



UNIVERSIDADE FEDERAL DO RIO GRANDE - FURG
CENTRO DE CIÊNCIAS COMPUTACIONAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO
CURSO DE MESTRADO EM ENGENHARIA DE COMPUTAÇÃO

Dissertação de Mestrado

3D Reconstruction and Elevation Angle Estimation on Underwater Sonar Data

Guilherme Correa de Oliveira

Master's dissertation presented to the Graduate Program in Computation of the Federal University of Rio Grande - FURG, as a partial requirement for the degree of Master in Computer Engineering

Advisor: Prof. Dr. Paulo Lilles Jorge Drews Jr.
Co-Advisor: Prof. Dr. Matheus Machado dos Santos

Rio Grande, 2025

O48d Oliveira, Guilherme Correa de
3D reconstruction and elevation angle estimation on underwater sonar data / Guilherme Correa de Oliveira. – 2025.
45 f.

Dissertação (Mestrado) – Universidade Federal do Rio Grande – Programa de Pós-Graduação em Computação, 2025.

Orientador: Dr. Paulo Lilles Jorge Drews Jr.
Coorientador: Dr. Matheus Machado dos Santos.

1. Computação. 2. Robótica. 3. Exploração subaquática. 4. Sonares. 4. Reconstrução 3D. I. Drews Jr., Paulo Lilles Jorge. II. Santos, Matheus Machado dos. III. Título.

CDU 004

Catalogação na fonte: Bibliotecário Clériston Ribeiro Ramos CRB10/1889

Universidade Federal do Rio Grande
Centro de Ciências Computacionais
Programa de Pós-Graduação em Computação
Curso de Mestrado em Engenharia de Computação

DISSERTAÇÃO DE MESTRADO

3D Reconstruction and Elevation Angle Estimation on Underwater Sonar Data

Guilherme Correa de Oliveira

Banca examinadora:

gov.br Documento assinado digitalmente
ALESSANDRO DE LIMA BICHO
Data: 15/09/2025 23:03:23-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Alessandro Lima Bicho

gov.br Documento assinado digitalmente
ANTONIO WILSON VIEIRA
Data: 08/09/2025 16:23:27-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Antônio Wilson Vieira

gov.br Documento assinado digitalmente
MARCELO DE GOMENSORO MALHEIROS
Data: 08/09/2025 16:46:45-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Marcelo de Gomensoro Malheiros

Assinado de forma digital por 9f328126-831c-4231-9cfe-86c078dd05e7
Dados: 2025.09.08 16:10:49 -03'00'

Prof. Dr. Paulo Lilles Jorge Drews Júnior
Orientador

RESUMO

OLIVEIRA, Guilherme Correa de. **3D Reconstruction and Elevation Angle Estimation on Underwater Sonar Data**. 2025. 45 f. Master's Dissertation Programa de Pós-Graduação em Computação. Universidade Federal do Rio Grande - FURG, Rio Grande.

A exploração subaquática possui diversas características referentes ao seu meio que limitam o uso de câmeras ópticas, tornando os sonares de imageamento uma alternativa viável. No entanto, as imagens de sonar são inerentemente ambíguas e ruidosas, o que complica sua interpretação para a reconstrução 3D. Embora o aprendizado de máquina possa mitigar esses problemas, sua aplicação é restringida pela escassez de *datasets*. Esta dissertação aborda esses desafios ao introduzir uma metodologia baseada em aprendizado profundo para corrigir a ambiguidade de imagens de sonar, estimando o ângulo de elevação para a realização de reconstrução 3D.

Uma das principais contribuições deste trabalho é o desenvolvimento de um *dataset* chamado: *Synthetic Enclosed Echoes (SEE)*, o qual é constituído por tanto dados sintéticos quanto reais, criado em um ambiente de simulação de alta fidelidade que replica um tanque de testes físico. Para processar esses dados, propomos o ElevateNET-R, uma rede neural baseada em regressão e adaptada para prever o ângulo de elevação por pixel a partir de uma única imagem de sonar 2D.

Experimentos quantitativos demonstram que o modelo ElevateNET-R proposto supera consistentemente os métodos existentes na literatura, incluindo abordagens clássicas e outros modelos baseados em aprendizado. Além disso, a eficácia da metodologia foi validada em um experimento de simulação para o real (*sim-to-real*), no qual o modelo, treinado exclusivamente com o conjunto de dados sintético SEE, realizou com sucesso a reconstrução 3D de dados de sonar do mundo real. As principais contribuições são a disponibilização pública do conjunto de dados SEE e seu ambiente de simulação para fomentar futuras pesquisas e a validação bem-sucedida de uma rede baseada em regressão para a correção da ambiguidade em sonares.

Palavras-chave: Imagens de Sonar, Reconstrução 3D, Robótica Subaquática.

ABSTRACT

OLIVEIRA, Guilherme Correa de. **A Integrated Approach to Sonar Data Analysis: 3D Reconstruction and Ambiguity Correction**. 2025. 45 f. Master's Dissertation Programa de Pós-Graduação em Computação. Universidade Federal do Rio Grande - FURG, Rio Grande.

Underwater exploration is hindered by environmental challenges that limit the use of optical cameras, making imaging sonars a viable alternative for imaging. However, sonar images are inherently ambiguous and noisy, which complicates their interpretation for 3D reconstruction. While machine learning can mitigate these issues, its application is restricted by the scarcity of suitable datasets. This dissertation addresses these challenges by introducing a deep learning-based methodology to correct sonar image ambiguity by estimating the elevation angle for 3D reconstruction.

A primary contribution of this work is the development of the Synthetic Enclosed Echoes (SEE) dataset, a new, comprehensive collection of annotated synthetic and real-world sonar data, created within a high-fidelity simulation of a physical test tank. To process this data, a new methodology called ElevateNET-R is proposed, which is a regression-based neural network adapted to predict the per-pixel elevation angle from a single 2D sonar image.

Quantitative experiments demonstrate that the proposed ElevateNET-R model consistently outperforms existing methods from the literature, including classical approaches and other learning-based models. Furthermore, the effectiveness of the methodology was validated in a sim-to-real experiment, where the model, trained exclusively on the synthetic SEE dataset, successfully performed 3D reconstruction on real-world sonar data. The primary contributions are the public release of the expansive SEE dataset and its simulation environment to foster further research, as well as the successful validation of a regression-based network for correcting sonar ambiguity.

Keywords: Sonar Image, 3D Reconstruction, Underwater Robotics.

FIGURES LIST

1 A scheme of an acoustic image of imaging sonar [32] 18
2 The 3D field of view of an imagin sonar [5] 18
3 A comprehensive overview of the proposed sim-to-real methodology for 3D reconstruction 21
4 Aquatec Tank Facilities 23
5 Synthetic Enclosed Echoes: a new dataset of synthetic and real-world sonar data for a closed underwater environment 24
6 Geometric primitives used in the dataset. 25
7 Propeller models used in the dataset. 26
8 Anchors models used in the dataset. 26
9 Construction assets used in the dataset. 26
10 Proposed scenarios for the dataset and their respective way-points . . 27
11 The two ground truth formats used in this study. 11a The 2D elevation mask provides a per-pixel ground truth for the network. 11b The 3D ground truth point cloud serves as the reference for geometric accuracy evaluation. 29
12 Modified versions of the Blue ROV2 used during the real data colection 30
13 Real sonar data in the Polar form (Azimuth x Radius), the top image is the raw image and the bottom is the image after the echo intensity filter 31
14 The ElevateNET-R reconstruction pipeline. A single sonar image is fed into the regression network, which predicts an elevation angle for each pixel. This elevation map is then used to project the sonar returns into 3D space, generating the final point cloud of the object. . 32
15 3D recosntruction ground truth 34
16 3D reconstruction using a classic approach, where the estimation of the $\phi = 0$. 34
17 3D recosntruction using the Neusis [28] 35
18 3D reconstruction using the ElevateNET [4]. 35
19 3D reconstruction using the algorithm proposed in this work. 36
20 3D reconstruction using the algorithm proposed in this work, where the reconstructed objects were excluded from the training dataset . . 36
21 3D reconstruction using the algorithm proposed in this work, where the multiview data is the training dataset. 37

22 Qualitative comparison of reconstruction results on real-world sonar data. 22a The experimental setup, showing the plastic pipe target within the indoor tank. 22b 3D reconstruction generated by the model trained with the ElevateNET-R single-view model. 22c 3D reconstruction generated by the model trained with a multiview data strategy. In both 22b and 22c, the red highlighted region indicates the segmented plastic pipe, which is the primary object of interest. 38

TABLE LIST

1 Table of works related to 3D reconstruction using acoustic images . . 16

2 Root mean square (RMS) and mean Hausdorff distance errors for the proposed experiments. 39

LIST OF ABBREVIATIONS AND ACRONYMS

AUVs	Autonomous Underwater Vehicles
CAD	Computer Aided Design
CNN	Convolutional Neural Networks
DVL	Doppler Velocity Log
FLS	Forward Looking Sonars
FURG	Federal University of Rio Grande
GPS	Global Positioning System
IMU	Inertial measurement unit
NAUTEC	Intelligent robotics and automation group
ROVs	Remotely Operated Vehicles
RSfM	Refractive Structure from Motion
SfM	Structure from Motion

SUMMARY

1 Introduction **9**
1.1 Contextualizing 9
1.2 Problem Definition 10
1.3 Objectives 10
1.4 Organization of the Text 11

2 Literature Review **12**
2.1 3D Reconstruction of Optical Images 12
2.2 3D Reconstruction of Acoustic Images 13

3 Theoretical Background **17**

4 Proposed Methodology **20**
4.1 Simulation 20
4.2 Dataset Proposal 22
4.2.1 Syntethic Data Collection 25
4.3 3D Reconstruction of the Sonar Data 30
4.4 Evaluation Metrics 32

5 Results **33**

6 Conclusion **40**

References **41**

1 INTRODUCTION

The use of 3D data can be very useful for a wide variety of applications, such as building maps, performing inspections, and environmental monitoring, among others [24]. The workflow known as 3D reconstruction is based on using sensors to collect the characteristics of the environment and targets, and stitching algorithms to structure what is being observed, thereby obtaining the characteristics of the scene and target, and restoring them [23].

Significant advancements have been made in underwater 3D reconstruction using optical sensors [14]. However, optical sensors face inherent limitations in underwater environments, such as rapid attenuation of light-wave energy, high turbidity, and low-light conditions [9, 10]. Consequently, these limitations often hinder the ability of optical sensing to meet the demands of real-world applications [13].

On the other hand, the use of acoustic sensors can become an excellent alternative for applications in real environments [8], given that sonar wave propagation in water has the characteristics of low loss, strong diffraction ability, long propagation distance, and little influence on the water quality conditions [7, 21]. With this, it is possible to achieve better results in more complex underwater environments, even in the absence of light sources, compared to optical sensors. Therefore, underwater 3D reconstruction based on sonar images has a good research prospect. However, sonar also has the disadvantages of low resolution and difficult data extraction, due to the ambiguity in the sonar images and the inability to provide accurate color information.

1.1 Contextualizing

One of the biggest challenges of working with sonar images obtained by active acoustic sensors, for example, imaging sonars, is the inherent effects of this type of sensor, such as ambiguity, reverberation, and other noises present in this type of sensor [35]. In turn, these problems inherent to the sensor can be minimized through computational strategies, such as data filtering and post-processing. Currently, there is no definitive solution to these problems, and a significant demand exists for research on these topics.

Characterizing sonar noise is a complex task, and even when modeled, it imposes significant computational overhead [30]. Furthermore, modeling certain effects, such as the sensor's ambiguity, presents considerable difficulties. An alternative approach to noise mitigation involves employing learning-based methods, which effectively model and abstract these intricate effects. However, a major limitation of learning algorithms is their substantial data requirement. Existing sonar datasets are notably scarce, primarily due to the high cost of sonar equipment and the logistical and infrastructural complexities associated with data collection.

Nevertheless, synthetic data for training learning methods is gaining prominence, as demonstrated in [43], where a purely synthetic dataset was used for monocular camera depth estimation. This concept of synthetic image data can be extended to sonar data, provided a reliable sonar simulation is available to accurately replicate the effects and nuances observed in real acoustic images.

1.2 Problem Definition

The 3D reconstruction of submerged structures is critical for applications like infrastructure inspection, but it is often impeded by low-light and high-turbidity environments where optical cameras are ineffective. While imaging sonars serve as a viable alternative in these challenging conditions, they introduce a fundamental obstacle to 3D reconstruction: inherent data ambiguity.

The central problem is that an imaging sonar scans a three-dimensional space defined by range (r), azimuth (θ), and elevation (ϕ). However, it projects this information onto a two-dimensional image, discarding the crucial elevation angle (ϕ) for each acoustic return. This loss of the elevation dimension makes it impossible to determine the true 3D coordinates of a point from a single 2D sonar image, presenting the primary barrier to 3D reconstruction.

Furthermore, while deep learning methods offer a promising pathway to resolve this ambiguity by estimating the missing elevation angle, their development is severely constrained. These algorithms require large, accurately annotated datasets for training, yet there is a pronounced scarcity of public sonar datasets specifically designed for 3D reconstruction tasks. This data bottleneck significantly hinders progress in the field.

1.3 Objectives

In response to the problem of data ambiguity in sonar imaging, this work's primary aim is to develop and validate a comprehensive deep learning methodology for 3D reconstruction. The following high-level objectives guide the research:

- To develop and contribute a novel, large-scale annotated sonar dataset to address the

resource scarcity in the field and enable robust training and evaluation of learning-based 3D reconstruction models.

- To design and implement a deep learning framework to estimate the elevation angle from a single 2D sonar image.
- To validate the proposed methodology by quantitatively demonstrating its performance against state-of-the-art methods and qualitatively confirming its effectiveness through sim-to-real transfer experiments on real-world sonar data.

1.4 Organization of the Text

The organization of this work is given as follows: Chapter 2 presents a literature review of relevant works on 3D reconstruction of cameras and sonar, and brings a deep evaluation of sonar datasets for 3D reconstruction, followed by Chapter 3 where a Theoretical Background about the main characteristics of imaging sonars is presented. The presentation of simulators, the simulation environment, and the methodology proposal are provided in Chapter 4. All the results obtained with this work, along with the discussion, are presented in Chapter 5. The conclusions of this work are presented in Chapter 6.

2 LITERATURE REVIEW

There is a large literature on 3D reconstruction, among which the most commonly used sensors for this activity are cameras, lidars, sonars, and radars. In this work, since the operation of radars and lidars in an underwater environment is very limited and the nuances of the environment can also lead to different methodologies, since obtaining data in aquatic environments is a more complex activity that can lead to drastic changes in methods, with this in mind, this work focuses mainly on reviewing work on camera and sonar applications.

The largest scope of application for imaging radars today is in their use for autonomous cars, as evident in the papers [36], [19]. As with sonars, there is a high level of ambiguity associated with the use of radar images, and many techniques for processing these images are similar. Still, due to the great difference in medium, radars will not be explored further in this work.

2.1 3D Reconstruction of Optical Images

Underwater 3D reconstruction using cameras is a reasonably mature area of research, and one of the biggest advantages of using optical sensors is that they are relatively inexpensive and effective under certain conditions. Structure from Motion (SfM) is a method used to perform 3D reconstruction of underwater scenes using cameras. The strategy was originally described by Longuet-Higgins in [20]. This technique is based on performing 3D reconstruction using a monocular camera, where images of a given scene are triangulated to determine the relative position of the camera and its movements.

Due to the refraction characteristics of the aquatic environment, more modern techniques such as Refractive Structure from Motion (RSfM) are employed, where this method utilizes the refraction index of the water to estimate the camera's position more accurately. In this regard, it is worth highlighting the work by Chadebecq et al. [3], in which an RSfM framework was developed for a camera looking through a thin refractive interface to refine an initial estimate of the relative camera pose.

Another strategy used for 3D reconstruction is photometric stereo. These methods

have been well studied and show very promising results. Still, their performance can vary considerably according to some particularities of an underwater environment, such as light scattering, refraction, and energy attenuation [42].

In photometric stereo 3D reconstruction, it is worth highlighting some notable works, such as Iao et al. [15], where a multi-spectral photometric stereo approach was proposed. This method used linear iterative clustering segmentation to solve the problem of multi-color scene reconstruction. In the work by Fan et al. [11], the combination of underwater photometric stereo and underwater laser triangulation is considered. It was used to overcome the large shape-recovery imperfections and improve underwater photometric stereo performance.

2.2 3D Reconstruction of Acoustic Images

The problem of 3D reconstruction using sonar images remains open due to various issues, including noise and ambiguities in sonar. Among the various sonar models, Forward-Looking Sonars (FLS) are one of the best alternatives for 3D imaging and reconstruction. This work will focus on this specific type of sonar due to its characteristics and advantages associated with imaging and reconstruction activities.

In this context, we can highlight the work of Guerneve et al. [12], which uses a BlueView imaging sonar to perform 3D reconstruction, proposing a methodology that reconstructs the target without knowing its shape a priori and uses a simplified mathematical model of the sonar and some filters to perform the reconstruction from a series of collected images. Other works, such as Rahman et al. [29], propose a setup that uses a stereo camera, an inertial sensor, a depth sensor, and a mechanical scanning profiling sonar to perform contour-based reconstruction.

McConnell et al. [25] propose a novel strategy to deal with the ambiguity associated with the sonar's elevation angle, using two FLS sonars, one positioned vertically and the other horizontally, thus generating overlapping fields of view, making it possible to create a matching of points and try to correct ambiguity errors. Kim et al. [17] propose a methodology for generating 3-dimensional maps from a single sonar installed on an autonomous underwater vehicle (AUV). The proposal is based on two main processes: 2D image reconstruction and point cloud generation. The fusion of these two methods generates a 3D map.

The use of learning methods is growing in prominence, and in the context of 3D reconstruction for sonar imaging, this is no different. These learning methods can be divided into two main groups: those based on supervised learning and those based on self-supervised learning. The main difference between these methods lies in the presence of ground truth, with supervised methods utilizing data that includes ground truth and self-supervised methods not requiring it, as seen in works based on supervised learning. In

the paper Sung et al. [37] propose a sonar-based underwater object classification method by reconstructing an object's 3D geometry. In this work, a point cloud is generated from sonar images. Then, a neural network is used to predict the class of the object partially seen in the point cloud. The proposed framework will only successfully 3D reconstruct an object that matches the trained classes.

Debortoli et al. [4] propose an application of a neural network to address the ambiguity problem in sonar images. Their strategy employs a modified U-Net architecture [31], utilizing an elevation map as ground truth. Debortoli in this work leverages synthetic data to create a small dataset, with the network initially trained on this synthetic data generated using a simulator developed by Cerqueira et al. [2]. The synthetic dataset comprises 8,454 images of spheres, cylinders, and cubes. Subsequent fine-tuning is performed using real-world data, consisting of 4,667 sonar images obtained with a Tritech Gemini 720i imaging sonar from custom-built concrete targets in the shape of spheres, cylinders, and cubes. The proposed methodology employed a very small sample of objects, resulting in an overfit approach for 3D reconstruction.

Wang et al. [39, 38] proposed a novel ground truth representation that generates a pseudo-front-depth map. This approach aims to perform 3D reconstruction through supervised learning. They created a small simulated dataset using a Blender-based simulator. The dataset comprises an artificial environment, terrain, and spheres, with 3D CAD models serving as ground truth. The real-world dataset was acquired in a large-scale water tank using a Sound Metrics ARIS EXPLORER 3000 imaging sonar, replicating the simulated objects and scenarios. This methodology requires that the real objects to be constructed have their front-view image previously collected in simulation, making it difficult to execute in different scenarios.

Within the domain of self-supervised learning, Qadri et al. [28] presented a technique for 3D reconstruction based on a neural network architecture named NEUSIS. Their work involved creating both simulated and real-world datasets. The simulated dataset, acquired using HoloOcean [26], comprises sonar images of two objects, each in three distinct configurations. The real-world datasets were collected using a Sound Metrics DIDSON imaging sonar to capture data from a test structure submerged in a test tank. Three datasets were captured, corresponding to the sonar's feasible elevation apertures (1°, 14°, and 28°). However, the method proposed by the author has difficulties reconstructing multiple objects in the scene and requires a long training time, thus limiting its operation to more specific cases.

Wang et al. [40] employed a self-supervised learning methodology, utilizing the same dataset as in [39] and [38]. However, sensor position and motion information were incorporated as inputs for each sonar image, enabling self-supervised learning. The methodology was validated using both synthetic and real-world data. Despite its contributions, this work has two primary limitations: the long training durations characteristic of self-

supervised methodologies and a dependency on auxiliary data beyond sonar imagery, which poses challenges for reproducibility.

Given these works, the table 1 summarizes and best exemplifies the work related to 3D reconstruction using acoustic sensors.

In summary, while significant strides have been made in employing metemathical and learning-based methods for sonar image processing, the literature reveals a notable disparity in dataset availability. As highlighted by [1] in their survey, most publicly available sonar datasets are primarily geared towards object classification tasks. Conversely, datasets specifically designed for 3D reconstruction from sonar imagery remain severely limited. This scarcity underscores the need for dedicated datasets, such as the SEE dataset proposed in this work, which is more complete by providing a greater variety of objects and more detailed feature information, to facilitate advancements in 3D reconstruction methodologies for underwater environments.

Table 1: Table of works related to 3D reconstruction using acoustic images

3D reconstruction of acoustic images	Mathematical methods for 3D reconstruction	Learning algorithms for 3D reconstruction	Simulated experiments	Field experiments	Tank experiments	Single sonar	Multiple sonars	Cameras and Sonar
Guerneve, T. 2015 [12]	X		X	X		X		
Rahman, S. 20219 [29]	X			X				X
McConnell, J. 2020 [25]	X			X	X		X	
Kim, B. 2021 [17]	X				X	X		
Sung, M. 2020 [37]		X	X	X		X		
Westman, E. 2020 [41]	X			X	X			X
DeBortoli. 2019 [4]		X	X		X	X		
Wang, Y. 2022 [38] [39]		X	X		X	X		
Qadri, M. 2023 [28]		X	X		X	X		

3 THEORETICAL BACKGROUND

Building on the discussion in the previous sections, this work focuses on the application of imaging sonars. This chapter will discuss and present the main characteristics of these sensors, their operation, the primary sources of noise, and the challenges associated with working with acoustic data.

Imaging sonars are active devices that generate acoustic pulses to produce images. The waves propagate through the medium until they reach an obstacle or are completely absorbed by the environment. When a wave hits an object, part of its wave energy is absorbed and part is reflected. The reflected part that returns to the sonar is measured by its hydrophones. The wave's travel back and forth is known as a ping.

The hydrophones capture acoustic signals, which are then processed and organized based on their direction of arrival θ_{bin} and distance traveled r_{bin} to the reflecting object. This discretization process utilizes bins with an angular width of $\Delta\theta_{beam}$ (beamwidth) and a range interval of $\Delta\rho_{bin}$ (range resolution).

Signals returning from the same direction constitute a beam. A sonar ping generates a fan-shaped image I(B, b), as depicted in Figure 1, where I(B, b) represents the intensity of beam B at bin b. This image represents a range and bearing (angular variation) determined by the sonar's design, which includes factors such as acoustic frequency and the number of hydrophones. The maximum imaging range can be adjusted by modifying the ping duration.

An intrinsic problem in the generation Sonar images is the inhomogeneous resolution, in the figure 1 it can be clearly seen that the more distant bins have a larger area when compared to the bins closer to the sensor, this difference means that more distant bins have a lower effective resolution, given that each bin represents one pixel in an image and thus objects further away from the sonar have a lower resolution.

Although the image obtained by the sonar represents a 2D plane, its acoustic beams have an elevation angle, so that the effective field of view of a sonar is a 3-dimensional space. As a result of this simplification of dimensions, it becomes highly ambiguous, as the elevation information of each bin is suppressed during image formation [16]. The true field of view of a sonar can be understood by two angles: horizontal field of view

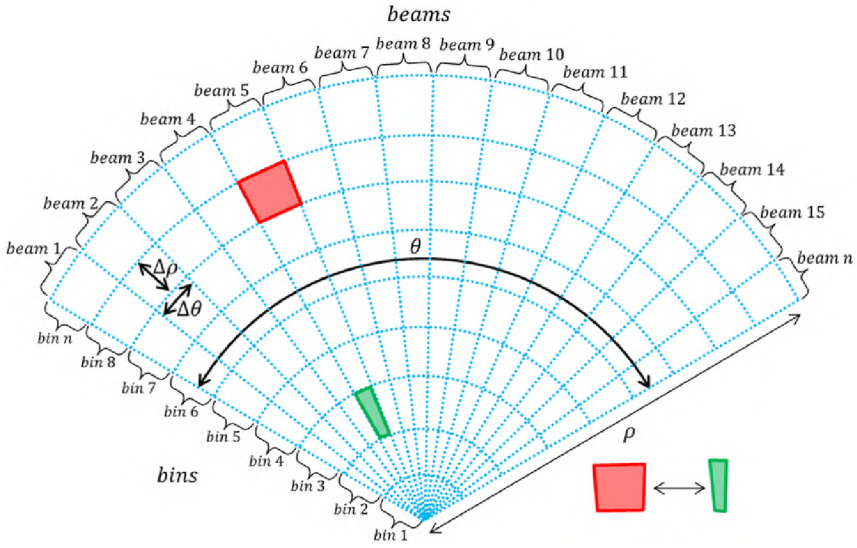


Fig 1: A scheme of an acoustic image of imaging sonar [32]

and vertical field of view, which are respectively referred to as θ and ϕ, or azimuth and elevation. The image 2 exemplifies the 3D space seen by the sonar.

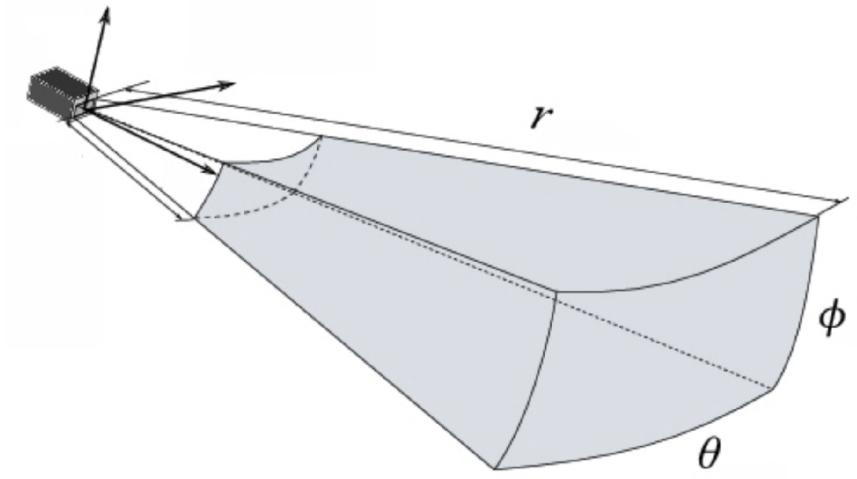


Fig 2: The 3D field of view of an imagin sonar [5]

Besides problems directly related to the acquisition of sonar data, images can also be affected by the environment or the method of collection. For example, in sonar images, deformations may occur due to the sensor's movement during the collection process. This effect, known as acoustic distortion, causes objects in the image to appear flatter. There is also a non-uniformity associated with the image pixels, which can occur due to specific water conditions, such as interference between the sonar hydrophones [33].

Acoustic sensors are often affected by environmental reverberations. The pulses emitted are frequently reverberated by objects in the environment, causing echoes to be detected by the sonar, which can generate phantom objects in the sonar image. Additionally, shadow effects occur when an object completely blocks sound from passing through, resulting in empty spots in the sonar images. In addition, this type of sensor is usually

sensitive to noise due to the signal-to-noise ratio, which is accentuated by the reflection of other waves on the surface of the water, from mutual interference causing speckle noise, from underwater engines of nearby surface vehicles, or other acoustic sensors[6].

The estimation of the suppressed elevation angle in the sonar image will be the primary focus of this work, since once this information is discovered, 3D reconstruction becomes a more trivial problem, and the other noises associated with acoustic data can still be mitigated by filters or even corrected during the methodology used to estimate the elevation angle.

4 PROPOSED METHODOLOGY

Based on the discussions in previous chapters, this work aims to achieve the proposed objectives by employing a hybrid strategy that combines development in simulated and physical environments. Taking work [28] as a starting point, where learning methods are used to carry out 3D reconstruction, and simulators are used to build a dataset that will serve as a training base for the methods developed at the end of the study. With this as a foundation, this work aims to develop a similar strategy. Still, with a stronger focus on the data, the primary objective of the work is to develop a diverse dataset comprising a broader range of sonar images in various scenarios. Throughout this chapter, each part of the proposed methodology will be described, from the choice of simulator to the proposed dataset and the proposed experiments.

The Figure 3 provides a comprehensive overview of the proposed sim-to-real methodology for 3D reconstruction. The workflow is divided into two distinct stages:

1. A simulation environment where raw sonar images and their corresponding ground truth are generated. This paired data is used to train the ElevateNet R network to predict an elevation map from a single sonar image, which is then used to create an initial 3D reconstruction.

2. Inference on real data where the model, trained exclusively on synthetic data, is then applied to real-world sonar images collected from a physical experiment. This inference step generates multiple elevation predictions from the real sensor data, which are subsequently fused to produce the final 3D reconstruction of the real-world scene.

4.1 Simulation

Simulators have emerged as tools for robotics research, mitigating the substantial material and logistical costs associated with practical experiments, particularly underwater robotics. Modern simulators offer comprehensive modeling of vehicle and environmental dynamics, along with sophisticated sensor models, enabling the acquisition of high-fidelity data and facilitating translation from simulated to real-world applications.

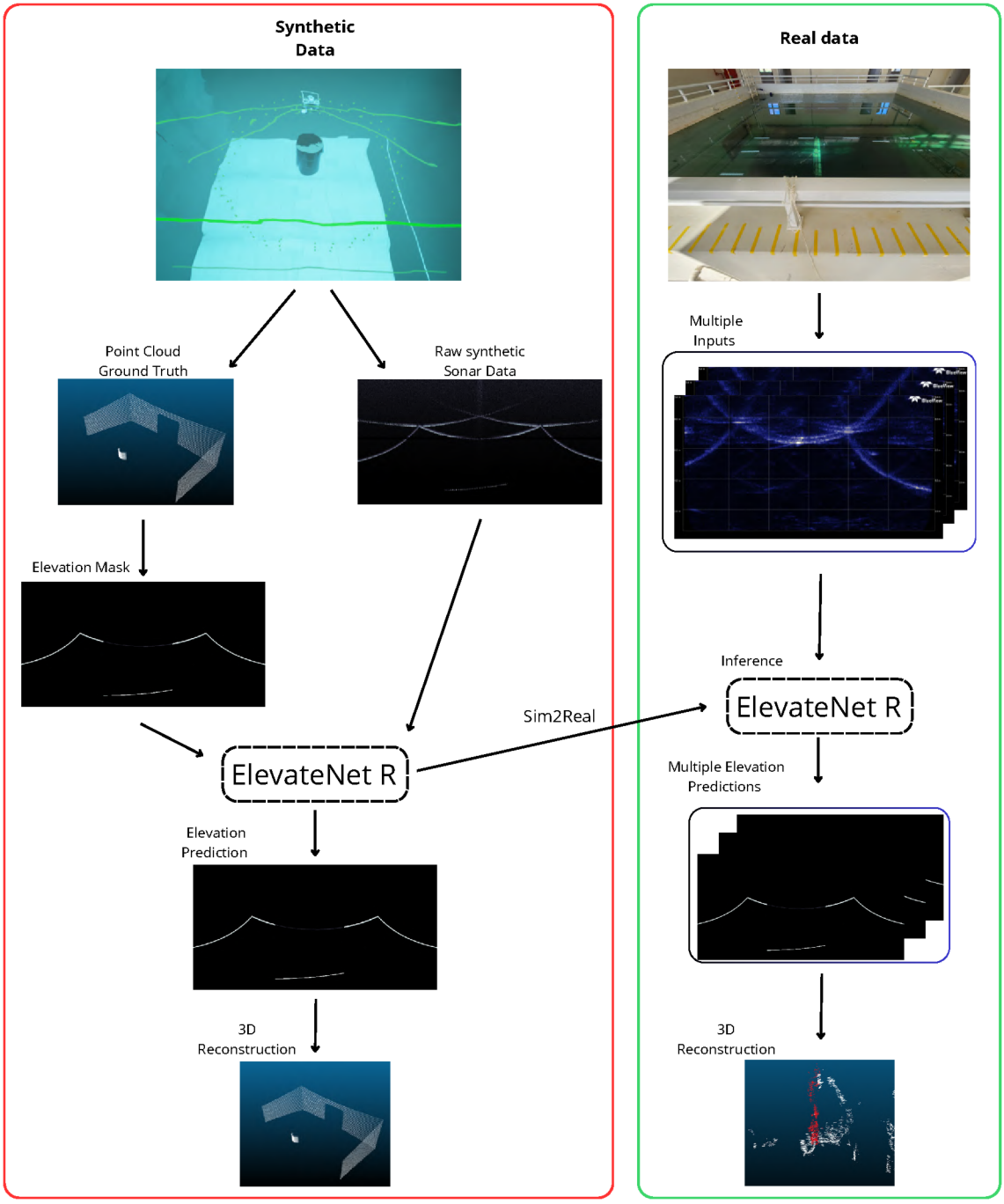


Fig 3: A comprehensive overview of the proposed sim-to-real methodology for 3D reconstruction

Among the available simulation platforms, Gazebo [18] stands out as an open-source simulator that is continuously enhanced by the community and features numerous vehicle and sensor simulation packages. It incorporates the UUV Simulator [22] package, which provides detailed simulations of remotely operated vehicles (ROVs), water dynamics, ocean currents, and sonar simulation [2].

HoloOcean [26] presents another viable simulator option, an open-source underwater simulation platform built on the Unreal Engine. This relatively new simulator incorpo-

rates models for underwater vehicles, including ROVs and AUVs, and a suite of sensors such as Inertial Measurement Units (IMUs), Global Positioning Systems (GPS), cameras, Doppler Velocity Logs (DVLs), and imaging sonar models. HoloOcean's sonar simulation replicates a diverse range of acoustic phenomena within the environment by utilizing octrees to describe the environment and subsequently performing calculations that simulate sound interaction with each environmental particle, resulting in one of the most realistic sonar simulations currently available [27].

For this study, HoloOcean [26] [27] was selected as the simulation platform. Several factors influenced this decision, including the high fidelity of its sonar imaging simulations, the ongoing updates and development of the simulator, its extensive expandability as an open-source project, and its implementation within the Unreal Engine, which enables high-quality visual effects and the creation of large-scale scenarios with manageable computational demands. The primary drawback associated with HoloOcean's sonar simulation is its potential computational cost, which can vary depending on the resolution and noise levels applied.

The purpose of using a simulator is to develop strategies that transition from simulated to real environments, where it is easy to obtain highly reliable synthetic data. This enables the reduction of real missions and experiments, thereby significantly reducing costs and accelerating development. To this end, a simulated environment was developed that represents one of the facilities of the Federal University of Rio Grande (FURG), specifically the tank located in the Aquatec building, as shown in Figure 4. The tank has dimensions of 7 meters in width, 7 meters in length, and 4.5 meters in depth. The materials and dimensions of this tank were carefully considered to ensure it could be integrated into the simulator.

4.2 Dataset Proposal

Having defined the simulator and the environment that will serve as a reference, another requirement of this work is a dataset with ground-truth information that can be used as a training base for learning methods. To this purpose, some key topics have been defined to guide the construction and generation of this data set:

- The data set needs to be composed of sonar data that has ground truth
- The simulation environment must be expandable and replicable
- The simulation environment must be able to simulate different sonar configurations
- The environment must have a range of objects, from simple structures to more complex structures that need to be minimally relevant to an underwater environment.



Fig 4: Aquatec Tank Facilities

- The simulation environment must provide a variety of scenarios for a closed environment.
- Data collection must be structured and have position reference information for each image, as well as images from different angles.

To fulfill the demands of the dataset, a world was built using the Unreal Editor to contain a representation of the tank, and a series of objects were added, including geometric solids, helices, anchors, and underwater civil construction materials such as trusses and concrete blocks. A key consideration for the objects was their dimensions, which needed to be centered within the tank and allow a vehicle to navigate around the target while collecting data without colliding with the objects.

Due to limitations in the simulator, such as the inability to dynamically position objects for interpretation by the sonar, it was necessary to develop a world in which the objects to be observed were statically pre-positioned. As a result, four different worlds were built, each with 64 tanks, 40 of which are filled with simulation props. The difference between the worlds is in the location of the props, which are floating, positioned at the bottom of the tank, positioned floating near a wall, and positioned at the bottom near a wall. The aim of creating these worlds is to expand the possibilities for locating the same object, from simple scenarios to more challenging ones, thereby expanding the variety of

simulated data. Figure 5 illustrates the scenarios and data obtained in this work, which comprise the Synthetic Enclosed Echoes (SEE) dataset.

This figure 5 illustrates the composition of the proposed dataset, which contains both synthetic and real-world components. The synthetic data, shown in the top panel, is organized into four distinct scenarios. For each scenario, the dataset provides multiple data representations: the visual context from the simulation environment, the raw sonar output in a fan-shaped Cartesian image, the processed polar image, often used as network input, and the corresponding ground-truth 3D point cloud. The bottom panel presents a parallel example from real-world data, including the physical experiment scenario and its resulting sonar images, which highlight the dataset's utility for sim-to-real validation.

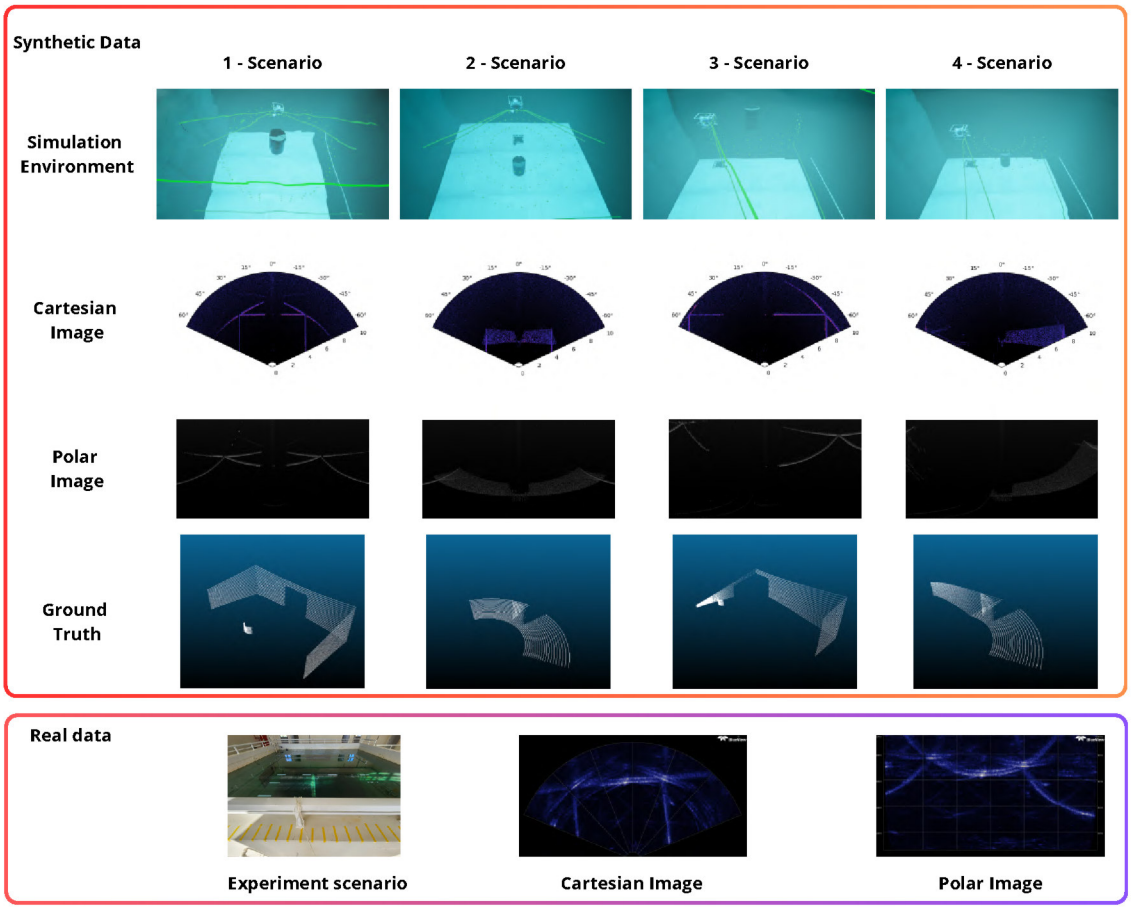


Fig 5: Synthetic Enclosed Echoes: a new dataset of synthetic and real-world sonar data for a closed underwater environment

The dataset proposed in this work is based on a collection of structures commonly found in underwater scenarios, such as barrels, trusses, and other structures widely used in submerged constructions. The Computer-Aided Design (CAD) files of these objects will serve as ground truth, and the same objects will be instantiated within the simulated environment with their geometry and materials duly described. The idea is that for each CAD file, there will be an associated point cloud as well as sonar images with different

noise intensities. Figure 6 illustrates the sixteen geometric primitives present in the dataset, Figure 7 exemplifies the eight models of propellers present in the dataset, the Figure 8 shows the eight models of anchors used in the dataset, and Figure 9 represents all the construction materials present in the dataset. With all these assets, the dataset comprises forty different objects in four distinct scenarios, resulting in 160 different combinations.

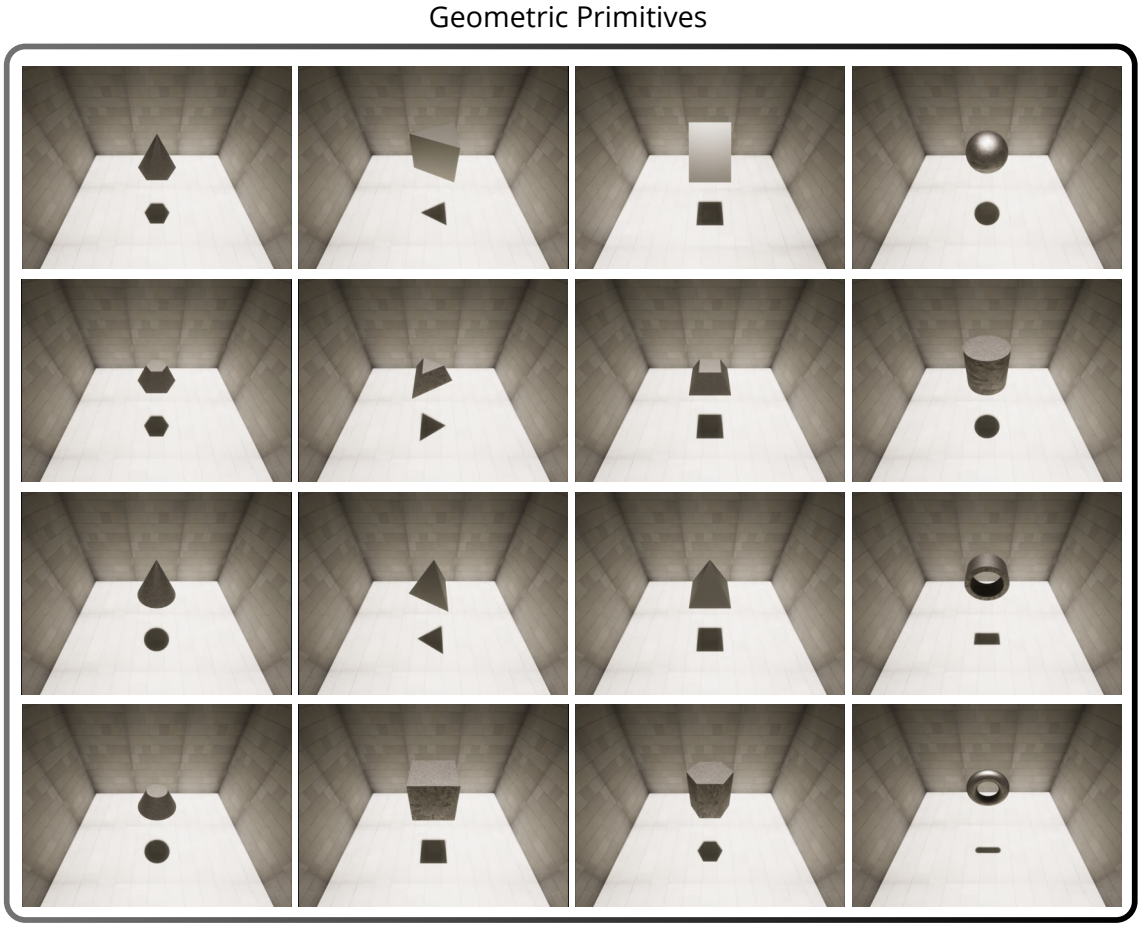


Fig 6: Geometric primitives used in the dataset.

4.2.1 Syntethic Data Collection

Once the simulation worlds were completed, it was necessary to define a methodology for data collection so that the data would be standardized and replicable across different scenarios. The strategy adopted was to create a circular trajectory around the target, with waypoints defined every 10º. The vehicle always navigates, keeping its center aligned with the object so that it never leaves the sonar's field of vision.

The simulated sonar was configured to emulate the BlueView P900, which was used in our real experiments. It operates with a 130° horizontal and 20° vertical field of view, a maximum range of 10 meters, and an output resolution of 768 angular beams by 394 range bins. To further bridge the reality gap, the simulator's noise characteristics were manually tuned to match the output of the physical sensor visually.

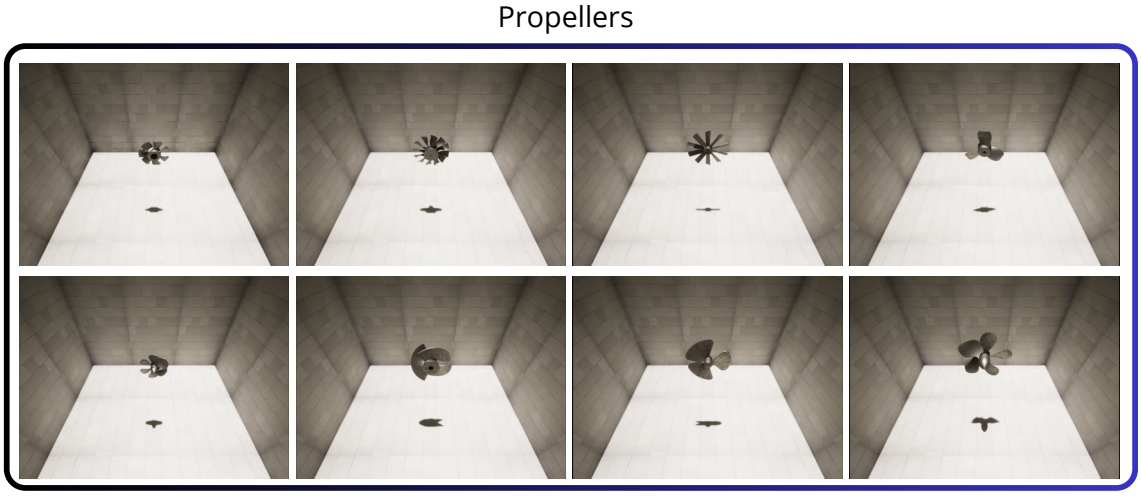


Fig 7: Propeller models used in the dataset.

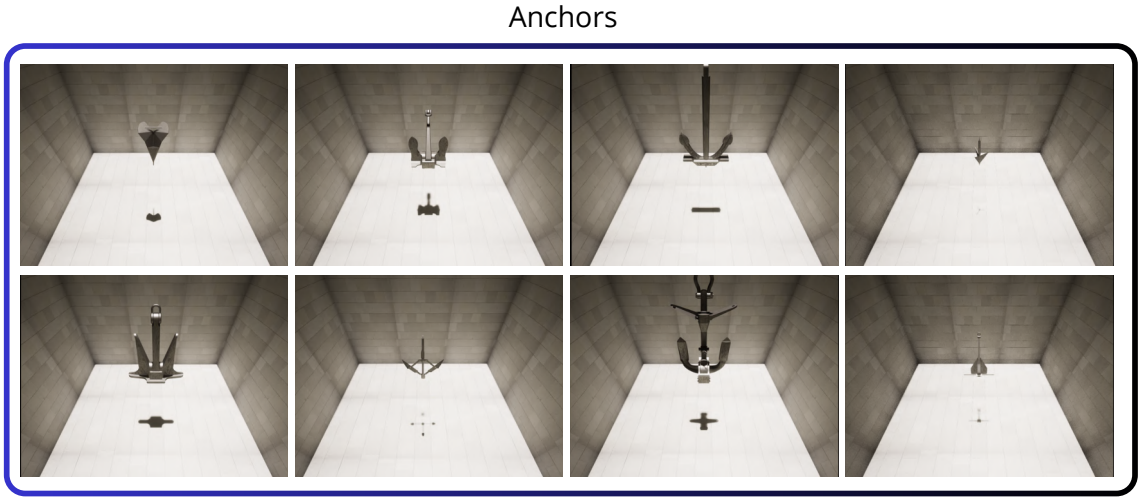


Fig 8: Anchors models used in the dataset.

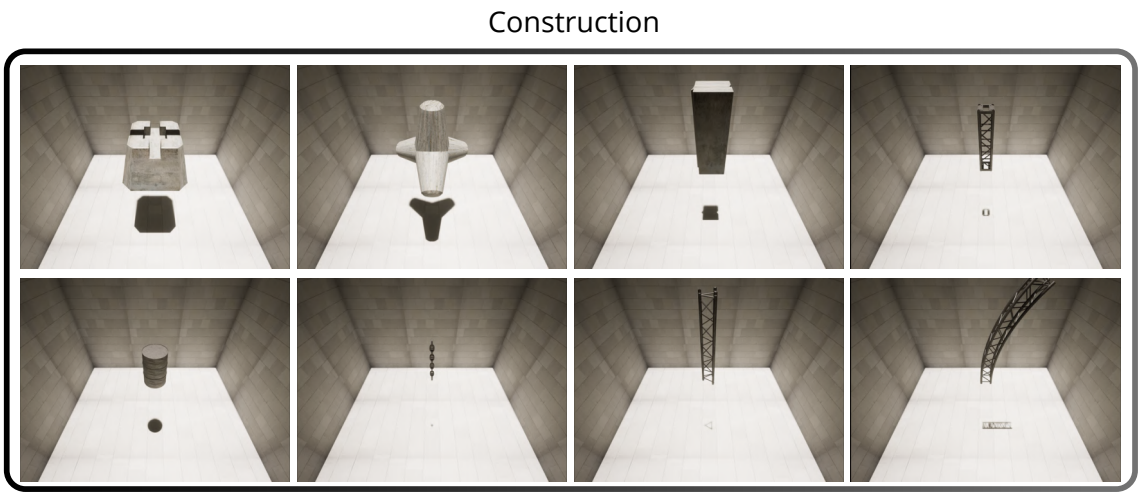


Fig 9: Construction assets used in the dataset.

A group of four scenarios was defined to ensure greater coverage of the objects' placement. The first scenario involves the object floating and positioned in the center of the tank. To collect the data, a ring path was defined, where the height of the vehicle varies as it completes a lap around the object. This circle has a radius of 2 meters, and further circles are created spaced 0.3 meters apart until the entire height of the object has been covered.

For the second scenario, the object is fixed to the bottom of the tank. For this scenario, the trajectory to be covered is circular. Still, without varying the height, by varying the radius of the circles, since the height of the reference points is always 1 meter above the object to be observed, the sonar is rotated by $-45°$ in the pitch for this scenario.

The third scenario is similar to the first one, with the key difference being that the object is positioned close to a wall of the tank. A similar trajectory to the first scenario is defined, albeit with a semi-circular path. The fourth scenario combines the third and second scenarios, where the object is positioned at the bottom of the tank and close to the wall, creating a challenging scenario for 3D reconstruction tasks. The data is collected in the same pattern as in the second scenario, but in a semi-circular trajectory. Figure 10 summarizes the four projected scenarios and the proposed paths; the green spots in the illustration indicate the waypoints that the vehicle will cover during data collection.

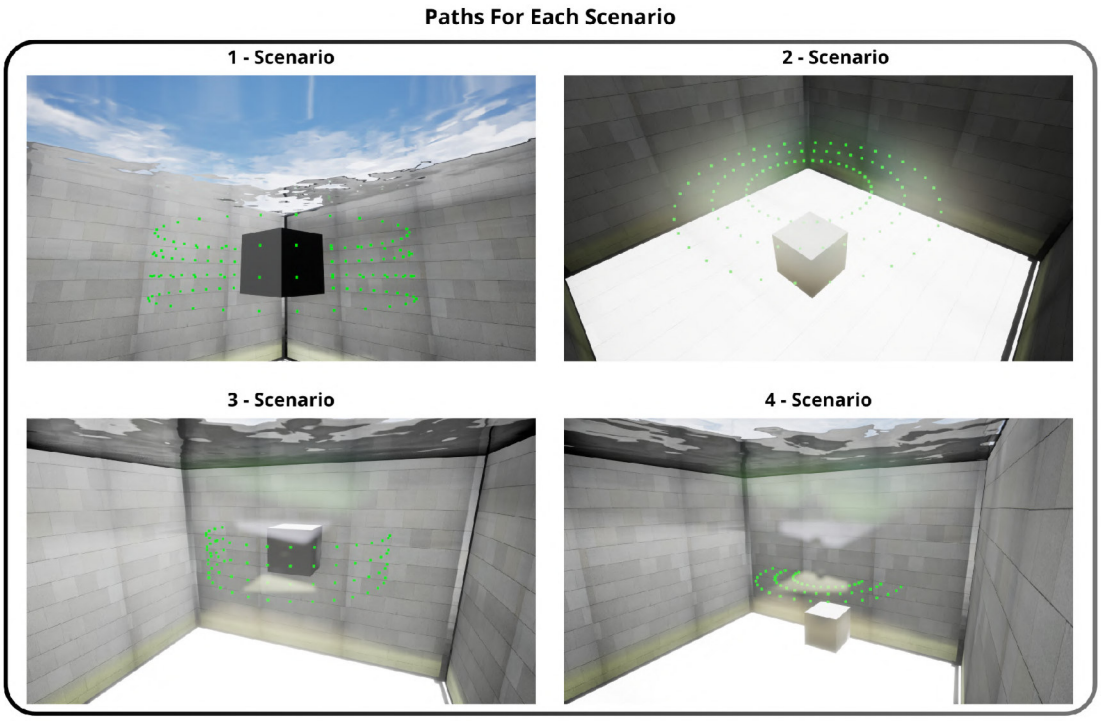


Fig 10: Proposed scenarios for the dataset and their respective way-points

4.2.1.1 Ground Truth

The ground truth generation for the sonar data is based on a simulated Lidar array designed to match the sonar's field of view, as shown in Figure 2. This array is constructed by placing a virtual laser rangefinder at each discrete azimuthal beam angle (θ) of the sonar. This line of sensors is then replicated for every one-degree increment along the elevation axis (ϕ), forming a comprehensive grid of virtual sensors.

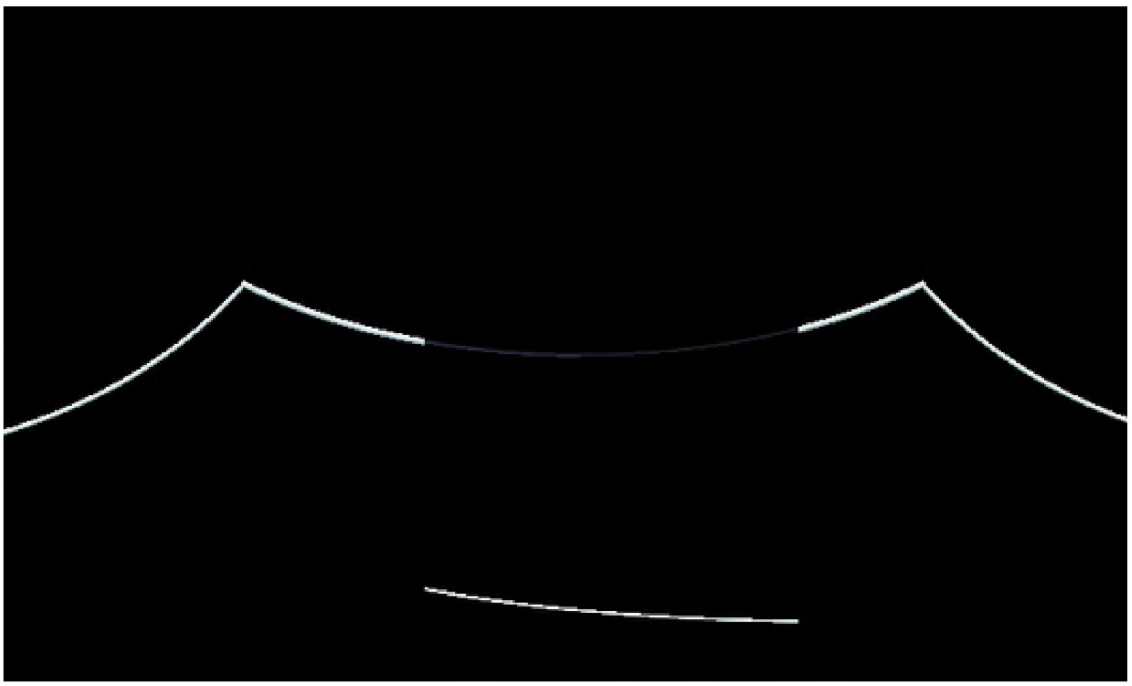
The data from this sensor grid is then transformed into a 2D ground truth map. This map uses the same range and azimuth coordinate system as the sonar image. The crucial ground truth information, the true elevation angle (ϕ) of each return, that is a value between 0 and 20, is encoded as the intensity value of the corresponding pixel. The final output is an elevation mask that is dimensionally identical to the sonar image, where each pixel's value directly provides its ground truth elevation. The figure 11 exemplifies the ground truth data used in this work.

4.2.1.2 Single-View

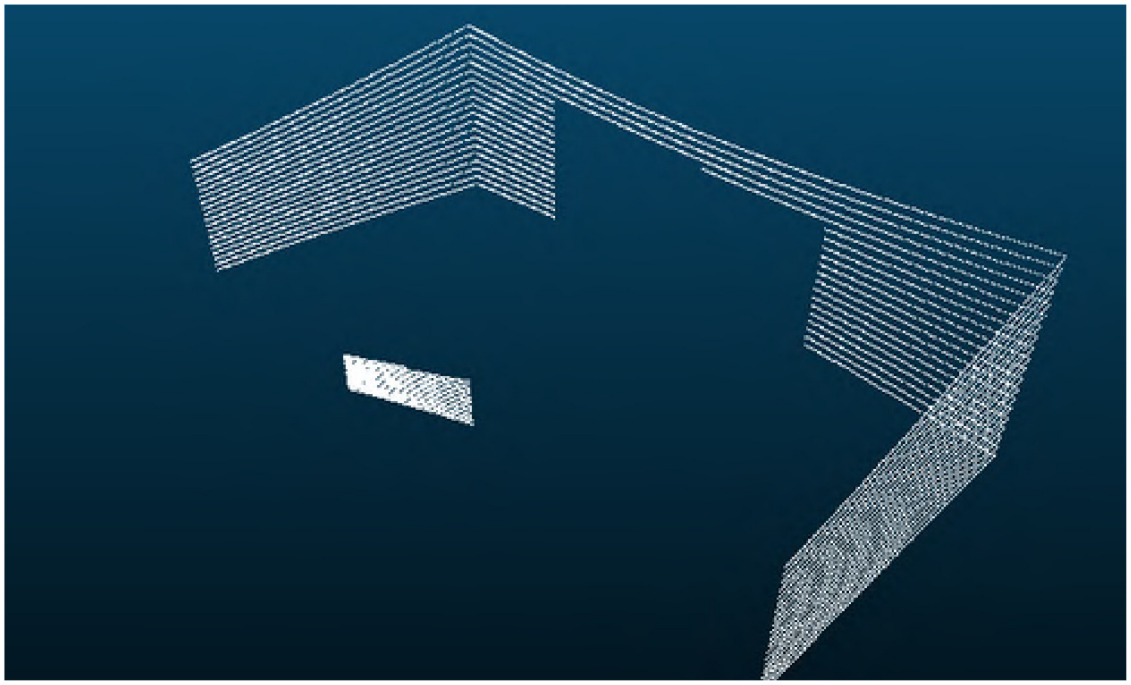
In the Single-View strategy, data is acquired from a predefined set of fixed viewpoints distributed around the object. These viewpoints are positioned within a single horizontal plane and are oriented towards the center of the target geometry. This configuration represents a conventional approach in scenarios utilizing vehicle-mounted sonar sensors, where data acquisition typically occurs at a constant altitude (or depth) and with a fixed sensor orientation. However, being restricted to a single plane of observation, this strategy has inherent limitations in resolving vertical features and capturing data from occluded regions.

4.2.1.3 Multi-View

The Multi-View strategy enhances the previous approach by introducing multi-perspective data acquisition through systematic variation of the sensor's pitch angle at each viewpoint. Specifically, data is collected at three distinct inclinations (-10°, 0°, and 10°). This technique is designed to mitigate the inherent spatial ambiguity of imaging sonars, as the introduction of pitch variations creates local parallax between adjacent views—a mechanism exploited in multi-view stereo methods to infer the elevation angle. For a conclusive validation of the methodology's generalization performance, we used real-world images from a BlueView P900 sonar, which our synthetic data was designed to emulate. Following a sim-to-real approach, the models were trained entirely on synthetic data and then tested on the real-world data. This experiment is designed to determine which input data representation yields the most accurate 3D reconstructions of previously unseen, real-world scenes. Real-world images were collected in a controlled experiment using a sonar-equipped ROV. A plastic pipe (0.2 m diameter) was used as the reconstruction target. The experiment took place in a 7 x 7 x 5 m indoor tank, as shown in Figure



(a) Elevation mask style ground truth



(b) Point cloud style ground truth

Fig 11: The two ground truth formats used in this study. 11a The 2D elevation mask provides a per-pixel ground truth for the network. 11b The 3D ground truth point cloud serves as the reference for geometric accuracy evaluation.

4, mirroring the simulation environment. The ROV performed a vertical trajectory while recording sonar images, vehicle orientation (roll, pitch, yaw), and depth from an acoustic altimeter. Figure 12 shows the vehicle used for the data acquisition.



Fig 12: Modified versions of the Blue ROV2 used during the real data colection

The collected sonar images were synchronized with the vehicle's navigation data. A pre-processing pipeline was applied to the real-world images to address discrepancies with the synthetic data. The raw acoustic images have a 16-bit echo resolution, whereas the simulated images are 8-bit. Therefore, the real images were first down-sampled to 8-bit resolution. Secondly, to mitigate the higher noise levels present in the real data, a filtering step was performed to remove low-intensity echoes that would otherwise appear as noise points in the image. The effect of this intensity filtering is illustrated in Figure 13, which shows a real sonar image before and after pre-processing.

4.3 3D Reconstruction of the Sonar Data

A fundamental challenge in 3D reconstruction from imaging sonar is resolving the elevation angle (ϕ) for each acoustic return, as the sensor typically provides only range and azimuth data. Conventional methods address this ambiguity with a simplifying planar assumption, where all returns are presumed to lie on a single plane at $\phi = 0°$. Under this assumption, each sonar image is treated as a 2D slice of the environment. By aggregating multiple, overlapping 2D scans from different viewpoints, a complete 3D model of the scene can then be progressively constructed. After that, we use a polar-to-Euler, Equation 1, angle transformation to put the sonar data in a 3D plane.

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} r\cos\theta\cos\phi \\ r\sin\theta\cos\phi \\ r\sin\phi \end{bmatrix} \tag{1}$$

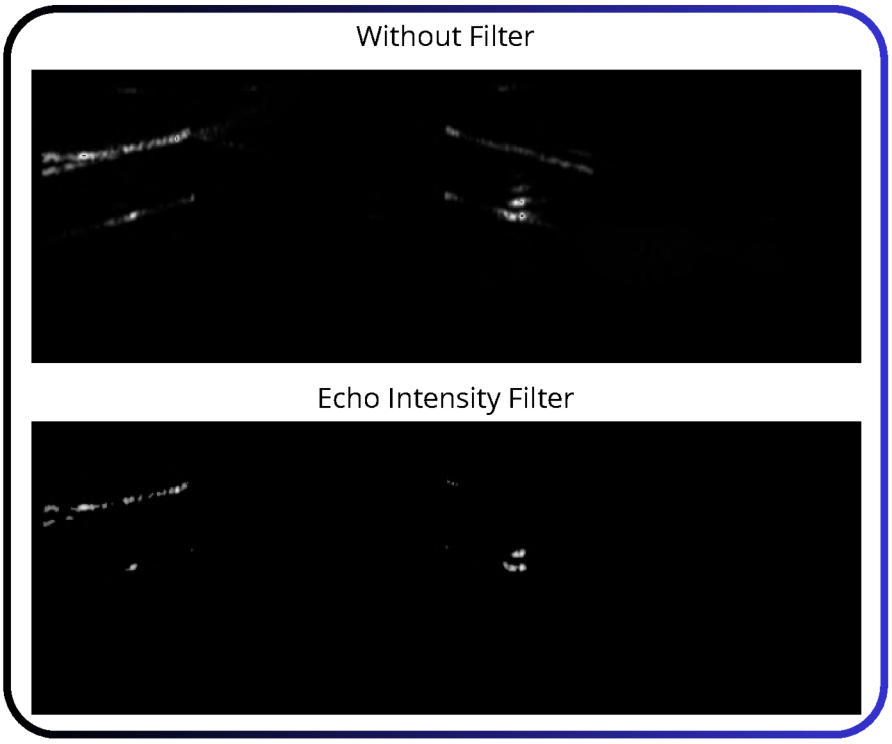


Fig 13: Real sonar data in the Polar form (Azimuth x Radius), the top image is the raw image and the bottom is the image after the echo intensity filter

The neural network is employed to estimate the elevation angle, thereby enabling more precise 3D reconstructions from sonar data. The proposed methodology builds upon the network architecture introduced by deBortoli et al. [4], adapting it into a regression framework to achieve improved performance in elevation angle prediction. The 3D reconstruction pipeline is illustrated in Figure 14.

The ElevateNET-R model is built upon a U-Net architecture, a type of encoder-decoder network well-suited for image-to-image tasks. The encoder progressively downsamples the input sonar image through a series of convolutional layers to capture high-level contextual features. The decoder then symmetrically upsamples these features, using skip connections to merge them with high-resolution features from corresponding encoder layers. This process allows the network to preserve precise spatial details in the final output.

For this work, the architecture was specifically adapted for a regression task by modifying the final output layer to predict a continuous value—the elevation angle—for each pixel. The network was trained using the Mean Squared Error (MSE) loss function, which effectively minimizes the average squared difference between the predicted elevation map and the ground-truth elevation mask, thereby optimizing the model for geometric accuracy. The model was trained for 300 epochs with a batch size of 10. We used the RMSprop optimizer with an initial learning rate of 10^{-5} and a weight decay of 10^{-8}.

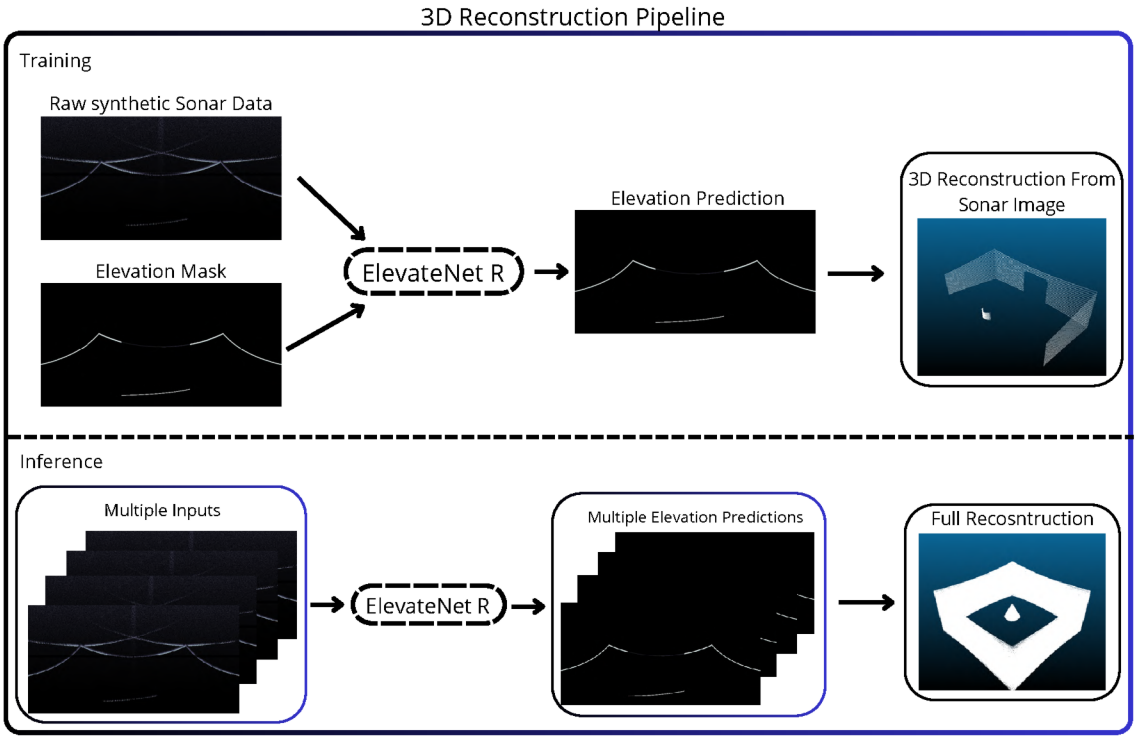


Fig 14: The ElevateNET-R reconstruction pipeline. A single sonar image is fed into the regression network, which predicts an elevation angle for each pixel. This elevation map is then used to project the sonar returns into 3D space, generating the final point cloud of the object.

4.4 Evaluation Metrics

The primary metric employed to validate the experiments was the Hausdorff Distance [34], chosen to evaluate the final 3D geometry. This metric measures the greatest distance from a point in one point cloud to the nearest point in the other, making it sensitive to shape deviations. A lower Hausdorff Distance indicates a more accurate reconstruction, with an ideal value of 0.

To enable a fair and reliable comparison, the ground truth (GT) was generated using a simulated laser sensor configured with the same geometry as the sonar. By casting beams across the same field of view (along the ϕ angle) without the characteristic ambiguity of the sonar, this approach produces a high-fidelity 3D reference. This GT serves as the basis for computing the Hausdorff Distance, ensuring that discrepancies measured by the metric reflect only the differences between the reconstructed geometry and the true object shape.

5 RESULTS

A series of experiments was conducted to evaluate current 3D reconstruction methods. Three distinct objects, a cone, a propeller, and a tetrapod, were selected for this comparative analysis across all four simulation scenarios. Figure 15 illustrates the ground truth 3D reconstruction. Figure 16 shows the 3D reconstruction when using a mathematical approach, where the estimation of $\phi = 0$ is considered, Figure 17 represents the results when using the Neusis methodology [28], Figure 18 illustrates the 3D reconstruction using the methodology proposed in [4], the results obteineds with the proposed methodoly can be see in the figure 19.

To evaluate the generalization capabilities of the proposed methodology, a new training set was created by removing the objects to be reconstructed from the original training batch. The result of this experiment can be seen in the Figure 20. To finalize the 3D reconstruction using the synthetic data, the proposed methodology was trained on a new batch of data, employing the multiview strategy; the results are shown in Figure 21.

It was observed that Neusis [28] failed to produce any discernible 3D reconstructions. The training of this method was performed on a GPU cluster comprising two NVIDIA Tesla V100 GPUs with 32GB of VRAM, enabling the simultaneous training of two models. The training duration ranged from 20 to 35 hours. It is hypothesized that Neusis's poor performance may be attributed to its original design, which does not account for environments containing multiple objects.

The classical approach exhibited significant limitations, yielding the sparsest point clouds. ElevateNET [4] successfully reconstructed the tank walls and bottom but produced suboptimal results for the primary objects of interest.

The proposed methodology outperformed ElevateNET in all scenarios, achieving superior reconstruction quality. While these results could indicate overfitting, the performance of ElevateNET-R*, where the reconstructed objects were excluded from the training dataset, demonstrated the proposed method's ability to learn generalized features, surpassing the original ElevateNET.

While the multi-view strategy was expected to improve reconstruction by providing additional geometric constraints, our results show it performed worse than the single-view

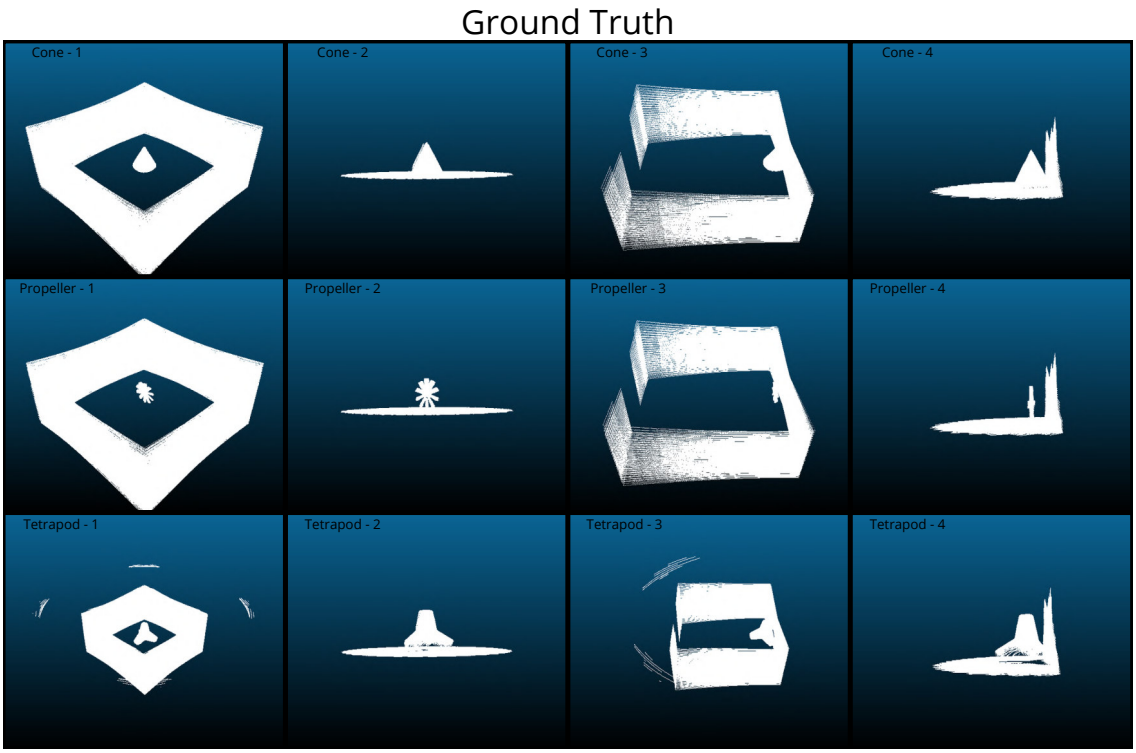


Fig 15: 3D recosntruction ground truth

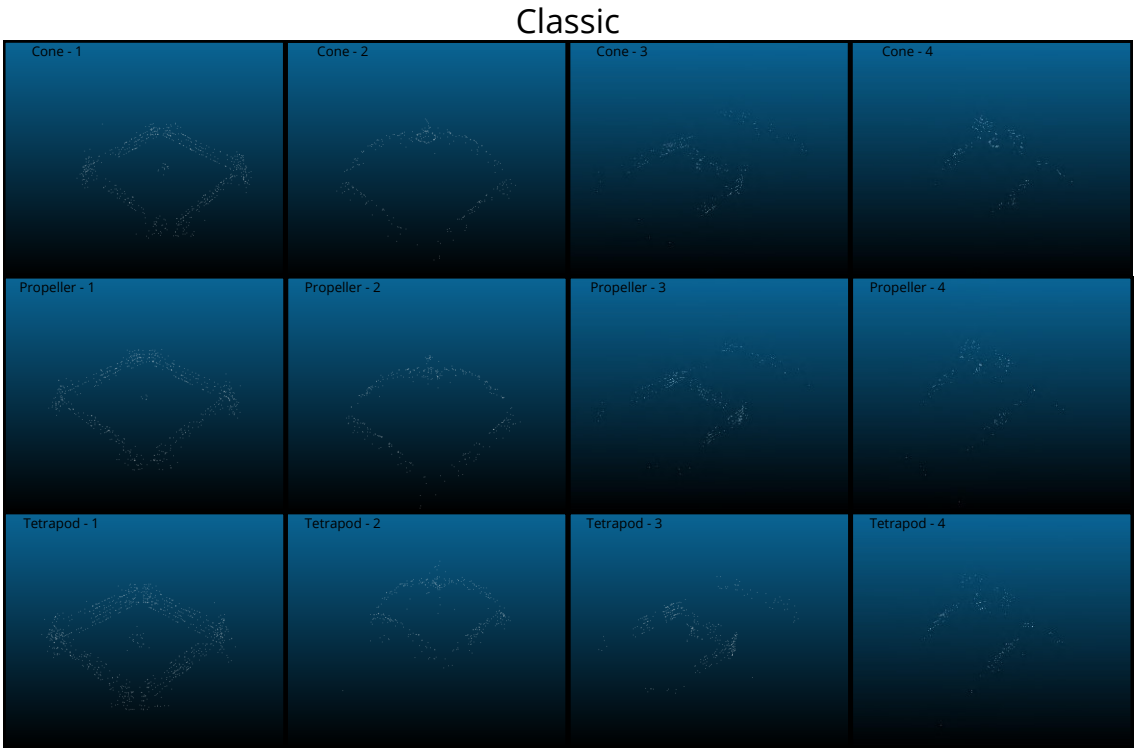


Fig 16: 3D reconstruction using a classic approach, where the estimation of the $\phi = 0$.

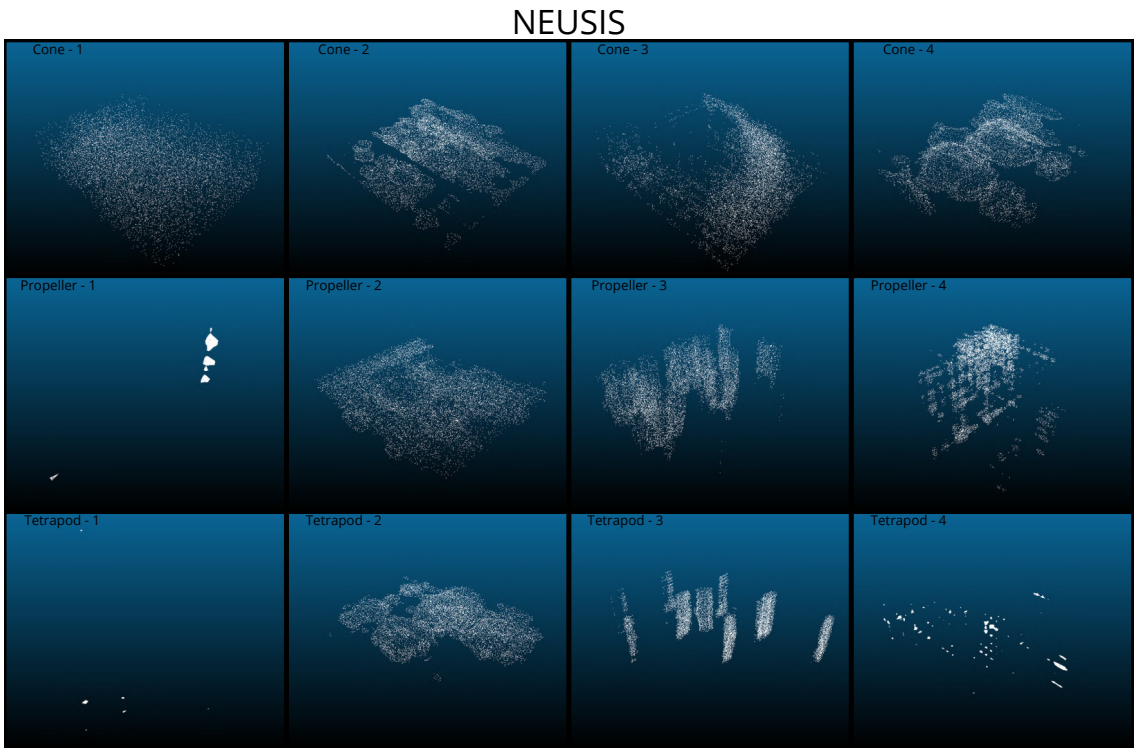


Fig 17: 3D recosntruction using the Neusis [28]

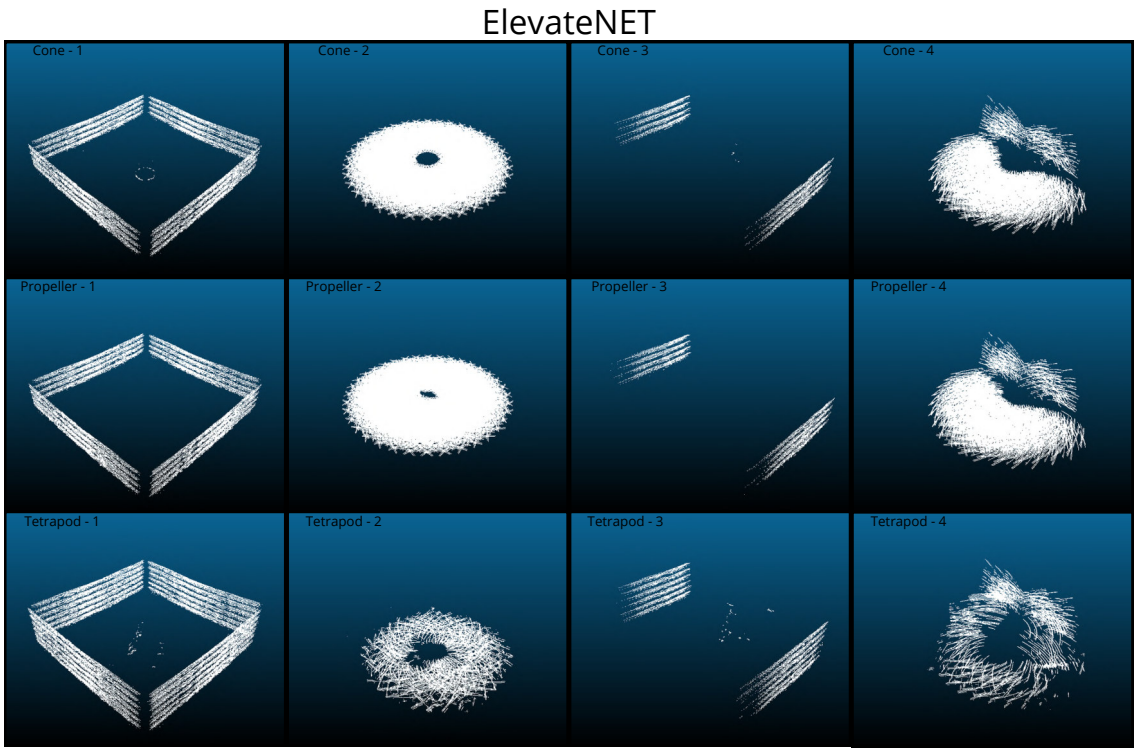


Fig 18: 3D reconstruction using the ElevateNET [4].

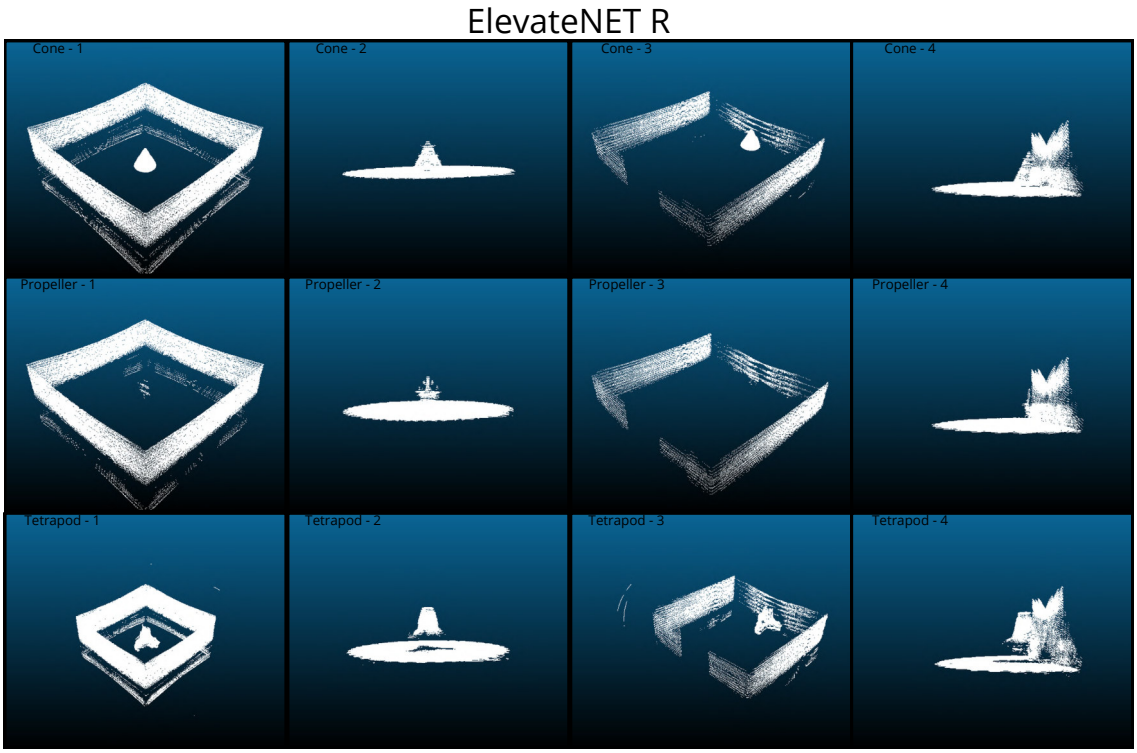


Fig 19: 3D reconstruction using the algorithm proposed in this work.

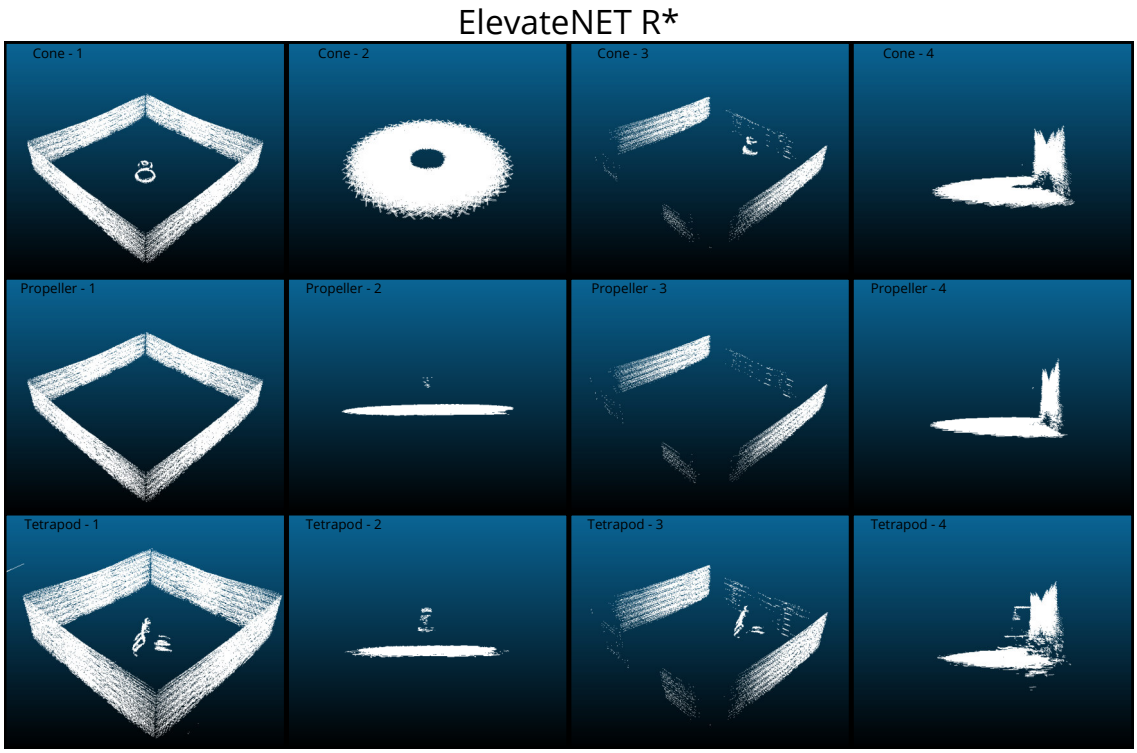


Fig 20: 3D reconstruction using the algorithm proposed in this work, where the reconstructed objects were excluded from the training dataset

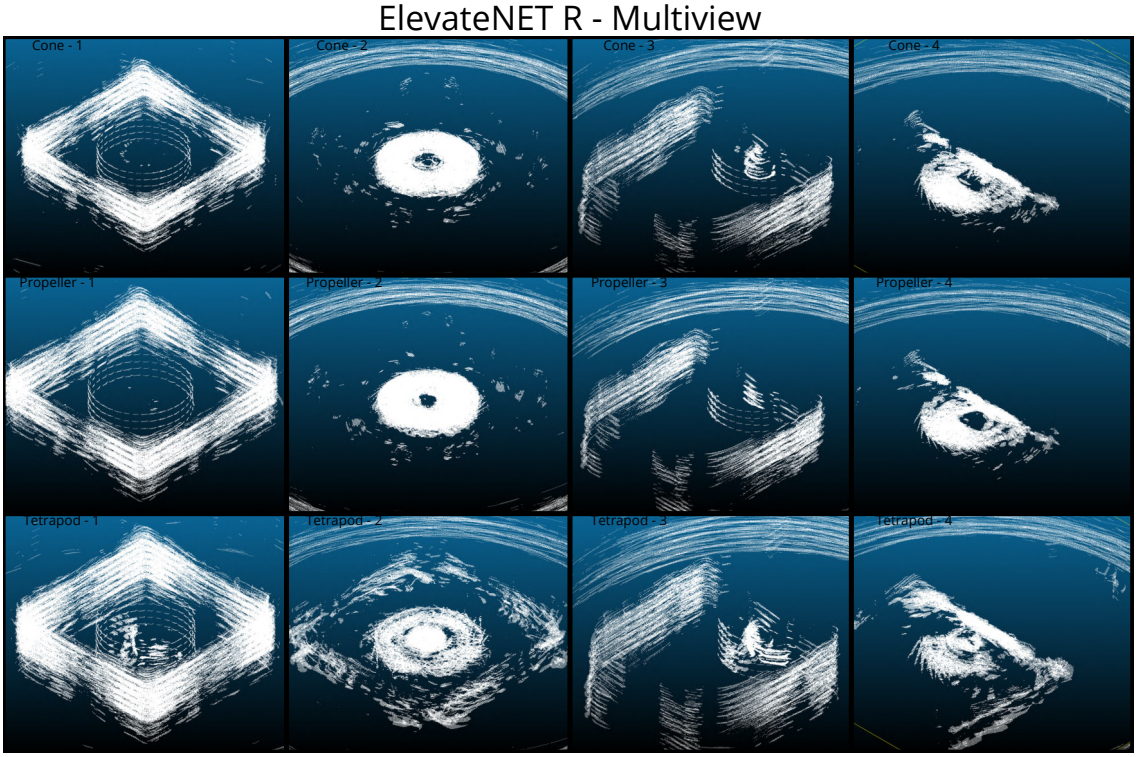


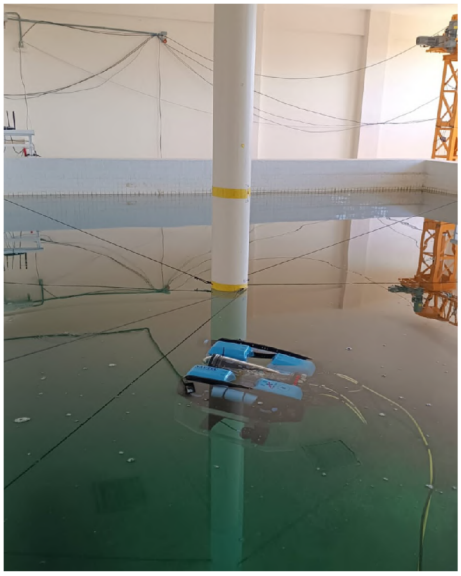
Fig 21: 3D reconstruction using the algorithm proposed in this work, where the multiview data is the training dataset.

approach. This counterintuitive outcome is likely due to the introduction of more complex and non-stationary acoustic noise in the multi-view data. Capturing images from multiple pitch angles within the enclosed tank environment resulted in a higher degree of noise, including inconsistent acoustic shadowing and more varied multipath reverberations from the tank walls, floor, and water surface. This complex noise profile appears to have hindered the network's ability to converge on the underlying object geometry, making it more difficult to distinguish true structural features from acoustic artifacts compared to the more consistent noise patterns present in the single-view dataset.

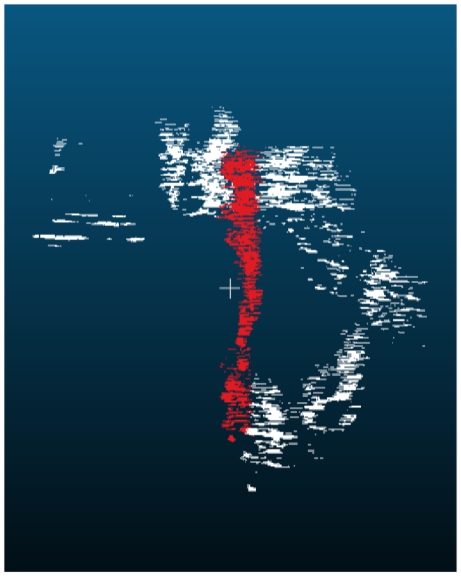
For the proposed methodology, the training time ranges from 8 to 12 hours, using a computer equipped with an NVIDIA RTX 4090 graphics card featuring 24 GB of VRAM, and the inference time ranges from 4 to 5 seconds.

The Hausdorff [34] distance, both in terms of mean and root mean square, was employed for quantitative comparison. Table 2 summarizes the numerical results.

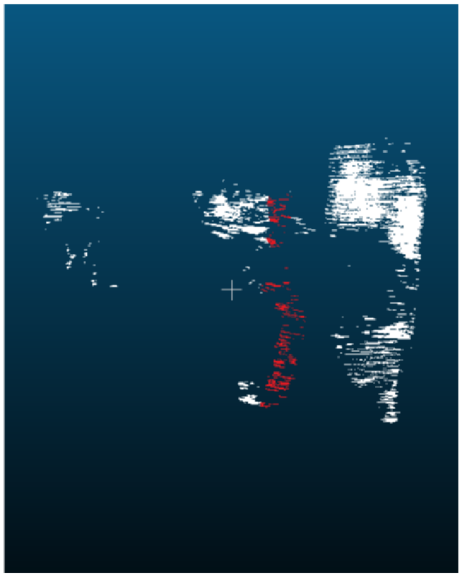
In addition to the quantitative analysis, we conducted a qualitative evaluation on real-world data to assess the sim-to-real transfer performance. For this experiment, we selected the best-performing models from our simulation experiments, the models trained with the cylindrical and mesh-based coverage strategies, which yielded the lowest Hausdorff Distance scores. These pre-trained models were then used to perform inference on the real sonar images. The resulting elevation maps were subsequently converted into 3D point clouds, providing a qualitative basis for comparison, as illustrated in the Figure 22.



(a) Experimental setup.



(b) reconstruction using ElevateNET-R single-view model.



(c) Reconstruction using ElevateNET-R multiview model.

Fig 22: Qualitative comparison of reconstruction results on real-world sonar data. 22a The experimental setup, showing the plastic pipe target within the indoor tank. 22b 3D reconstruction generated by the model trained with the ElevateNET-R single-view model. 22c 3D reconstruction generated by the model trained with a multiview data strategy. In both 22b and 22c, the red highlighted region indicates the segmented plastic pipe, which is the primary object of interest.

Table 2: Root mean square (RMS) and mean Hausdorff distance errors for the proposed experiments.

Objects	Classic		Neusis		ElevateNET		ElevateNET-R		ElevateNET-R*		ElevateNET-R Multi View	
	Mean	RMS	Mean	RMS	Mean	RMS	Mean	RMS	Mean	RMS	Mean	RMS
Cone - 1	0.3230307	0.4283095	1.8332017	2.6673460	0.2001159	0.4088612	**0.0356741**	**0.0783030**	0.1419549	0.3244358	0.1742769	0.3054860
Cone - 2	0.8455840	0.9647435	4.9558206	5.2943482	0.0343334	0.1298238	**0.0075557**	**0.0120230**	0.0383856	0.1434662	1.8985670	2.17293610
Cone - 3	0.7858377	1.9948458	3.1134795	4.0114953	0.4760223	0.7963873	**0.0752777**	**0.1940996**	0.1429360	0.2886642	1.2716735	2.4733807
Cone - 4	0.8617679	1.6550872	4.6956369	4.9297677	0.0429019	0.1117694	**0.0092025**	**0.0143699**	0.0337111	0.1044244	1.8324844	2.4293188
Propeller -1	0.3401036	0.4329429	4.1635870	5.3039222	0.2886401	0.6733100	**0.0599430**	**0.1385479**	0.2320398	0.6221491	0.1637850	0.2650502
Propeller - 2	0.8788012	0.9843889	5.2432549	5.6456313	0.0237024	0.1043392	**0.0072436**	**0.0117776**	0.0114937	0.0356924	1.8546513	2.12243933
Propeller - 3	0.6979312	1.8573711	1.7457104	2.3335265	0.8988169	1.4825769	**0.0959434**	**0.2313184**	0.1516724	0.3000660	1.3292969	2.53782966
Propeller - 4	0.7691845	1.3556808	4.6402903	5.2186675	0.0321321	0.0856037	**0.0098836**	**0.0174073**	0.0198051	0.0636210	1.8333883	2.42492189
Tetrapod - 1	0.3041520	0.3940989	2.2253484	2.6596196	0.1524595	0.3428520	**0.0268286**	**0.0774420**	0.1051520	0.2741911	0.1539010	0.2665516
Tetrapod - 2	0.5601162	0.6656648	4.8643184	5.1939374	0.0813025	0.1790147	**0.0119844**	**0.0288320**	0.0553840	0.1078482	2.04852269	2.3134872
Tetrapod - 3	0.7292655	1.9210931	2.8957832	3.8481003	0.3059703	0.5775840	**0.0483592**	**0.1291978**	0.1167501	0.2524158	1.18635537	2.4105384
Tetrapod - 4	0.8544794	1.6274831	5.6930161	5.9554727	0.1251040	0.2134584	**0.0171282**	**0.0403382**	0.0474862	0.0977827	1.86757190	2.49604351

6 CONCLUSION

The major contribution of this work lies in the dataset and, above all, the simulation environment, which enables the simulation of different types of sonar and various structures. This work successfully addressed the challenge of 3D reconstruction from ambiguous sonar imagery by developing and validating a deep learning-based methodology for estimating elevation angles. To achieve this, a high-fidelity simulation environment was created in HoloOcean to mirror a physical test tank. This environment facilitated the development of the Synthetic Enclosed Echoes (SEE) dataset, a core contribution that features a diverse collection of annotated synthetic sonar images. The final component was the ElevateNET-R, a learning-based model adapted from an existing architecture into a regression framework to predict the per-pixel elevation angle from a single sonar image.

The proposed ElevateNET-R model was evaluated and validated against its objectives. In quantitative comparisons with methods from the literature—including a classical planar assumption, Neusis [28], and the original ElevateNet [4], the proposed approach consistently demonstrated superior reconstruction quality. Furthermore, the methodology's ability to generalize from simulation to the real world was confirmed through a sim-to-real experiment. The model, trained exclusively on synthetic data, successfully performed inference and 3D reconstruction on real-world data collected with a BlueView P900 sonar, validating the effectiveness of the approach.

The primary contributions of this work are twofold: the creation of a comprehensive and expandable dataset with its simulation environment, and the successful application of a regression-based network for sonar ambiguity correction. As a next step, the SEE dataset and the simulation environment will be publicly released to foster further research and collaboration within the underwater robotics community. This will enable more extensive studies and support the development of more robust and generalized methods for processing diverse sonar data in the future.

REFERENCES

[1] Aubard, M., Madureira, A., Teixeira, L., and Pinto, J. (2024). Sonar-based deep learning in underwater robotics: Overview, robustness and challenges. *arXiv preprint arXiv:2412.11840*.

[2] Cerqueira, R., Trocoli, T., Neves, G., Joyeux, S., Albiez, J., and Oliveira, L. (2017). A novel gpu-based sonar simulator for real-time applications. *Computers & Graphics*, 68:66–76.

[3] Chadebecq, F., Vasconcelos, F., Lacher, R., Maneas, E., Desjardins, A., Ourselin, S., Vercauteren, T., and Stoyanov, D. (2020). Refractive two-view reconstruction for underwater 3d vision. *International Journal of Computer Vision*, 128:1101–1117.

[4] DeBortoli, R., Li, F., and Hollinger, G. A. (2019). Elevatenet: A convolutional neural network for estimating the missing dimension in 2d underwater sonar images. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8040–8047. IEEE.

[5] Dos Santos, M., Ribeiro, P. O., Núñez, P., Drews-Jr, P., and Botelho, S. (2017a). Object classification in semi structured enviroment using forward-looking sonar. *Sensors*, 17(10):2235.

[6] Dos Santos, M., Ribeiro, P. O., Núñez, P., Drews-Jr, P., and Botelho, S. (2017b). Object classification in semi structured enviroment using forward-looking sonar. *Sensors*, 17(10):2235.

[7] Dos Santos, M. M., De Giacomo, G. G., Drews-Jr, P. L., and Botelho, S. S. (2022). Cross-view and cross-domain underwater localization based on optical aerial and acoustic underwater images. *IEEE RA-L*, 7(2):4969–4974.

[8] dos Santos, M. M., de Oliveira Evald, P. J. D., De Giacomo, G. G., Drews-Jr, P. L. J., and da Costa Botelho, S. S. (2023). A probabilistic underwater localisation based on cross-view and cross-domain acoustic and aerial images. *Journal of Intelligent & Robotic Systems*, 108(3):34.

[9] Drews-Jr, P., do Nascimento, E., Moraes, F., Botelho, S., and Campos, M. (2013). Transmission estimation in underwater single images. In *IEEE ICCVw*.

[10] Drews-Jr, P. L. J., Nascimento, E. R., Botelho, S. S. C., and Campos, M. F. M. (2016). Underwater depth estimation and image restoration based on single images. *IEEE CG&A*, 36(2):24–35.

[11] Fan, H., Qi, L., Chen, C., Rao, Y., Kong, L., Dong, J., and Yu, H. (2021). Underwater optical 3-d reconstruction of photometric stereo considering light refraction and attenuation. *IEEE Journal of Oceanic Engineering*, 47(1):46–58.

[12] Guerneve, T. and Petillot, Y. (2015). Underwater 3d reconstruction using blueview imaging sonar. In *OCEANS 2015-Genova*, pages 1–7. IEEE.

[13] Guth, F., Silveira, L., Botelho, S., Drews-Jr, P., and Ballester, P. (2014). Underwater slam: Challenges, state of the art, algorithms and a new biologically-inspired approach. In *IEEE BioRob*, pages 981–986.

[14] Hu, K., Wang, T., Shen, C., Weng, C., Zhou, F., Xia, M., and Weng, L. (2023). Overview of underwater 3d reconstruction technology based on optical images. *Journal of Marine Science and Engineering*, 11(5):949.

[15] Jiao, H., Luo, Y., Wang, N., Qi, L., Dong, J., and Lei, H. (2016). Underwater multispectral photometric stereo reconstruction from a single rgbd image. In *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, pages 1–4. IEEE.

[16] Johannsson, H., Kaess, M., Englot, B., Hover, F., and Leonard, J. (2010). Imaging sonar-aided navigation for autonomous underwater harbor surveillance. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4396–4403. IEEE.

[17] Kim, B., Joe, H., and Yu, S.-C. (2021). High-precision underwater 3d mapping using imaging sonar for navigation of autonomous underwater vehicle. *International Journal of Control, Automation and Systems*, 19(9):3199–3208.

[18] Koenig, N. and Howard, A. (2004). Design and use paradigms for gazebo, an open-source multi-robot simulator. In *2004 IEEE/RSJ international conference on intelligent robots and systems (IROS)(IEEE Cat. No. 04CH37566)*, volume 3, pages 2149–2154. IEEE.

[19] Liu, J., Zhao, Q., Xiong, W., Huang, T., Han, Q.-L., and Zhu, B. (2023). Smurf: Spatial multi-representation fusion for 3d object detection with 4d imaging radar. *IEEE Transactions on Intelligent Vehicles*.

[20] Longuet-Higgins, H. C. (1981). A computer algorithm for reconstructing a scene from two projections. *Nature*, 293(5828):133–135.

[21] Machado Dos Santos, M., De Giacomo, G. G., Drews-Jr, P. L. J., and Botelho, S. S. C. (2020). Matching color aerial images and underwater sonar images using deep learning for underwater localization. *IEEE RA-L*, 5(4):6365–6370.

[22] Manhães, M. M. M., Scherer, S. A., Voss, M., Douat, L. R., and Rauschenbach, T. (2016). Uuv simulator: A gazebo-based package for underwater intervention and multi-robot simulation. In *OCEANS 2016 MTS/IEEE Monterey*, pages 1–8. IEEE.

[23] Massot-Campos, M. and Oliver-Codina, G. (2015). Optical sensors and methods for underwater 3d reconstruction. *Sensors*, 15(12):31525–31557.

[24] Maurell, I. P., dos Santos, M. M., de Oliveira Evald, P. J. D., Justo, B. H., Arigony-Neto, J., Vieira, A. W., Botelho, S. S., and Drews, P. L. (2022). Volume change estimation of underwater structures using 2-d sonar data. *IEEE Sensors Journal*, 22(23):23380–23392.

[25] McConnell, J., Martin, J. D., and Englot, B. (2020). Fusing concurrent orthogonal wide-aperture sonar images for dense underwater 3d reconstruction. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1653–1660. IEEE.

[26] Potokar, E., Ashford, S., Kaess, M., and Mangelson, J. G. (2022a). Holoocean: An underwater robotics simulator. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 3040–3046. IEEE.

[27] Potokar, E., Lay, K., Norman, K., Benham, D., Neilsen, T. B., Kaess, M., and Mangelson, J. G. (2022b). Holoocean: Realistic sonar simulation. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8450–8456. IEEE.

[28] Qadri, M., Kaess, M., and Gkioulekas, I. (2023). Neural implicit surface reconstruction using imaging sonar. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1040–1047. IEEE.

[29] Rahman, S., Li, A. Q., and Rekleitis, I. (2019). Contour based reconstruction of underwater structures using sonar, visual, inertial, and depth sensor. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8054–8059. IEEE.

[30] Ribeiro, P. O., dos Santos, M. M., Drews-Jr, P. L., and Botelho, S. S. (2017). Forward looking sonar scene matching using deep learning. In *IEEE ICMLA*, pages 574–579.

[31] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer.

[32] Santos, M. M., Zaffari, G. B., Ribeiro, P. O., Drews-Jr, P. L., and Botelho, S. S. (2019a). Underwater place recognition using forward-looking sonar images: A topological approach. *Journal of Field Robotics*, 36(2):355–369.

[33] Santos, M. M., Zaffari, G. B., Ribeiro, P. O., Drews-Jr, P. L., and Botelho, S. S. (2019b). Underwater place recognition using forward-looking sonar images: A topological approach. *Journal of Field Robotics*, 36(2):355–369.

[34] Serra, J. (1998). Hausdorff distances and interpolations. *Computational Imaging and Vision*, 12:107–114.

[35] Stewart, J. L. and Westerfield, E. C. (1959). A theory of active sonar detection. *Proceedings of the IRE*, 47(5):872–881.

[36] Sun, Y., Huang, Z., Zhang, H., Cao, Z., and Xu, D. (2021). 3drimr: 3d reconstruction and imaging via mmwave radar based on deep learning. In *2021 IEEE International Performance, Computing, and Communications Conference (IPCCC)*, pages 1–8. IEEE.

[37] Sung, M., Kim, J., Cho, H., Lee, M., and Yu, S.-C. (2020). Underwater-sonar-image-based 3d point cloud reconstruction for high data utilization and object classification using a neural network. *Electronics*, 9(11):1763.

[38] Wang, Y., Ji, Y., Liu, D., Tsuchiya, H., Yamashita, A., and Asama, H. (2021). Elevation angle estimation in 2d acoustic images using pseudo front view. *IEEE Robotics and Automation Letters*, 6(2):1535–1542.

[39] Wang, Y., Ji, Y., Tsuchiya, H., Asama, H., and Yamashita, A. (2022). Learning pseudo front depth for 2d forward-looking sonar-based multi-view stereo. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8730–8737. IEEE.

[40] Wang, Y., Ji, Y., Wu, C., Tsuchiya, H., Asama, H., and Yamashita, A. (2023). Motion degeneracy in self-supervised learning of elevation angle estimation for 2d forward-looking sonar. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6133–6140. IEEE.

[41] Westman, E., Gkioulekas, I., and Kaess, M. (2020). A volumetric albedo framework for 3d imaging sonar reconstruction. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9645–9651. IEEE.

[42] Woodham, R. J. (1980). Photometric method for determining surface orientation from multiple images. *Optical engineering*, 19(1):139–144.

[43] Yang, L., Kang, B., Huang, Z., Zhao, Z., Xu, X., Feng, J., and Zhao, H. (2024). Depth anything v2. *arXiv:2406.09414*.