Original article

Journal of High School Science

Advancing Seagrass semantic segmentation with SAM2 models

Huang S

Submitted: January 1, 2025, Revised: version 1, February 24, 2025, version 2, April 28, 2025, version 3, May 15, 2025
Accepted: May 22, 2025

Abstract

Seagrass ecosystems in shallow waters are important for assessing environmental factors such as pollution and global warming, because they are highly sensitive to changes in water quality and temperature. This paper investigates advanced deep learning techniques for semantic segmentation of Seagrass images collected from shallow water areas, focusing on fine tuning the Segment Anything Model 2 (SAM2) and SAM2-UNet architectures. We conducted extensive experiments on five SAM2 models to optimize segmentation performance on two subsets of data obtained at different depths of the ‘Looking for Seagrass’ dataset. Additionally, we performed an ablation study using the best SAM2 model with a lesser number of annotated training set images. Furthermore, we analyzed the inference time of the different models to provide insights for further model selection. The experiment results demonstrated the effectiveness of the SAM2 models, with the highest Intersection over Union (IoU) of 0.923 and 0.935, and Pixel Accuracy (PA) of 0.967 and 0.972 in two sub-datasets; obtained using the SAM2-UNet model; outperforming the method provided in the original dataset paper. The ablation study using a lesser number of annotated images in the training dataset provided similar accuracy as the second-best performing model. The total processing time was the least for the SAM2-UNet model. This work also highlights the potential of real-time Seagrass monitoring using drone and underwater camera footage.

Keywords

Seagrass, Seagrass coverage, Seagrass segmentation, Segment Anything Model, U-Net, Deep Learning, Fine tuning, Semantic segmentation, Shallow water ecosystem, Marine ecology

Siyuan Huang, Piedmont Hills High School, 1377 Piedmont Rd, San Jose, CA 95132, USA.
phoenixemerald16@gmail.com

https://doi.org/10.64336/001c.138532

1. Introduction

Seagrass plays a crucial role as a spawning ground and habitat for various marine species while also absorbing carbon dioxide, a key contributor to global warming. The growing interest in understanding the interaction between Seagrass vegetation and hydrodynamic processes, as well as its contribution to coastal protection, has led to an increasing number of experimental, field, and numerical studies (1). Seagrass ecosystems in shallow waters are particularly important for assessing environmental factors such as pollution and global warming, as they are highly sensitive to changes in water quality and temperature. These areas act as early indicators of environmental stress, making them crucial for monitoring ecosystem health and informing conservation efforts. Effective management of Seagrass ecosystems requires a thorough understanding of their distribution and extent, alongside continuous monitoring of changes in both their coverage and condition (2).

Before conducting quantitative analysis of Seagrass, semantic segmentation must first be performed. Traditional machine learning techniques, such as maximum likelihood classification, K-means clustering, support vector machines (SVM), random forest (RF), and convolutional neural networks (CNNs) (3), have been employed for this purpose. Recently, deep learning methods have been increasingly used for Seagrass segmentation. Some studies have combined traditional machine learning techniques with deep learning methods, wherein images are divided into patches before applying neural networks for classification of Seagrass (4). Fully Convolutional Neural Networks (FCNNs) have been applied in semi-supervised settings to map Seagrass and other coastal features from optical drone images, achieving state-of-the-art results with only subsets of labeled data (5, 6).

Several studies have advanced the application of deep learning techniques in Seagrass segmentation. Mehrubeoglu et al. (7) utilized pre-trained ResNet-18 and ResNet-50 models for Seagrass blade surface composition classification. Raine et al. (8) pioneered the use of transfer learning-based VGG-16 and ResNet-50 models for multi-species Seagrass identification from underwater images. Noman et al. (9) proposed a transfer learning-based EfficientNet-B5 classifier, achieving high accuracies on the Deep Seagrass dataset. In another study, Noman et al. (10) developed a two-stage semi-supervised deep learning model using an EfficientNet-B5 backbone network for automatic labeling of large datasets. Weidmann et al. (11) applied various deep learning models, including DeepLabv3+, FCN, dilNet, and U-Net, to the Looking for Seagrass dataset. Chen et al. (12) implemented an encoder-decoder-based segmentation model for Seagrass mapping. Recently, Savinov (13) explored the application of Neural Radiance Fields (NeRFs) to Multibeam Echosounder Systems (MBES) data for unsupervised Seagrass segmentation. Additionally, Noman et al. (14) proposed the BAOS-CNN, a novel deep neuroevolution algorithm for multispecies Seagrass detection, achieving state-of-the-art accuracy on the Deep Seagrass dataset. Chen et al. (15) introduced YOLO-C, a model combining U-Net and YOLOX (an improved version of YOLO) for precise semantic segmentation of seabed sediments, including Seagrass. Roelfsema et al. (16) developed a deep learning-based subtidal Seagrass detector, which, while not explicitly using SegNet or YOLOv5, reflects the growing trend of applying advanced deep learning techniques to Seagrass detection and classification.

However, the aforementioned methods often compromise efficiency and accuracy, making them challenging to implement in real-time systems. Recently, the Segment Anything Model (SAM) has been widely used in image segmentation tasks due to its zero-shot capabilities (17). To improve mask prediction quality for objects with complex structures, several adapters have been developed. HQ-SAM enhances mask details by integrating mask decoder features with ViT features (18). PA-SAM optimizes prompt features using a prompt-driven adapter to extract detailed information about the network's focal area (19). These adapters preserve SAM's original architecture while adding minimal additional parameters. Building upon these advancements, SAM2 extends segmentation capabilities to video, treating images as single-frame videos. It introduces a streaming memory architecture that processes video frames sequentially, enabling real-time, interactive applications (20). SAM2 outperforms its predecessor in both accuracy and speed, achieving better video segmentation with fewer interactions and faster image segmentation. SAM2-UNet adopts SAM2's hierarchical backbone as the encoder and uses a U-shaped design for the decoder, which is applicable in camouflaged object detection, salient object detection, and marine animal segmentation (21).

Our work focuses on finetuning SAM2 and SAM2-UNet for Seagrass semantic segmentation, which demonstrates efficiency and enhanced accuracy in the analysis of Seagrass habitats.

2. Methods

Seagrass images present unique challenges for segmentation due to their complex visual characteristics. These images often contain hundreds or thousands of individual Seagrass leaves, which may belong to different species and may frequently overlap as well. The visual complexity is further compounded by environmental factors such as turbidity, water column discoloration, varying depths, lighting conditions, shadows, and camera blur. Additionally, Seagrass instances can exhibit different visual characteristics based on their growth stage, health, and location (22). Our approach concentrated on Seagrass in shallow waters, since these regions serve as early indicators of environmental stress, playing a key role in monitoring ecosystem health and guiding conservation efforts.

SAM2 (Segment Anything Model 2) and SAM2-UNet are promising approaches for Seagrass segmentation due to their ability to process complex, overlapping instances and adapt to varying visual conditions. SAM2's architecture allows it to process images sequentially, which can be used for analyzing underwater footage of Seagrass meadows (20). The model's real-time performance and ability to process dynamic content make it suitable for interactive applications in marine ecosystem monitoring. SAM2-UNet, by combining the strengths of SAM2 and U-Net architectures, can potentially provide more precise segmentation of intricate Seagrass structures while maintaining the flexibility to adapt to diverse underwater conditions (7). These models' capacity for zero-shot generalization is particularly valuable in Seagrass segmentation, where the visual appearance can vary significantly across different environments and species.

2.1 Segment Anything Model v2 (SAM2)

SAM2's architecture builds upon its predecessor - SAM (17), by introducing several key components to process video data efficiently, hence the investigation with

SAM2 for Seagrass segmentation can be easily extended to real-time Seagrass detection with drones or underwater cameras (20). The memory mechanism that includes a memory encoder, memory bank, and memory attention module are designed to process temporal dependencies in videos. The occlusion Head can process scenarios where objects are not visible, predicting the likelihood of object occlusion.

Our work focuses on the precise segmentation of Seagrass images. As illustrated in Figure 1, we fine tuned the mask decoder of the SAM2 model to enhance its capability in accurately delineating Seagrass regions.

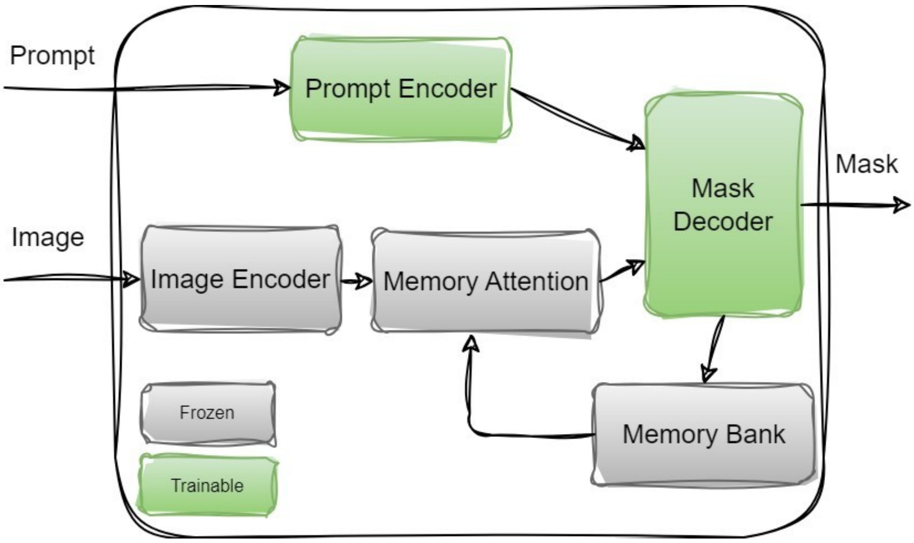

Figure 1. SAM2 fine-tuning architecture.

In our training approach for the SAM2 model, we employed a composite loss function that combined binary cross-entropy loss and IoU loss. This combination leveraged the strengths of both loss functions to enhance segmentation performance. The binary cross-entropy loss exceled at capturing fine-grained pixel-level details, while the IoU loss effectively evaluated the overall segmentation quality by measuring the overlap between predicted and ground truth masks. The synergy of these loss components enabled the model to simultaneously learn precise pixel-wise classifications and accurate global object shapes. The total loss function is defined as $L_{total}=\lambda_1 L_{BCE}+\lambda_2 L_{IoU}$ (eq 1), Where, L_{BCE} is the Binary Cross-Entropy loss, $L_{IoU}$is the Intersection over Union loss and λ_1 and $\lambda_{12}$are weighting factors to balance the contribution of each loss component The Binary Cross-Entropy loss can be defined as

$$L_{BCE}=\frac{-1}{N}\sum_{i=1}^{N}\left[y_i\log\left(\hat{y}_i\right)+\left(1-y_i\right)\log\left(1-\hat{y}_i\right)\right] \quad \text{(eq 2)}$$

Where y_i is the ground truth label and $\hat{y}_i$ is the predicted probability for the i-th pixel, and N is the total number of pixels. The IoU loss is typically defined as $L_{IoU}=1-\frac{\left|Y \cap \hat{Y}\right|}{\left|Y \cup \hat{Y}\right|}$ (eq 3), Where Y is the ground truth mask and $\hat{Y}$ is the predicted mask. This comprehensive loss function guides the SAM2 model to achieve high-quality Seagrass segmentation by optimizing both local and global aspects of the segmentation task.

2.2 SAM2-UNet

The SAM2-UNet architecture represents an innovative approach to image segmentation, combining the strengths of the Segment Anything Model 2 with the classic U-Net (23) design. SAM2-UNet leverages the powerful Hiera backbone from SAM2 as its encoder, a choice that brings significant advantages. Building upon this robust encoder, SAM2-UNet incorporates a decoder that follows the classic U-Net design. This decoder consists of three blocks, each featuring two sets of convolutions, batch normalization, and ReLU activation layers (21). This classic structure has proven its effectiveness time and again in various segmentation tasks.

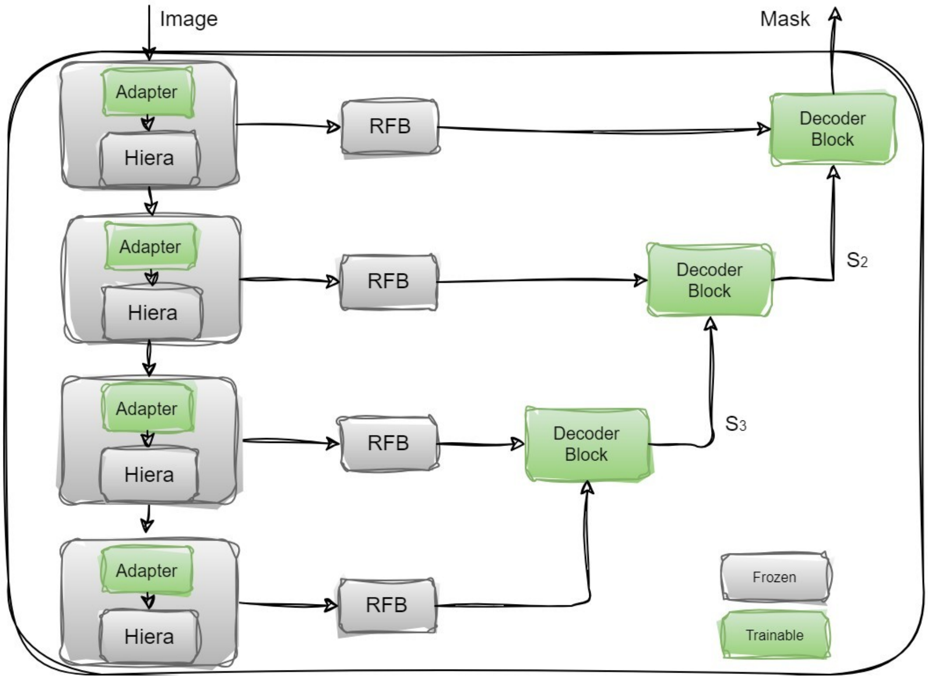

Figure 2. SAM2-UNet fine-tuning architecture.

As illustrated in Figure 2, to further enhance the model's capabilities, the architects introduced Receptive Field Blocks (RFBs) into the mix. These RFBs are placed after the extraction of encoder features, serving a dual purpose. First, they reduce the number of channels to a manageable 64, and second, they work to enhance the lightweight features, contributing to the model's overall efficiency and effectiveness.

One of the most innovative aspects of the SAM2-UNet is the inclusion of adapters within the encoder. These adapters are very effective when it comes to fine tuning the model. By enabling parameter-efficient fine-tuning, it is possible to train the model even on devices with limited memory resources. This feature significantly broadens the potential applications and accessibility of the model.

In our approach, we fine tuned the adapters positioned before each multi-scale block of the Hiera architecture (24), and the decoder blocks. Each adapter in our framework was structured as follows: a linear layer for downsampling, followed by a GeLU activation function, then another linear layer for upsampling, and a final GeLU activation. This design allowed for efficient parameter adaptation while maintaining the pre-trained model's core capability. The Decoder block in the model was composed of two sequential 'Conv-BN-ReLU' combinations. Here, 'Conv' refers to a 3x3 convolution layer, 'BN' denotes batch normalization, and 'ReLU' is the rectified linear unit activation function. This sequence was followed by an upscaler, which was responsible for increasing the spatial resolution of the features from the lower decoder blocks. This architecture enabled the model to effectively reconstruct detailed segmentation masks from the encoded features, progressively refining the output through the decoder stages.

We trained the model with similar weighted IoU loss and BCE loss as described in Equation 1. Since the SAM2-UNet contains three decoder blocks $S_i(i \in 1,2,3)$, the loss function is used for all three outputs. The total loss function is formulated as $L_{total} = \sum_{i=1}^{3} L(G_i, S_i)$ (eq 4),

Where G_i is the downsampled ground truth corresponding to the resolution to each S_i.

3. Experiment

3.1 Dataset

We used the 'Looking for Seagrass' dataset (25) to train our models. This dataset provides a valuable resource for studying Seagrass coverage in underwater environments. Collected along Croatia's Adriatic coast, the dataset comprises 12,682 images of the seabed captured at various depths using a diving robot. Of these, 6,036 images have been meticulously annotated with pixel-level polygon masks, offering precise ground truth data for Seagrass coverage. This comprehensive dataset enables researchers to develop and evaluate advanced image processing and machine learning techniques for automatic Seagrass detection and coverage estimation.

To enhance the robustness and generalization capability of the models, we employed various data augmentation techniques. These included geometric transformations such as random flipping (horizontal and vertical), rotation, and scaling. In addition, we applied color space augmentations, including adjustments to brightness, contrast, and hue, to simulate different lighting conditions and water turbidity levels. We also introduced image blurring and degradation. These augmentation techniques not only increased the effective size of our training dataset but also helped our models become more resilient to variations in image orientation, scale, and underwater lighting conditions, which are common challenges in marine image analysis.

Our research focused on shallow water Seagrass, hence we selected two annotated subsets from the 'Looking for Seagrass' dataset, stratified by depth: (a) Shallow coastal subset 00d01, which was captured at depths < 1 meter and comprised of 890 (9,790, after data augmentation) annotated images, and (b) Intermediate depth subset 01d02, which was captured at depths ranging from 1 to 2 meters and comprised of 2,522 (27,742, after data augmentation) annotated images. Both the subsets were split into training (70%), test (15%), and evaluation (15%) sets to facilitate model training, performance assessment, and generalization evaluation, respectively.

We also performed an ablation study to ascertain if we could achieve similar performance with a lesser number of annotated images in the training set. All the images in the training set for all the studies were annotated. As an example, for the subset 01d02 in our original experiment, the number of images used for training, evaluation and testing were 19420 (27742*0.7), 4161 (27742*0.15) and 4161 (27742*0.15) respectively for the original experiment and 13871 (27742*0.5), 4161 (27742*0.15) and 4161 (27742*0.15) respectively for the ablation study.

3.2 Experimental setup

In our experimental setup, we trained five different SAM2 models: SAM2-Large, SAM2-Base-Plus, SAM2-Small, SAM2-Tiny, and SAM2-UNet. For the first four models, we employed the AdamW optimizer with a batch size of 4, an initial learning rate of $1x10^{-5}$, and a weight decay of $4x10^{-5}$. The SAM2-UNet model was trained separately using the AdamW optimizer with a batch size of 4, a higher initial learning rate of $3x10^{-4}$, and a weight decay of $5x10^{-4}$ to account for its unique architecture.

We used two subsets of the 'Looking for Seagrass' dataset, designated as subset 01d02 and subset 00d01. All the 5 models were trained with these 2 subsets. Among the 5 models, the SAM2-UNet employed an image resolution of 352x352, while the remaining 4 SAM2 models utilized images of resolution 1024x1024. All images needed to be resized when loading from the original dataset, which contained images with a resolution of 1920x1080.

All training procedures were performed on NVIDIA A100 GPUs, leveraging its high-performance computing capabilities to efficiently process image data and complex model architectures. Throughout the training, we closely monitored key performance metrics and adjusted hyperparameters as necessary to ensure optimal convergence and generalization of the models.

3.3 Metrics

Intersection over Union (IoU)

Also known as the Jaccard Index, IoU measures the overlap between the predicted segmentation and the ground truth.

$$IoU = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A \vee +B| - |A \cap B|} \quad \text{(eq 5)}$$

Where, A is the predicted segmentation mask, B is the ground truth segmentation mask. $|A \cap B|$ represents the area of overlap between the predicted and ground truth masks and $|A \cup B|$ represents the total area covered by both the predicted and ground truth masks.

Pixel Accuracy

The ratio of correctly classified pixels to the total number of pixels, given by $PA = \frac{TP+TN}{TP+TN+FP+FN}$ (eq 6), where TP, TN, FP and FN are True Positive, True Negative, False Positive and False Negative respectively.

Positive Predictive Value (PPV)

Also known as Precision, it measures the proportion of correctly identified positive pixels among all pixels classified as positive, given by $PPV = \frac{TP}{TP+FP}$ (eq 7).

True Positive Rate (TPR)

Also known as Recall, it quantifies all relevant positive pixels from the entire set of actual positive pixels, given by $TPR=\frac{TP}{TP+FN}$ (eq 8).

Dice Similarity Coefficient (DSC)
Also known as the F1 score, DSC is similar to IoU but less penalizing for under- and over-segmentation, given by $DSC=\frac{2*TP}{2*TP+FP+FN}$ (eq 9).

4. Results

4.1 Quantitative results

We ran the metrics calculations as defined in Section 3.3 on the various models. The results are shown in Table 1. The SAM2-UNet model consistently outperformed the other variants across both datasets and evaluation metrics. For example, on the 01d02 subset, the SAM2-UNet model achieved the highest Intersection over Union (IoU) score of 0.923, accompanied by a Dice Similarity Coefficient (DSC) of 0.960 and a Pixel Accuracy (PA) of 0.967. The second-best model, SAM2-L, could only achieve an IoU of 0.871, a DSC of 0.931, and a PA of 0.941. This represented increases of 6%, 3.1% and 2.8% for the IoU, DSC and PA respectively. Note that this represents the percent improvement of the SAM2-UNet over the second-best performing model. The SAM2-UNet model outperformed all the other SAM2 models, as well as the baseline method from the original 'Looking for Seagrass' dataset, which achieved a PA of 0.945 (17). Although the baseline method used the entire dataset, the PA for the SAM2-UNet method for both subsets was > 0.945, with the average being 0.967; representing an increase of 2.3% for the PA. It is hence possible to conclude that combining the UNet architecture into the SAM2 framework enhanced segmentation performance, making the SAM2-UNet model a promising candidate for reliable and accurate Seagrass image segmentation tasks.

Despite the SAM2 Base Plus model being about twice the size of its small and tiny counterparts, and the large version being three times the Base Plus, their performance metrics were closer than anticipated. The difference in the averages of the (L, B+ models) versus the (S, T) models for IoU, PA and DSC were |0.022|, |0.0125|, and |0.013| respectively. The larger models thus achieved marginally better results with the largest percent increase of 2.6% for the IoU. These findings suggest that for real-time Seagrass monitoring with limited hardware resources, selecting a smaller (SAM2-only) model would not necessarily significantly affect performance. This may allow for efficient deployment of smaller models in resource-constrained environments without a substantial decrease in accuracy.

Table 1. Performance comparison of the different models on the 01d02 and 00d01 subsets.

	01d02 Subset					00d01 Subset				
Model	IoU	PA	PPV	TPR	DSC	IoU	PA	PPV	TPR	DSC
SAM2-L	0.871	0.941	0.918	0.944	0.931	0.916	0.963	0.948	0.965	0.956
SAM2-B+	0.867	0.940	0.916	0.942	0.929	0.908	0.959	0.943	0.961	0.952
SAM2-S	0.856	0.933	0.908	0.937	0.922	0.905	0.958	0.941	0.960	0.950
SAM2-T	0.838	0.923	0.896	0.928	0.912	0.901	0.956	0.938	0.958	0.948
SAM2-UNet	0.923	0.967	0.952	0.968	0.960	0.935	0.972	0.960	0.973	0.966

4.2 Qualitative results

We visualized the Seagrass segmentation results for subsets 00d01 and 00d02, as shown in Figures 3 and 4, respectively. The SAM2-UNet model demonstrated the highest accuracy across all the 5 SAM2 models, as shown in the last column of the Figures. We also observed that SAM2-UNet produced much smoother edges in the masks. This effect can be primarily attributed to the image processing pipeline. Initially, the training images are downsampled from 1920x1080 to 352x352 pixels. Subsequently, the generated 352x352 mask is upsampled back to 1920x1080. During this process, high-frequency noise is effectively filtered out, resulting in a smoother final output.

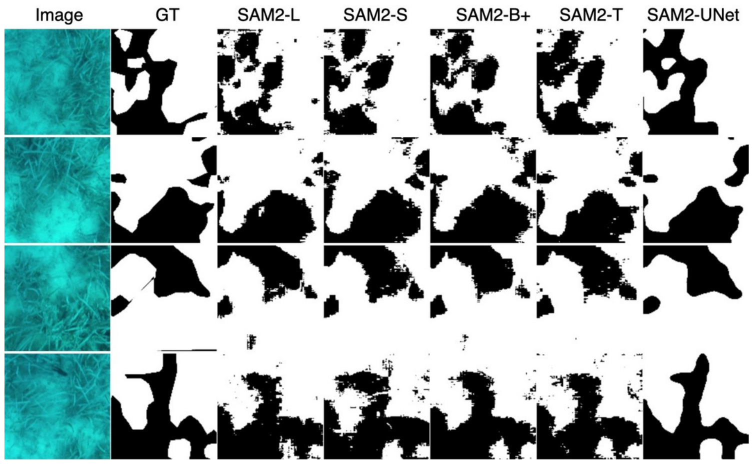

Figure 3. Visualization results on 00d01 subset.

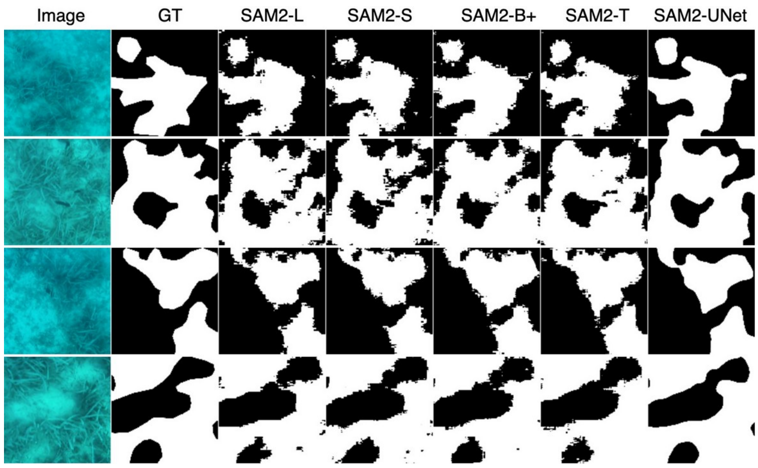

Figure 4. Visualization results on 01d02 subset.

4.3 Inference time comparison

The speed of the models was measured by performing time consumption tests on the 5 models, measuring their specific resolution generation and resizing time. The inference process consisted of three steps: downsampling the image, generating the mask, and upsampling the mask. The time consumption for each of these steps, is shown in Table 2.

Table 2 shows that that SAM2-UNet required the least total processing time, at 44.22 ms, to predict a mask of the same size as the input images (1920x1080). Focusing only on inference time, the SAM2 tiny version processed 1024x1024 resolution images the fastest, while SAM2-UNet took longer at 35.40 ms for 352x352 images. This test provided insights for model selection if system latency were to be a critical factor, given specific performance requirements and hardware constraints.

Table 2. Inference time comparison of different models in milliseconds

	SAM2-T	SAM2-S	SAM2-B+	SAM2-L	SAM2-UNet
Downsampling	14.65	14.65	14.65	14.65	8.52
Inference	34.51	38.04	40.26	48.58	35.40
Upsampling	0.22	0.22	0.22	0.22	0.30
Total Time	49.38	52.91	55.13	63.45	44.22

4.4 Ablation study

To evaluate whether the models could use less annotated images and still provide similar accuracy, we performed an ablation study using 50/15/15 split for the datasets, which means among the original 70% annotated images used in the regular experiment in Section 3.1, only 50% were used for training and the rest 20% were excluded in the experiment. We used the same hyperparameters and environment settings for this study. The results of this experiment are summarized in Table 3. It is evident that, despite the lesser number of training images, the models still showed comparable results to those from the original experiment shown in Table 1, albeit with a decline in performance. Specifically, the SAM2-UNet model performed largely equivalent to the second-best performing model (SAM2-L) in the original experiment. Pixel Accuracy, however, did not show any significant change. These findings suggest that the methods presented in this work are robust and require relatively lesser amounts of training data for effective convergence and generalization. This is especially relevant since a large amount of time and resources are spent in manual annotation of images. Achieving similar accuracy with a lesser number of manually annotated images hence represents a significant improvement in resource and time allocation with faster availability of results.

Table 3. Ablation study with reduced training dataset

	01d02 Subset					00d01 Subset				
Model	IoU	PA	PPV	TPR	DSC	IoU	PA	PPV	TPR	DSC
SAM2-L	0.852	0.931	0.906	0.935	0.920	0.901	0.956	0.937	0.958	0.948
SAM2-B+	0.855	0.932	0.908	0.936	0.922	0.896	0.954	0.935	0.956	0.945
SAM2-S	0.842	0.925	0.899	0.930	0.914	0.898	0.955	0.936	0.957	0.946
SAM2-T	0.803	0.902	0.872	0.911	0.891	0.890	0.951	0.931	0.953	0.942
SAM2-UNet	0.904	0.958	0.940	0.959	0.950	0.913	0.962	0.946	0.963	0.955

5. Conclusion

In this study, we successfully trained and evaluated the SAM2 and SAM2-UNet models using the 'Looking for Seagrass' dataset, outperforming the baseline method of 0.945 Pixel Accuracy (17). We also compared the accuracy of these models using a lesser number of annotated images in the training set and found this accuracy for the SAM2-UNet model to be ≥ to the second-best performing SAM2-L model in the original study. The inference times of these different models were measured to help select the optimal model for different use cases. Our experiments demonstrated the potential of these advanced models for accurate and efficient Seagrass detection in underwater environments, especially in shallow water, thus enabling real-time Seagrass monitoring. This capability could significantly enhance marine ecological research and conservation efforts by providing rapid, accurate assessments of Seagrass coverage and distribution.

A limitation of our methods was that they were designed specifically for images from shallow water areas and are hence not applicable to aerial images. In future, our research will continue to focus on fine-tuning SAM2 variant models to develop a more generalized segmentation model. We plan to incorporate additional Seagrass datasets from diverse marine environments including aerial images to enhance the model's versatility and robustness. Our ultimate aim is to create a generic Seagrass model capable of zero-shot semantic segmentation. This ongoing work has the potential to revolutionize Seagrass monitoring and conservation efforts globally, and provide marine ecologists and environmental researchers with tools for assessing and protecting these important coastal habitats.

Acknowledgment

I would like to thank Professor Bo Yang from the Department of Environmental Studies at UC Santa Cruz for his mentorship and support throughout this research project. His guidance and constructive feedback significantly shaped my work. I truly appreciate the time he invested in my development as a researcher.

6. References

1. Ondiviela B, Losada IJ, Lara JL, Maza M, Galv´an C, Bouma TJ, et al. The role of Seagrass es in coastal protection in a changing climate. *Coastal Engineering*, 87,158–168, 2014. https://doi.org/10.1016/j.coastaleng.2013.11.005

2. Maxwell SP, Eklof SJ, Katwijk MM, O'Brien KR, Torre-Castro M, Bostrom C, et al. The fundamental role of ecological feedback mechanisms for the adaptive management of Seagrass ecosystems–a review. *Biological Reviews*, 92(3),1521–1538, 2017. https://doi.org/10.1111/brv.12294

3. Jeon E, Kim S, Park S, Kwak J, Choi I. Semantic segmentation of Seagrass habitat from drone imagery based on deep learning: A comparative study. *Ecological Informatics*, 66,101430, 2021. https://doi.org/10.1016/j.ecoinf.2021.101430

4. Burguera A. Segmentation through patch classification: A neural network approach to detect posidonia oceanica in underwater images. *Ecological Informatics*, 56,101053, 2020. https://doi.org/10.1016/j.ecoinf.2020.101053

5. Hobley B, Arosio R, French G, Bremner J, Dolphin T, Mackiewicz M. Semi-supervised segmentation for coastal monitoring Seagrass using rpa imagery. *Remote Sensing*, 13(9), 2021. https://www.mdpi.com/2072-4292/13/9/ 1741

6. Wang S, Mizuno K, Tabeta S, Terayama K, Sakamoto S, Sugimoto Y, et al. An efficient segmentation method based on semi-supervised learning for seafloor monitoring in pujada bay, Philippines. *Ecological Informatics*, 78,102371, 2023. https://doi.org/10.1016/j.ecoinf.2023.102371

7. Mehrubeoglu M, Vargas I, Huang C, Cammarata K. Segmentation of Seagrass blade images using deep learning. Proc. SPIE 11736, Real-Time Image Processing and Deep Learning 2021, 1173605, 2021. https://doi.org/10.1117/12.2587057

8. Raine S, Marchant R, Moghadam P, Maire F, Kettle B, Kusy B. Multi-species Seagrass detection and classification from underwater images. *2020 Digital Image Computing: Techniques and Applications (DICTA)*, Melbourne, Australia, 2020, pages 1-8. https://doi.org/10.1109/DICTA51227.2020.9363371

9. Noman MK, Islam SM, Abu-Khalaf, J, Jalali SM, Lavery, P. Improving accuracy and efficiency in Seagrass detection using state-of-the-art AI techniques. *Ecological Informatics*, 76,102047, 2023. https://doi.org/10.1016/j.ecoinf.2023.102047

10. Noman MK, Islam SM, Abu-Khalaf J, Lavery P. Multi-species Seagrass Detection Using Semi-supervised Learning. *International Conference on Image and Vision Computing New Zealand (IVCNZ),* 2021, pp. 1-6, https://doi.org/10.1109/IVCNZ54163.2021.9653222

11. Islam, KA. Deep Learning Approaches for Seagrass Detection in Multispectral Imagery. *Dissertation*, Electrical & Computer Engineering, Old Dominion University, 2021. https://doi.org/10.25777/gct9-yr76

12. Chen J, Sasaki J. Mapping of subtidal and intertidal Seagrass meadows via application of the Feature Pyramid Network to unmanned aerial vehicle orthophotos. *Remote Sensing*, 13(23), 4880, 2021. https://doi.org/10.3390/rs13234880

13. Savinov CA. Seagrass segmentation with multispectral MBES using implicit neural representations. Master’s thesis, KTH Royal Institute of Technology, Stockholm, SE, 2024. https://kth.diva-portal.org/smash/record.jsf?pid=diva2%3A1901059&dswid=-3139

14. Noman MK, Islam SM, Jafar Jalali SM, Abu-Khalaf J, Lavery P. Baos-cnn: A novel deep neuroevolution algorithm for multispecies Seagrass detection. *Plos one,*19(6), e0281568, 2024. https://doi.org/10.1371/journal.pone.0281568

15. Chen X, Shi P, and Hu Y. A precise semantic segmentation model for seabed sediment detection using yolo-c. *Journal of Marine Science and Engineering*, 11(7),1475, 2023. https://doi.org/10.3390/jmse11071475

16. Roelfsema C. Subtidal Seagrass detector: development of a deep learning model to detect Seagrass in underwater imagery. *Frontiers in Marine Science*, 10,1197695, 2023. https://doi.org/10.3389/fmars.2023.1197695

17. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, et al. Segment anything. *arXiv preprint arXiv:2304.02643*, 2023. https://doi.org/10.48550/arXiv.2304.02643

18. Ke L, Ye M, Danelljan M, Liu Y, Tai YW, Tang CK, et al. Segment anything in high quality. *arXiv preprint arXiv:2306.01567*, 2023. https://doi.org/10.48550/arXiv.2306.01567

19. Xu H, Zhang Y, Zhang J, Zhao R, Shan L, Zhang B. Prompt adapter: Prompt-based adapter for segment anything model. *arXiv preprint arXiv:2305.15036*, 2023. https://doi.org/10.48550/arXiv.2408.08870

20. Feichtenhofer C, Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, et al. Sam 2: Segment anything in images and videos. *arXiv preprint arXiv:2408.00714*, 2024. https://doi.org/10.48550/arXiv.2408.00714

21. Xiong X, Wu Z, Tan S, Li W, Tang F, Chen Y, et al. SAM2-UNet: Segment anything 2 makes strong encoder for natural and medical image segmentation. *arXiv preprint arXiv:2408.08870*, 2024. https://doi.org/10.48550/arXiv.2408.08870

22. Raine S, Maire F, Suenderhauf N, Fischer T. Reducing label dependency for underwater scene understanding: A survey of datasets, techniques and applications. *arXiv preprint arXiv.2411.11287*, 11 2024. https://doi.org/10.48550/arXiv.2411.11287

23. Ronneberger O, Fischer P, and Brox T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computerassisted intervention*, pages 234–241. Springer, 2015. https://doi.org/10.1007/978-3-319-24574-4_28

24. Ryali C, Hu YT, Bolya D, Wei C, Fan H, Huang PY, et al. Hiera: A hierarchical vision transformer without the bells-andwhistles. In *Proceedings of Machine Learning Research*, 202, 2023. http://dx.doi.org/10.48550/arXiv.2306.00989

25. Reus G, Moller T, J¨ager J, Schultz ST, Kruschel C, Hasenauer J, et al. Looking for Seagrass: Deep learning for visual coverage estimation. In *2018 OCEANS - MTS/IEEE Kobe Techno-Oceans (OTO)*, pages 1–6, May 2018. https://doi.org/10.1109/OCEANSKOBE.2018.8559302