SJER Scientific Journal of Engineering Research

futuristech PT. Teknologi Futuristik Indonesia

Date of publication April 18, 2026, date of current version April 18, 2026.
e-ISSN: 3109-1725
Digital Object Identifier 10.64539/sjer.v2i2.2026.454

Article

Parameter-Efficient Fine-Tuning for Sonar Shipwreck Segmentation: A Seed Averaged Study with SegFormer and LoRA

Shehan Maxwell Beruwalage[1,*], Chunyong Yin[2], Muhammad Raza[1], Deshan Sachintha Kannangara[1], Sachini Amani Hendavitharana[3]

[1] School of Computer Science and Technology, Nanjing University of Information Science and Technology, Nanjing, 210044, China; e-mail: Shehanbbn@gmail.com (S. M. Beruwalage), Mrazabng125@gmail.com (M. Raza), deshansachinth777x@gmail.com (D. S. Kannangara).
[2] School of Cyber Science and Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China; e-mail: Yinchunyong@hotmail.com (C. Yin).
[3] School of Artificial Intelligence, Nanjing University of Information Science and Technology, Nanjing, 210044, China; e-mail: sachinivitharana90@gmail.com (S. A. Hendavitharana).
* Correspondence Author

The authors received no financial support for the research, authorship, and/or publication of this article.

Abstract: Accurate segmentation of shipwreck targets in sonar imagery is important for underwater archaeology, marine monitoring, and search operations, but the task remains difficult because labeled sonar masks are scarce and full adaptation of transformer models can be computationally expensive. This study evaluates whether parameter-efficient fine-tuning can provide a practical alternative for binary sonar shipwreck segmentation. Using SegFormer-B0 initialized from a pretrained checkpoint, three adaptation strategies were compared under a consistent protocol: full fine-tuning of all model parameters (FullFT), training only the segmentation head (Head-only), and LoRA-based adaptation of selected linear layers together with head training (LoRA-A+Head). Models were selected by the best validation epoch and evaluated on a held-out test set. Across three random seeds, FullFT achieved the best performance, with a Dice score of 0.614 ± 0.008 and IoU of 0.487 ± 0.007. LoRA-A+Head achieved a Dice score of 0.546 ± 0.010 and IoU of 0.401 ± 0.008 while updating only 1.57% of the parameters, whereas Head-only reached 0.494 ± 0.010 Dice and 0.354 ± 0.008 IoU. These results show a clear accuracy efficiency trade off, full fine-tuning gives the highest accuracy, whereas LoRA-A+Head offers a practical option when reducing the number of updated parameters is important. The findings support the use of parameter-efficient adaptation for sonar segmentation in compute-limited settings.

Keywords: Parameter-efficient fine-tuning; Sonar shipwreck segmentation; SegFormer-B0; LoRA; Model efficiency; Dice; IoU; Training efficiency; Segmentation accuracy.

1. Introduction

Sonar shipwreck segmentation is an important task in marine exploration, Deep Sea rescue operations, underwater archaeology, and ocean monitoring, where identifying submerged wreck structures helps researchers analyze maritime history and underwater environments. The task focuses on generating pixel level segmentation masks that accurately separate shipwreck objects from the surrounding seabed and underwater terrain.

Sonar imagery is difficult to analyze because it often contains noise, shadow effects, irregular textures, and low contrast between objects and background. These factors make segmentation a challenging computer vision problem compared to natural image segmentation tasks [1], [2]. In addition, high-quality labeled datasets are limited because creating ground-truth segmentation masks for sonar images requires expert knowledge and extensive manual annotation effort [3]. Data collection is also expensive and hazardous due to severe weather conditions, heavy storms, and near-zero visibility at depth. Due to these limitations, segmentation models must be able to perform well even when training data is relatively small or imbal-

anced. Deep learning techniques have significantly improved image segmentation performance in many domains, including medical imaging, satellite imagery, and underwater analysis.

In recent years, transformer-based architectures have become highly effective for semantic segmentation because they capture global context and long-range dependencies within images [4]. One notable architecture is SegFormer, which combines hierarchical feature extraction with transformer-based attention mechanisms to produce accurate segmentation results with efficient computation [5]. Several studies have demonstrated that transformer-based models outperform many traditional convolutional neural network (CNN) approaches in complex segmentation tasks, including maritime and sonar imaging applications [5]. Despite these advantages, adapting these models to specialized tasks such as shipwreck segmentation can still be computationally demanding. A key limitation of full fine-tuning is that it requires updating all model parameters, a process commonly referred to as Full Fine-Tuning (FullFT).

While FullFT often provides the best performance, it requires significant computational resources, including GPU memory, training time, and energy consumption [6], [7]. This challenge is further compounded when working with large transformer models containing millions of parameters that must be updated during training. Research has shown that the cost of training and adapting such models can limit their use in smaller research environments or projects with restricted hardware resources [8]. Consequently, researchers have started exploring alternative methods that reduce training cost while maintaining strong performance. Parameter-Efficient Fine-Tuning (PEFT) methods have emerged as a promising approach to address the limitations of full model training. Rather than updating all parameters, PEFT methods modify only a small portion of the model while keeping most pretrained weights unchanged. One of the most widely used PEFT techniques is Low-Rank Adaptation (LoRA), which introduces additional low-rank matrices into selected layers of the model [9]. This method allows models to adapt to new tasks with significantly fewer trainable parameters while preserving most of the original model knowledge [10]. Previous studies have demonstrated that PEFT approaches can achieve competitive results while reducing computational cost and memory requirements [11]. In addition, LoRA-based methods have shown promising results in applications related to underwater imaging and maritime analysis [1].

Despite growing interest in PEFT methods, their application in sonar-based segmentation tasks remains limited. Many previous studies have focused on general image segmentation or object detection tasks, rather than domain-specific challenges such as sonar shipwreck detection [12]. Furthermore, there is limited empirical analysis comparing PEFT methods with full fine-tuning approaches in sonar imagery datasets. As a result, the trade-off between segmentation accuracy and computational efficiency in this domain is not yet well understood [13].

This study aims to evaluate the effectiveness of different model adaptation strategies for binary sonar shipwreck segmentation using the SegFormer-B0 architecture. Three training strategies are investigated:

- Full Fine-Tuning (FullFT), where all model parameters are updated during training.
- Head-only tuning, where only the segmentation head is trained while the backbone remains frozen.
- LoRA-A+Head adaptation, where LoRA modules are applied to selected layers while also training the segmentation head.

To ensure reproducibility and reliable evaluation, the study follows a structured experimental protocol. Each adaptation method is trained using three different random seeds (123, 456, and 789) to measure stability across experiments [14]. The best-performing model is selected based on validation performance, and final results are reported using a held-out test dataset. Performance is evaluated using segmentation metrics such as Dice coefficient and Intersection over Union (IoU), which are commonly used in segmentation research [7]. This research is guided by the following key questions:

1) How much segmentation performance is affected when replacing full fine-tuning with parameter-efficient fine-tuning methods for sonar shipwreck segmentation?
2) What is the trade-off between computational efficiency and segmentation accuracy among FullFT, Head-only, and LoRA-based adaptation strategies?

The contributions of this study are as follows:

- Provides a direct benchmark of FullFT, Head-only, and LoRA-A+Head for binary sonar shipwreck segmentation using SegFormer-B0.
- Uses a reproducible evaluation protocol with three random seeds, validation-based checkpoint selection, and final reporting on a held-out test set.
- Reports both segmentation accuracy and practical efficiency indicators, including Dice, IoU, trainable-parameter percentage, peak VRAM, and time per epoch.
- Presents quantitative and qualitative analyses that clarify the trade-off between accuracy and parameter efficiency in sonar segmentation.

Together, these contributions provide a compute-aware benchmark for adapting transformer models to sonar shipwreck segmentation.

The remainder of this paper is organized as follows. Section 2 reviews related studies. Section 3 describes the dataset, model architecture, adaptation strategies, and

evaluation protocol. Section 4 presents the results and discussion. Section 5 concludes the paper and outlines limitations and future work.

2. Related Work

2.1. Deep Learning & Transformer-Based Models for Sonar Image Segmentation

Sonar image analysis has gained increasing attention in recent years due to its importance in underwater exploration, marine monitoring, and shipwreck detection. Traditional image processing methods often struggle with sonar data because sonar images contain noise, shadows, and low contrast, making object boundaries difficult to detect accurately [1]. Recent research has shown that deep learning models can significantly improve segmentation performance in sonar-based applications compared to conventional techniques [15]. several studies have applied convolutional neural networks (CNNs) and deep learning approaches to detect underwater structures & seabed anomalies. However, many of these approaches rely heavily on large labeled datasets, which are often unavailable in real-world sonar datasets [13].

Transformer architectures have recently emerged as powerful models for computer vision tasks, including semantic segmentation. Unlike CNN-based models, transformers are capable of capturing global contextual information across an entire image, which improves segmentation accuracy in complex scenes [4]. The SegFormer architecture has been proposed as an efficient transformer-based model designed specifically for semantic segmentation tasks [5]. SegFormer combines hierarchical feature representation with lightweight attention mechanisms, allowing it to perform well while maintaining computational efficiency. These advances suggest that transformer models may also be effective for sonar shipwreck segmentation tasks.

2.2. Challenges in Fine-Tuning Large Models

Despite their strong performance, transformer models are often large and require significant computational resources to fine-tune on new datasets. Full fine-tuning involves updating all model parameters during training, which can be computationally expensive and time-consuming [6]. This problem becomes more severe in research environments with limited hardware resources or when working with small datasets. Several studies have highlighted that training large models can increase GPU memory usage, training time, and energy consumption [16]. These challenges motivate the development of more efficient training strategies that can reduce computational requirements without sacrificing model performance.

2.3. Parameter-Efficient Fine-Tuning (PEFT)

Parameter-Efficient Fine-Tuning (PEFT) techniques have recently been introduced to address the limitations of full model training. instead of updating the entire model, PEFT methods modify only a small subset of parameters, making training faster and more efficient. One of the most widely used PEFT techniques is Low-Rank Adaptation (LoRA), which introduces trainable low-rank matrices into selected layers of the model [9]. LoRA allows models to adapt to new tasks while keeping most pretrained weights unchanged, reducing memory and computation requirements. Research has shown that LoRA and similar approaches can achieve competitive performance compared to full fine-tuning while significantly reducing the number of trainable parameters [11], [17]. These methods have been successfully applied in various domains, including natural language processing, computer vision, and underwater imaging tasks [18].

2.4. PEFT in Sonar Applications & Research Gap

Although PEFT techniques have been widely studied in general machine learning tasks, their application in marine and sonar imaging is still limited. Some recent studies have explored efficient model adaptation strategies for maritime object detection and underwater imaging systems [12]. other research has investigated deep learning approaches for sonar segmentation, but most of these studies rely on conventional fine-tuning methods rather than parameter-efficient approaches [14]. Additionally, there is limited empirical comparison between different adaptation strategies in sonar datasets, particularly when evaluating both accuracy and efficiency [19].

Based on the existing literature, several key limitations can be identified Limited research specifically focusing on transformer-based segmentation for sonar shipwreck detection. Lack of comparative studies evaluating PEFT methods in sonar datasets. Insufficient analysis of the trade-off between segmentation performance and computational efficiency in underwater imaging models. Addressing these gaps is important for developing efficient and practical segmentation models that can be used in real-world marine research environments. This study builds upon previous research in sonar image segmentation and parameter-efficient fine-tuning methods. Unlike earlier works, this research provides a direct comparison between three adaptation strategies,

- Full Fine-Tuning (FullFT).
- Head-only training.
- LoRA-based adaptation combined with head training.

The study focuses on evaluating both segmentation accuracy and computational efficiency, which is a key factor for real-world deployment of deep learning models in underwater exploration tasks.

3. Proposed Method

This section describes the methodology used to evaluate parameter-efficient fine-tuning strategies for sonar

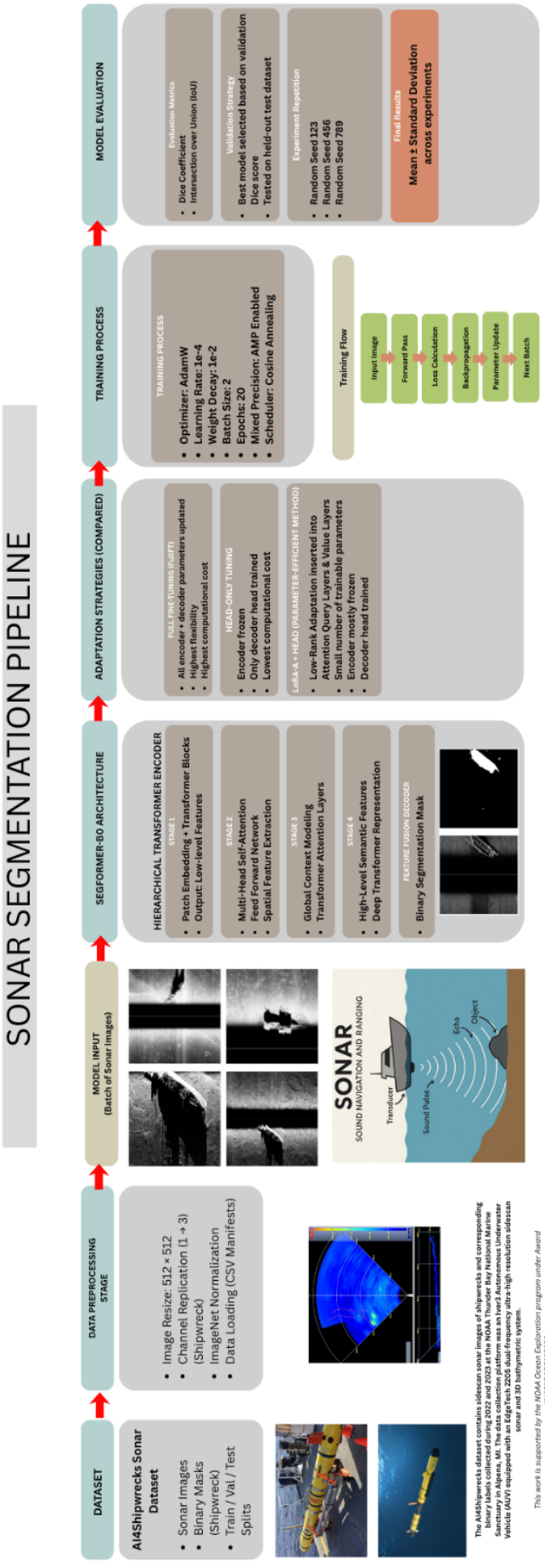


Figure 1. Overview of the proposed framework based on the SegFormer-B0 architecture with different fine-tuning strategies.

shipwreck segmentation. The approach is based on a transformer segmentation architecture and compares different adaptation strategies under a consistent experimental setup. The overall objective is to analyze how efficiently pretrained transformer models can be adapted to sonar imagery with limited training data while maintaining segmentation performance. The overall framework of the proposed sonar segmentation pipeline is illustrated in Figure 1.

3.1. Model Architecture

The backbone used in this study is the SegFormer-B0 model, a transformer-based architecture designed for semantic segmentation tasks. SegFormer combines hierarchical transformer encoders with a lightweight decoding head, enabling efficient segmentation without heavy computational overhead [5]. Transformer-based segmentation models have demonstrated strong performance in various computer vision applications due to their ability to capture global contextual information across an image [4]. The SegFormer-B0 model used in this work is initialized from the pretrained checkpoint nvidia/segformer-b0-finetuned-ade-512-512, which was originally trained on the ADE20K dataset for multi-class semantic segmentation. Since the task in this study involves binary segmentation (shipwreck vs background), the final classification layer of the model is modified to output a single-channel segmentation mask. Each pixel prediction represents the probability that the pixel belongs to a shipwreck object. Using pretrained transformer models has been shown to improve segmentation performance when labeled datasets are limited, particularly in specialized domains such as sonar imaging and underwater object detection [20]. The detailed architecture of the SegFormer-B0 model is presented in Figure 2.

3.2. Adaptation Strategies

To evaluate the trade-off between computational efficiency and segmentation performance, three adaptation strategies are investigated in this study.

3.2.1. Head-only Tuning

The Head-only tuning strategy freezes the pretrained backbone and updates only the segmentation decoder head. This significantly reduces the number of trainable parameters and computational cost while still allowing the model to adapt to the new segmentation task. Previous studies have shown that freezing backbone layers while training task-specific heads can provide a lightweight baseline when computational resources are limited [8], [16]. However, since the encoder features remain unchanged, the model may struggle to capture domain-specific characteristics present in sonar imagery.

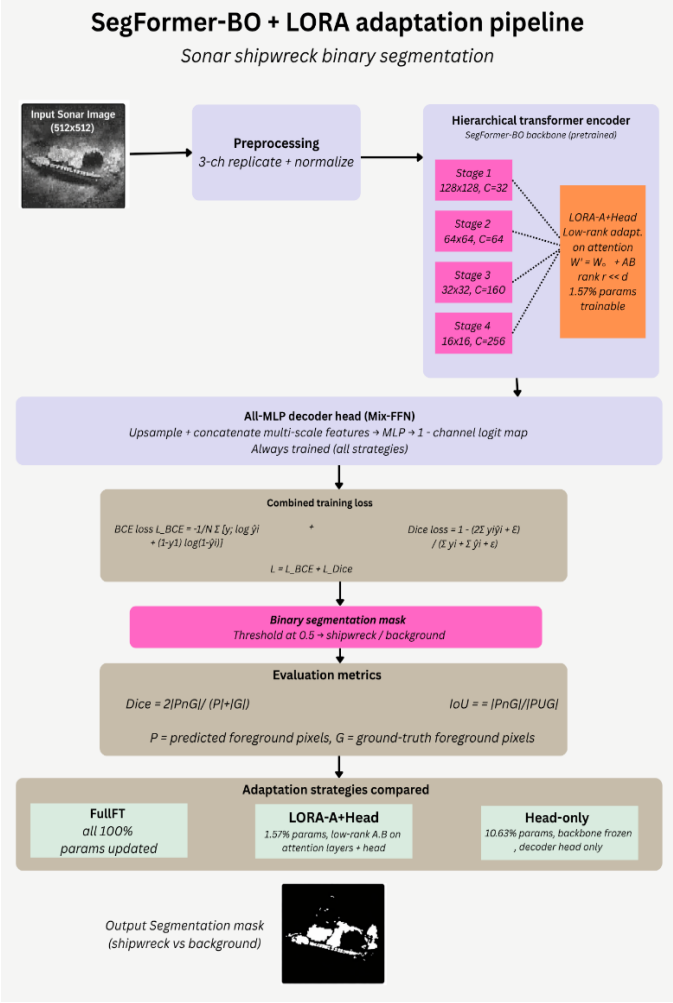


Figure 1. Architecture of the proposed SegFormer-B0 adaptation pipeline for binary sonar shipwreck segmentation.

3.2.2. Full Fine-Tuning (FullFT)

In the Full Fine-Tuning approach, all model parameters are updated during training, as illustrated in Figure 1. This includes both the transformer encoder layers and the decoder head responsible for generating segmentation masks. Full fine-tuning is commonly used when adapting pretrained models to new tasks, as it allows the model to fully adjust to domain-specific features [5]. Although this method typically provides the highest segmentation accuracy, it also requires substantial computational resources due to the large number of trainable parameters in transformer-based models [11]. Therefore, FullFT serves as a performance benchmark for comparison with more efficient adaptation strategies.

The figure illustrates the end-to-end pipeline, beginning with sonar image preprocessing (512×512 resizing, 3-channel replication, and ImageNet normalization), followed by the hierarchical transformer encoder comprising four feature extraction stages at progressively reduced spatial resolutions. The LoRA-A+Head adaptation module is shown alongside the encoder attention layers, applying low-rank matrix decomposition ($W' = W_0 + AB$) to update only 1.57% of parameters. Multi-scale features (F1–F4) are passed to the All-MLP decoder head, which produces a single-channel logit map thresholded at 0.5 for binary mask prediction. The combined training loss ($L = L_{BCE} + L_{Dice}$) and evaluation metrics (Dice and IoU) are defined within the figure. The legend at the bottom distinguishes the three adaptation strategies compared in this study: FullFT (100% parameters), LoRA-A+Head (1.57%), and Head-only (10.63%).

3.2.3. LoRA-A+Head (Parameter-Efficient Fine-Tuning)

Low-Rank Adaptation (LoRA) is a parameter-efficient fine-tuning technique that introduces additional low-rank matrices into selected linear layers of a pretrained model [11]. Instead of updating the full model, LoRA modifies only a small subset of parameters, reducing training cost while maintaining performance. in this study, LoRA modules are applied to the attention layers of the SegFormer encoder, while the decoder head is also trained simultaneously. This strategy enables the model to adapt to sonar-specific features while keeping most of the pretrained weights frozen. Parameter-efficient fine-tuning methods such as LoRA have been shown to achieve competitive results with significantly fewer trainable parameters compared to full fine-tuning approaches [11].

3.2.4. Evaluation Metric Equations

Segmentation performance was evaluated using the Dice coefficient and Intersection over Union (IoU). Let Pdenote the set of predicted foreground pixels and Gdenote the set of ground-truth foreground pixels. The Dice coefficient is defined as:

$$Dice = \frac{2 \mid P \cap G \mid}{\mid P \mid + \mid G \mid} \quad (1)$$

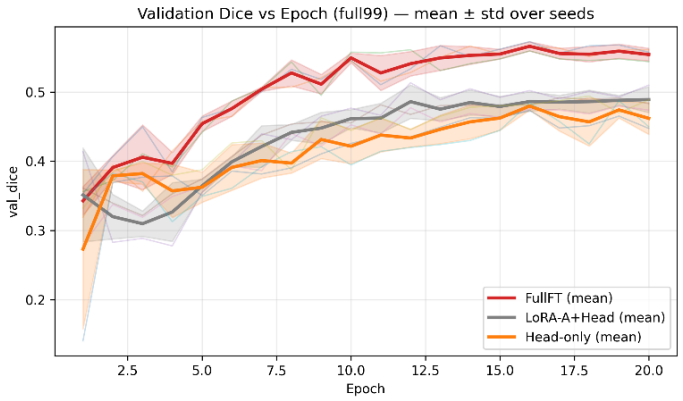
and IoU is defined as:

$$IoU = \frac{\mid P \cap G \mid}{\mid P \cup G \mid} \quad (2)$$

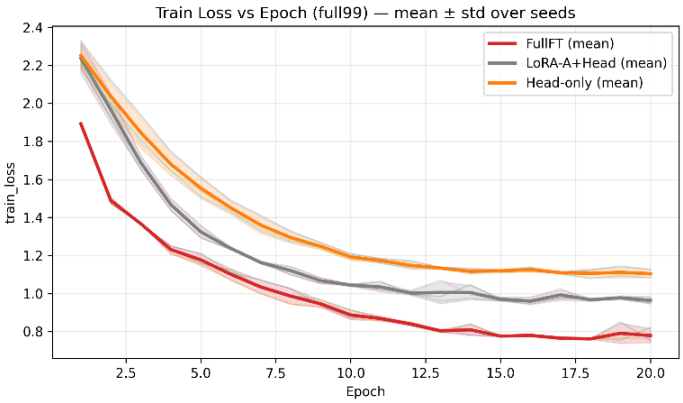
where $|\cdot|$ denotes the number of foreground pixels. For numerical stability in implementation, a small constant εcan be added to the numerator and denominator when needed. Predictions were thresholded at 0.5 to obtain binary masks. In addition, samples with empty ground-truth masks were excluded from metric aggregation under the EMPTY_POLICY = ignore setting to avoid inflating performance with trivial empty cases.

The training objective combines binary cross entropy loss and Dice loss to balance pixel-wise classification and region overlap quality:

$$L = L_{BCE} + L_{Dice} \quad (3)$$



(a)



(b)

Figure 3. Training and validation curves across epochs (validation Dice and train loss).

where,

$$L_{\mathrm{BCE}} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)] \quad (4)$$

and

$$L_{\mathrm{Dice}} = 1 - \frac{2\sum_{i=1}^{N} y_i\,\hat{y}_i + \varepsilon}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \hat{y}_i + \varepsilon} \quad (5)$$

Here, $y_i$is the ground-truth label for pixel i, $\hat{y}_i$is the predicted probability, $N$is the number of pixels, and $\varepsilon$is a small constant for numerical stability. This combined objective was used for all compared methods.

3.3. Dataset

The dataset used in this study is the AI4Shipwrecks dataset, which consists of sonar images annotated with binary masks representing shipwreck regions. images of shipwrecks and corresponding binary labels collected during 2022 and 2023 at the NOAA Thunder Bay National Marine Sanctuary in Alpena, MI. The data collection platform was an Iver3 Autonomous Underwater Vehicle (AUV) equipped with an EdgeTech 2205 dual-frequency ultra-high resolution sidescan sonar and 3D bathymetric system. The labels were compiled from reference labels created by experts in marine archaeology. [21] To ensure reproducibility, the dataset is divided into three subsets. Training split, Validation split, Test split. Images with empty ground-truth masks are excluded from evaluation using the "EMPTY_POLICY = ignore" setting to maintain consistent performance measurement.

3.4. Data Preprocessing

Before we conducted the training, several preprocessing steps are applied to the dataset to ensure compatibility with the SegFormer model. First, all images are resized to 512 × 512 pixels to match the input resolution expected by the model. Second, since sonar images are typically single-channel, they are replicated into three channels to align with models pretrained on RGB datasets such as ImageNet. Finally, image normalization is applied using ImageNet mean and standard deviation values, which is a common preprocessing approach for pretrained vision models [1].

Figure 3 illustrate the optimization behavior of the compared methods across training epochs. This figure shows the overall training dynamics, including both optimization behavior (train loss) and performance progression (validation Dice) across epochs. FullFT consistently reached the highest validation Dice and IoU values, which is consistent with its superior test performance. LoRA-A+Head followed a similar trend but converged to a lower plateau, indicating that the parameter-efficient updates captured part, but not all, of the domain adaptation gained by full fine-tuning. Head-only plateaued earlier and at a lower level, suggesting that restricting optimization to the decoder head limited the model's ability to learn sonar-specific representations. The curves also show relatively consistent trends across seeds, which supports the claim of stable optimization under the fixed experimental split.

3.5. Training and Evaluation Protocol

A consistent training and evaluation protocol is used to ensure fair comparison among the different adaptation strategies. Models are trained for 20 epochs with a batch size of 2. The "AdamW optimizer" is used with a learning rate of 1e-4 and weight decay of 1e-2. A cosine annealing learning rate scheduler is applied during training. Mixed-precision training using Automatic Mixed Precision (AMP) is enabled to improve training efficiency and reduce GPU memory usage. Model selection is based on the best validation Dice score, and the final performance is evaluated on a held-out test dataset. Segmentation performance is measured using Dice coefficient and Intersection over Union (IoU). These metrics are widely used in segmentation research for evaluating model performance [7]. we ensure reproducibility of each experiment using three repeated random seeds such as "123", "456", "789". Results are reported as mean ± standard deviation across the three runs.

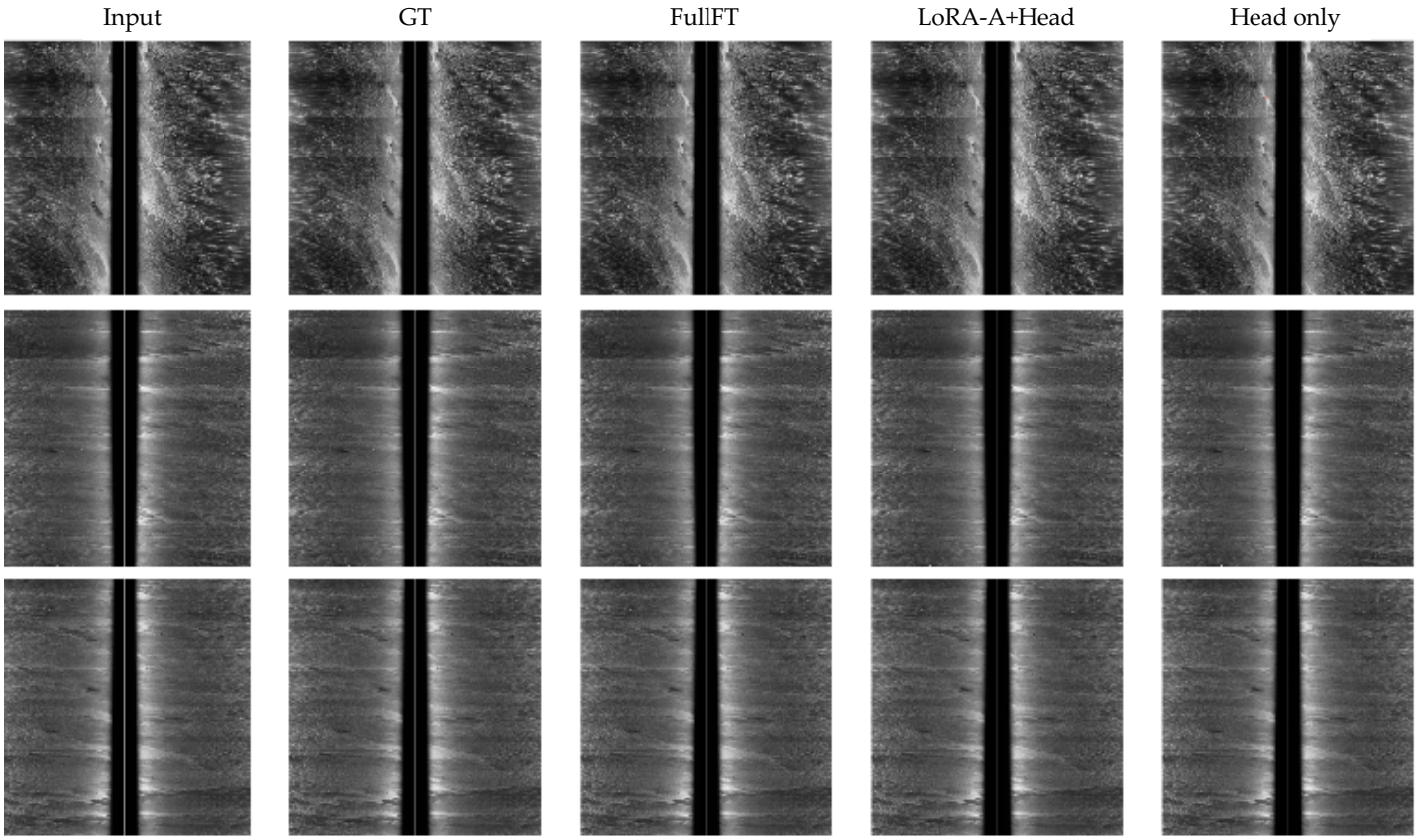


Figure 4. Failure cases showing segmentation challenges in low-contrast sonar images.

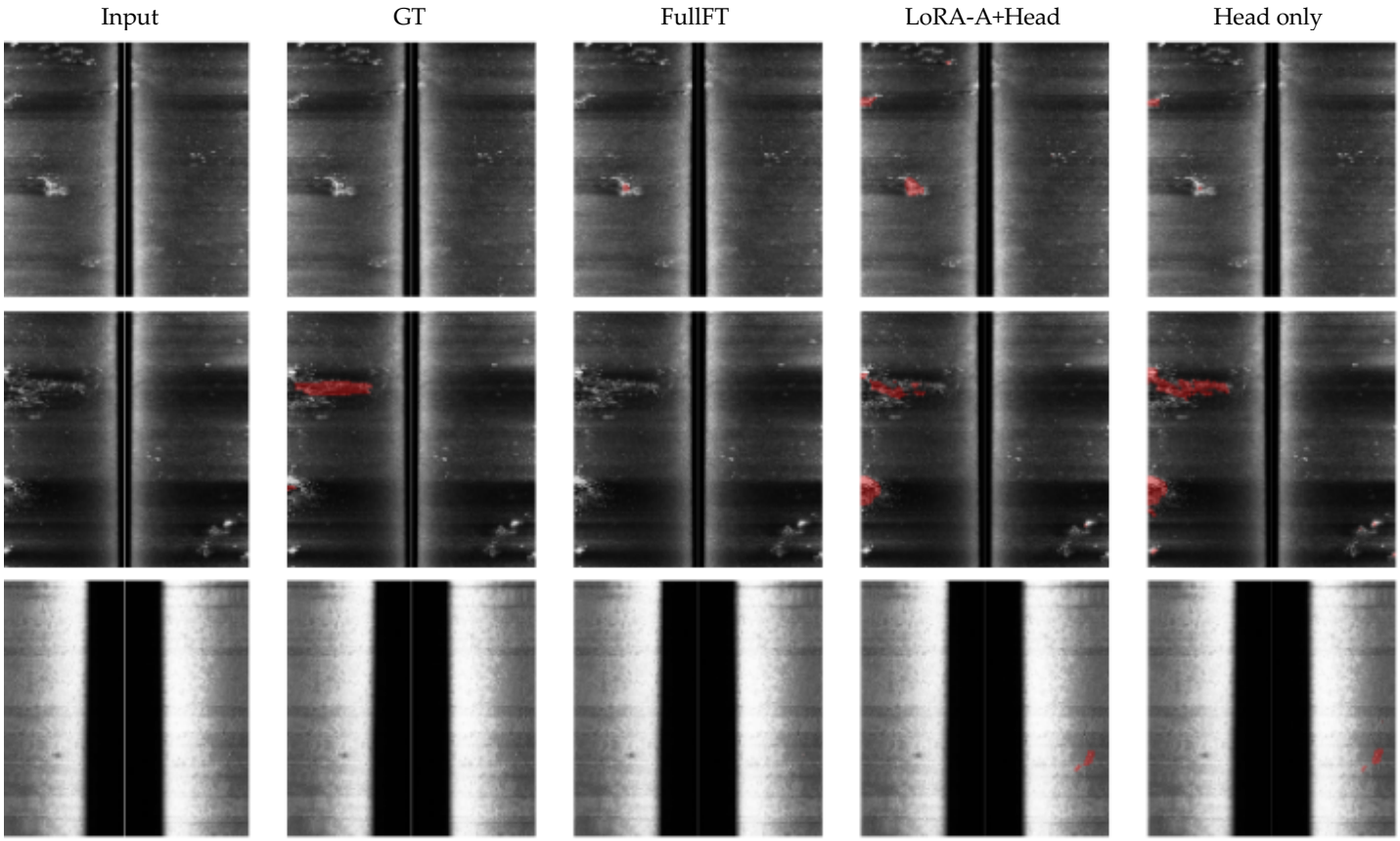


Figure 5. Example segmentation results comparing FullFT, LoRA-A+Head, and Head-only methods on selected test images. Qual panel good.

The qualitative examples shown in Figure 5 further clarify the differences among the three adaptation strategies. In representative high-contrast cases, all methods were able to identify the shipwreck region reasonably well, although FullFT generally produced cleaner and more spatially continuous masks. LoRA-A+Head often

Table 1. Seed-averaged test performance (mean ± std over three seeds) and efficiency metrics for the compared adaptation strategies.

Method	Seeds	Dice (mean ± std)	IoU (mean ± std)	Trainable (%)	VRAM (GB)	Time/Epoch (s)
FullFT	3	0.614 ± 0.008	0.487 ± 0.007	100	0.89	7.8
LoRA-A+Head	3	0.546 ± 0.010	0.401 ± 0.008	1.57	0.84	8.0
Head-only	3	0.494 ± 0.010	0.354 ± 0.008	10.63	0.65	7.1
LoRA-A only(f)	3	0.295 ± 0.034	0.198 ± 0.022	N/A	N/A	N/A

Table 2. Additional seed-averaged accuracy comparison including the LoRA-A only(f) ablation.

Method	Seeds	Dice (mean ± std)	IoU (mean ± std)
FullFT	3	0.567 ± 0.006	0.458 ± 0.007
LoRA-A+Head	3	0.495 ± 0.017	0.376 ± 0.015
Head-only	3	0.485 ± 0.011	0.362 ± 0.009
LoRA-A only(f)	3	0.295 ± 0.034	0.198 ± 0.022

preserved the main wreck structure but showed occasional boundary fragmentation or partial omission in difficult regions. Head-only was more prone to under-segmentation and boundary loss, especially when the sonar target exhibited weak contrast or irregular texture. In failure cases, all methods struggled when the target boundary was faint or the seabed pattern introduced strong visual ambiguity, as illustrated in Figure 4. These observations are consistent with the quantitative results and suggest that low-contrast sonar scenes remain a major source of error for all compared strategies.

The combined training loss is defined in Section 3.2.4 (Equation 6) and applied consistently across all compared methods.

$$L = L_{Bce} + L_{Dce} \quad (6)$$

4. Experimental Results and Discussion

This section presents the quantitative and qualitative results for evaluating the three adaptation strategies Full Fine-Tuning (FullFT), Head only Tuning, and LoRA-A+Head for sonar shipwreck segmentation using SegFormer-B0 Results are reported on the full99 dataset split, as well as few-shot settings, with all experiments conducted using three random seeds (123, 456, 789). Metrics include Dice coefficient, Intersection over Union (IoU), trainable parameters, VRAM usage, and time per epoch.

4.1. Main quantitative results (FULL99 Split)

4.1.1. Validation Performance

These results reveal a clear trade-off between computational efficiency and segmentation performance across the three adaptation strategies. FullFT achieves the highest Dice and IoU on validation, providing the strongest segmentation accuracy. LoRA-A+Head provides a reasonable trade-off between accuracy and efficiency, updating only 1.57% of the parameters. Head-only performs slightly worse than LoRA-A+Head but remains a computationally efficient baseline. LoRA-A only (head frozen) performs poorly, emphasizing the necessity of training the decoder head alongside the encoder.

The held-out test performance and efficiency metrics are summarized in Table 1. FullFT achieved the strongest segmentation accuracy, with a Dice score of 0.614 ± 0.008 and IoU of 0.487 ± 0.007, showing that full adaptation of the pretrained model remained the most effective option in this dataset. LoRA-A+Head achieved the second-best performance, with 0.546 ± 0.010 Dice and 0.401 ± 0.008 IoU, and outperformed Head-only by a clear margin. This indicates that limited encoder adaptation through low-rank updates is more effective than restricting learning to the segmentation head alone. At the same time, the measured VRAM and training-time differences among methods were modest in this setup, so the main efficiency benefit of LoRA-A+Head should be interpreted as reduced trainable-parameter budget rather than a large reduction in runtime. For completeness, Table 2 provides an additional accuracy comparison including LoRA-A only(f), which yielded substantially lower accuracy (Dice 0.295 ± 0.034, IoU 0.198 ± 0.022). Because the corresponding efficiency metrics were not recorded, its trainable-parameter, VRAM, and time-per-epoch fields are reported as N/A.

4.1.2. Test Performance

Seed-averaged test results were evaluated using test_pairs.csv with best-validation checkpoints. Table 1 shows that FullFT achieved the highest test performance, indicating that updating the full backbone remained the most effective strategy for adapting SegFormer-B0 to the sonar domain. LoRA-A+Head ranked second on both Dice and IoU and clearly outperformed Head-only, which suggests that limited encoder adaptation is more effective than training only the decoder head. LoRA-A only(f), as also reflected in Table 2, remained well below the other methods, confirming that decoder-head adaptation is essential for this task.

However, the efficiency advantage of LoRA-A+Head was expressed mainly in the number of trainable parameters rather than large reductions in runtime or VRAM. Head-only required the least memory and shortest epoch time, but this came with the lowest segmentation accuracy among the three primary methods. Overall, the results indicate that the choice of adaptation strategy depends on whether the priority is maximum accuracy, low memory usage, or reduced parameter updates.

Table 3. Detailed per-seed Dice/IoU stability across seeds.

FullFT	LoRA-A+Head	Head-only
Seed123: Dice 0.6123, IoU 0.4842	Seed123: Dice 0.5372, IoU 0.3936	Seed123: Dice 0.4953, IoU 0.3560
Seed456: Dice 0.6081, IoU 0.4812	Seed456: Dice 0.5561, IoU 0.4090	Seed456: Dice 0.5034, IoU 0.3614
Seed789: Dice 0.6230, IoU 0.4947	Seed789: Dice 0.5439, IoU 0.3995	Seed789: Dice 0.4839, IoU 0.3454

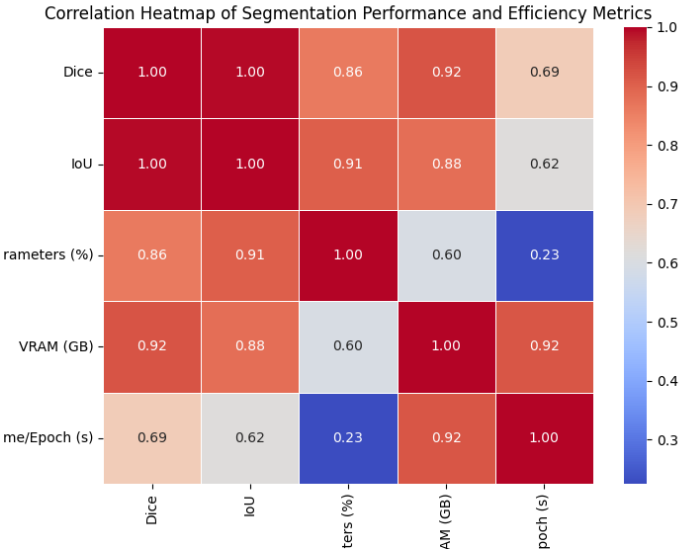


Figure 6. Correlation heatmap of segmentation performance and efficiency metrics across the compared adaptation strategies.

Table 4. Best-validation checkpoints for the three primary compared methods.

Method	Seed	Best Epoch	IoU (mean ± std)
FullFT	123	12	0.561
FullFT	456	14	0.566
FullFT	789	16	0.573
LoRA-A+Head	123	17	0.492
LoRA-A+Head	456	12	0.513
LoRA-A+Head	789	19	0.480
Head-only	123	19	0.487
Head-only	456	18	0.494
Head-only	789	16	0.473

Table 5. Few-shot validation performance. Head-only and LoRA-A only(f) were not evaluated in the few-shot setting and are therefore reported as N/A.

Method	10-shot Dice	10-shot IoU	25-shot Dice	25-shot IoU
FullFT	0.428	0.331	0.454	0.336
LoRA-A+Head	0.438	0.340	0.434	0.335
LoRA-B	0.211	0.137	0.235	0.151

4.1.3. Per-Seed Test Results

Detailed per-seed Dice and IoU results are reported in Table 3. Notably, the low standard deviation across the three random seeds indicates stable optimization behavior under a fixed train/validation/test split. However, because the data split was not varied, this analysis reflects sensitivity to training stochasticity rather than full robustness to data variability. Broader robustness assessment would require repeated experiments with alternative dataset splits or external sonar datasets.

To provide an additional descriptive view of the relationship between segmentation quality and efficiency, Figure 6 presents a correlation heatmap computed from the reported evaluation metrics across the compared adaptation strategies. Figure 6 provides a descriptive summary of how the reported evaluation metrics vary across the compared methods. As expected, Dice and IoU show a very strong positive relationship because both measure segmentation overlap quality. The heatmap also shows that VRAM usage is positively associated with segmentation performance in this small comparison, reflecting that the strongest-performing setting, FullFT, also required the largest resource allocation. By contrast, time per epoch and trainable-parameter percentage do not show equally strong alignment with accuracy, which is consistent with the main results indicating that the advantage of LoRA-A+Head lies primarily in reducing the parameter-update budget rather than in producing large reductions in runtime or memory. Because this analysis is based on a small number of compared methods, it should be interpreted as an exploratory visualization rather than as a statistical claim.

4.1.4. Best on Validation Checkpoints

Per-seed checkpoint details were retained only for FullFT, LoRA-A+Head, and Head-only; the LoRA-A only(f) ablation is therefore summarized separately in Table 2. FullFT, LoRA-A+Head, and Head-only models were selected based on highest validation Dice, as summarized in Table 4.

4.2. Training dynamics and Learning Curves

The few-shot experiments, summarized in Table 5, provide an initial view of adaptation behavior under limited training data. In the 10-shot setting, LoRA-A+Head slightly outperformed FullFT on validation metrics, whereas in the 25-shot setting the two methods showed very similar performance. By contrast, LoRA-B remained clearly weaker in both settings. Head-only and LoRA-A only(f) were not evaluated in the few-shot setting and are therefore marked as N/A in Table 5. Because these experiments were limited in scope and reported only on the validation split, they should be interpreted as exploratory rather than definitive. Nevertheless, the results suggest that parameter-efficient adaptation can remain competitive

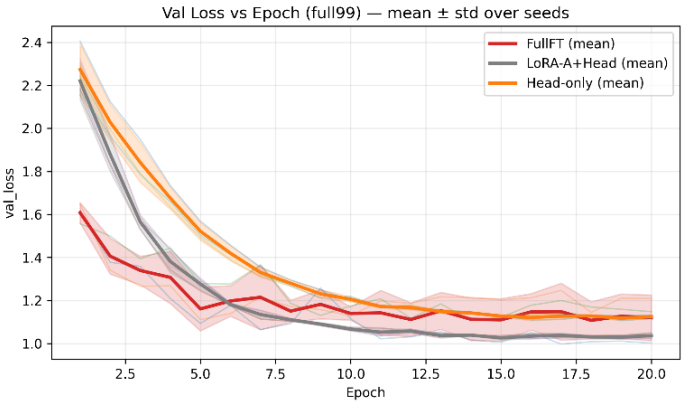


Figure 7. Validation loss trends across epochs.

when the number of labeled training samples is small. Learning curves visualize model convergence and stability across epochs. The validation loss trends across training epochs are illustrated in Figure 7, providing a more detailed view of generalization behavior.

FullFT converges to the highest Dice and IoU values consistently. LoRA-A+Head shows slightly slower convergence but remains competitive throughout training. Head-only plateaus early, achieving lower maximum metrics. In qualitative analysis, FullFT generally produces the most accurate masks, with the best boundary continuity. LoRA-A+Head maintains competitive segmentation quality despite its minimal trainable parameter count. Head-only is prone to errors in low-contrast regions but remains computationally lightweight.

4.3. Few-shot Experiments

In the few-shot experiments, the primary goal was to test how well the model could adapt to sonar shipwreck segmentation with limited training data. To this end, we evaluated the performance of the model using 10-shot and 25-shot settings, where the training data was intentionally restricted to only a few samples. As shown in Table 5, LoRA-A+Head performed competitively with FullFT in these low-data scenarios, achieving Dice scores of 0.438 and 0.434 for the 10-shot and 25-shot settings, respectively, compared with FullFT values of 0.428 and 0.454. LoRA-B remained clearly weaker in both settings. Head-only and LoRA-A only(f) were not evaluated in this few-shot experiment and are therefore reported as N/A in Table 5. These findings provide an initial indication that parameter-efficient adaptation can remain competitive when labeled training data are scarce, although broader few-shot evaluation is still needed.

4.4. Comparison with Prior Studies

When comparing our results with prior work in the field of sonar shipwreck segmentation and parameter-efficient fine-tuning (PEFT), our findings show consistent trends with previous studies while offering some novel insights. Prior studies, such as [9], have demonstrated the efficiency of LoRA-based methods in reducing the number of trainable parameters while retaining competitive accuracy. In our study, we found that LoRA-A+Head offered a good trade-off between accuracy and efficiency, with only 1.57% of parameters being trainable while maintaining strong segmentation performance (Dice: 0.546 ± 0.010, IoU: 0.401 ± 0.008), aligning with similar PEFT strategies reported in [5]. Moreover, our research builds on earlier investigations into the performance of transformer-based models for sonar segmentation, like those seen in [22], by incorporating more advanced adaptation strategies and comparing them against a well-established baseline (FullFT). Notably, our Head-only adaptation strategy, while computationally efficient, falls short in segmentation quality compared to LoRA-A+Head, which highlights the crucial role of encoder feature adaptation for this task. The comparative analysis with other PEFT techniques such as AdapterFusion [17] and quantized LoRA variants [10] further reinforces our conclusion that LoRA-A+Head strikes the best balance for resource-constrained applications without compromising segmentation accuracy too much.

Compared with earlier sonar segmentation studies that primarily emphasized architecture design or conventional fine-tuning, the present work focuses on the adaptation strategy itself. The main contribution is therefore not a new backbone, but a controlled comparison between full fine-tuning, head-only tuning, and LoRA-based parameter-efficient adaptation under the same dataset split and evaluation protocol. The results show that full fine-tuning remains the strongest option in absolute accuracy, while LoRA-A+Head offers a more parameter-efficient alternative with moderate loss in segmentation quality. This positions the study as a compute-aware benchmark for sonar shipwreck segmentation rather than as a claim of state-of-the-art performance across all sonar segmentation methods.

5. Analysis and Interpretation

5.1. Key Findings

Experiments highlight a clear trade-off between segmentation accuracy and computational efficiency across the evaluated adaptation strategies. As for Full Fine-Tuning (FullFT) Achieves the highest segmentation accuracy with a test Dice score of 0.614 ± 0.008 and IoU of 0.487 ± 0.007. Benefits from full access to model capacity, learning both high-level and low-level features requires all model parameters to be trained, resulting in higher computational cost and memory usage (0.89 GB VRAM, 7.8 s/epoch). Produces most contiguous and accurate segmentation masks in qualitative results and handles low-contrast wreck boundaries better than other methods.

LoRA-A+Head offers a strong accuracy–efficiency balance, achieving Dice 0.546 ± 0.010 and IoU 0.401 ± 0.008. Only 1.57% of model parameters are trainable, making it

computationally efficient (0.84 GB VRAM, 8.0 s/epoch). Qualitative results show that LoRA-A+Head approximates FullFT in well-defined wreck scenarios but struggles in low-contrast areas. It is suitable for deployment in resource-constrained environments and aligns with prior research on LoRA-based PEFT efficiency [9], [11].

Head-only is the most computationally efficient strategy, training only 10.63% of parameters with the lowest VRAM usage (0.65 GB) and fastest per-epoch runtime (7.1 s). Test accuracy is lowest among methods (Dice 0.494 ± 0.010, IoU 0.354 ± 0.008), reflecting the inability to adapt encoder features, which are critical for sonar-specific feature extraction. It is suitable for extremely resource-limited scenarios but is not recommended for high-accuracy segmentation tasks.

LoRA-A only (head frozen) performs poorly (Dice 0.295 ± 0.034, IoU 0.198 ± 0.022), demonstrating the importance of adapting the decoder head in PEFT strategies. This near-zero test performance confirms that task-specific adaptation requires both encoder and decoder adjustments. The low standard deviation across random seeds (123, 456, 789) indicates consistent and reproducible performance, reinforcing the reliability of the conclusions drawn from the quantitative results.

5.2. Interpretation of Results

FullFT outperforms other methods by leveraging full model capacity, capturing both low-level sonar features and high-level structural patterns. This aligns with literature showing full fine-tuning of transformer models yields the best segmentation performance [9]. LoRA-A+Head provides a practical alternative to FullFT for situations with limited computational resources, achieving competitive accuracy while substantially reducing trainable parameters. This reflects previous findings that LoRA-based PEFT can maintain performance with fewer updates [11]. Head-only is efficient but unable to fully adapt the encoder, resulting in weaker feature extraction.

This mirrors prior studies showing that decoder-only adaptation is insufficient for domain-specific tasks [7]. LoRA-A only (head frozen) confirms the critical role of the decoder head in task-specific segmentation. Without decoder adaptation, the model fails to generalize to sonar shipwrecks, supporting prior PEFT findings [13].

Regarding efficiency, LoRA-A+Head trains only 1.57% of parameters with moderate VRAM usage and runtime, balancing efficiency and performance and making it suitable for on-board deployment in underwater vehicles. Head-only is the fastest and least memory-intensive strategy, but its lower accuracy makes it appropriate only for high-speed, low-resource applications. FullFT achieves the highest accuracy at the cost of greater computational resources, making it best suited for offline, high-precision tasks where hardware constraints are not a concern. LoRA-A only (head frozen) is not practical due to poor segmentation performance despite its efficiency benefits. From a qualitative standpoint, high-contrast sonar images are well-segmented by all methods, with FullFT providing the best boundary continuity and precision. LoRA-A+Head is slightly less precise but still produces high-quality masks. In failure cases, low-contrast or noisy sonar images reveal boundary fragmentation in LoRA-A+Head and Head-only, highlighting the need for enhanced low-contrast feature handling in PEFT methods.

Although the Head-only setting may appear conceptually simpler (training only the decoder head), the number of trainable parameters depends on the size of the segmentation head relative to the low-rank matrices inserted in the LoRA-A+Head method. In our implementation, the trainable decoder head accounted for a larger fraction of parameters than the selected low-rank adapter matrices in LoRA-A+Head. This explains why Head-only ended up with a higher trainable-parameter percentage (10.63%) than LoRA-A+Head (1.57%). We have thoroughly rechecked the parameter counting procedure and confirmed that the reported values are accurate.

As for These findings have several practical implications. LoRA-A+Head offers efficient and practical segmentation for resource-limited marine applications, with potential utility in underwater archaeology, marine monitoring, and search-and-rescue operations, where processing speed and computational efficiency are critical for timely decision-making. The choice of PEFT method ultimately depends on the accuracy-efficiency trade-off relevant to the deployment context, guiding strategy selection for embedded systems or large-scale analyses.

LoRA-A+Head achieved intermediate performance between FullFT and Head-only. It improved substantially over Head-only while requiring far fewer trainable parameters than FullFT. In this experimental setting, the main advantage of LoRA-A+Head lies in reducing the parameter-update budget rather than in large reductions in wall-clock training time or VRAM usage.

5.3. Limitations & Future Work

This study is limited to SegFormer-B0; future work should explore larger transformer models such as SegFormer-B2/B5. Results are based on the AI4Shipwrecks dataset, and generalizability to other sonar datasets remains untested. LoRA-A+Head and Head-only both struggle with low-contrast sonar data, suggesting that enhanced feature extraction or targeted data augmentation strategies are needed. Future directions include evaluating LoRA-A+Head on larger backbones (SegFormer-B2/B5) to improve segmentation while retaining parameter efficiency, applying contrast enhancement and domain-specific pretraining for low-contrast scenarios, incorporating post-processing steps such as morphological operations or conditional random fields (CRFs) to refine predicted boundaries, and testing all methods on additional sonar datasets

collected from diverse environments to assess broader generalizability.

6. Conclusion

This study evaluated three adaptation strategies for binary sonar shipwreck segmentation using SegFormer-B0: FullFT, Head-only, and LoRA-A+Head. Full fine-tuning achieved the best performance on the held-out test set, with a Dice score of 0.614 ± 0.008 and IoU of 0.487 ± 0.007. LoRA-A+Head ranked second, with 0.546 ± 0.010 Dice and 0.401 ± 0.008 IoU, while updating only a small fraction of model parameters. Head-only required the least memory and shortest training time per epoch, but it produced the lowest segmentation accuracy among the three main methods. These results show that the main benefit of LoRA-A+Head in this setup is parameter efficiency rather than a large reduction in runtime. The findings therefore support FullFT when the goal is maximum accuracy and support LoRA-A+Head when reducing the number of updated parameters is more important. Because the experiments were limited to SegFormer-B0 and a fixed dataset split, future work should include additional sonar datasets, alternative splits, and larger backbones to examine the generality of these conclusions.

7. Declarations

7.1. Author Contributions

Shehan Maxwell Beruwalage: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing - Original Draft, Visualization; **Chunyong Yin:** Supervision, Writing - Review & Editing, Project administration; **Muhammad Raza:** Software, Investigation; **Deshan Sachintha Kannangara:** Resources, Investigation; **Sachini Amani Hendavitharana:** Writing - Review & Editing, Validation.

7.2. Institutional Review Board Statement

Not applicable. This study did not involve human participants, human data, or animal subjects.

7.3. Informed Consent Statement

Not applicable.

7.4. Data Availability Statement

The data presented in this study are openly available. The AI4Shipwrecks dataset used in this research was collected during 2022–2023 surveys at the NOAA Thunder Bay National Marine Sanctuary and is publicly accessible at: https://umfieldrobotics.github.io/ai4shipwrecks/.

7.5. Acknowledgment

The authors would like to thank colleagues and friends who provided support during experimentation, model testing, and technical discussions. Special thanks are given to the researchers and contributors who developed and released the AI4Shipwrecks dataset, which made this study possible.

7.6. Conflicts of Interest

The authors declare no conflicts of interest.

8. References

[1] D. R. Blidberg, "The development of autonomous underwater vehicles," *Ocean Engineering*, 2001. https://www.researchgate.net/publication/247835516.

[2] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015. https://doi.org/10.1007/978-3-319-24574-4_28.

[3] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015. https://doi.org/10.1038/nature14539.

[4] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin Transformer: Hierarchical vision transformer using shifted windows," in *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021. https://doi.org/10.1109/ICCV48922.2021.00986.

[5] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, P. Luo, "SegFormer: Simple and efficient design for semantic segmentation with transformers," *arXiv preprint arXiv:2105.15203*, 2021. https://doi.org/10.48550/arXiv.2105.15203.

[6] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016. https://synapse.koreamed.org/pdf/10.4258/hir.2016.22.4.351.

[7] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Fourth International Conference on 3D Vision (3DV)*, 2016. https://doi.org/10.1109/3DV.2016.79.

[8] S. Kaplan, J. McCandlish, T. Henighan, T. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu, and D. Amodei, "Scaling laws for neural language models," *arXiv preprint arXiv:2001.08361*, 2020. https://doi.org/10.48550/arXiv.2001.08361.

[9] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "LoRA: Low-rank adaptation of large language models," *arXiv preprint arXiv:2106.09685*, 2022. https://doi.org/10.48550/arXiv.2106.09685.

[10] T. Dettmers, A. Pagnoni, A. Holtzman, L. Zettlemoyer, "QLoRA: Efficient finetuning of quantized LLMs," *arXiv preprint arXiv:2305.14314*, 2023. https://doi.org/10.48550/arXiv.2305.14314.

[11] N. Houlsby, A. Giurgiu, S. Jastrzebski, B. Morrone, Q. De Laroussilhe, A. Gesmundo, M. Attariyan, and S. Gelly, "Parameter-efficient transfer learning for NLP," *arXiv preprint arXiv:1902.00751*, 2019. https://doi.org/10.48550/arXiv.1902.00751.

[12] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2021. https://doi.org/10.48550/arXiv.2010.11929.

[13] J. Lei, H. Wang, L. Fan, Q. Gu, S. Rong, and H. Zhang, "SonarNet: Global feature based hybrid attention network for side scan sonar image segmentation," *Remote Sensing*, vol. 17, no. 14, 2450, 2025. https://doi.org/10.3390/rs17142450.

[14] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. https://doi.org/10.1109/CVPR.2015.7298965.

[15] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84-90, 2017. https://doi.org/10.1145/3065386.

[16] J. Dean, G. Corrado, R. Monga, K. Chen, M. Devin, M. Mao, A. Senior, P. Tucker, K. Yang, Q. Le, and A. Ng, "Large scale distributed deep networks," in *Advances in Neural Information Processing Systems*, 2012. https://proceedings.neurips.cc/paper_files/paper/2012/hash/6aca97005c68f1206823815f66102863-Abstract.html.

[17] J. Pfeiffer, A. Kamath, A. Rücklé, K. Cho, I. Gurevych, "AdapterFusion: Non-destructive task composition for transfer learning," in *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pp. 487–503, 2021. https://doi.org/10.18653/v1/2021.eacl-main.39.

[18] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM*, vol. 58, no. 3, 2011. https://doi.org/10.1145/1970392.1970395.

[19] A. Shorten and T. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 6, 2019. https://doi.org/10.1186/s40537-019-0197-0.

[20] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. https://doi.org/10.1109/CVPR.2017.106.

[21] 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*, vol. 44, no. 3, pp. 341–354, 2025. https://umfieldrobotics.github.io/ai4shipwrecks/.

[22] P. Zeng, Y. Chen, W. Zhang, X. Zhang, and Y. Chen, "Multi beam sonar target segmentation based on BS UNet," *Electronics*, vol. 13, no. 14, 2841, 2024. https://doi.org/10.3390/electronics13142841.