Journal of
*Marine Science
and Engineering*

MDPI

Article

WaterSAM: Adapting SAM for Underwater Object Segmentation

Yang Hong, Xiaowei Zhou *, Ruzhuang Hua, Qingxuan Lv and Junyu Dong *

School of Computer Science and Technology, West Coast Campus, Ocean University of China,
No. 1299 Sansha Road, Binhai Street, Huangdao District, Qingdao 266100, China;
hongyang@stu.ouc.edu.cn (Y.H.)
* Correspondence: zhouxiaowei@ouc.edu.cn (X.Z.); dongjunyu@ouc.edu.cn (J.D.)

Abstract: Object segmentation, a key type of image segmentation, focuses on detecting and delineating individual objects within an image, essential for applications like robotic vision and augmented reality. Despite advancements in deep learning improving object segmentation, underwater object segmentation remains challenging due to unique underwater complexities such as turbulence diffusion, light absorption, noise, low contrast, uneven illumination, and intricate backgrounds. The scarcity of underwater datasets further complicates these challenges. The Segment Anything Model (SAM) has shown potential in addressing these issues, but its adaptation for underwater environments, AquaSAM, requires fine-tuning all parameters, demanding more labeled data and high computational costs. In this paper, we propose WaterSAM, an adapted model for underwater object segmentation. Inspired by Low-Rank Adaptation (LoRA), WaterSAM incorporates trainable rank decomposition matrices into the Transformer's layers, specifically enhancing the image encoder. This approach significantly reduces the number of trainable parameters to 6.7% of SAM's parameters, lowering computational costs. We validated WaterSAM on three underwater image datasets: COD10K, SUIM, and UIIS. Results demonstrate that WaterSAM significantly outperforms pre-trained SAM in underwater segmentation tasks, contributing to advancements in marine biology, underwater archaeology, and environmental monitoring.

Keywords: underwater object segmentation; underwater image; Segment Anything Model (SAM)

check for updates

Citation: Hong, Y.; Zhou, X.; Hua, R.; Lv, Q.; Dong, J. WaterSAM: Adapting SAM for Underwater Object Segmentation. *J. Mar. Sci. Eng.* **2024**, *12*, 1616. https://doi.org/10.3390/jmse12091616

Academic Editor: Marco Cococcioni

Received: 11 August 2024
Revised: 28 August 2024
Accepted: 9 September 2024
Published: 11 September 2024

1. Introduction

Image segmentation has emerged as a critical technique, enabling machines to interpret and understand visual information. It involves partitioning an image into meaningful segments, allowing for the identification of distinct objects and regions within the visual data. Among the various types of image segmentation, object segmentation plays a vital role by focusing on detecting and delineating individual objects within an image. For instance, in robotic vision, object segmentation allows robots to identify and manipulate objects with precision. In the field of augmented reality, it enhances the ability to overlay digital content seamlessly onto real-world objects.

With the rapid advancement of deep learning, the performance of object segmentation has seen remarkable improvements. However, underwater object segmentation still faces significant hurdles, posing a critical challenge for numerous underwater visual applications. Unlike images captured in general settings, underwater images are marred by unique complexities. The ocean environment introduces underwater turbulence diffusion, intense light absorption and scattering, various types of noise, low contrast, uneven illumination, monotonous color palettes, and intricate backgrounds [1]. Moreover, the scarcity of underwater datasets further exacerbates these issues, complicating efforts to improve segmentation accuracy. Addressing these challenges is crucial for advancements in marine biology, underwater archaeology, and environmental monitoring, paving the way for more accurate and efficient underwater exploration and analysis.

Despite the critical importance of this field, research dedicated to underwater object segmentation remains limited. With the emergence of the foundation model, i.e., the Segment Anything Model (SAM) [2], improving the performance of underwater object segmentation with limited labelled data becomes realistic. AquaSAM [3] represents the pioneering effort to adapt the Segment Anything Model (SAM) to the underwater domain, aiming to achieve universal image segmentation in this challenging environment. However, AquaSAM fine-tunes all the parameters of SAM, which needs more labelled data for better performance and a high computation cost.

In this paper, we propose an adapted segment anything model for underwater object segmentation, named WaterSAM. Inspired by LoRA [4], WaterSAM adapts SAM to underwater scenarios by injecting trainable rank decomposition matrices into each layer of the Transformer architecture. Specifically, WaterSAM adds trainable rank decomposition matrices into the image encoder of the original SAM, which enhances the feature extraction ability of the image encoder and helps extract robust image features from underwater images. Compared with fine-tuning all the parameters of SAM, WaterSAM has only 6.7% of all the parameters to be trained while keeping the parameters of pre-trained SAM unchanged, which greatly reduces the number of trainable parameters and the computation cost. Furthermore, by injecting trainable rank decomposition matrices into WaterSAM, WaterSAM can efficiently capture downstream task-specific information with fewer labelled data.We validated our proposed model on three underwater image datasets: COD10K, SUIM, and UIIS. The results on these datasets demonstrate that our model significantly outperforms pre-trained SAM alone in underwater segmentation tasks, making an important contribution to the field of underwater segmentation and related tasks.

In summary, our contributions are as follows:

- We propose WaterSAM, an adapted version of the Segment Anything Model (SAM) specifically designed to address the unique challenges of underwater object segmentation.
- We collect and process three underwater image datasets (COD10K, SUIM, UIIS) to enhance their suitability for evaluating underwater segmentation performance.
- WaterSAM significantly reduces the number of trainable parameters to just 6.7% of the original model, achieving strong performance in underwater segmentation tasks. Experimental results on these datasets demonstrate the effectiveness of WaterSAM, offering lower computational costs and efficient training with fewer labeled data.

2. Background

In this section, we review the related work to our WaterSAM. It includes the image segmentation and foundation model adaptation.

2.1. Image Segmentation

Image segmentation is a classical computer vision task that divides an image into specific, unique areas to highlight objects of interest and provide information for object detection and related tasks. Over the years, numerous methods for image segmentation have been developed, including classic segmentation methods, co-segmentation methods, and deep learning-based semantic segmentation. Classical image segmentation methods aim to divide an image into segments or regions based on features such as color, brightness, or texture. Co-segmentation, on the other hand, involves the simultaneous segmentation of multiple images to identify and extract the same or similar objects across these images. However, due to the limitations in performance of these traditional methods, deep learning-based approaches are now commonly applied for more accurate and effective image segmentation.

2.1.1. Segmentation Methods Based on Deep Learning

FCN: The Fully Convolutional Network (FCN) [5], was introduced by Long et al. in 2015. It transformed image segmentation into an end-to-end image processing problem by replacing fully connected layers with convolutional layers. The key innovation of FCN is its ability to handle images of any size, making it a foundational model for modern deep neural networks in semantic segmentation.

YOLO: YOLO [6] (You Only Look Once) is a prediction method based on global image information and functions as an end-to-end object detection system, which can also be employed for segmentation tasks. YOLO divides images into grids and predicts the bounding boxes and categories of objects within each grid cell. The latest version, YOLOv8, builds upon the YOLO family, incorporating the experiences of previous versions while introducing innovative features and improvements to enhance performance and flexibility.

Mask R-CNN: Mask R-CNN [7] is a state-of-the-art algorithm for object segmentation, offering capabilities in target detection, object segmentation, and key point detection. Notable for its high speed and accuracy, Mask R-CNN builds on the foundations of two classical algorithms: Faster R-CNN for object detection and FCN (Fully Convolutional Networks) for semantic segmentation. Faster R-CNN provides efficient and precise target detection, while FCN excels in semantic segmentation tasks.

U-Net: U-Net [8] is a convolutional neural network designed for biomedical image segmentation. It builds upon a fully convolutional neural network with architectural modifications and extensions that enable it to achieve more accurate segmentations with fewer training images. Beyond biomedical segmentation, the U-Net architecture has been applied in diffusion models for iterative image denoising and serves as a foundation for many contemporary image generation models.

2.1.2. Underwater Image Segmentation Models

Underwater segmentation technology is pivotal in marine science, robotics, and computer vision for identifying and classifying objects or regions in underwater images. The task is particularly challenging due to poor visibility and color distortion caused by light absorption and scattering. This section provides an academic review of classical and contemporary techniques in underwater image segmentation.

Initially, traditional computer vision techniques were employed for underwater image segmentation, but they struggled with the unique characteristics of underwater environments and did not achieve high-precision results. Zhang et al. [9] proposed a locally adaptive color correction method based on the principle of minimum color loss and a fusion strategy guided by the maximum attenuation map, effectively minimizing color loss by accounting for different color channels' distinct attenuation characteristics. Similarly, Li et al. [10] introduced an underwater color image segmentation method that achieves high segmentation accuracy by dynamically estimating the optimal weights for fusing the RGB channels, resulting in a grayscale image with high foreground-background contrast.

Subsequently, researchers turned to machine learning methods, which demonstrated superior performance in extracting complex features from underwater images. Efforts have been directed towards addressing color distortion issues and employing deep learning techniques to enhance segmentation performance. Drews-Jr et al. [11] pioneered the use of a convolutional neural network (CNN) for underwater image segmentation in natural settings, pretraining the network on non-underwater images and fine-tuning it with a smaller dataset of manually labeled underwater images. Similarly, Arain et al. [12] presented methods for improving underwater obstacle detection by integrating sparse stereo point clouds with monocular semantic image segmentation. Their approach enhanced obstacle detection, effectively rejected transient objects such as fish, and improved range estimation compared to using feature-based sparse and dense stereo point clouds alone.

2.2. Foundation Model Adaption

Foundation model fine-tuning involves further training a pre-trained foundation model with domain-specific datasets. This process aims to optimize the model's performance on specific tasks, enabling better adaptation to and completion of tasks within a particular domain. Fine-tuning is an efficient way to enhance model performance, as it allows larger models to achieve more customized functionality. While large-scale models possess formidable capabilities, their efficacy may vary across specialized domains. However, through fine-tuning, these models can be meticulously tailored to meet the specific demands and nuances of a designated domain. This section introduces some classic foundation model fine-tuning methods.

Parameter-Efficient Fine-Tuning (PEFT) Methods

Fine-tuning a foundation model typically involves adjusting all layers and parameters to suit a specific task, using smaller learning rates and task-specific data. This process leverages the shared features of the pre-trained model but often requires substantial computational resources. In contrast, Parameter-Efficient Fine-Tuning (PEFT) technology allows models to adapt swiftly to new tasks, even in resource-constrained environments, by capitalizing on the knowledge embedded within pre-trained models. PEFT enhances model performance, reduces training duration, and lowers computational costs, making deep learning research more accessible. PEFT methods include LoRA, QLoRA, and Adapter Tuning.

LoRA: Low-Rank Adaptation (LoRA) [4] is a technique for fine-tuning large pre-trained language models, such as GPT-3 or BERT. It introduces small, low-rank matrices at crucial layers of the model, enabling fine-tuning without significant modifications to the entire model structure. This approach allows effective model fine-tuning while preserving its original performance level and minimizing additional computational burden.

QloRA: Quantized Low-Rank Adaptation (QLoRA) [13] is an efficient model fine-tuning technique that combines the principles of LoRA with deep quantization technology. QLoRA integrates quantization techniques, quantized operations, and fine-tuning stages. This approach significantly reduces memory and computational requirements in large-scale models, facilitating deployment and training in resource-constrained environments.

Adapter Tuning: Similar to LoRA, adapter tuning [14] aims to enable a pre-trained model to adapt to new tasks while keeping the original parameters unchanged. This method involves inserting small neural network modules, known as "adapters", between each layer or selected layers of the model. These adapters are trainable, whereas the parameters of the original model remain fixed.

3. Preliminary and Methodology

In this section, we provide a detailed overview of our proposed WaterSAM model. Since it is built upon the SAM model, we begin with a review of the SAM model's image encoder, which we have adapted. Next, we offer a concise introduction to Low-Rank Adaptation (LoRA) [15]. Finally, we explain how LoRA is integrated into the image encoder.

3.1. Segment Anything Model

To enhance the performance of the SAM model in underwater regions, we utilize SAM as the backbone and leverage the knowledge it has learned. As illustrated in Figure 1, SAM comprises a prompt encoder, an image encoder, and a lightweight mask decoder. It employs a Transformer-based architecture, with the image encoder built on Vision Transformer (ViT) to extract image embeddings.

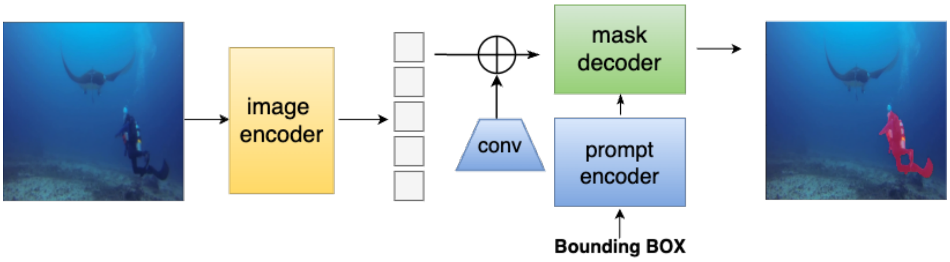

Figure 1. The network architecture of SAM.

SAM divides the input image into fixed-size blocks, linearly transforms each block into a vector representation, and adds positional coding to these vectors to retain positional information. This sequence of embedded vectors is then processed through a multi-layer standard Transformer encoder, which includes a multi-head self-attention mechanism and a feedforward neural network. Finally, a classification header processes the output sequence to complete the image classification task.

The image encoder processes an input image of 1024 × 1024 pixels and outputs a 64 × 64 feature map. We keep the weights of the pre-trained prompt encoder and mask decoder frozen to avoid substantial computational overhead. During training, we primarily fine-tune the image encoder and use bounding boxes as prompts to assist the model in achieving better segmentation performance. With the aim to improve SAM performance in the underwater scenario, we use SAM as the backbone and leverage the knowledge learned from it.

3.2. Low-Rank Adaptation

Low-Rank Adaptation (LoRA) is a parameter-efficient fine-tuning technique designed to perform an implicit low-rank transformation of the weight matrix in a foundational model. The core idea of LoRA is to approximate the incremental parameters of full-parameter fine-tuning in large language models (LLMs) with fewer training parameters. This results in efficient fine-tuning that uses less memory.

In contrast to full fine-tuning, where the model starts with its pre-trained weights and undergoes iterative gradient updates to optimize the conditional language modeling objective, LoRA significantly reduces the number of trainable parameters required for downstream tasks. It achieves this by employing low-rank approximation training with smaller matrices, while keeping the original LLM parameters frozen.

The LoRA technique is represented by the equation:

$$W_0 + \Delta W = W_0 + BA \tag{1}$$

where W_0 represents the original parameters, ΔW represents the change in parameters, and B and A are smaller matrices used for the low-rank approximation. The schematic diagram of the principle is shown in Figure 2.

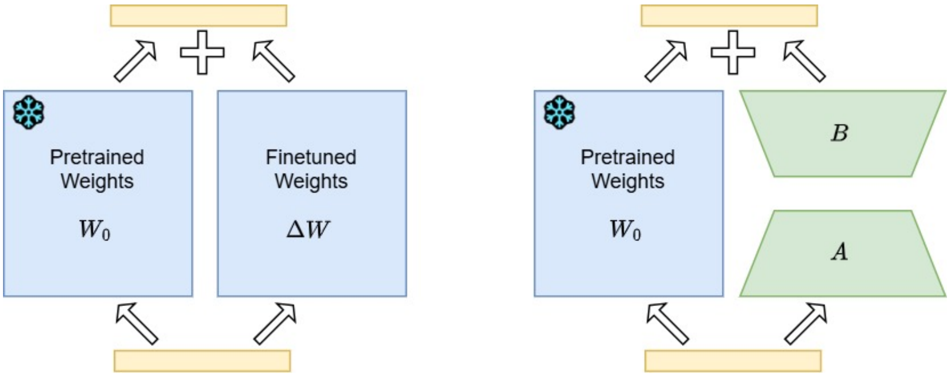

Figure 2. Illustration of the low-rank adaptation module.

3.3. Adapted Image Encoder in WaterSAM

To adapt SAM for underwater segmentation tasks, we add an adaption module to the image encoder of SAM. The adaption is developed based on LoRA, which is added to attention mechanisms of the image encoder. In this section, we will present our approach and the underlying principles.

The structure of the trainable LoRA and its application to the self-attention mechanism in the image encoder, which is based on Vision Transformer (ViT), are visualized in Figure 3. In our method, we apply LoRA to the query (Q) and value (V) layers. By modifying the query layer, the model can influence its information selection process, while adjusting the value layer allows the model to govern how it processes and utilizes the selected information.

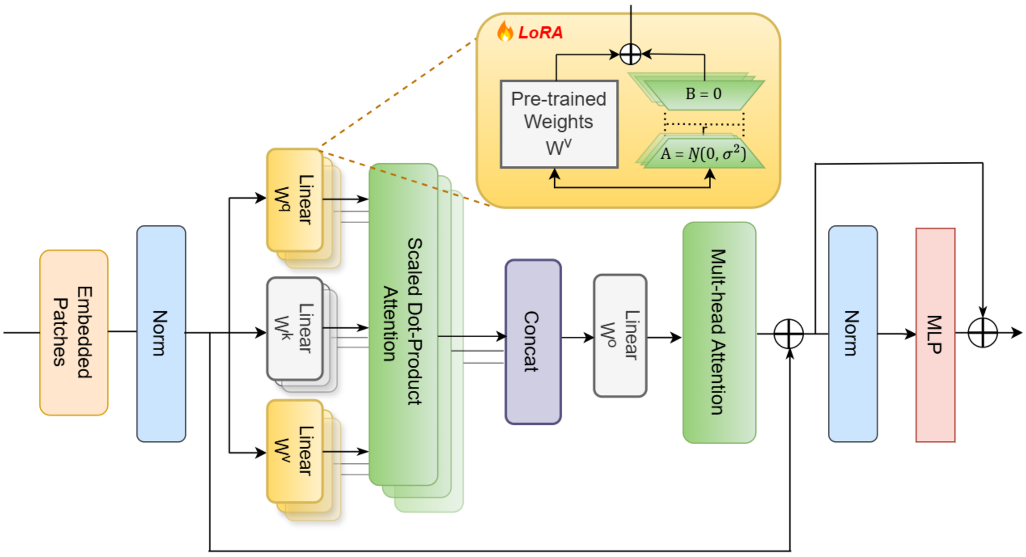

Figure 3. Architecture of image encoder in WaterSAM: designing LoRA in the attention mechanism in the image encoder.

The number of trainable parameters is determined by the rank r and the shape of the original weights, given by $|\Theta| = 2 \times L^{\text{LoRA}} \times d_{\text{model}} \times r$, where L^{LoRA} represents the number of weight matrices to which we apply LoRA, r represents the rank of weight matrices. This approach allows for efficient fine-tuning with a significantly reduced number of parameters, enhancing the encoder's adaptability without the need for extensive computational resources. For the prompt encoder and mask decoder in WaterSAM, we keep them the same as the those in the pre-trained SAM.

4. Experiment

4.1. Experimental Environment

Our experiments were conducted on the online GPU cloud computation platform, AutoDL[1], which provided a flexible and scalable environment to meet our computational needs. The configurations used in our experiments are as follows: Python 3.8.19, CUDA 10.0, CPU 20 vCPU Intel(R) Xeon(R) Platinum 8457C, Ubuntu 18.04.6 LTS, GPU NVIDIA L20 with 48 GB GPU Memory.

4.2. Datasets

We evaluated our proposed model on three public datasets: COD10K, SUIM and UIIS.

COD10K: The COD10K [16] dataset comprises 10,000 images, including 5066 camouflaged, 3000 background, and 1934 non-camouflaged images. These images were collected from various photography websites. For our study, we extracted all figures from the aquatic category and utilized the object masks for our training and testing, which included 759 training figures and 474 testing figures.

SUIM: The SUIM [17] encompasses seven categories, containing 1525 paired samples for training and validation, and 110 paired samples for benchmark evaluation. Each image within the dataset may include various species. We observed that SAM excels at segmenting individual entities but has limitations with large continuous areas such as corals and seawater. To enhance SAM's capability in segmenting distinct underwater entities, we selected three specific categories for our study: Human divers, Robots (AUVs/ROVs/instruments), and Fish and vertebrates. Consequently, we processed images containing multiple categories into binary object-mask pairs. The training set comprises 1466 pairs, while the test set contains 120 pairs.

UIIS: The UIIS [18] dataset is the first general Underwater Image Instance Segmentation (UIIS) dataset containing 4628 images across seven categories with pixel-level annotations for underwater instance segmentation tasks. The dataset is provided in a JSON file format. Before usage, we converted it to a structure analogous to that of the SUIM dataset. Subsequently, we performed two sets of experiments: the first utilizing the entire UIIS dataset, and the second focusing specifically on the categories of fish, human divers, and robots.

4.3. Parameter Setting

When training on the datasets, we experimented with multiple values of rank to optimize the balance between computational efficiency and performance. Ultimately, we found that setting the rank to 64 provided the optimal balance across various datasets. Specifically, for the UIIS dataset, we also experimented with training using a rank of 128 (denoted as UIIS**). This higher rank value did improve the segmentation accuracy, but it also doubled the time required for each epoch, resulting in significantly higher computational costs.

4.4. Experiment Outcomes and Analysis

4.4.1. Quantative Results

In our data evaluation, we utilized the mean intersection over union (mIoU) and overall accuracy (OA) to measure the degree of region overlap and boundary agreement between the ground truth and segmentation results. We performed multiple comparisons between our WaterSAM model and the pretrained SAM (ViT-B) model across three datasets. As shown in the Table 1, our WaterSAM demonstrated superior performance for underwater segmentation tasks across all datasets.

For the COD10K dataset, the mIoU metric experienced the most significant increase, rising from 16.98% to 84.96%, marking a 400% improvement. The OA metric also saw a substantial rise from 29.47% to 96.09%.

In the SUIM * dataset, WaterSAM achieved a notable increase in mIoU from 38.41% to 90.47%, which corresponds to a 136% improvement. The OA also improved considerably from 61.98% to 98.08%, a 58% enhancement.

Regarding the UIIS dataset, the mIoU improved from 46.64% to 84.24%, an 81% increase, while the OA increased from 64.58% to 93.16%, reflecting a 44% improvement. For the UIIS * subset, the improvements were even more pronounced, with the mIoU rising from 50.35% to 94.38%, an 87% increase, and the OA increasing from 79.92% to 99.31%, a 24% improvement. Additionally, for the UIIS ** dataset, WaterSAM maintained a high performance with an mIoU of 94.49% and an OA of 99.34%, showcasing its robustness even with challenging categories like reefs and the sea floor.

Table 1. Results comparison between SAM and our WaterSAM across three datasets.

Dataset	Mean IoU Score			Overall Accuracy		
	WaterSAM	**SAM**	**Improve (%)**	**WaterSAM**	**SAM**	**Improve (%)**
COD10K	84.96	16.98	400	96.09	29.47	226
SUIM *	90.47	38.41	136	98.08	61.98	58
UIIS	84.24	46.64	81	93.16	64.58	44
UIIS *	94.38	50.35	87	99.31	79.92	24
UIIS **	94.49	-	-	99.34	-	-

* Segmentation based solely on object categories with a learning rate of 0.001. ** Training with a learning rate of 0.0005, which yields the best results.

4.4.2. Qualitative Results

As shown in Figure 4, we observed that the pretrained SAM (ViT-B) model frequently struggles to accurately delineate object boundaries, sometimes failing to detect the boundaries altogether. This suggests that one of the major challenges for SAM in underwater object segmentation is the indistinct separation of objects in underwater environments. Accurately segmenting targets with weak boundaries is particularly difficult.

Additionally, SAM's segmentation performance is significantly hampered when the target object is large, has a complex shape, or has a color tone similar to the background. Conversely, SAM may exhibit lower accuracy or produce anomalies when segmenting objects with clear boundaries, especially when the surrounding objects also show good contrast, such as in the segmentation of multiple fish within a single image.

These observations indicate that WaterSAM enhances SAM's ability to tackle more challenging segmentation tasks. Specifically, WaterSAM demonstrates improved robustness in suboptimal environments with indistinct object boundaries, particularly in underwater image segmentation. Furthermore, WaterSAM shows a significant improvement in performing segmentation tasks involving small entities.

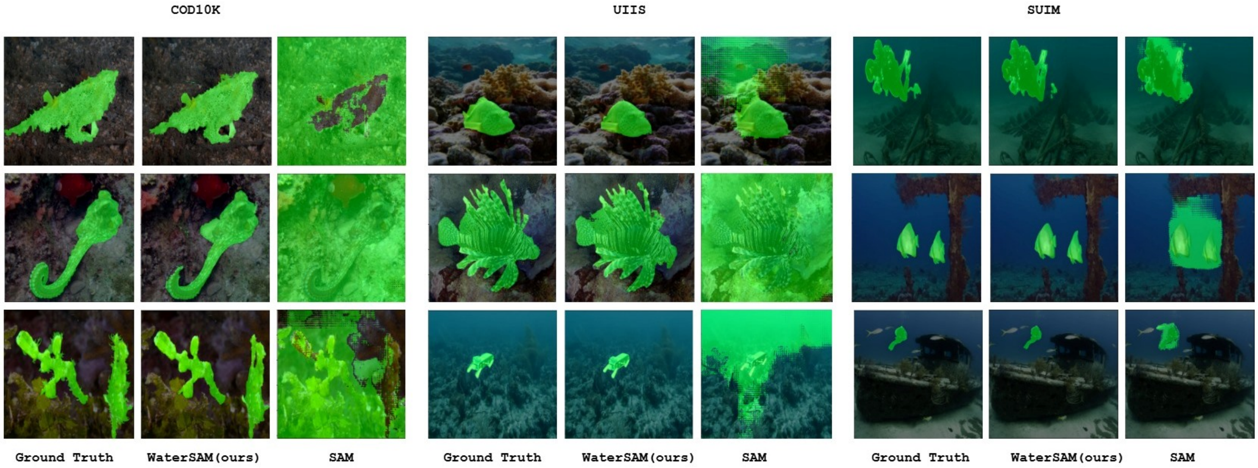

Figure 4. Visualization of segmentation results of WaterSAM in comparison to SAM.

4.4.3. Comparison with Existing Methods

For underwater image segmentation, the existing work AquaSAM utilizes the original SAM model without any modifications. It relies on fine-tuning all the parameters of SAM, which demands substantial computational resources and a large amount of labeled data. In contrast, our proposed WaterSAM requires training only 6.7% of the SAM model's parameters, making it more resource-efficient. We compare the performance of WaterSAM and AquaSAM using the same dataset, SUIM. We report the mean IoU score of two models in the Table 2. As shown in the table, WaterSAM achieves more accurate segmentation results than AquaSAM, with an improvement of approximately 20%. This advantage stems from the fact that AquaSAM demands more labeled data for training, but the SUIM dataset contains only 1500 training images. This highlights WaterSAM's advantage in adapting SAM for downstream tasks.

Table 2. A comprehensive comparison of WaterSAM and AquaSAM on SUIM dataset.

	WaterSAM	**AquaSAM**
mean IoU	98.08	76.64

5. Conclusions

In conclusion, this paper presents WaterSAM, an adapted model for underwater object segmentation based on the Segment Anything Model (SAM). Our comprehensive evaluations across multiple underwater datasets, including COD10K, SUIM, and UIIS, demonstrate WaterSAM's significant improvements in segmentation performance. By integrating trainable rank decomposition matrices into the Transformer's layers, WaterSAM effectively reduces computational costs while maintaining high accuracy. This advancement is particularly notable in challenging underwater environments, where traditional models struggle with poor visibility and complex backgrounds. The results highlight WaterSAM's potential for enhancing applications in marine biology, underwater archaeology, and environmental monitoring. As the development of SAM, the new iteration, SAM2 [19], performs better in image segmentation and can be used for video segmentation. For future work, we will adapt SAM2 for underwater video segmentation tasks.

Author Contributions: Conceptualization, Y.H. and J.D.; Methodology, Y.H. and X.Z.; Ruzhuang H.J.D.; Software, Y.H. and R.H.; Formal analysis, X.Z.; investigation, X.Z.; Resources, Q.L.; Data curation, Q.L.; Writing—original draf, Y.H.; Writing—review & editing, X.Z.; Visualization, Y.H. and R.H.; Supervision, X.Z. and J.D.; Project administration, J.D. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Postdoctoral Fellowship Program of CPSF under Grant GZC20241614 as well as Sanya Science and Technology Special Fund under Grant 2022KJCX92.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All the data used in this paper are publicly available.

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

Note

[1] https://www.autodl.com/home (accessed on 8 September 2024).

References

1. Jian, M.; Liu, X.; Luo, H.; Lu, X.; Yu, H.; Dong, J. Underwater image processing and analysis: A review. *Signal Process. Image Commun.* **2021**, *91*, 116088. [CrossRef]
2. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.Y.; et al. Segment anything. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 4015–4026.
3. Xu, M.; Su, J.; Liu, Y. Aquasam: Underwater image foreground segmentation. In Proceedings of the International Forum on Digital TV and Wireless Multimedia Communications, Beijing, China, 21–22 December 2023; Springer: Berlin/Heidelberg, Germany, 2023; pp. 3–14.
4. Hu, E.J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; Chen, W. Lora: Low-rank adaptation of large language models. *arXiv* **2021**, arXiv:2106.09685.
5. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
6. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
7. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
8. Siddique, N.; Paheding, S.; Elkin, C.P.; Devabhaktuni, V. U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access* **2021**, *9*, 82031–82057. [CrossRef]
9. Zhang, T.; Xia, Y.; Feng, D.D. A deformable cosegmentation algorithm for brain MR images. In Proceedings of the 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, CA, USA, 28 August–1 September 2012; IEEE: Washington, DC, USA, 2012; pp. 3215–3218.
10. Li, Z.; Chen, J. Superpixel segmentation using linear spectral clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1356–1363.
11. Drews, P., Jr.; Souza, I.D.; Maurell, I.P.; Protas, E.V.; Botelho, S.S.C. Underwater image segmentation in the wild using deep learning. *J. Braz. Comput. Soc.* **2021**, *27*, 1–14.
12. Arain, B.; McCool, C.; Rigby, P.; Cagara, D.; Dunbabin, M. Improving underwater obstacle detection using semantic image segmentation. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019; IEEE: Washington, DC, USA, 2019; pp. 9271–9277.
13. Xu, Y.; Xie, L.; Gu, X.; Chen, X.; Chang, H.; Zhang, H.; Chen, Z.; Zhang, X.; Tian, Q. Qa-lora: Quantization-aware low-rank adaptation of large language models. *arXiv* **2023**, arXiv:2309.14717.
14. Chen, T.; Zhu, L.; Ding, C.; Cao, R.; Wang, Y.; Li, Z.; Sun, L.; Mao, P.; Zang, Y. SAM Fails to Segment Anything?—SAM-Adapter: Adapting SAM in Underperformed Scenes: Camouflage, Shadow, Medical Image Segmentation, and More. *arXiv* **2023**, arXiv:2304.09148.

15. Wang, X.; Ye, F.; Zhang, Y. Task-Aware Low-Rank Adaptation of Segment Anything Model. *arXiv* **2024**, arXiv:2403.10971.
16. Fan, D.P.; Ji, G.P.; Sun, G.; Cheng, M.M.; Shen, J.; Shao, L. Camouflaged object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 2777–2787.
17. Islam, M.J.; Edge, C.; Xiao, Y.; Luo, P.; Mehtaz, M.; Morse, C.; Enan, S.S.; Sattar, J. Semantic segmentation of underwater imagery: Dataset and benchmark. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 25–29 October 2020; IEEE: Washington, DC, USA, 2020; pp. 1769–1776.
18. Lian, S.; Li, H.; Cong, R.; Li, S.; Zhang, W.; Kwong, S. Watermask: Instance segmentation for underwater imagery. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 1305–1315.
19. Ravi, N.; Gabeur, V.; Hu, Y.T.; Hu, R.; Ryali, C.; Ma, T.; Khedr, H.; Rädle, R.; Rolland, C.; Gustafson, L.; et al. Sam 2: Segment anything in images and videos. *arXiv* **2024**, arXiv:2408.00714.

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.