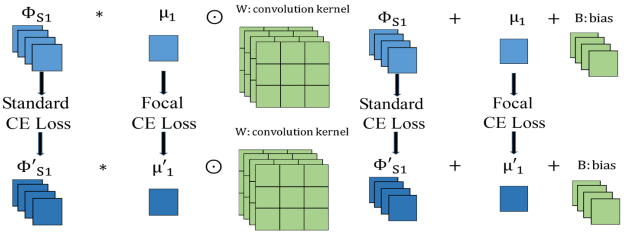


Figure 3. A2S: Adaptive Scale and Shift on network parameters.

resulting in more stable adaptation under strong noise and limited supervision. The A2S module works by decomposing training into pretraining and meta-learning phases. Firstly, a pre-training phase is performed on a large-scale dataset to obtain generalizable feature representations. This serves as the initialization for the segmentation model. Then, transfer learning is applied by fine-tuning the model on the target dataset to improve its performance on sonar image segmentation tasks, as shown in Eq. (2):

$$[\Theta;\theta] = [\Theta;\theta] - \alpha\nabla\mathcal{L}_D([\Theta;\theta]) \quad (2)$$

where Θ is the backbone feature extractor, θ is the concatenation of encoder and classifier parameters, $\mathcal{L}_D$ is the loss function, and α is the optimized learning rate. The formulation of the loss function is as follows:

$$\mathcal{L}_D([\Theta;\theta]) = \frac{1}{|D|}\sum_{(x,y)\in D} L(f_{[\Theta;\theta]}(x), y) \quad (3)$$

Here x denotes a given picture, y denotes the label of the corresponding input picture, $f_{[\Theta;\theta]}$ is the segmentation function, the corresponding output is the segmentation result of the input picture x, and L is the cross-entropy loss calculation function. The pretrained segmenter can be obtained using using the gradient descent method as described in Eq. (3).

Post-pre-training, feature extraction parameters are frozen while the classifier part is retrained. The meta-training phase is divided into two stages: a test task $T^{(te)}$ and a training task $T^{(tr)}$, where the outer loop requires scaling and shifting of the feature extractor, and the inner loop requires fine-tuning of the classifier and updating the adaptive factors. During the outer loop stage, the learnable parameters $\Phi_{S_{\{1,2\}}}$ are assigned to each set of convolution and bias, here the learnable parameter Φ_{S_1} of the convolution kernel is initialized to 1, while the learnable parameter Φ_{S_2} of bias is initialized to 0. In the outer loop phase, a temporary classifier $\theta^{'}$ is first set up, and the update formula for $\theta^{'}$ is as follows:

$$\theta^{'} \leftarrow \theta - \beta\nabla_\theta\mathcal{L}_{T^{(tr)}}([\Theta;\theta], \Phi_{S_{\{1,2\}}}) \quad (4)$$

Here θ denotes the current classifier parameters, β denotes the learning rate of the outer loop, and $\mathcal{L}_{T^{(tr)}}$ denotes the

standard cross-entropy loss on the training task, as shown in Eq. (5).

$$\mathcal{L}_{T^{(tr)}}([\Theta;\theta],\Phi_{S_{\{1,2\}}}) = \frac{1}{T^{(tr)}}\sum_{(x,y)\in T^{(tr)}} L(f_{[\Theta;\theta;\Phi_{S_{\{1,2\}}}]}(x),y) \quad (5)$$

Here $\Phi_{S_{\{1,2\}}}$ denote the inner-loop hyperparameters. We optimize the temporary classifier $\theta^{'}$ using the loss of the training task $T^{(tr)}$, and update the hyperparameters $\Phi_{S_{\{1,2\}}}$ with the adaptive factor $\mu_{1,2}$:

$$\Phi_{S_i} = \Phi_{S_i} - \eta\nabla_{\Phi_{S_i}}\mathcal{L}_{T^{(te)}}([\Theta;\theta^{'}],\Phi_{S_{\{1,2\}}}) \quad (6)$$

$$\mu_i = \mu_i - \eta\nabla_{\mu_i}\mathcal{L}_{FL}([\Theta;\theta^{'}],\mu_{\{1,2\}}) \quad (7)$$

where Φ_{S_i} is the learnable parameter in the convolution or bias, η is the inner-loop learning rate, and $\mathcal{L}_{T^{(te)}}$ denotes the standard cross-entropy loss computed by the classifier $\theta^{'}$ on the test task. These updates aim to dynamically adjust the convolution kernel and bias to adapt to task-specific distributions. To address class imbalance in sonar image segmentation, we incorporate the focal loss, the focal loss function $\mathcal{L}_{FL}$ as follows:

$$\mathcal{L}_{FL}([\Theta;\theta],\mu_{\{1,2\}}) = -\frac{1}{N}\sum_{i=0}^{N}(1-p_i)^{\gamma}\log(p_i) + p_i^{\gamma}\log(1-p_i) \quad (8)$$

where N is the number of samples, p is the probability of accurate prediction, and $\gamma > 0$ is an adjustable factor. In the inner loop, Φ_{S_i} is modulated by the adaptive factors through scaling and shifting operations, the convolution kernel and bias are adjusted, the original convolution kernel is used to perform the Hadamard product operation with Φ_{S_1}, and the original bias is used to add with Φ_{S_2}, as shown in Figure 3.

$$A2S(X;W,b;\Phi_{S_{\{1,2\}}};\mu_{\{1,2\}}) = W\odot(\Phi_{S1}*\mu_1))X + (b+(\Phi_{S2}+\mu_2)) \quad (9)$$

where X is the input feature map, W is the convolution kernel, and b is bias. The operator $\odot$ denotes the Hadamard product (element-wise multiplication), while $*$ represents scalar multiplication, used here to scale the learnable parameters. Specifically, the convolution kernel W is adaptively rescaled via an element-wise Hadamard product with the scaled parameter $\Phi_{S_1} * \mu_1$, where Φ_{S_1} is a learnable mask and μ_1 is a tunable scaling factor. Similarly, the bias b is adaptively shifted by the sum $(\Phi_{S_2} + \mu_2)$, allowing for flexible adjustment of the bias term. The final update formula for the classifier θ is obtained in the inner loop as follows:

$$\theta = \theta - \eta\nabla_{\theta}\mathcal{L}_{T^{(te)}}([\Theta;\theta^{'}],\Phi_{S_{\{1,2\}}}) \quad (10)$$

where $\mathcal{L}_{T^{(te)}}$ denotes the loss computed by the classifier $\theta^{'}$ on the test task, and $\Phi_{S_{\{1,2\}}}$ represents the updated scaling and shifting hyperparameters. The desired meta-classifier can be obtained by utilizing adaptive shifting and scaling through the above steps.

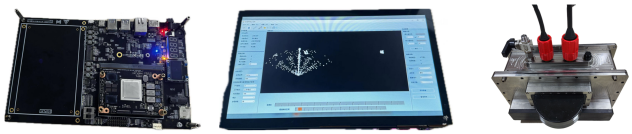
During inner-loop adaptation, only the task-specific A2S parameters are updated, while the backbone parameters remain fixed. In the outer loop, backbone parameters and CSSAG parameters are optimized using query set gradients. Different loss functions are applied to different parameter groups according to their optimization objectives. Specifically, the shared backbone parameters are optimized using cross-entropy loss to ensure stable convergence across tasks, while focal loss is applied to task-adaptive parameters to emphasize hard-to-segment regions and alleviate class imbalance commonly observed in sonar images.

Although A2S shares conceptual similarity with FiLM [20] and MTL, its role is not to replace feature modulation, but to provide a lightweight and stable parameter adaptation mechanism suitable for few-shot sonar segmentation and edge deployment. This enables the model to adjust task-specific representations at the parameter level, enhancing generalization with fewer updates.

3.3. Image Segmentation Hardware Device for Edge Computing

This section presents the design and implementation of the hardware architecture tailored for few-shot sonar image segmentation in an edge computing environment. The system integrates an underwater sonar device with an embedded host computer connected via Gigabit Ethernet, enabling real-time on-site processing.

The embedded host computer integrates an Atlas 200 processor, an AI chip optimized for edge computing, and a touch screen interface, serving as the core computing platform. It receives video streams or images from the underwater sonar device, executes real-time segmentation inference using the MetaSegNet model, and displays the segmentation results locally on the touch screen. This design ensures low-latency processing and reduces the need for bandwidth-intensive data transmission to remote servers.



(a) (b) (c)

Figure 4. The overview of edge computing platform. (a) Signal processing board. (b) Embedded host computer. (c) Underwater sonar equipment.

Figure 4 illustrates the system components: (a) the signal processing board built around the Atlas 200 edge AI

processor, (b) the embedded host computer with integrated display, and (c) the underwater sonar device responsible for beamforming and echo acquisition. The embedded host's Gigabit Ethernet interface receives sonar data, which it parses and preprocesses before feeding into the segmentation network. The underwater sonar unit, depicted in Figure 4(c), consists of transmitting and receiving transducers integrated within a watertight cable assembly, facilitating reliable underwater operation and data acquisition.

The segmentation model is pre-trained offline on a server and exported in ONNX format. It is then converted to the Atlas 200-compatible OM format via the ATC tool for edge deployment. During inference, the embedded host normalizes input images, runs them through MetaSegNet, and renders the segmentation outputs on the screen.

By leveraging the Atlas 200 edge AI module, the system achieves fast inference, low power consumption, and high portability—ideal for field-deployed, real-time sonar image segmentation applications.

4. Experiments

4.1. Qiandao Lake Test

To evaluate the real-world performance of the segmentation model, we conducted field experiments at Qiandao Lake using a self-developed forward-looking sonar device, shown in Figure 1(a). The device operates at 350 kHz with a 25 m detection range, 135° field of view, and 512 beams. During vessel movement, it captured one image every three seconds, with targets such as the vessel and float visible, under significant and irregular noise interference.

Due to the limited availability of labeled data, we used feature maps of piers and dikes as the training set, selecting five images per class as the support set. Figure 1(b)(c) illustrates the segmentation results, demonstrating that the proposed method performs robustly even under poor image quality and limited supervision, showing promise for future few-shot segmentation in sonar imagery.

Device	Inference Time(ms)	FPS (Hz)	Power (W)	FPS/Power (Hz/W)
X86 CPU (i9-10900X)	58.2	17	42	0.404
GPU (2080Ti)	**8.3**	**119**	194	0.613
Atlas 200DK (Ascend 310)	151	6.6	**8.3**	**0.795**

Table 1. Performance Comparison of Different Hardware Platforms.

To evaluate the effectiveness of deploying the proposed segmentation algorithm in an edge computing scenario, we conducted comparative experiments on three hardware platforms: CPU, GPU, and the Atlas 200-based embedded device. We measured inference time, power consumption, and calculated the frame rate per watt (PFS). As shown in Table 1, the Atlas 200 platform achieved a processing speed of 6.6 frames per second with the lowest power consumption per frame, demonstrating superior energy efficiency.

These results validate the suitability of the proposed model for edge deployment. When integrated into the self-developed device equipped with the Atlas 200 edge AI processor, the model maintains high inference performance while significantly reducing power requirements. This confirms the practicality of real-time, low-power sonar image segmentation in maritime edge environments using embedded devices.

4.2. Datasets

Forward-looking sonar using echo imaging inevitably encounters scattering noise. This paper employs a data enhancement method to improve training within the network. In this study, the original sonar image undergoes preprocessing, and an adaptive three-dimensional block-matched filter is applied to remove the scattering noise from the image, which was ultimately used in our method and baseline methods. The denoising results are depicted in Figure 5.

Pre-training Dataset used in this study is the publicly available COCO-Stuff 10K dataset v1.1, available at https://github.com/yaoxinthu/cocostuff, which was adopted for the pre-training experiments. It contains 10,000 complex natural images, each annotated with five labeled objects.

Marine Dataset is a collection of sonar images captured using an ARIS Explorer 3000 forward-looking sonar device, available at http://www.soundmetrics.com. Video data were sampled at 1 frame per second and manually annotated using LabelMe. Each image is 360×360 pixels, and we extracted corresponding types of underwater targets with typical representativeness, including barrels, divers, sharks, propellers, salmon. For computational efficiency, all images were resized to 224×224 pixels.

4.3. Comparative Experiments on Segmentation

Few-shot segmentation experiments on the sonar image dataset compared MAML, FSS, VAT, MTL, and the proposed method, using single labeled images as the support set. As shown in Table 2, the proposed method achieves the highest average mIoU, improving performance by 4.6%, 3.9%, 1.6%, and 4.0% compared to MAML, FSS, VAT, and MTL, respectively.

Although transformer-based few-shot segmentation methods have achieved remarkable results in large-scale datasets, their heavy parameterization and dependence on abundant data make them less suitable for limited sonar

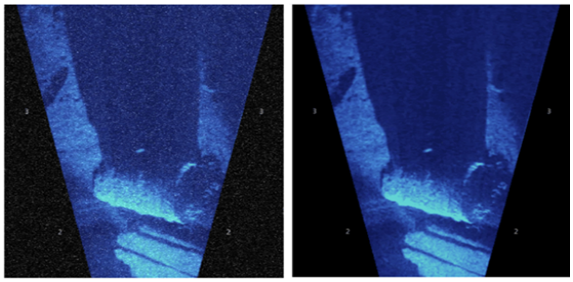


Figure 5. Preprocessing results of sonar images. (a) Pre-filter. (b) Filtered.

samples. Hence, we focus on meta-learning-based convolutional frameworks for better adaptability under low-data conditions. Following common practice in few-shot segmentation, we report mIoU as the primary evaluation metric. For forward-looking sonar images, mIoU is particularly sensitive to clutter-induced false positives and missed detections of small targets, which aligns with the key challenges addressed in this work.

Segmentation results for all methods are visualized in Figure 6. Because there may be echoes with the same intensity between different locations of different objects in the sonar image and the few-shot network needs to categorize and learn the features of different types of objects and then fit them to the query set, the segmentation results usually have false detection regions. Compared to other few-shot methods, the proposed algorithm demonstrates fewer misdetected regions.

Method	MAML [7]	FSS [15]	VAT [10]	MTL [28]	**Ours**
Barrel	72.9%	66.8%	72.4%	82.7%	**83.5%**
Divers	49.8%	58.0%	**64.7%**	53.3%	53.8%
Shark	44.0%	62.1%	60.2%	46.3%	**62.4%**
Propeller	83.4%	85.9%	90.0%	89.0%	**90.5%**
Salmon	**74.8%**	54.6%	51.1%	56.3%	57.7%
mIoU	64.9%	65.6%	67.9%	65.5%	**69.5%**

Table 2. Marine sonar image segmentation results

4.4. Ablation Experiment

CSSAG and A2S address different sources of performance degradation in few-shot sonar segmentation. CSSAG operates at the feature level by selectively gating encoder representations before skip fusion, which helps suppress clutter-induced activations and preserves fine-grained spatial details, particularly benefiting small or weakly bounded targets under strong acoustic noise.In contrast, A2S performs task-specific modulation at the parameter level by adapting convolutional weights and biases using scale-and-shift coefficients. This enables rapid model adaptation to varying target appearances and background statistics across few-shot tasks.

Therefore, CSSAG mainly improves spatial feature qual-

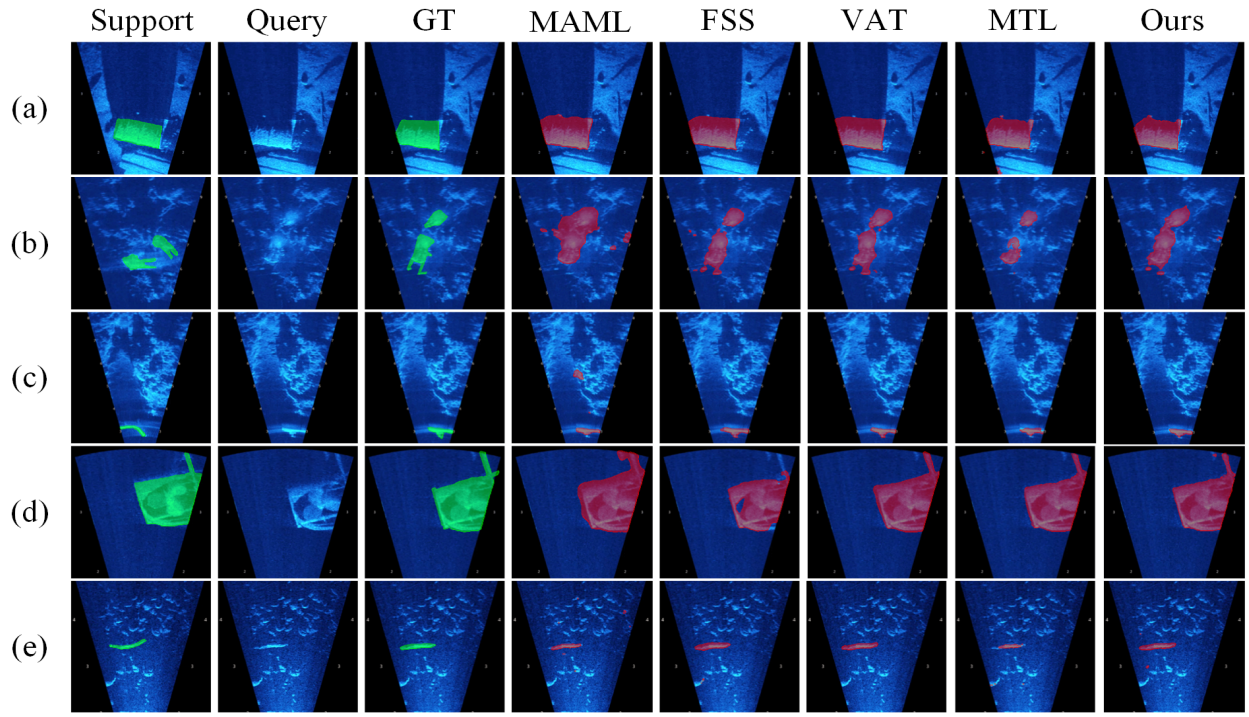


Figure 6. Segmentation results of marine sonar images. (a) Barrel. (b) Divers. (c) Shark. (d) Propeller. (e) Salmon.

ity and noise robustness, while A2S enhances task-level generalization through efficient parameter adaptation. This complementary behavior is also reflected in the ablation results in Table 3, where CSSAG contributes more to structural accuracy, while A2S improves cross-task stability. Their combination allows MetaSegNet to jointly handle scale variation and noise interference in forward-looking sonar imagery.

To validate the effectiveness and generalization of the CSSAG and A2S modules, we integrated them into the MTL network for comparison. As shown in Table 3, integrating A2S and CSSAG into the MTL backbone yields consistent improvements, and their joint deployment achieves the highest performance, confirming their complementary contributions.

Backbone Feature	Methods	mIoU In Water Tank
U-Net	MTL	65.5%
U-Net	MTL + A2S	66.8%
U-Net + CSSAG	MTL	68.3%
U-Net + CSSAG	MTL + A2S	**69.5%**

Table 3. Segmented performance metrics for each model

5. Conclusion

This paper proposes a few-shot sonar image segmentation method based on meta-learning, an attention mechanism, and adaptive factors. The model enhances the baseline segmentation backbone by integrating attention mechanisms into skip connections and preserving multi-scale features tailored to front-view sonar images, thereby improving segmentation accuracy. A meta-learning module with an adaptive factor is incorporated during training using a composite loss function, enabling the model to better adapt to few-shot segmentation tasks. Experimental results on datasets collected with the edge sonar equipment demonstrate that the proposed algorithm achieves superior accuracy and stronger adaptability, effectively meeting the specific requirements of sonar image segmentation. Furthermore, the algorithm has been deployed on the Atlas 200 embedded system platform, providing an efficient and cost-effective solution for practical sonar image processing and segmentation. Future work will explore larger-scale and more diverse sonar datasets to further evaluate the robustness of MetaSegNet under complex underwater environments. We hope this work inspires and assists researchers in the field of sonar image analysis.

References

[1] A. Abu and R. Diamant. Unsupervised local spatial mixture segmentation of underwater objects in sonar images. *IEEE Journal of Oceanic Engineering*, 44(4):1179–1197, 2019. 1

[2] A. Antoniou, A. Storkey, and H. Edwards. Data augmentation generative adversarial networks, 2018. 2

[3] Q. Cai, Y. Pan, T. Yao, C. Yan, and T. Mei. Memory matching networks for one-shot image recognition. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4080–4088, 2018. 2

[4] W. Cai, J. Zhu, and M. Zhang. From classical approach to deep learning: A review on underwater target segmentation with sonar image. *Neurocomputing*, 637:130087, 2025. 1

[5] H. Cheng, Y. Wang, H. Li, A. C. Kot, and B. Wen. Disentangled feature representation for few-shot image classification. *IEEE Transactions on Neural Networks and Learning Systems*, 35(8):10422–10435, 2024. 2

[6] J. Dong, N. Wang, H. Fang, Y. Shen, B. Li, D. Di, and K. Zhai. Cbam-optimized automatic segmentation and reconstruction system for monocular images with asphalt pavement potholes. *IEEE Transactions on Intelligent Transportation Systems*, 25(8):10313–10330, 2024. 4

[7] C. Finn, P. Abbeel, and S. Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, pages 1126–1135. JMLR.org, 2017. 3, 7

[8] I. D. Gerg and V. Monga. Deep multi-look sequence processing for synthetic aperture sonar image segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–15, 2023. 2

[9] M. Heinrich and P. K. LeHardy. Record breaking deep ocean salvage operations. In *OCEANS 2021: San Diego – Porto*, pages 1–6, 2021. 1

[10] S. Hong, S. Cho, J. Nam, and S. Kim. Cost aggregation is all you need for few-shot segmentation, 2021. 3, 7

[11] T. Hospedales, A. Antoniou, P. Micaelli, and A. Storkey. Meta-learning in neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9):5149–5169, 2022. 3

[12] Y.-K. Hsieh, J.-W. Hsieh, Y.-Y. Chen, and Y.-C. Tseng. Strengthening spatial relations to multi-scale features for few-shot learning. In *2023 5th International Conference on Computer Communication and the Internet (ICCCI)*, pages 56–59, 2023. 2

[13] S. T. Jose, O. Simeone, and G. Durisi. Transfer meta-learning: Information- theoretic bounds and information meta-risk minimization. *IEEE Transactions on Information Theory*, 68(1):474–501, 2022. 3

[14] S. Li, J. Ma, Y. Wu, Z. Xiang, S. Bian, and G. Zhai. Sss small target detection via combining weighted sparse model with shadow characteristics. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–11, 2023. 1

[15] X. Li, T. Wei, Y. P. Chen, Y.-W. Tai, and C.-K. Tang. Fss-1000: A 1000-class dataset for few-shot segmentation. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2866–2875, 2020. 3, 7

[16] J. Liu, F. Chao, and C.-M. Lin. Task augmentation by rotating for meta-learning. *ArXiv*, abs/2003.00804, 2020. 2

[17] S. Mehta and M. Rastegari. Mobilevit: Light-weight, general-purpose, and mobile-friendly vision transformer. *ArXiv*, abs/2110.02178, 2021. 4

[18] W. Meng, T. Zhou, L. Jiang, and X. Li. Research on ocean acoustic signal processing based on particle swarm optimization algorithm. In *Proceedings of the 2024 8th International Conference on Electronic Information Technology and Computer Engineering (EITCE '24)*, pages 487–492, New York, NY, USA, 2025. Association for Computing Machinery. 1

[19] J. Pan, A. Bulat, F. Tan, X. Zhu, L. Dudziak, H. Li, G. Tzimiropoulos, and B. Martinez. Edgevits: Competing light-weight cnns on mobile devices with vision transformers, 2022. 4

[20] E. Perez, F. Strub, H. de Vries, V. Dumoulin, and A. Courville. Film: visual reasoning with a general conditioning layer. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*, AAAI'18. AAAI Press, 2018. 5

[21] D. Połap, N. Wawrzyniak, and M. Włodarczyk-Sielicka. Side-scan sonar analysis using roi analysis and deep neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–8, 2022. 1

[22] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing. 2

[23] V. Sandfort, K. Yan, P. J. Pickhardt, and R. M. Summers. Data augmentation using generative adversarial networks (cyclegan) to improve generalizability in ct segmentation tasks. *Scientific Reports*, 9(1):16884, 2019. 2

[24] A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, and T. Lillicrap. One-shot learning with memory-augmented neural networks. *arXiv preprint arXiv:1605.06065*, 2016. 2

[25] A. V. Sethuraman, A. Sheppard, O. Bagoren, C. Pinnow, J. Anderson, T. C. Havens, and K. A. Skinner. Machine learning for shipwreck segmentation from side scan sonar imagery: Dataset and benchmark. *The International Journal of Robotics Research*, 44(3):341–354, Mar. 2025. 1

[26] C. Shen, L. Shen, M. Li, and M. Yu. Epl-uflsid: Efficient pseudo labels-driven underwater forward-looking sonar images object detection. In *Proceedings of the 32nd ACM International Conference on Multimedia*, MM '24, pages 4349–4357, New York, NY, USA, 2024. Association for Computing Machinery. 1

[27] D. Stewart, A. Kreulach, S. F. Johnson, and A. Zare. Image-to-height domain translation for synthetic aperture sonar. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–13, 2023. 2

[28] Q. Sun, Y. Liu, T.-S. Chua, and B. Schiele. Meta-transfer learning for few-shot learning. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 403–412, 2019. 3, 7

[29] Y.-C. Sun, I. D. Gerg, and V. Monga. Iterative, deep, and unsupervised synthetic aperture sonar image segmentation. In *OCEANS 2021: San Diego – Porto*, pages 1–5, 2021. 1

[30] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. Torr, and T. M. Hospedales. Learning to compare: Relation network for few-shot learning. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1199–1208, 2018. 2

[31] Y. Tian, L. Lan, and L. Sun. A review of sonar image segmentation for underwater small targets. In *Proceedings of the 2020 International Conference on Pattern Recognition and Intelligent Systems*, PRIS '20, New York, NY, USA, 2020. Association for Computing Machinery. 1

[32] Z. Wang, J. Guo, L. Zeng, C. Zhang, and B. Wang. Mlffnet: Multilevel feature fusion network for object detection in sonar images. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–19, 2022. 2

[33] Z. Wang, S. Zhang, W. Huang, J. Guo, and L. Zeng. Sonar image target detection based on adaptive global feature enhancement network. *IEEE Sensors Journal*, 22(2):1509–1530, 2022. 1

[34] C. Xu, R. Qian, H. Fang, X. Ma, W. I. Atlas, J. Liu, and M. A. Spoljaric. Salina: Towards sustainable live sonar analytics in wild ecosystems. In *Proceedings of the 22nd ACM Conference on Embedded Networked Sensor Systems*, SenSys '24, pages 68–81, New York, NY, USA, 2024. Association for Computing Machinery. 2

[35] T. Yan, Z. Wan, X. Deng, P. Zhang, Y. Liu, and H. Lu. Mas-sam: Segment any marine animal with aggregated features. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI-24)*, IJCAI '24, pages 761–769, 2024. 1

[36] H. Yang, J. Cao, W. Li, S. Wang, H. Li, J. Guan, and S. Zhou. Spatial-temporal data mining for ocean science: Data, methodologies and opportunities. *ACM Transactions on Knowledge Discovery from Data*, July 2025. Just Accepted. 1

[37] R. Yang, X. Xu, X. Li, L. Wang, and F. Pu. Learning relation by graph neural network for sar image few-shot learning. In *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, pages 1743–1746, 2020. 2

[38] S.-W. Yang, L.-H. Shen, H.-H. Shuai, and K.-T. Feng. Cmaf: Cross-modal augmentation via fusion for underwater acoustic image recognition. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 20(5):124:1–124:25, January 2024. 1

[39] Y. Yu, J. Zhao, C. Huang, and X. Zhao. Treat noise as domain shift: Noise feature disentanglement for underwater perception and maritime surveys in side-scan sonar images. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–15, 2023. 1

[40] L. Zhang, B. Jindal, A. Alaa, R. Weinreb, D. Wilson, E. Segal, J. Zou, and P. Xie. Generative ai enables medical image segmentation in ultra low-data regimes. *Nature Communications*, 16(1):6486, 2025. 2

[41] P. Zhang, J. Tang, H. Zhong, M. Ning, D. Liu, and K. Wu. Self-trained target detection of radar and sonar images using automatic deep learning. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2022. 2

[42] D. Zhao, W. Mao, P. Chen, Y. Hu, H. Liang, Y. Dang, R. Liang, and X. Guo. A distributed and parallel accelerator design for 3-d acoustic imaging on fpga-based systems. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 43(5):1401–1414, 2024. 1

[43] D. Zhao, H. Zhou, P. Chen, Y. Hu, W. Ge, Y. Dang, and R. Liang. Design of forward-looking sonar system for real-time image segmentation with light multiscale attention net. *IEEE Transactions on Instrumentation and Measurement*, 73:1–17, 2024. 1

[44] L. Zhou, R. Gao, and J. Wang. A self-supervised, few-shot semantic segmentation study based on mobilevit model structure. In *2023 IEEE International Conference on Control, Electronics and Computer Technology (ICCECT)*, pages 917–921, 2023. 3

[45] X. Zhou, C. Yu, X. Yuan, Y. Wu, H. Feng, and C. Luo. Nonlinear intensity sonar image matching based on deep convolution features. In *OCEANS 2022 - Chennai*, pages 1–5, 2022. 1

[46] J. Zhu, W. Cai, M. Zhang, and M. Pan. Sonar image coarse-to-fine few-shot segmentation based on object-shadow feature pair localization and level set method. *IEEE Sensors Journal*, 24(6):8346–8360, 2024. 3