Change Detection in Synthetic Aperture Sonar Imagery Using the Segment Anything Model

William Hedlund[1], Per Abrahamsson[1], and Louise Rixon Fuchs[3]

[1]Saab AB, SE-581 88 Linköping, Sweden

William Hedlund, Bröderna Ugglas gata, SE-581 88 Linköping, Sweden
william.hedlund3@saabgroup.com

Abstract: *Recent advancements in artificial intelligence have led to the development of foundation models, which are large, general-purpose models capable of performing a wide range of tasks across domains. This paper investigates the use of the Segment Anything Model (SAM), a foundation model for image segmentation, for change detection in synthetic aperture sonar (SAS) imagery. The method is entirely training-free and applies SAM to segment bi-temporal SAS images, which are then compared to identify changes. Three prompting strategies are evaluated: two use grids of points with varying spacings, and one uses a change map generated by a traditional method. The approach is tested on both real and synthetically generated datasets, with the synthetic data created through image composition to simulate changes. Results show that the change-map and fine-grid methods achieve similar precision and recall, while the coarse grid results in more missed detections but higher precision. On real data, the SAM-based approach outperforms a traditional log-ratio method by improving both precision and recall, reducing false alarms caused by misregistration and environmental noise. These results demonstrate SAM's adaptability to new domains and its potential as a viable, training-free alternative for SAS change detection, supporting broader applications of deep learning in marine research and environmental monitoring.*

Keywords: *change detection, synthetic aperture sonar, segmentation, Segment Anything Model*

1. INTRODUCTION

Change detection (CD) refers to the task of identifying differences between two observations of the same scene taken at different times. In the underwater domain, this is typically performed by comparing synthetic aperture sonar (SAS) images collected by autonomous underwater vehicles (AUVs) on repeat-pass surveys. As AUV and SAS technologies continue to improve in coverage speed and resolution, the volume of collected data increases significantly, exceeding the capacity of manual review and demanding reliable automated processing techniques [1]. A key application of SAS-based surveying is mine reconnaissance, where regular repeat-pass imaging in high-traffic areas enables detection of newly introduced or moved objects. Automated change detection in such contexts supports maritime safety by aiding early threat identification.

Early methods for CD in SAS imagery, matched new detections to known historical objects. Modern approaches enhance this by co-registering repeated SAS images, analysing differences in backscatter intensity to produce coherence maps [1]. These methods are sensitive to co-registration errors, environmental changes, and threshold tuning, limiting their robustness in complex, cluttered underwater environments. Furthermore, they operate solely on pixel-level intensity comparisons, without leveraging higher-level contextual information that human operators often use to distinguish meaningful changes from irrelevant noise.

In other remote sensing domains such as satellite and aerial imagery, deep learning has surpassed traditional methods for change detection [2]. However, deep learning models typically require large annotated datasets, which are costly or impractical to obtain in SAS applications due to operational constraints and the sensitive nature of much sonar data. Deep learning has been applied to SAS imagery primarily for target classification, seabed characterization, and image enhancement [3, 4, 5]. Despite these advances, the application of deep learning for change detection in SAS imagery remains limited. This highlights a need for methods that can leverage deep learning's capabilities while addressing the challenges of scarce labelled data.

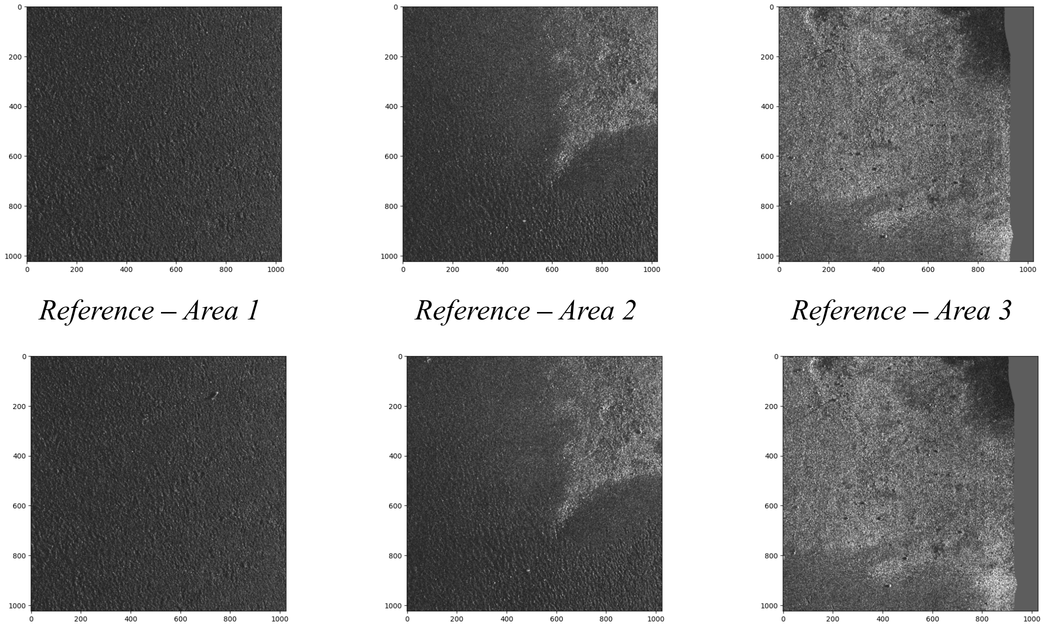
In recent years, models trained on massive datasets have given rise to foundation models, which are large, pre-trained models with capabilities spanning a wide range of tasks. These models exhibit strong generalization and often require little or no fine-tuning to adapt to new domains. One such model is the Segment Anything Model (SAM) [6], which performs image segmentation by using different types of input prompts such as points, boxes, or masks to guide and condition the model's behaviour to focus on specific regions or features in the image. SAM has shown adaptability across domains, including medical imaging, microscopy, and satellite imagery. This raises the question of whether these capabilities can also benefit the domain of sonar imagery and change detection, where the visual modality differs significantly from natural images.

2. METHOD

The work of this paper focuses on leveraging a foundation model to perform training-free change detection in SAS imagery. To evaluate the results, the proposed method will be compared to a traditional algorithmic log-ratio approach described in Section 3. Prerequisites include data collection and preprocessing of the images, these components are briefly described in this section.

2.1. DATA COLLECTION AND PREPARATION

The data used in this study was collected using Saab's AUV62-MR system, an autonomous underwater platform equipped with dual SAS sensors. The surveys were conducted over two separate missions with a one-day interval. Between passes, several objects were manually introduced to simulate real changes. These included a rock, a metal pipe, and a bundle of rope, placed in diverse environments such as textured sand and gravel. The SAS system produced high-resolution imagery at approximately 9 cm^2 per pixel, as seen in Figure 1.



Reference – Area 1 *Reference – Area 2* *Reference – Area 3*

Sensed – Area 1 *Sensed – Area 2* *Sensed – Area 3*

Figure 1: Areas where real change is introduced. The images are approximately 30×30 m. In each case, the changed object is visible in the image: top right for Area 1, top left for Area 2, and top left for Area 3.

Because real change data is costly and time-consuming to collect, a synthetic dataset was also constructed. This involved segmenting real objects from prior missions and inserting them into unchanged image backgrounds using cut-and-paste augmentation [7]. This approach enables controlled evaluation of the change detection method across a wider variety of object types and scenes.

Co-registration is essential to correct for navigational drift and ensure spatial alignment between bi-temporal images. We adopted a multistage registration approach based on the methods described in [8, 9]. First, SIFT keypoints are extracted and matched, with RANSAC applied to reject outliers and achieve a coarse global alignment. This is followed by localized refinement using normalized cross-correlation in the Fourier domain. The final transformation is computed on a per-tile basis, enabling sub-pixel precision and allowing compensation for local distortions caused by terrain variation.

3. TRADITIONAL LOG-RATIO-BASED CHANGE DETECTION

Given two co-registered SAS intensity images $I_1, I_2 \in \mathbb{R}^{H\times W}$, the *log-ratio operator* is defined to highlight regions of change while reducing the influence of multiplicative noise:

$$D(x,y) = \left| \log\left(\frac{I_1(x,y)}{I_2(x,y)}\right) \right| \tag{1}$$

The resulting *difference image* D is transformed using pixel-level saliency [10, 11] as:

$$\text{Sal}(a_m) = \sum_{n=0}^{255} f_n \left| a_m - a_n \right| \tag{2}$$

where f_n is the frequency of pixel intensity a_n in the image histogram, and $|a_m - a_n|$ is the absolute intensity difference. This saliency transformation emphasizes perceptually significant regions, reducing speckle noise. The saliency-enhanced image is then thresholded using *Otsu's method* [12], which determines the optimal threshold T_{Otsu} by maximizing inter-class variance. The final binary *change map* CM is defined such that $CM(x,y) = 1$ if $D(x,y) > T_{\text{Otsu}}$, and $CM(x,y) = 0$ otherwise. To reduce noise and speckle artifacts, a median filter is applied to the resulting binary map CM.

4. SAM-BASED OBJECT-LEVEL CHANGE DETECTION

In contrast to pixel-based differencing, the proposed method performs change detection at the object level using SAM. Each image I_1 and I_2 is segmented independently. We evaluate three prompting strategies: (1) a **sparse grid** of point prompts with 32-pixel spacing, (2) a **dense grid** of point prompts with 16-pixel spacing, and (3) prompts derived from the traditional method, using the **change map as input** to focus on more salient regions of the image.

SAM returns a set of binary object masks for each image, denoted as $\mathcal{S}_1 = \{M_1^i\}$ and $\mathcal{S}_2 = \{M_2^j\}$. Change detection is performed by comparing these masks using the Intersection over Union (IoU), defined as $\text{IoU}(M_1^i, M_2^j) = \frac{|M_1^i \cap M_2^j|}{|M_1^i \cup M_2^j|}$. An object M_1^i is marked as changed if its maximum IoU with any object in the other image is below 0.5. Symmetrically, unmatched masks from $\mathcal{S}_2$ are also considered changes. To reduce false positives, segmented objects with fewer than 100 pixels are discarded. Morphological operations are then applied to refine the boundaries of the masks.

5. PERFORMANCE METRICS FOR CHANGE DETECTION

Evaluation of change detection is performed per pixel, as each pixel is classified as either changed or unchanged. Due to the inherent class imbalance where changed pixels are much less frequent, metrics such as: precision, recall, and F1 score are used. Precision is defined as $\text{Precision} = \frac{TP}{TP+FP}$, recall as $\text{Recall} = \frac{TP}{TP+FN}$, and the F1 score as $\text{F1 Score} = \frac{2\cdot\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}}$. While overall accuracy is frequently reported, it can be misleading for imbalanced datasets. The F1 score provides a more balanced evaluation by combining both precision and recall.

6. RESULTS

Evaluation is performed on two datasets: synthetically generated data for quantitative comparison of prompt-methods on the SAM-based method, and qualitative comparison on the real data comparing the traditional method and selected prompt-method.

6.1. QUANTITATIVE EVALUATION – SYNTHETIC DATA

As shown in Table 1, the method using a change map as input achieves the highest F1 score (0.73), indicating a good balance between precision (0.68) and recall (0.78). The dense grid method (16-pixel spacing) follows closely with an F1 score of 0.71, with notably higher recall (0.82) but lower precision (0.62), suggesting more detected changes but also more false positives. The sparse grid method (32-pixel spacing) achieves the highest precision (0.71), indicating fewer false positives, but has the lowest recall (0.49), resulting in many missed changes and the lowest F1 score (0.58). This shows that the prompt method has a significant impact on the results.

Point Prompt Method	**Precision**	**Recall**	**F1 Score**
Change map	0.68	0.78	0.73
Sparse grid (32 px)	0.71	0.49	0.58
Dense grid (16 px)	0.62	0.82	0.71

Table 1: Performance metrics on synthetic data for different SAM prompt strategies.

6.2. QUALITATIVE EVALUATION – REAL DATA

For this qualitative comparison, we focus on the method that uses the change map from the traditional approach, as it demonstrated the best trade-off between precision and recall in the quantitative evaluation. The traditional method was tuned on synthetic data using a median blur kernel size of 3×3 and a minimum region size of 5×5. A larger blur kernel removed thin objects entirely, while omitting it resulted in excessive false alarms. SAM was configured with a confidence threshold of 0.8.

The resulting change maps are shown in Figure 2. In Area 1, the SAM-based method captures both the changed object and its shadow completely, which is desirable for accurate interpretation. The predicted change map includes three false alarms, which were found to be caused by shadows resulting from an offset in the collection path rather than true scene changes. In contrast, the traditional method produces a fragmented and incomplete mask of the object, with numerous false alarms distributed across the image. In Area 2, the SAM-predicted mask contains no false alarms and clearly highlights the relevant change, in contrast to the traditional method, where the change is difficult to perceive due to false alarms.

While grid-based prompt strategies were also evaluated, they are not included in this qualitative comparison. These methods generally showed higher recall but introduced more false alarms, consistent with the quantitative results. These observations reinforce that prompt selection is a critical factor in SAM-based change detection. Point prompts targeting salient regions

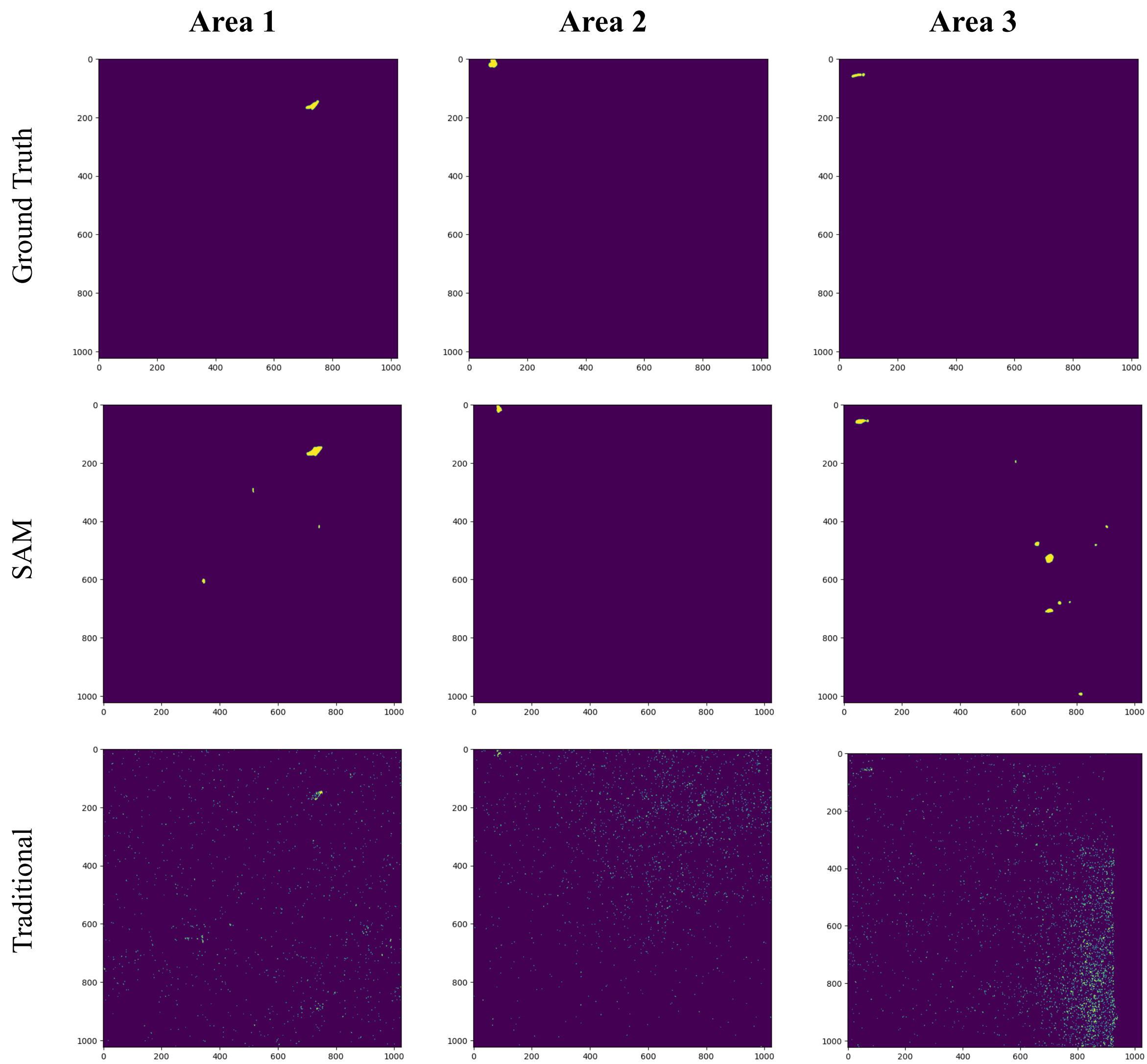


Figure 2: Comparison of change detection methods across three test areas. Columns correspond to Area 1, Area 2, and Area 3. Rows show ground truth (top), SAM-based predictions (middle), and traditional method results (bottom).

offer a strong balance between precision and recall, improving change mask quality while reducing false detections.

7. CONCLUSION

This work demonstrates the potential of adapting foundation models, specifically SAM, for change detection in SAS imagery. Despite the domain mismatch between natural and sonar images, the results show that SAM can be effectively applied using selected point prompts, particularly when guided by salient regions from a change map. This indicates a promising approach for low-data scenarios where training deep models is impractical. However, the real-world dataset used in this study is limited in size, making it difficult to draw definitive conclusions.

Future work could focus on developing methods that incorporate semantic understanding and are robust to variations in survey trajectories and viewing angles. Such advancements would further support the practical deployment of change detection systems in complex underwater environments.

REFERENCES

[1] Crawford, A., Hansen, R., Johnson, S., King, J., Midtgaard, Ø., Sæbø, T., Salas, J., Shouldice, E., Smith, S., Sternlicht, D., and Woods, D.: "Improvements to Automated Change Detection Tools for SAS Images," in *Proceedings of OCEANS 2022*, IEEE, 2022.

[2] Shafique, A., Cao, G., Khan, Z., Asad, M., and Aslam, M.: "Deep Learning-Based Change Detection in Remote Sensing Images: A Review," *Remote Sensing*, vol. 14, no. 4, pp. 871, 2022.

[3] Williams, D.P.: "Acoustic-Color-Based Convolutional Neural Networks for UXO Classification with Low-Frequency Sonar," in *Proceedings of the Underwater Acoustics Conference*, 2019.

[4] Sun, Y.-C., Gerg, I.D., and Monga, V.: "Iterative, Deep Synthetic Aperture Sonar Image Segmentation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022.

[5] Fuchs, L.R., Larsson, C., and Gällström, A.: "Deep Learning-Based Technique for Enhanced Sonar Imaging," in *Proceedings of the 5th Underwater Acoustics Conference & Exhibition (UACE)*, Hersonissos, Crete, Greece, pp. 1021–1028, 2019.

[6] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., Dollár, P., and Girshick, R.: "Segment Anything," arXiv:2304.02643, 2023. `https://arxiv.org/abs/2304.02643`

[7] Ghiasi, G., Cui, Y., Srinivas, A., Qian, R., Lin, T.-Y., Cubuk, E.D., Le, Q.V., and Zoph, B.: "Simple Copy-Paste Is a Strong Data Augmentation Method for Instance Segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2918–2928, 2021.

[8] G-Michael, T., Marchand, B., Tucker, J.D., Marston, T.M., Sternlicht, D.D., and Azimi-Sadjadi, M.R.: "Image-Based Automated Change Detection for Synthetic Aperture Sonar by Multistage Coregistration and Canonical Correlation Analysis," *IEEE Journal of Oceanic Engineering*, vol. 41, no. 3, pp. 592–612, 2016.

[9] Wang, V.T. and Hayes, M.P.: "Synthetic Aperture Sonar Track Registration Using SIFT Image Correspondences," *IEEE Journal of Oceanic Engineering*, vol. 42, no. 4, pp. 901–913, 2017.

[10] Zhai, Y. and Shah, M.: "Visual Attention Detection in Video Sequences Using Spatiotemporal Cues," in *Proceedings of the 14th ACM International Conference on Multimedia*, pp. 815–824, 2006.

[11] Zhang, Y., Wang, S., Wang, C., Li, J., and Zhang, H.: "SAR Image Change Detection Using Saliency Extraction and Shearlet Transform," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 12, pp. 4701–4710, 2018.

[12] Otsu, N.: "A Threshold Selection Method from Gray-Level Histograms," *Automatica*, vol. 11, pp. 23–27, 1975.