3D Reconstruction of Underwater Structures using Forward-Scan Sonar and Gaussian Splatting

Vallabh Deogaonkar
Department of Ocean Engineering
IIT Madras
Chennai, India
0000-0002-5217-4657

Mohammed Ibrahim M
Department of Ocean Engineering
IIT Madras
Chennai, India
ibrahim252.iitm@gmail.com

Bokinala Samyak Kumar
Department of Ocean Engineering
IIT Madras
Chennai, India
na23b007@smail.iitm.ac.in

Aatmaj Bhayani
Department of Ocean Engineering
IIT Madras
Chennai, India
na22b018@smail.iitm.ac.in

Abhilash Somayajula
Department of Ocean Engineering
IIT Madras
Chennai, India
0000-0002-5654-4627

***Abstract*—3D reconstruction of underwater structures is important for inspection and maintenance but due to turbidity and low-visibility environments optical sensors fail to provide reliable data. Acoustic imaging with multi-beam sonars provides reliable data, but requires expensive hardware and produces dense measurements. In contrast, forward-scan sonars are low-cost and acquire data sequentially as sparse range–intensity measurements. But this sparseness in the data leads to strong elevation ambiguity and makes direct 3D reconstruction ill-posed. Existing Gaussian splatting methods are mainly designed for imaging sonars or multi-modal setups and do not handle single-beam data effectively. This study proposes a Gaussian splatting framework for forward-scan sonar. Each sonar ping is modeled as an elevation-ambiguous arc, and multiple views are combined to initialize and optimize 3D Gaussian primitives. An elevation-aware densification strategy helps resolve geometric ambiguity through multi-view consistency. The proposed method is validated on simulated data, showing accurate and physically consistent 3D reconstructions using low-cost forward-scan sonar systems.**

***Index Terms*—Gaussian Splatting, Forward-scan sonar, 3D Reconstruction, Sparse Acoustic Sensing, Sonar imaging**

I. Introduction

3D reconstruction of underwater structures is critical for inspection, maintenance, and monitoring of infrastructure such as pipelines, bridge piers, and offshore platforms. However, underwater environments often suffer from turbidity, low illumination, and light scattering, which significantly degrade the performance of optical sensing systems. As a result, acoustic sensing, particularly imaging sonar, has become the primary modality for perception in such conditions.

Imaging sonars provide dense range–azimuth measurements and have been widely used for underwater reconstruction and mapping tasks. However, these systems are often expensive and complex, limiting their deployment in cost-sensitive and small-scale platforms. In contrast, forward-scan sonars offer a compact and low-cost alternative. Unlike imaging sonars, forward-scan sonars acquire data sequentially as sparse range–intensity measurements, typically corresponding to a single beam at a time. This fundamental difference introduces significant challenges for 3D reconstruction. Each measurement provides only partial spatial information and suffers from severe elevation ambiguity, making the reconstruction problem inherently ill-posed.

Recent advances in neural rendering, including Neural Radiance Fields (NeRF) [1] and 3D Gaussian Splatting [2], have demonstrated strong capabilities for novel view synthesis and 3D reconstruction. These approaches have been extended to sonar imaging, enabling reconstruction from multi-beam sonar data [3], [4]. More recently, Gaussian splatting–based methods such as SonarSplat [5] have shown improved efficiency and realism in sonar image synthesis.

Despite these advances, existing methods are primarily designed for imaging sonars or multi-modal sensing setups [6], where dense observations are available. These assumptions do not hold for forward-scan sonar, where measurements are sparse, sequential, and ambiguous along the elevation dimension. As a result, directly applying existing frameworks leads to poor geometric consistency and unreliable reconstructions.

This study addresses these challenges by proposing a Gaussian splatting framework tailored for forward-scan sonar data. Instead of treating sonar returns as discrete points, each measurement is modeled as an elevation-ambiguous arc. These arcs are accumulated across multiple viewpoints to initialize a set of 3D Gaussian primitives. The scene is then optimized by rendering synthetic sonar observations and minimizing the discrepancy with measured data. To resolve elevation ambiguity, multi-view consistency is leveraged and the geometry is refined.

The proposed approach is validated on both simulated datasets. The results demonstrate that accurate and physically consistent 3D reconstruction can be achieved using low-cost forward-scan sonar systems, highlighting their potential for practical underwater inspection applications.

II. Related Work

A. Sonar-Based 3D Reconstruction

Underwater 3D reconstruction using sonar has been widely studied due to the limitations of optical sensing in turbid environments. Early approaches rely on geometric modeling and sensor fusion techniques, combining multiple sonar modalities to recover 3D structure [7]. Multi-beam and forward-looking imaging sonars provide dense range–azimuth measurements, enabling reconstruction through point cloud aggregation or probabilistic modeling. More recent work has explored forward-looking sonar systems for reconstruction by introducing mechanical scanning or additional sensing mechanisms to recover elevation information [8]. While these approaches improve reconstruction quality, they often require specialized hardware or additional motion constraints, limiting their applicability in practical deployments.

B. Neural Rendering for Sonar

Neural rendering methods such as NeRF [1] have revolutionized 3D reconstruction by modeling scenes as continuous volumetric functions. These methods have been extended to sonar imaging by incorporating acoustic image formation models. Neusis [3] introduces a NeRF-based framework for reconstructing objects from imaging sonar, while subsequent works improve efficiency and scalability using techniques such as differentiable space carving [4]. Other approaches extend NeRF to account for sonar-specific properties such as acoustic propagation and noise [9]. While these methods demonstrate promising results, they rely on computationally expensive volumetric rendering and assume relatively dense observations, which are not available in forward-scan sonar.

C. Gaussian Splatting for Underwater Reconstruction

3D Gaussian Splatting has emerged as an efficient alternative to NeRF, enabling real-time rendering and fast optimization [2]. This approach represents scenes as collections of 3D Gaussian primitives, which are rasterized to generate images. Recent work has adapted Gaussian splatting to underwater environments, incorporating physics-aware models for light attenuation and noise [10]. In the context of sonar, SonarSplat [5] introduces a Gaussian splatting framework for imaging sonar, modeling acoustic reflectivity and azimuth streaking. Similarly, fusion-based approaches such as Z-Splat [6] combine sonar and camera data to improve reconstruction quality. Recent extensions also address sonar noise modeling within Gaussian splatting frameworks [11].

However, these methods are primarily designed for imaging sonar or multi-modal setups and assume dense range–azimuth measurements. They do not explicitly address the challenges of forward-scan sonar, where data is acquired sequentially and lacks elevation information.

D. Limitations of Existing Methods

Despite significant progress, existing approaches rely on dense observations or additional sensing modalities to resolve ambiguity. Forward-scan sonar, by contrast, provides sparse and sequential measurements, where each observation corresponds to a set of possible 3D locations along an elevation arc. This makes reconstruction fundamentally under-constrained.

To the best of our knowledge, there is limited work that directly addresses 3D reconstruction from forward-scan sonar within a neural rendering or Gaussian splatting framework. This motivates the need for a method that explicitly models the ambiguity of forward-scan measurements and resolves it through multi-view consistency.

III. Problem Formulation

A forward-scan sonar measurement can be represented as an image defined in polar coordinates. Specifically, a sonar image is denoted as $I(r, \theta)$, where r represents the range from the sensor and θ represents the bearing (azimuth) angle. Each pixel in this image encodes the intensity of the acoustic return received from a particular range and bearing direction. Intuitively, the sonar image describes how strongly the emitted acoustic wave is reflected back to the sensor from different directions and distances in the environment.

Unlike conventional optical images, which are formed by perspective projection, sonar images are formed through a time-of-flight measurement process. As a result, the image lies in the (r, θ) domain rather than Cartesian coordinates, and each measurement corresponds to a set of possible 3D points in space rather than a unique location.

Forward-scan sonar (FSS) differs fundamentally from imaging sonar in the way measurements are acquired. Instead of producing dense range–azimuth images, FSS systems acquire sparse range-intensity measurements sequentially, typically corresponding to a single beam at a given time step. This measurement process introduces significant ambiguity in the underlying 3D geometry.

A. Forward Model

Let the 3D scene be represented by a function $G(\mathbf{x})$ that encodes the acoustic reflectivity of points $\mathbf{x} \in \mathbb{R}^3$. Given a sonar pose, a forward-scan sonar measurement corresponds to acoustic returns along a beam direction in which the sonar is facing. FSS operates by emitting an acoustic wavefront and measuring the intensity of the returned signal as a function of range and bearing. For a fixed bearing angle, the sonar transmits energy that propagates outward and interacts with the scene. The received signal at a particular range corresponds to the aggregated reflections from all points that lie at that range along the beam.

A key characteristic of this sensing process is that the forward scan sonar does not directly resolve elevation. As a result, all points that share the same range (r) and bearing (θ), but differ in elevation (ϕ), contribute to the measurement at the same point in the resulting image. This measurement process can be viewed abstractly as a forward operator that maps a 3D scene to a lower-dimensional set of sonar observations. Importantly, this mapping is many-to-one and multiple distinct 3D configurations can produce identical sonar measurements.

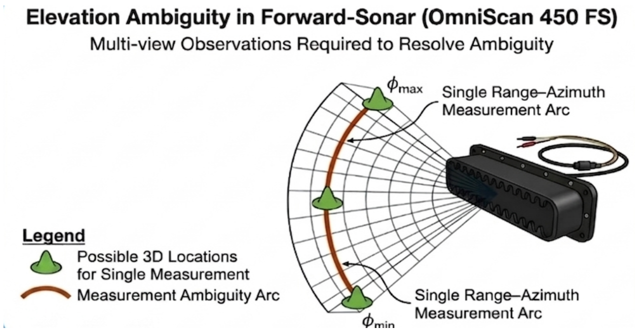


Fig. 1: Illustration of elevation ambiguity in forward-scan sonar. A single range–azimuth measurement corresponds to an arc of possible 3D locations along the elevation dimension. Any point along this arc can produce an identical measurement, making the inverse problem ill-posed. Multi-view observations are required to resolve this ambiguity.

This forward operator can be abstracted as follows:

$$I = \mathcal{F}(G) \tag{1}$$

where $\mathcal{F}$ maps the 3D scene (G) to sonar measurements (I).

B. Elevation Ambiguity

A key challenge arises from the integration over elevation ϕ. A single measurement (r, θ) does not correspond to a unique 3D point, but rather to a set of points lying along an *elevation arc*:

$$\mathcal{A}(r,\theta) = \{\mathbf{x} \in \mathbb{R}^3 \mid \|\mathbf{x}\| = r,\ \angle(\mathbf{x}) = \theta,\ \phi \in [\phi_{\min}, \phi_{\max}]\} \tag{2}$$

Thus, any point along this arc can produce an identical measurement as shown in Figure 1. This leads to inherent ambiguity in mapping sonar measurements to 3D structure.

C. Inverse Problem

The goal of 3D reconstruction is to recover the scene G from measurements I, i.e.,

$$G = \mathcal{F}^{-1}(I) \tag{3}$$

However, the characteristics of the inverse function $(\mathcal{F}^{-1})$ make the problem ill-posed due to:

1) Non-uniqueness: Multiple 3D configurations can produce identical sonar measurements due to elevation ambiguity.
2) Instability: Variations in measurement accuracy can lead to significant variations in the reconstructed geometry.
3) Incomplete observations: Each measurement provides only partial information about the scene.

As a result, directly inverting $\mathcal{F}$ is not feasible.

D. Optimization-Based Reconstruction

To address this, the reconstruction process is formulated as an optimization problem. The goal of this optimization process is to obtain the scene representation $\hat{G}$ that, when passed through the forward model, best explains the observed measurements :

$$\hat{G} = \arg\min_{G}\ \mathcal{L}(\mathcal{F}'(G), I) \tag{4}$$

where $\mathcal{L}$ is a loss function measuring the discrepancy between predicted and observed sonar measurements. In practice, the forward operator $\mathcal{F}$ is approximated using a tractable rendering model $\mathcal{F}'$, which simplifies the underlying acoustic wave propagation process.

In this work, Gaussian based scene representation is adopted and multi-view observations are leveraged to resolve elevation ambiguity. By aggregating measurements across different viewpoints, the feasible set of solutions is progressively constrained, enabling recovery of a consistent 3D structure.

IV. Simulation Model and Data Collection

A. Forward-Scan Sonar Simulation Model

To evaluate the proposed reconstruction framework in a controlled setting, a physics-inspired forward-scan sonar simulator that generates synthetic sonar images from 3D geometry is used. The simulator follows the general sonar image formation model used in prior work [3], while adopting a computationally efficient ray-based approximation.

A forward-scan sonar is a time-of-flight sensor that measures acoustic returns in a polar image defined over range and azimuth. The resulting sonar image is represented as $I \in \mathbb{R}^{N_r \times N_\theta}$, where N_r and N_θ denote the number of range bins and azimuth beams, respectively.

The continuous sonar image formation process can be expressed as an integration over range, azimuth, and elevation:

$$I(r_i,\theta_j) = \int_{\phi_{\min}}^{\phi_{\max}} \int_{\theta_j^-}^{\theta_j^+} \int_{r_i^-}^{r_i^+} E_e\, T(r,\theta,\phi)\, \sigma(r,\theta,\phi)\, dr\, d\theta\, d\phi \tag{5}$$

where E_e is the emitted acoustic intensity, $T(\cdot)$ denotes transmittance, and $\sigma(\cdot)$ represents the reflectivity of the scene. Here $\theta_j^- = \theta_j - \epsilon_a$, $\theta_j^+ = \theta_j + \epsilon_a$, $r_j^- = r_j - \epsilon_r$ and $r_j^+ = r_j + \epsilon_r$ where ϵ_a and ϵ_r are the azimuth and range resolutions respectively

In practice, evaluating this integral directly is computationally expensive. Therefore, it is approximated in the forward model using discrete sampling within each beam. For each azimuth beam θ_j, finite set of elevation angles $\{\phi_k\}$ and azimuth offsets $\{\theta_l\}$ are sampled within the beam width. For each sampled direction, a ray is cast into the scene to compute intersections with the geometry.

Let r_{hit} denote the distance to the closest intersection point along a sampled ray. The contribution of this ray to the sonar image is assigned to the corresponding range bin r_i if $r_{hit} \in [r_i - \epsilon_r, r_i + \epsilon_r]$. The discretized approximation of the sonar image can then be written as:

$$I(r_i,\theta_j) \approx \sum_{k,l} w(\theta_l,\theta_j)\, \alpha(r_{hit})\, \Delta\phi\, \Delta\theta \tag{6}$$

where $w(\cdot)$ is a beam pattern weighting function and $\alpha(r_{hit})$ represents the range-dependent contribution.

In this study, the beam pattern is modeled as a Gaussian:

$$w(\theta_l, \theta_j) = \exp\left(-\frac{1}{2}\left(\frac{\theta_l - \theta_j}{\epsilon_a}\right)^2\right) \tag{7}$$

The range contribution is approximated using a simplified model:

$$\alpha(r) = E_e \, T \, \sigma \, \Delta r \tag{8}$$

where T and σ are treated as constants for computational efficiency. This corresponds to a linear approximation of the range integral used in the continuous formulation. This modeling choice accounts for the gain adjusted outputs which sonars provide as their measurements.

The final sonar image is obtained by accumulating contributions from all sampled rays across all beams. This formulation provides a tractable approximation of the sonar image formation process while capturing key physical characteristics such as beam spread and range-dependent accumulation.

B. Data Collection Strategy

To generate diverse multi-view sonar observations, an automated data collection pipeline that samples sensor poses around the scene is implemented. The sensor is constrained to lie on a predefined sampling plane (e.g., XY, XZ, or YZ), while the remaining two coordinates are sampled within specified bounds. This enables controlled coverage of the scene from multiple viewpoints.

At each sampled position, the sensor is oriented to face the centroid of the structure, ensuring that the region of interest remains within the field of view. To increase viewpoint diversity, small random perturbations are added to the sensor orientation. This results in a distribution of poses that vary in both position and viewing direction. For each pose, a sonar image is generated using the simulation model, and the corresponding sensor pose is recorded. The dataset thus consists of paired sonar images and poses, which are used for reconstruction. All captures, including sensor intrinsics and extrinsics, are stored for reproducibility.

V. 3D Reconstruction Methodology

A. Gaussian Scene Representation

The 3D scene is represented as a set of N Gaussian primitives:

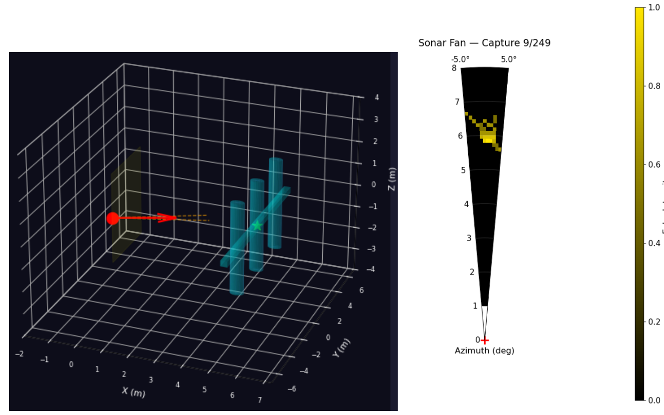
$$\mathcal{G} = \{\mathcal{G}_i\}_{i=1}^{N} \tag{9}$$

where each Gaussian $\mathcal{G}_i$ is parameterized by its mean $\boldsymbol{\mu}_i \in \mathbb{R}^3$, covariance $\boldsymbol{\Sigma}_i \in \mathbb{R}^{3\times3}$, opacity $\alpha_i \in [0,1]$, and reflectivity $\rho_i \in \mathbb{R}^+$.

The covariance is represented using anisotropic scaling and rotation:

$$\boldsymbol{\Sigma}_i = \mathbf{R}_i \, \mathbf{S}_i^2 \, \mathbf{R}_i^\top \tag{10}$$

where $\mathbf{R}_i$ is derived from a quaternion and $\mathbf{S}_i$ is a diagonal scale matrix. This representation enables continuous volumetric modeling while remaining efficient for rendering and optimization.



(a) 3D structure used in simulation (b) FSS image

Fig. 2: Example of simulated data. The left image shows the 3D fence structure used in the simulation, while the right image shows the corresponding forward-scan sonar image in (r, θ) coordinates. The sonar image demonstrates how the 3D geometry is projected into the sonar domain with significant elevation ambiguity.

B. Gaussian Initialization from Sonar Observations

Unlike imaging-based methods, forward-scan sonar measurements do not directly correspond to unique 3D points. To initialize the Gaussian set, the sonar measurements are first converted into 3D point hypotheses by sampling along the elevation dimension. For each active sonar pixel (r, θ), a set of candidate 3D points are generated along the elevation arc:

$$\mathbf{x}(r, \theta, \phi_k), \quad \phi_k \in [\phi_{\min}, \phi_{\max}] \tag{11}$$

where $\{\phi_k\}$ are uniformly sampled elevation angles.

To reduce ambiguity, multi-view consistency is enforced by retaining only those points that are observed from multiple viewpoints. Specifically, points are clustered within a spatial radius and retained if they are supported by at least M distinct views. This results in a filtered set of points:

$$\mathcal{P} = \{\mathbf{x}_i \mid \text{observed in } \geq M \text{ views}\} \tag{12}$$

Each retained point initializes a Gaussian with position $\boldsymbol{\mu}_i = \mathbf{x}_i$, reflectivity proportional to the measured intensity, and initial opacity:

$$\alpha_i = \sqrt{\rho_i} \tag{13}$$

Additionally, each Gaussian is assigned a set of *view tags* indicating the sonar scans in which it contributes. This enables selective optimization based on observation support.

C. Sonar Rendering with Gaussian Splatting

Given a set of Gaussians and a sensor pose, a synthetic sonar image is rendered by projecting the 3D Gaussian primitives into the sonar image space and accumulating their contributions along each beam.

1) World-to-Sensor Transformation: Each Gaussian mean $\boldsymbol{\mu}_i \in \mathbb{R}^3$ is defined in the world coordinate frame. To render from a given sensor viewpoint, the Gaussians are transformed into the sensor coordinate frame using a homogeneous transformation matrix:

$$\mathbf{T} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix} \in \mathbb{R}^{4\times 4} \tag{14}$$

where $\mathbf{R} \in \mathbb{R}^{3\times 3}$ is the rotation matrix representing the sensor orientation, and $\mathbf{t} \in \mathbb{R}^3$ is the translation vector representing the sensor position in the world frame.

Using homogeneous coordinates, the transformed Gaussian mean is given by:

$$\boldsymbol{\mu}_i^{(s)} = \mathbf{T} \begin{bmatrix} \boldsymbol{\mu}_i \\ 1 \end{bmatrix} \tag{15}$$

The covariance is transformed using only the rotational component:

$$\boldsymbol{\Sigma}_i^{(s)} = \mathbf{R}\,\boldsymbol{\Sigma}_i\,\mathbf{R}^\top \tag{16}$$

where $\boldsymbol{\mu}_i^{(s)} = (x_i, y_i, z_i)^\top$ denotes the Gaussian mean in the sensor frame.

2) Cartesian-to-Spherical Transformation: To map the Gaussian distribution from Cartesian coordinates to sonar measurement space, the mean and covariance in sensor frame is transformed into spherical coordinates:

$$\begin{aligned} r &= \sqrt{x^2 + y^2 + z^2}, \\ \theta &= \arctan 2(y, x), \\ \phi &= \arctan 2(z, \sqrt{x^2 + y^2}) \end{aligned} \tag{17}$$

Following SonarSplat [5], this nonlinear transformation is approximated using a first-order linearization around the Gaussian mean. The covariance is propagated using the Jacobian of the transformation:

$$\boldsymbol{\Sigma}_i^{(r\theta\phi)} = \mathbf{J}_i\,\boldsymbol{\Sigma}_i^{(s)}\,\mathbf{J}_i^\top \tag{18}$$

The Jacobian $\mathbf{J}_i \in \mathbb{R}^{3\times 3}$ is given by:

$$\mathbf{J}_i = \begin{bmatrix} \frac{x}{r} & \frac{y}{r} & \frac{z}{r} \\ -\frac{y}{x^2+y^2} & \frac{x}{x^2+y^2} & 0 \\ -\frac{xz}{r^2\sqrt{x^2+y^2}} & -\frac{yz}{r^2\sqrt{x^2+y^2}} & \frac{\sqrt{x^2+y^2}}{r^2} \end{bmatrix} \tag{19}$$

This linearization enables efficient propagation of uncertainty into sonar measurement space while avoiding nonlinear integration.

3) Projection to Image Space: The sonar image is defined over discrete range and azimuth bins. Let H denote the number of range bins and W the number of azimuth beams.

The continuous spherical coordinates are mapped to pixel coordinates as:

$$u_r = \frac{r - r_{\min}}{\epsilon_r}, \quad u_\theta = \frac{\theta + \text{FOV}/2}{\text{FOV}}(W-1) \tag{20}$$

where:

- $r_{\min}$ is the minimum sensing range,
- ϵ_r is the range resolution,
- FOV is the azimuth field of view.

The projected Gaussian mean in image space is:

$$\mathbf{u}_i = (u_r, u_\theta)^\top \tag{21}$$

The covariance is reduced to the (r, θ) subspace:

$$\boldsymbol{\Sigma}_i^{(r\theta)} = \boldsymbol{\Sigma}_i^{(r\theta\phi)}[0:2, 0:2] \tag{22}$$

To convert to pixel units, a scaling matrix is applied:

$$\mathbf{D} = \begin{bmatrix} \frac{1}{\epsilon_r} & 0 \\ 0 & \frac{1}{\epsilon_\theta} \end{bmatrix} \tag{23}$$

where $\epsilon_\theta = \frac{\text{FOV}}{W-1}$.

The pixel-space covariance is:

$$\boldsymbol{\Sigma}_i^{(pix)} = \mathbf{D}\,\boldsymbol{\Sigma}_i^{(r\theta)}\,\mathbf{D}^\top \tag{24}$$

4) Gaussian Splatting: Let $\mathbf{p} = (r_p, \theta_p)$ denote a pixel location in the sonar image grid. Each Gaussian contributes to pixel $\mathbf{p}$ via a 2D Gaussian kernel:

$$f_i(\mathbf{p}) = \exp\left(-\frac{1}{2}(\mathbf{p} - \mathbf{u}_i)^\top \boldsymbol{\Sigma}_i^{-1} (\mathbf{p} - \mathbf{u}_i)\right) \tag{25}$$

where:

- $\mathbf{u}_i$ is the projected mean of Gaussian i,
- $\boldsymbol{\Sigma}_i$ is the covariance in pixel space.

5) Volumetric Accumulation and Transmittance: Following a volumetric rendering formulation, Gaussians are sorted along each beam in increasing order of range r_i.

The opacity contribution of Gaussian i at pixel $\mathbf{p}$ is:

$$\alpha_i(\mathbf{p}) = \alpha_i\, f_i(\mathbf{p}) \tag{26}$$

where α_i is the opacity of the i-th Gaussian. The transmittance up to Gaussian i is defined as:

$$T_i(\mathbf{p}) = \prod_{j<i} (1 - \alpha_j(\mathbf{p})) \tag{27}$$

The rendered intensity at pixel $\mathbf{p}$ is:

$$\hat{I}(\mathbf{p}) = \sum_i T_i(\mathbf{p})\,\alpha_i(\mathbf{p})\,\rho_i \tag{28}$$

where ρ_i denotes the reflectivity of Gaussian i.

6) Approximate Transmittance Computation: Computing per-pixel transmittance for all Gaussians can be computationally expensive. To improve efficiency, the transmittance per Gaussian by averaging is approximated over its effective support region.

Let $\mathcal{N}_i$ denote the set of pixels where $f_i(\mathbf{p})$ is non-negligible. The approximate transmittance is:

$$\bar{T}_i = \frac{1}{|\mathcal{N}_i|} \sum_{\mathbf{p}\in\mathcal{N}_i} T_i(\mathbf{p}) \tag{29}$$

The rendered image is then computed as:

$$\hat{I}(\mathbf{p}) = \sum_i \bar{T}_i\,\alpha_i\,\rho_i\,f_i(\mathbf{p}) \tag{30}$$

This approximation preserves the ordering effects of occlusion while significantly reducing computational cost.

7) View-Dependent Subset Selection: To improve computational efficiency and reduce spurious gradients, only a subset of Gaussians contributes to each rendering pass. Specifically, each Gaussian is associated with a set of views in which it is observed, and only those Gaussians relevant to the current view are considered during rendering. This results in a view-dependent sparse splatting process, which significantly stabilizes optimization.

D. Optimization Objective

The reconstruction is formulated as an optimization problem:

$$\hat{\mathcal{G}} = \arg\min_{\mathcal{G}} \sum_k \mathcal{L}(\hat{I}_k, I_k) \tag{31}$$

where $\hat{I}_k$ and I_k denote predicted and observed sonar images for view k. The reconstruction loss is calculated as the ℓ_1 norm between the predicted and the observed image.

$$\mathcal{L}_{\text{recon}} = \sum_k \|\hat{I}_k - I_k\|_1 \tag{32}$$

1) Opacity and Scale Regularization: In addition to the reconstruction loss, $\mathcal{L}_{\text{recon}}$, regularization terms are used to minimize the number of Gaussians required to represent the underlying 3D geometry. This is achieved by including regularization terms to reduce the opacity and size of the Gaussians and then pruning the Gaussians whose opacity or distance from the surrounding Gaussians reduces below a specified threshold value.

2) Elevation-Ray Regularization: When only the image reconstruction loss is used, the optimizer may distribute Gaussian opacity along elongated tubes in 3D that project to the same sonar image, resulting in vertical smearing and floating Gaussian trails. To mitigate this ambiguity, an elevation-ray regularization term is introduced. This term is introduced to discretize not only the azimuth dimension, but also the elevation dimension, and encourage the Gaussian mass along each discretized (θ, ϕ) ray to remain compact in range.

For each Gaussian, the mean position transformed into the sensor frame is first converted to spherical coordinates. The azimuth coordinate is discretized using the same field of view and beam spacing as the renderer:

$$b_{\theta,i} = \left\lfloor \frac{\theta_i + \text{FOV}/2}{\text{FOV}} (W-1) \right\rfloor \tag{33}$$

Similarly, the elevation coordinate is discretized into bins of width $\Delta\phi$:

$$b_{\phi,i} = \left\lfloor \frac{\phi_i - \phi_{\min}}{\Delta\phi} \right\rfloor \tag{34}$$

All Gaussians sharing the same pair of indices (b_θ, b_ϕ) are considered to lie on the same acoustic ray. Within each ray, the Gaussians are sorted by increasing range. For a Gaussian k on a particular ray, a scalar score is computed:

$$\ell_k = n_k \, f_k \tag{35}$$

where n_k is the number of views in which the Gaussian is observed and f_k is an image-consistency weight derived from the measured sonar intensity at the Gaussian's projected range–azimuth pixel. Gaussians supported by more views and stronger sonar returns therefore receive larger scores.

The scores are converted into a normalized mass distribution along the ray using a softmax:

$$p_k = \frac{\exp(\ell_k)}{\sum_j \exp(\ell_j)} \tag{36}$$

A compact reconstruction along the ray is encouraged using two terms. First, the entropy of the mass distribution is minimized:

$$\mathcal{H}_{\text{ray}} = -\sum_k p_k \log(p_k + \epsilon) \tag{37}$$

where ϵ is a small constant for numerical stability. Minimizing this entropy encourages the Gaussian mass to concentrate around a small number of locations instead of being spread across many depths. Second, the weighted variance of the Gaussian depths along the ray is minimized. Let

$$\bar{r} = \sum_k p_k r_k \tag{38}$$

denote the weighted mean range of the Gaussians on the ray. The depth variance is then

$$\text{Var}_r = \sum_k p_k (r_k - \bar{r})^2 \tag{39}$$

The regularization loss for a single ray is considered as

$$\mathcal{L}_{\text{ray}} = \mathcal{H}_{\text{ray}} + \lambda_{\text{depth}} \, \text{Var}_r \tag{40}$$

where λ_{depth} controls the strength of the depth compactness term.

The final elevation-ray regularization is obtained by averaging over all rays containing at least two Gaussians:

$$\mathcal{L}_{\text{elev}} = \frac{1}{N_{\text{ray}}} \sum_{m=1}^{N_{\text{ray}}} \mathcal{L}_{\text{ray}}^{(m)} \tag{41}$$

The overall training objective is modified as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{recon}} + \lambda_\alpha \|\alpha\|^2 + \lambda_s \|\mathbf{S}\|^2 + \lambda_{\text{elev}} \mathcal{L}_{\text{elev}} \tag{42}$$

where α denotes the opacity regularization and $\mathbf{S}$ denotes the scale regularization. Thus regularization suppresses diffuse vertical trails and encourages the Gaussians to converge onto compact, physically plausible structures that are consistent across multiple views and with the observed sonar intensities.

E. Opacity Gradient Alignment

Additionally an opacity update strategy is used that aligns opacity changes with the reconstruction loss gradient. To aid 3D geometry shaping with the gaussians, instead of only relying on standard gradient descent, the gradient of the reconstruction loss with respect to opacity is also computed:

$$g_i = \frac{\partial \mathcal{L}_{\text{recon}}}{\partial \alpha_i} \tag{43}$$

Opacity is then updated multiplicatively after the optimization step as:

$$\alpha_i \leftarrow \alpha_i \cdot \exp(-\eta g_i) \tag{44}$$

where η is the hyperparameter to control the strength of opacity gradient alignment. To suppress Gaussians that consistently cause poor reconstruction, an exponential moving average of the gradients is maintained and an additional decay is applied:

$$h_i^{(t)} = \beta h_i^{(t-1)} + (1-\beta)\max(g_i, 0) \tag{45}$$

$$\alpha_i \leftarrow \alpha_i \cdot \exp(-\lambda_h h_i) \tag{46}$$

where β controls the momentum of the contribution of the Gaussians which are causing poor reconstruction, λ_h is the decay rate factor for reducing the opacity of these poorly contributing Gaussians. This encourages removal of inconsistent Gaussians while preserving stable ones.

F. Density-Based Pruning

To remove floating or isolated Gaussians, a density-based pruning strategy is also introduced. For each Gaussian, the number of neighboring Gaussians within a radius r_d is computed:

$$n_i = |\{j \mid \|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\| < r_d\}| \tag{47}$$

Gaussians with fewer than a minimum number of neighbors are removed:

$$\mathcal{G} \leftarrow \{\mathcal{G}_i \mid n_i \geq N_{\min}\} \tag{48}$$

This effectively eliminates spurious isolated structures and improves reconstruction consistency.

G. Reconstruction-Aware Pruning Guard

To prevent removal of important Gaussians, the contribution of each Gaussian to the reconstruction loss is estimated using a first-order approximation:

$$s_i \approx \left| \frac{\partial \mathcal{L}_{\text{recon}}}{\partial \alpha_i} \cdot \alpha_i \right| \tag{49}$$

Gaussians with high impact scores are protected during pruning, ensuring that important structural elements are preserved even if they have low opacity or density initially during the training.

H. Discussion

The proposed formulation adapts Gaussian splatting to forward-scan sonar by explicitly accounting for sparse, ambiguous measurements. Key design choices such as elevation-based initialization, view-dependent gradient computation, and reconstruction-aware pruning enable stable optimization despite the ill-posed nature of the problem. Furthermore, collecting sonar data from diverse viewpoints also significantly improves reconstruction quality by reducing ambiguity and increasing multi-view consistency.

VI. Results and Discussion

In this section, two test cases from the simulation analysis are shown.

A. Case 1 : Reconstruction of a Fence Structure

The proposed method is implemented on a simple fence structure consisting of three vertical rods and a horizontal bar in the center as shown in Fig 2. This structure is chosen to discuss the various parameters used and highlight the challenges posed by thin, elongated geometries under forward-scan sonar measurements.

1) Data Collection and Preprocessing: Sonar data is collected from multiple viewpoints surrounding the structure. Specifically, data is acquired from all four sides of the scene, resulting in an initial set of poses. From these, a subset of 249 poses is selected by filtering out views that produce excessively sparse sonar returns. Each selected pose is associated with a sonar image, forming a dataset of paired observations used for reconstruction.

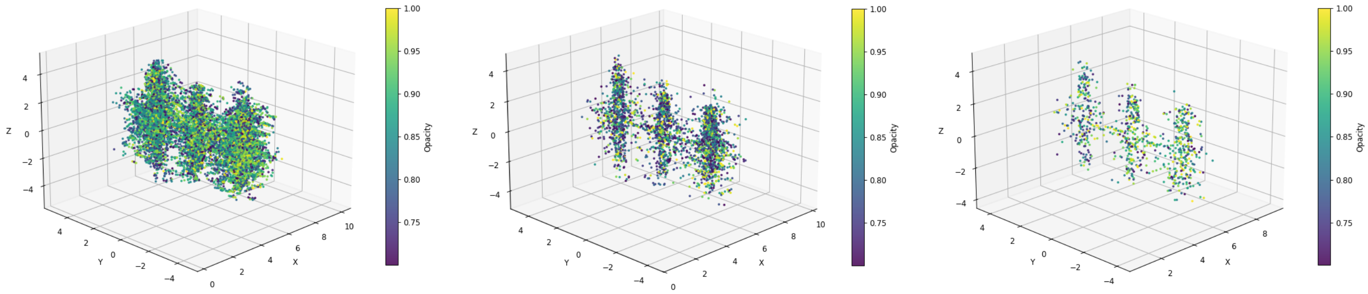
2) Gaussian Initialization: From the collected sonar data, an initial set of 3D point hypotheses is generated using the elevation-based densification approach described in Section V-B. This results in a total of 50,940 candidate points sampled along elevation arcs. To reduce ambiguity and remove spurious points, two filtering steps are applied. First an intensity thresholding is applied to remove low-confidence measurements and then multi-view consistency filtering is applied, which retains only points supported by multiple viewpoints. After filtering, a total of 38,024 points are retained and used to initialize the Gaussian set. The coordinates of these points define the Gaussian means, while opacity and reflectivity are initialized based on the corresponding sonar intensities, as described in Section V-B.

3) Optimization and Reconstruction: The initialized Gaussian representation is optimized using the Gaussian splatting framework described in Section V-C. The optimization process adjusts the Gaussian parameters to minimize the discrepancy between rendered and observed sonar images across all viewpoints.

4) Training Evolution: Fig. 3 illustrates the evolution of the Gaussian representation over the course of training. At early stages, the Gaussians are widely distributed due to elevation ambiguity. As optimization progresses, multi-view consistency and transmittance-based accumulation guide the Gaussians toward the true structure. By the final epoch, the reconstructed Gaussian distribution closely resembles the underlying fence geometry, with clear localization of the vertical rods and horizontal bar.

B. Case 2 : Reconstruction of a Truss Structure

The proposed method is further evaluated with a more complex truss structure as shown in Fig. 4. Unlike the fence structure, the truss contains a larger number of thin members, intersecting beams, and repeated patterns. Such structures are particularly challenging for forward-scan sonar reconstruction because multiple parts of the object may produce overlapping returns in the sonar image, increasing ambiguity along the elevation direction.



(a) Epoch 5 (b) Epoch 15 (c) Epoch 100

Fig. 3: Evolution of the Gaussian representation during optimization for the fence structure. Each plot is showing the opacity of the Gaussians at a particular epoch thresholded at 0.7. At early epochs, the Gaussians are widely distributed due to elevation ambiguity in the sonar measurements. As optimization progresses, the Gaussian set becomes increasingly concentrated around the true structure. By Epoch 100, the three vertical rods and horizontal connecting bar are clearly recovered and most of the spurious Gaussians are pruned.

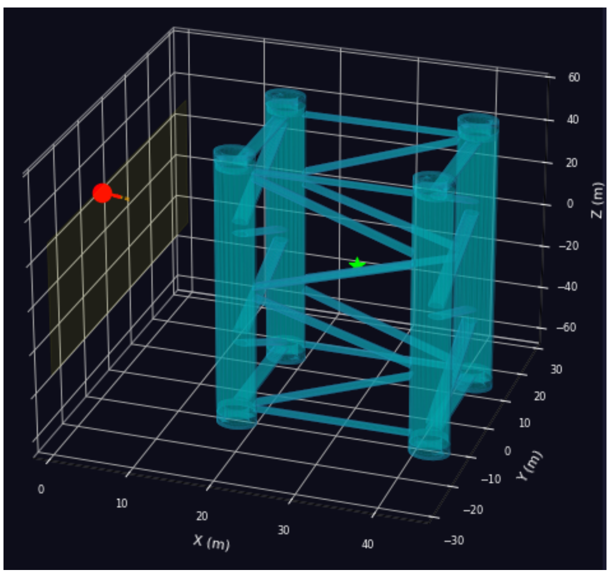


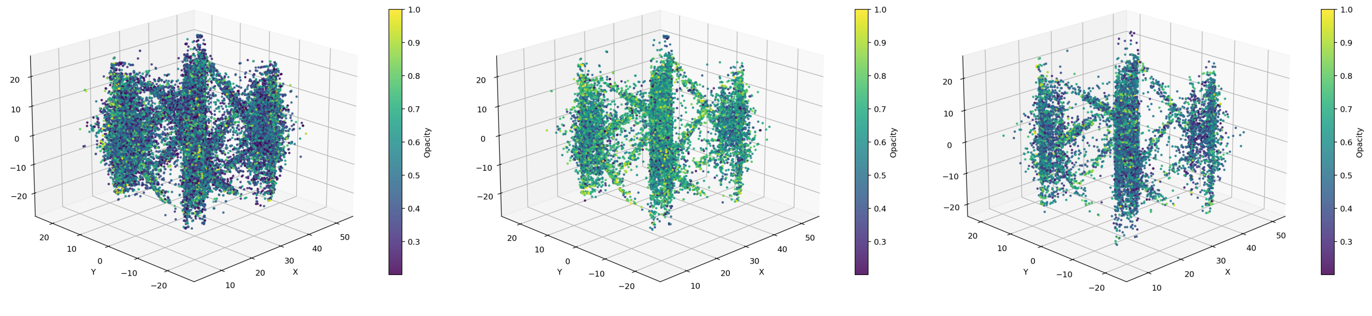
Fig. 4: Truss structure used in the second reconstruction experiment with a sampled location of the sonar sensor (red point). The structure contains multiple intersecting members and repeated geometric patterns, making the reconstruction problem significantly more challenging than the fence case.

1) Data Collection and Preprocessing: Due to the increased structural complexity, view diversity becomes significantly more important in this case. Data is therefore collected from all sides of the truss to ensure that every member of the structure is observed from multiple viewpoints. Without sufficient angular diversity, many of the individual beams become difficult to distinguish and can produce overlapping or incomplete reconstructions.

Similar to the previous case, an initial set of poses is generated around the structure and views producing excessively sparse sonar scans are discarded. The remaining poses and corresponding sonar images are used for reconstruction. A total of 886 viewpoints are selected for reconstructing of the truss structure.

2) Gaussian Initialization: The elevation-based densification approach described in Section V-B is applied to all selected sonar images to generate an initial set of 3D point hypotheses. Owing to the larger spatial extent and greater geometric complexity of the truss, this process produces a substantially larger number of candidate points than in the fence reconstruction case. The total number of candidate points generated before filtering is 811,690. Directly initializing all of these points as Gaussians would result in a prohibitively expensive optimization problem. Therefore, the same filtering procedure used in Case 1 is applied. First, low-intensity points are removed using intensity thresholding. Next, the multi-view consistency filter described in Section V-B is used to retain only points that are supported by observations from multiple viewpoints. After filtering, the number of retained Gaussians is reduced to 174,637. This initial pruning stage significantly reduces the computational cost while also improving the quality of the initialization by removing many of the spurious points arising from elevation ambiguity.

3) View-Dependent Gaussian Selection: Even after the initial filtering stage, the number of Gaussians in the truss reconstruction remains substantially larger than in the fence case. Rendering all Gaussians for every training image would therefore be computationally expensive and would introduce gradients from Gaussians that are not relevant to the current viewpoint. To address this issue, each Gaussian is associated with the set of views in which it is observed. During rendering, only the Gaussians whose view tags include the current image are selected for splatting. This view-dependent subset selection strategy, described in Section V-C7, greatly reduces the number of Gaussians considered in each rendering pass. In practice, this reduces the computational cost during training and leads to improved optimization stability by ensuring that each Gaussian only receives gradients from images in which it contributes to the reconstruction. This is especially important for the truss case, where many Gaussians correspond to



(a) Epoch 5 (b) Epoch 15 (c) Epoch 100

Fig. 5: Evolution of the Gaussian representation during optimization for the truss structure. Each plot shows the Gaussians with opacity greater than 0.2. Owing to the larger size and increased geometric complexity of the structure, the initial reconstruction contains a significantly larger number of ambiguous Gaussians. Over the course of optimization, these Gaussians progressively converge onto the truss members, while many spurious Gaussians are removed through pruning and view-consistent optimization.

different portions of the structure that are only visible from a limited set of viewpoints.

4) Optimization and Reconstruction: The filtered Gaussian set is optimized using the same Gaussian splatting framework as in Case 1. During training, the optimization progressively suppresses spurious Gaussians while strengthening Gaussians that consistently explain the sonar measurements across multiple viewpoints.

5) Training Evolution: The evolution of the reconstruction during training is shown in Fig. 5. Similar to the fence case, the initial Gaussian distribution is diffuse and contains many ambiguous points distributed along elevation arcs. As training proceeds, the Gaussians become increasingly concentrated along the true truss members. By the final epoch, the major structural components of the truss are clearly reconstructed, including the intersecting beams and repeated geometric features. The final reconstruction demonstrates that the proposed method is capable of recovering more complex underwater structures from forward-scan sonar measurements when sufficient multi-view coverage is available.

6) Quantitative Evaluation: To quantitatively evaluate the reconstruction quality, the final Gaussian means are compared against the ground-truth mesh of the corresponding structure. Following the evaluation protocol used in [5], the reconstructed Gaussian cloud is first aligned to the ground-truth mesh using Iterative Closest Point (ICP). The aligned Gaussian set is then cropped to the ground-truth bounding box and compared to the mesh using the RMS L_1 Chamfer Distance (CD-l_1) and Hausdorff Distance (HD), both expressed in meters.

The CD-l_1 metric measures the average bidirectional distance between the reconstructed points and the true surface, while the Hausdorff Distance measures the maximum mismatch between the two sets. Lower values therefore indicate better reconstruction quality.

For the fence structure from Case 1, the proposed method achieves a CD-l_1 of only 0.129 m. Since the rods are separated

TABLE I: Quantitative reconstruction accuracy for the simulated structures.

Structure	**CD-l_1 (m)**	**HD (m)**
Fence Structure	0.129	1.391
Truss Structure	0.740	10.181

by approximately 3 m and have a height of approximately 4 m, this corresponds to only a small fraction of the object dimensions. The reconstructed Gaussian cloud therefore closely matches the true geometry. The larger Hausdorff Distance of 1.391 m is primarily caused by a small number of remaining ambiguous Gaussians produced by the elevation uncertainty of the sonar measurements. These outliers do not significantly affect the overall structure but increase the maximum error.

In contrast, the truss structure from Case 2 presents a considerably more challenging reconstruction problem due to its larger size and the presence of many thin intersecting members. The proposed method achieves a CD-l_1 of 0.740 m, which remains small compared to the overall dimensions of the truss (approximately $20 \times 20 \times 40$ m). The reconstruction therefore captures the overall geometry and major truss members accurately. However, the Hausdorff Distance increases to 10.181 m, indicating that a few regions of the structure remain poorly reconstructed.

The degradation in performance for the truss structure can be attributed to two factors. First, the larger spatial extent of the structure increases the ambiguity introduced by the sonar elevation integration. Second, the thin members of the truss often generate similar sonar returns from multiple possible locations along the elevation arc. As a result, some Gaussian mass remains distributed away from the true structure even after optimization.

Fig. 6 further illustrates the reconstruction quality for the truss structure. The majority of the Gaussian means are concentrated around the true truss members, indicating that the optimization successfully recovers the dominant geometry

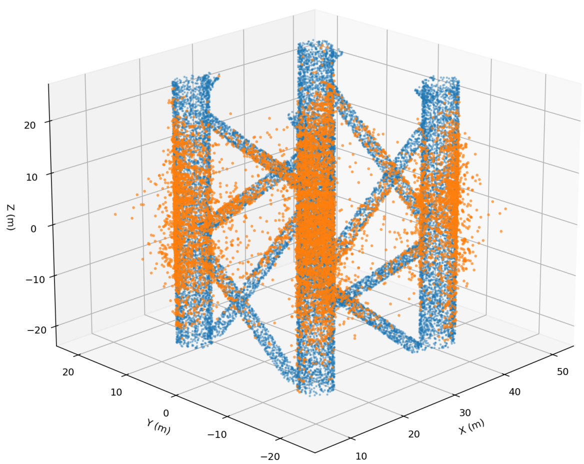


Fig. 6: Comparison between the reconstructed Gaussian means and the ground-truth truss geometry after alignment. The reconstructed Gaussian means are shown in orange and the ground-truth truss mesh samples are shown in blue. The overall structure and the major truss members are recovered accurately, although some residual Gaussian spread remains around the true members due to elevation ambiguity.

of the structure. Nevertheless, a noticeable spread of Gaussians remains around the vertical columns and diagonal members. This residual spread is consistent with the elevation ambiguity inherent in forward-looking sonar, where multiple 3D locations can produce similar observations in the (r, θ) image.

Overall, the results demonstrate that the proposed Gaussian splatting framework is able to recover both simple and complex structures from forward-looking sonar measurements. While the reconstruction quality decreases for larger and more geometrically complex scenes, the method remains capable of capturing the dominant structural elements with sub-meter average error.

VII. Conclusion and Future works

This work presents a Gaussian splatting–based framework for 3D reconstruction using forward-scan sonar, explicitly addressing the challenge of elevation ambiguity inherent in sparse range–azimuth measurements. By modeling each sonar return as an elevation-ambiguous arc and leveraging multi-view consistency, the proposed approach enables the recovery of geometrically consistent 3D structures from sequential sonar data.

The incorporation of elevation-aware initialization, view-dependent optimization, and regularization strategies allows the framework to suppress ambiguity and converge to physically meaningful reconstructions. Results on simulated structures demonstrate the capability of the method to recover both simple and complex geometries using low-cost forward-scan sonar systems. Overall, this study highlights the potential of combining physics-inspired modeling with Gaussian splatting for underwater reconstruction in sparse sensing scenarios, providing a foundation for further development toward real-world deployment.

Future work will focus on validating the proposed framework on real forward-scan sonar data collected from controlled underwater experiments, enabling quantitative evaluation under realistic noise and sparsity conditions. A systematic comparison with baseline methods, including NeRF-based sonar reconstruction, imaging sonar Gaussian splatting approaches, and traditional geometric techniques, will be conducted to assess improvements in accuracy and robustness.

Further extensions will incorporate more accurate acoustic modeling, such as range-dependent attenuation and angle-dependent reflectivity, along with potential sensor fusion using IMU or vision data. Finally, efforts will be directed toward reducing computational complexity and enabling real-time reconstruction, with the goal of deploying the system on autonomous underwater platforms and extending this work into a journal publication.

Acknowledgment

The author, Vallabh Deogaonkar (PMRF Research Scholar), gratefully acknowledges the support of the Prime Minister's Research Fellowship, India, for supporting the completion of this study.

References

[1] B. Mildenhall, P. Srinivasan, and M. Tancik, "Nerf: Representing scenes as neural radiance fields for view synthesis," in *ECCV*, 2020.
[2] B. Kerbl, G. Kopanas, and T. Leimkuhler, "3d gaussian splatting for real-time radiance field rendering," *ACM TOG*, 2023.
[3] M. Qadri and M. Kaess, "Neural implicit surface reconstruction using imaging sonar," *ICRA*, 2023.
[4] Y. Feng and W. Lu, "Differentiable space carving for 3d reconstruction using imaging sonar," *RA-L*, 2024.
[5] A. V. Sethuraman *et al.*, "Sonarsplat: Novel view synthesis of imaging sonar via gaussian splatting," *RA-L*, 2025.
[6] Z. Qu *et al.*, "Z-splat: Z-axis gaussian splatting for camera-sonar fusion," *TPAMI*, 2025.
[7] H. Joe, "Probabilistic 3d reconstruction using sonar fusion," *Sensors*, 2022.
[8] H. Zhi, "3d reconstruction using oscillatory forward-looking sonar," *JMSE*, 2025.
[9] X. Zhang, "Nerf-enhanced sonar detection," *Ocean Engineering*, 2026.
[10] W. Xing, "Uw-3dgs: Underwater 3d reconstruction with gaussian splatting," *arXiv*, 2025.
[11] S. Xu, "Nas-gs: Noise-aware sonar gaussian splatting," *arXiv*, 2026.