# Promptable sonar image segmentation for distance measurement using SAM.

TOLIE, H.F., REN, J., HASAN, M.J., KANNAN, S. and FOUGH, N.

2024

# Promptable Sonar Image Segmentation for Distance Measurement using SAM

Hamidreza Farhadi Tolie[1], Jinchang Ren[2], Md. Junayed Hasan[3], Somasundar Kannan[4], Nazila Fough[5]
*School of Computing, Engineering, and Technology*
*Robert Gordon University*, Aberdeen, UK
Emails: {h.farhadi-tolie, j.ren, j.hasan, s.kannan1, n.fough1}@rgu.ac.uk

***Abstract*—The subsea environment presents numerous challenges for robotic vision, including non-uniform light attenuation, backscattering, floating particles, and low-light conditions, which significantly degrade underwater images. This degradation impacts robotic operations that heavily rely on environmental feedback. However, these limitations can be mitigated using sonar imaging, which employs sound pulses instead of light. In this paper, we explore the use of small, affordable sonar devices for automatic target object localization and distance measurement. Specifically, we propose using a promptable image segmentation method to identify target objects within sonar images, leveraging its ability to identify connected components without requiring labeled datasets. Through laboratory experiments, we analyzed the usability of the Ping360 single-beam sonar and verified the effectiveness of our approach in the automatic identification and distance measurement of objects made from various materials. The collected raw and processed data alongside the source code of the proposed approach will be shared at https://github.com/hfarhaditolie/PSIS-ADM.**

***Index Terms*—Sonar image segmentation, distance measurement, Ping360, single-beam sonar**

## I. Introduction

In recent years, there has been increasing attention on the exploration of underwater environments. Traditionally, underwater related tasks were carried out by human divers, often resulted in physical and mental harm [1]. However, technological advancements have led to the development of various underwater vehicles aimed at delving into the depths of the sea [2]. These vehicles not only facilitate the investigation and monitoring of the underwater world but also perform robotic tasks such as installation, maintenance, and object retrieval. Typically, these vehicles or robots are remotely operated by humans who rely on visual data from the environment. Yet, underwater robotic vision faces challenges due to water turbidity, light attenuation, and scattering, severely limiting the usefulness of optical sensors in subsea applications. Although efforts have been made to address these limitations and improve visual data quality [3], they currently only provide insights for short-range operations.

Moreover, the emergence of three-dimensional (3D) vision has enabled precise distance measurements to obstacles or target objects, leading to accurate control and performance of robots [4]. Consequently, 3D optical cameras and stereo vision have been successfully developed and implemented across various industries [5]–[7]. However, the complexities mentioned earlier make the utilization of 3D vision challenging underwater. Conversely, acoustic and sound waves offer a viable alternative, easily transmitting through the underwater environment even in dark conditions, aiding in target/obstacle detection [8].

Sonar systems are broadly categorized into two main types based on the acoustic pulses they emit: single-beam and multi-beam. Single-beam systems emit a single beam of sound waves at various angles to generate an acoustic map. In contrast, multi-beam sonars emit multiple acoustic pulses simultaneously, allowing them to construct detailed and precise 3D maps of underwater objects. However, multi-beam sonar systems are often costly, leading to a preference for single-beam sonars (SBS) among diverse communities due to their affordability.

While SBS also provide depth information, they are subject to noise and shadowing zones. Moreover, unlike multi-beam sonars, single-beam systems lack object-specific details such as shape and dimensions, making it challenging to distinguish objects from background noise. Hence, our research aims to explore image and signal processing techniques to mitigate noise and shadowing effects, and to leverage semantic segmentation models for the automatic detection of target objects within sonar images, along with their distance from the sensor.

After reviewing the literature and doing experiments using a SBS, i.e., Ping360[1], it is found that existing methods for semantic segmentation cannot fully identify the objects due to the noise and shadowing zones present in these images. Thus, we have reported a search on finding the ideal parameters for SBS system to reduce the noise level and utilization of set of signal/image processing techniques to further denoise the image and eliminate the shadowing zones. Then, we were able to address the time-consuming and imprecise nature of manual analysis of SBS images by implementing a semantic segmentation method.

## II. Background

SBS are widely used for navigation, mapping and localization [9], in some cases even for detecting objects and obstacles underwater. This is because the sound waves can travel further in water compared with the electromagnetic

This work is partially supported by the SeaSense project, funded by the Net Zero Technology Centre, UK.

[1]https://bluerobotics.com/store/sensors-sonars-cameras/sonar/ping360-sonar-r1-rp/Ping360 Scanning Imaging Sonar

waves. In addition, they can operate under low-light and turbid conditions, where the optical cameras and human diver's eye tend to fail in such scenarios [10].

SBS systems send a narrow beam of sound waves into water and then listen back for echoes. The reflection is then recorded using a transducer and the process is repeated for a specified scanning angles usually with an interval of $1^{\circ}$. The transmitted sound waves are reflected when they hit an object or even a small particle, which makes the sonar images very noisy. In addition, in areas where sound waves are blocked or redirected, shadowing zones may form, hiding objects behind obstacles. Moreover, sound waves lose energy as they propagate through water due to absorption by the medium. In regions where acoustic energy is absorbed or attenuated, the intensity of the returning signals decreases, resulting in shadowing zones. This shadowing zones makes it very difficult to estimate how big the objects/obstacles are.

"Furthermore, our examination of SBS capabilities revealed that the shape, surface roughness, and material properties of objects significantly influence the resulting sonar image. This variation complicates the precise determination of object boundaries and the accurate estimation of object distances. Consequently, achieving an accurate and detailed interpretation of SBS images remains a significant challenge.

## III. Materials and Methods

In order to explore the capabilities and limitations of the SBS systems we have conducted in lab experiments using a Ping360 mechanical scanning sonar in a water tank. We have collected data with objects of various materials and shapes positioned at varying distances from the sensor. The collected data are then analysed to examine the ideal parameters for improved data acquisition. Finally, we have performed a set of image processing techniques to eliminate noise and shadowing zones and guide a deep learning-based semantic segmentation model to determine the location of the objects and provide distance measurements. The following subsections present the specification of the testing environment, SBS device, our observations, and the proposed methodologies.

### A. Experimental setup and data acquisition

As shown in Figure 1, the experiments took place within a glass water tank measuring $60 \times 60 \times 150$ cm (height $\times$ width $\times$ length), with tank walls 1 cm thick. The water level reached 28 cm in height. To enhance reflection quality attributable to the tank's glass structure, acoustic foams were affixed to its interior walls.

The data were acquired using Ping360, a mechanical SBS with capability of localising targets inspecting and tracking underwater structures or objects that reflect sound waves [11], via its publicly available API [12]. It has a scanning range (0.75 m – 50 m), scanning sector($0^{\circ}$ – $360^{\circ}$) and voltage gain level (low, medium and high). Note that, while the Ping360 is operating, it generates strong noise close to the sonar head due to the rotational motion of the transducer [11]. This results in a 0.25m of noise surrounding the sonar head.

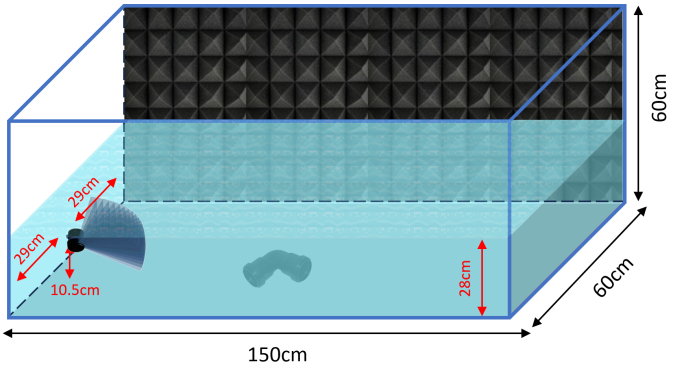


Fig. 1. A schematic diagram of the testing environment within a water tank.

For data collection, the sensor was oriented downward at a scanning angle of $90^{\circ}$, utilizing all available gain settings. Positioned squarely on the far left side of the tank, it maintained a consistent distance of 29 cm from both the left and right interior walls. Additionally, the sensor was set at an altitude of 10.5 cm above the tank bottom. The parameter setting used for data collection is specified in Table I.

TABLE I
Parameter setting of the Ping360

| **Range** | **Gain** | **Number of Samples** |
|---|---|---|
| 0-2m | low, medium, and high | 1200 |
| **Transmit duration** | **Transmit frequency** | **Speed of sound** |
| 16$\mu$s | 1000 kHz | 1500 m/s |

Of the parameters listed in Table I, the *number of samples* denotes the sampling rate per reflected signal, we have used the maximum value, i.e., 1200, to get the best resolution. Moreover, the *transmit duration* indicates how far the acoustic wave can travel before attenuating. We have empirically set this parameter to 16$\mu$s. Meanwhile, *transmit frequency* influences the system's resolution, with higher frequencies generally offering finer resolution, enabling the sonar to discern smaller objects. However, they come with shorter range capabilities, which are not a concern for this study as we intend to employ the sonar to support robotic vision for short-range operations like object grasping and retrieval. The *speed of sound* is another parameter that affects the propagation of acoustic signals and the interpretation of echo data and it depends on several factors, including temperature, pressure, and salinity [13]. We have set a commonly used approximate value for the speed of sound in freshwater at room temperature, i.e. $20^{\circ}$c, which is around 1500 meters per second (m/s).

### B. Promptable segmentation for distance measurement

The Ping360 sonar is equipped with a graphical interface that enables users to establish connections, view real-time data, and record sonar readings. Additionally, it features a distance axis for approximating the distance to target objects.

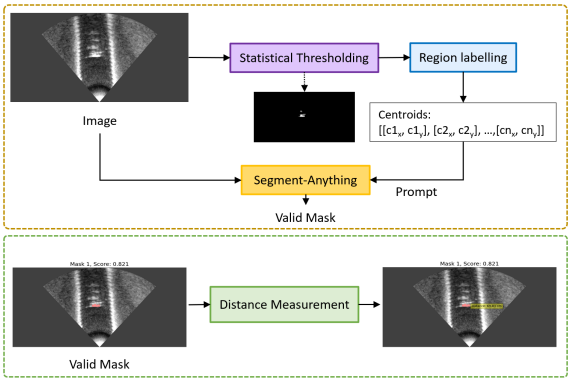


Fig. 2. General framework of the proposed methodology using SAM.

However, for precise distance measurements and to eliminate manual data interpretation, we propose to integrate artificial intelligence (AI) for automating distance estimation by automatically identifying the objects within the image. This AI-based approach can also facilitate the identification of multiple objects within the sonar image, providing distance measurements for each object.

Traditionally, object detection methods such as those described in [10], [14], [15] identify objects by drawing their bounding boxes, while segmentation models [16]–[18] pinpoint the exact object locations. However, these approaches rely on objects with distinct shapes, making them less suitable for SBSs, which primarily detect object presence rather than the shape information. Furthermore, due to limited data availability and the presence of noise and shadowing zones in underwater environments, training such models is exceptionally challenging. Therefore, we propose to utilize state-of-the-art promtable segmentation methods.

The framework of the proposed methodology is illustrated in Figure 2. As seen, using the statistical properties and region labelling of the sonar image we generate a segmentation prompt and then alongside the recorded sonar image, it is fed to the state-of-the-art Segment-Anything Model (SAM) [19] to approximately localise the object/target. Then, based on the generated mask, we measure the diagonal distance to the object by taking the sonar properties into account. The reason to use the SAM instead of simple region labelling for distance measurement is that while the region labeling step identifies regions of interest within the image, it may not capture the complete object boundary. For example, we have observed instances where a single object with partial color coating resulted in separate object identifications through region labeling. However, SAM excels in handling such complexities by considering the connectivity of regions, thereby facilitating more accurate object identification. The next subsections discuss the prompt generation process and distance measurement module.

### C. Prompt generation

In the experimental scenario in accordance with the previously mentioned experimental setup, each angle's data points were recorded as intensity values $I$ ranging from $0-255$, with each point representing a segment of the total scanned distance. For instance, with a maximum distance $D_{\max} = 2$ meters and 1200 *number of samples*, each sample point corresponds to approximately $0.00167$ meters. Initial filtering excluded unreliable data within 0.25 meters and beyond 1.40 meters (distance from centre of the sensor to the farthest wall of the tank), based on experimental setup constraints. Statistical thresholding was then applied within the region of interest (ROI), where $I$ is greater than or equal to $2 \times \mu + \sigma$ (based on empirical observation) were retained, ensuring the removal of noise. The mean intensity $\mu$ and standard deviation $\sigma$ for each angle are calculated as:

$$\mu = \frac{1}{N'} \sum_{i=1}^{N'} I_i \tag{1}$$

$$\sigma = \sqrt{\frac{1}{N'-1} \sum_{i=1}^{N'} (I_i - \mu)^2} \tag{2}$$

where $N'$ is the number of samples within the ROI.

Finally, the denoised data was converted from Cartesian to polar coordinates for analysis, as follows:

$$r = \sqrt{x^2 + y^2} \tag{3}$$

$$\theta = \arctan\left(\frac{y}{x}\right) \tag{4}$$

The SAM algorithm necessitates the pixel coordinates as input prompts. To achieve this, we employed Python's *scikit-image regionprops* function to identify potential regions within the filtered image. Subsequently, regions with an area of less than 600 pixels (determined empirically) were classified as shadowing zones or noise and thus eliminated. The central points of the remaining regions were then utilized as input prompts for the SAM algorithm.

### D. Distance measurement

To measure the distance from the sonar to the object, we used the detected masks within the image. First, using the mask coordinates, we generated a bounding box around the identified object. Next, we localized the center point of the closest edge along the x-axis (the bottom-most edge) and computed the distance as:

$$d = \sqrt{(X_c^o - X_c^s)^2 + (Y_c^o - Y_c^s)^2} \tag{5}$$

where $(X_c^o, Y_c^o)$ and $(X_c^s, Y_c^s)$ denote the center points of the object and the sensor, respectively.

Upon computing $d$, we determined the distance in terms of the number of pixels. To convert this distance value into centimeters, we divided $d$ by 6. Given the number of samples per angle (1200) and the maximum range set to 2 meters (200 centimeters (cm)), each centimeter is represented by 6 pixels in the captured sonar image. Thus, to compute the distance, we divided the computed $d$ value in Eq. 5 by 6.

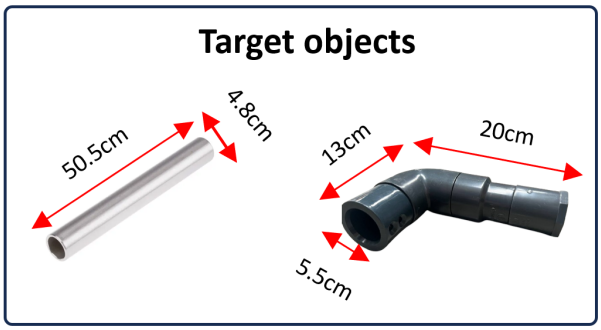


Fig. 3. Target objects used for the experiments along with their corresponding dimensions. The uniform pipe on the left is galvanized, while the bent pipe is made of PVC fabric.

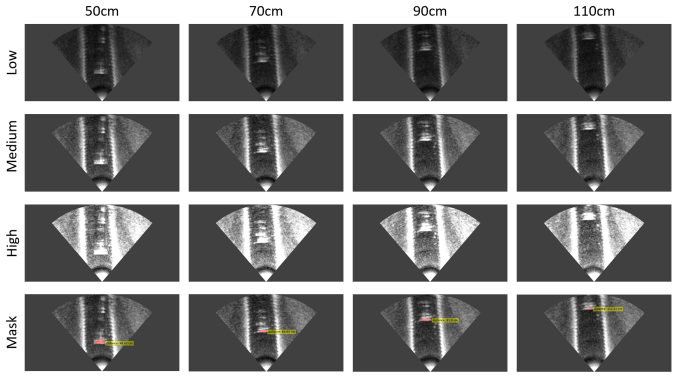


Fig. 4. Acquired sonar images using low, medium, and high gain setting with bent pipe at distances of 50, 70, 90, 110 centimeters and the generated mask image using SAM with the medium gain image as input.

## IV. In Lab Experiments

To verify the efficiency and effectiveness of the proposed approach, we first conducted experiments with a bent pipe, shown in Figure 3, to determine the best gain setting for the Ping360. The pipe was placed at diagonal distances of 50, 70, 90, and 110 cm from the sensor, and data were collected at three different gain levels: low, medium, and high. The acquired sonar images are shown in Figure 4. Based on the collected data, we observed the following:

- Low gain results have lower noise, but the reflection from the object is not strong enough for clear identification.
- High gain produces a noisy sonar image with strong object reflections; the shadowing zone also becomes more pronounced, leading to misidentification of objects.
- The most ideal images are acquired using the medium gain setting, resulting in more accurate object identification and distance measurement in our testing scenarios.
- The shadowing zone appears depending on the sensor's altitude from the tank's floor and the viewing angle (straight or tilted). At a low altitude of 10 cm with a straight field of view, an object placed at a distance of 50 cm produces more shadowing than at other distances.

Next, to verify the performance of the SAM in identification and distance measurements we have reported the measured distances for the bent pipe data collected using the medium gain setting and placed at different distance in Table II. Based on the results, the error in distance measurement varies between 0.17 and 1.33 cm, which shows almost a good accuracy in localising the object and measuring the distance. In addition to the bent pipe, we have also conducted experiments using an uniform pipe, shown in Figure 3. Similar to the bent pipe, the uniform one was also placed at diagonal distances of 50, 70, 90, 110 cm respectively and the measured results are reported in Table III. The results show an absolute error range of 0.33 to 1.0 cm.

TABLE II
Distance measurement results of bent pipe at various distances.

| **Ground truth (cm)** | 50 | 70 | 90 | 110 |
|---|---|---|---|---|
| **Measured distance (cm)** | 48.67 | 69.83 | 91.0 | 111.33 |
| **Absolute Error (cm)** | 1.33 | 0.17 | 1.0 | 1.33 |

TABLE III
Distance measurement results of uniform pipe at various distances.

| **Ground truth (cm)** | 50 | 70 | 90 | 110 |
|---|---|---|---|---|
| **Measured distance (cm)** | 51.0 | 70.67 | 90.33 | 111.0 |
| **Absolute Error (cm)** | 1.0 | 0.67 | 0.33 | 1.0 |

Overall, we observed the following:

- Although the ideal operating range of the Ping360 is 0.75 m to 50 m, it can still provide information for the objects located between 0.5 m to 0.75 m.
- Using a single Ping360 it is not possible to determine the exact shape or dimension of the object.
- Using the proposed approach on Ping360 recorded data, diagonal distance to the object can be automatically measured with a maximum error of 1.5 cm
- With a GPU of Tesla T4 in the Google Colab, the average processing time for each sonar image is 2.5 seconds.

## V. Conclusion

In conclusion, we have introduced a promptable image segmentation method applied to single-beam sonar images to identify target objects and automatically measure the diagonal distance from the sensor to the object. Utilizing the Ping360 single-beam sonar device, we collected data in a water tank with two target objects: a PVC bent pipe and a galvanized uniform pipe, placed at distances of 50, 70, 90, and 110 cm, under three different gain settings. Our proposed statistical thresholding technique effectively generated prompt images for the state-of-the-art Segment-Anything Model (SAM), enabling accurate identification of the target objects and measurement of distances. Throughout the in lab experiments, we have observed that the most ideal images are acquired using the medium gain setting, resulting in more accurate object

identification and distance measurement. Our experiments validated the model's accuracy and effectiveness. Future research could explore integrating single-beam sonar images with stereo vision to enhance the depth images acquired underwater.

## REFERENCES


[1] D. M. Barratt, P. G. Harch, and K. Van Meter, "Decompression illness in divers: a review of the literature," *The Neurologist*, vol. 8, no. 3, pp. 186–202, 2002.

[2] J. P. Ray, "Development of underwater robots for under water inspection and cleaning applications," 2023.

[3] H. F. Tolie, J. Ren, and E. Elyan, "Dicam: Deep inception and channel-wise attention modules for underwater image enhancement," *Neurocomputing*, vol. 584, p. 127585, 2024.

[4] V. Tadic, A. Toth, Z. Vizvari, M. Klincsik, Z. Sari, P. Sarcevic, J. Sarosi, and I. Biro, "Perspectives of realsense and zed depth sensors for robotic vision applications," *Machines*, vol. 10, no. 3, p. 183, 2022.

[5] N. Carey, J. Werfel, and R. Nagpal, "Fast, accurate, small-scale 3d scene capture using a low-cost depth sensor," in *2017 IEEE winter conference on applications of computer vision (WACV)*, pp. 1268–1276, IEEE, 2017.

[6] R. B. Rusu, Z. C. Marton, N. Blodow, M. Dolha, and M. Beetz, "Towards 3d point cloud based object maps for household environments," *Robotics and Autonomous Systems*, vol. 56, no. 11, pp. 927–941, 2008.

[7] E. El-Sayed, R. F. Abdel-Kader, H. Nashaat, and M. Marei, "Plane detection in 3d point cloud using octree-balanced density down-sampling and iterative adaptive plane extraction," *IET Image Processing*, vol. 12, no. 9, pp. 1595–1605, 2018.

[8] F. Ferreira, D. Machado, G. Ferri, S. Dugelay, and J. Potter, "Underwater optical and acoustic imaging: A time for fusion? a brief overview of the state-of-the-art," *OCEANS 2016 MTS/IEEE Monterey*, pp. 1–6, 2016.

[9] H. Martínez-Barberá, P. Bernal-Polo, and D. Herrero-Perez, "Sensor modeling for underwater localization using a particle filter," *Sensors*, vol. 21, no. 4, p. 1549, 2021.

[10] G. Neves, M. Ruiz, J. Fontinele, and L. Oliveira, "Rotated object detection with forward-looking sonar in underwater applications," *Expert Systems with Applications*, vol. 140, p. 112870, 2020.

[11] J. Hwang, N. Bose, B. Robinson, and W. Thanyamanta, "Sonar based delineation of oil plume proxies using an auv," *Int. J. Mech. Eng. Robot. Res*, vol. 11, pp. 207–214, 2022.

[12] Blue Robotics, "ping-python." https://github.com/bluerobotics/ping-python, 2024.

[13] J. Hwang, N. Bose, G. Millar, C. Bulger, and G. Nazareth, "Bubble plume tracking using a backseat driver on an autonomous underwater vehicle," *Drones*, vol. 7, no. 10, p. 635, 2023.

[14] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448, 2015.

[15] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, 2016.

[16] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pp. 740–755, Springer, 2014.

[17] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.

[18] Z. Chen, Y. Wang, W. Tian, J. Liu, Y. Zhou, and J. Shen, "Underwater sonar image segmentation combining pixel-level and region-level information," *Computers and Electrical Engineering*, vol. 100, p. 107853, 2022.

[19] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, *et al.*, "Segment anything," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4015–4026, 2023.