APERTURE SYNTHESIS WITH PHYSICS-INFORMED NEURAL NETWORKS

A. Xenaki   Science and Technology Organization, Centre for Maritime Research and Experimentation, NATO, La Spezia, 19126, Italy

A. Monti   Science and Technology Organization, Centre for Maritime Research and Experimentation, NATO, La Spezia, 19126, Italy

Y. Pailhas   Science and Technology Organization, Centre for Maritime Research and Experimentation, NATO, La Spezia, 19126, Italy

1   INTRODUCTION

Aperture synthesis is an advanced signal processing technique that enables cutting-edge imaging performance in sonar and radar systems, among various applications[1]. Synthetic aperture sonar (SAS) systems rely on the motion of an active sonar, typically mounted on an autonomous underwater vehicle, to synthesize an aperture that is much larger than the physical sonar antenna[2]. Coherent processing of the backscattered echoes from successive acoustic pulses (pings) along the motion trajectory results in underwater imaging and mapping with distinctively high resolution that is independent of range[3].

Sufficient spatial sampling of the synthetic aperture is necessary to achieve SAS reconstruction free of azimuthal ambiguities, which sets competing requirements on the physical size of the array, the imaging range, the ping repetition frequency and the speed of the platform[4]. In particular, spatial aliasing artifacts manifest in the reconstruction when the sampling distance d exceeds half the acoustic wavelength λ considering two-way propagation in active sensing[5], $d \geq \lambda/2$. Even though spatial aliasing is mitigated utilizing receiver arrays densely populated with spatially extended sensors, it can still induce distortions in the reconstructed SAS image[6,7].

This study proposes a method to alleviate the impact of potential spatial aliasing by interpolating between recordings of backscattered echoes. The proposed method is based on a physics-informed neural network (PINN), which learns a continuous representation of the underlying sound field from existing datapoints and wave propagation models. PINNs have emerged as the interface between machine learning (ML) methods and physics-based models. Specifically, PINNs are trained by fitting a relatively small amount of data combined with informative prior information in the form of physical laws from domain knowledge, resulting in robust generalization performance from limited data and incomplete models[8]. Notably, a PINN can be interpreted as an implicit neural representation of a function defined on a continuous domain. The representation is learned by optimizing the parameters of a deep neural network based on a few discrete observations and model-based constraints. The inductive bias introduced by the physics-informed constraints results in implicit neural representations that provide physically consistent predictions at any point in the continuous domain from noisy or missing data, e.g., interpolation[9]. Combining the versatility of neural networks as universal function approximators and smooth, periodic functions as non-linear activations[10], PINNs have the capacity to represent physical models, such as

partial differential equations (PDE).

Recently, implicit neural representations have been proposed for volumetric scattering field reconstruction in forward-looking sonar[11] and SAS imaging[12]. In these studies, a neural network is trained to infer the scatterers' distribution as a continuous function of spatial coordinates. The training involves formulating the backprojection processing step as an optimization problem. Specifically, the optimization objective involves finding the scatterers' distribution that, when forward propagated through a physical model, best fits the recorded data under some constraints such as scatterer sparsity and surface continuity. However, the neural backprojection is practically applicable for small-scale imagery and depends on the compromise between accuracy and efficient computational implementation of the forward propagation model[12]. Another group of studies has used PINNs for sound field reconstruction from a limited number of recordings of the room impulse response collected from linear[13] and planar microphone arrays[14]. In this context, a neural network implicitly represents the sound field as a continuous function of spatiotemporal coordinates and is trained based on limited data and the acoustic wave equation. Herein, we adapt such a neural sound field reconstruction method for interpolating the along-track matched-filtered recordings from a SAS system. Training the neural network with physics-informed constraints, results in SAS imaging free from spatial aliasing artifacts.

2 SYNTHETIC APERTURE SONAR IMAGING

Synthetic aperture sonar (SAS) combines coherently the backscattered echoes recorded with an active sonar, while it moves along a predefined trajectory[3]. Commonly, the synthetic aperture is formed along a linear trajectory and the antenna is focused towards broadside, i.e., in strip-map mode[5].

The active sonar transmits a short broadband pulse $q(t), t \in [0, \tau_q]$ of duration τ_q, referred to as a ping, and records the backscattered echoes repeatedly as the platform moves along the track. Monostatic transmission and reception is assumed with the phase center approximation (PCA)[15], which replaces each transmitter-receiver pair with a virtual transceiver located at the middle of the distance between them. The backscattered wave from a scatterer located at $\mathbf{r}_s$ is a replica of the transmitted pulse delayed by the travel time for the two-way distance between the virtual transceiver and the scatterer, multiplied by its scattering strength $s(\mathbf{r}_s)$. Hence, the sound pressure $p_{\text{rec}}(x, t)$ recorded at a receiver located at the along-track coordinate x is the superposition of the backscattered echoes from all the scatterers within the insonified volume $\mathcal{V}$,

$$p_{\text{rec}}(x,t) = \int_{\mathcal{V}} q\left(t - \frac{2\|x - \mathbf{r}_s\|_2}{c}\right) s(\mathbf{r}_s)\mathrm{d}\mathbf{r}_s. \tag{1}$$

For simplicity, amplitude scaling factors due to the shading function of the transceiver and the spherical spreading are incorporated to the scattering strength. To improve the range resolution, the recorded signal (1) is pulse-compressed with matched filtering by convolving the recording with a time-reversed replica of the transmitted pulse,

$$p(x,t) = q(-t) * \int_{\mathcal{V}} q\left(t - \frac{2\|x - \mathbf{r}_s\|_2}{c}\right) s(\mathbf{r}_s)\mathrm{d}\mathbf{r}_s = \int_{\mathcal{V}} \overline{q}\left(t - \frac{2\|x - \mathbf{r}_s\|_2}{c}\right) s(\mathbf{r}_s)\mathrm{d}\mathbf{r}_s, \tag{2}$$

where $\overline{q}(t) = q(-t) * q(t)$ denotes the cross-correlated pulse.

The sound pressure (2) in water satisfies the acoustic wave equation[16] which is described by the inhomogeneous PDE,

$$\nabla^2 p - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = -g(\mathbf{r}, t), \tag{3}$$

where $g(\mathbf{r}, t)$ describes an arbitrary spatiotemporal distribution of scattering sources.

3 PHYSICS-INFORMED NEURAL NETWORK MODEL

Consider the problem of determining a functional Φ on the continuous spatiotemporal domain, $x \in \mathbb{R}^D$, $t \in \mathbb{R}^+$, which satisfies a set of constraints that involve the evaluation of the function and its derivatives on k discrete collocation points,

$$\Phi : (x,t) \to \Phi(x,t) \;\; \text{subject to} \;\; \mathcal{C}_k\left(\Phi, \nabla_x\Phi, \nabla_t\Phi, \nabla_x^2\Phi, \nabla_t^2\Phi, \dots\right) = 0. \tag{4}$$

To solve this problem in the machine learning framework, the functional Φ is implicitly represented by a neural network $\mathcal{F}_\phi(x,t)$ with trainable parameters ϕ, which maps the input spatiotemporal coordinates (x,t) to the corresponding function value. After optimizing the network parameters ϕ on a training set defined by the constraints $\mathcal{C}_k$ at a discrete number of collocation points, the network can be used to predict (interpolate) the value of the implicit function at any point within the continuous spatiotemporal domain.

Such an implicit neural representation has the multi-layer perceptron (MLP) architecture shown in Fig. 1, i.e., is a deep neural network with L fully connected layers. Each layer, $\ell \in \{1, \cdots, L\}$, is described by the d_ℓ-dimensional vector $\mathbf{u}_\ell \in \mathbb{R}^{d_\ell}$ such that,

$$\mathbf{u}_\ell = \sigma_\ell\left(\omega_\ell\left(\mathbf{W}_\ell \cdot \mathbf{u}_{\ell-1} + \mathbf{b}_\ell\right)\right). \tag{5}$$

Specifically, the vector $\mathbf{u}_\ell$ results from an affine transformation of the input vector from the previous layer $\mathbf{u}_{\ell-1}$ defined by the weight matrix $\mathbf{W}_\ell \in \mathbb{R}^{d_\ell \times d_{\ell-1}}$ and the bias vector $\mathbf{b}_\ell \in \mathbb{R}^{d_\ell}$, followed by the (non-linear) activation function σ_ℓ. The weight matrices and bias vectors for all layers constitute the trainable parameters of the network, $\phi = \{\mathbf{W}_\ell, \mathbf{b}_\ell\}_{\ell \in \{1, \cdots, L\}}$, whereas ω_ℓ is a predefined parameter of the activation function. The trainable parameters are initialized as random samples from a uniform distribution and optimized progressively such that the network output best fits the set of constraints described by a loss function $\mathcal{L}$.
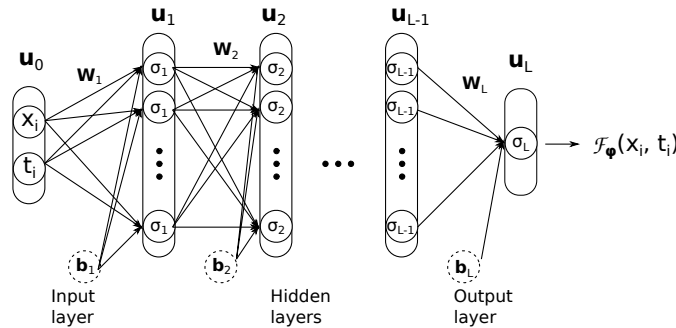
Figure 1: Schematic of a multi-layer perceptron architecture.

In particular, the parameters are updated with stochastic gradient descent over a number of iterations until convergence, such that at the ith iteration $\phi_i = \phi_{i-1} - \alpha\nabla_\phi\mathcal{L}(\phi)$, where α is the learning rate parameter. Typically, the loss function that drives the optimization of the network parameters and, consequently, the representation capacity of the network depends heavily on data. The distinctive characteristic of a PINN formulation is that the loss function comprises additional constraints imposed by a physical model, such that the space of possible solutions is restricted to fit both the available data and the underlying physics.

Herein, we propose a PINN model for aperture synthesis that aims to represent the matched-filtered sound pressure at any point along the SAS trajectory by ensuring that it fits the data recorded from a few

pings as expressed by the loss term,

$$\mathcal{L}_{\text{data}} = \frac{1}{N} \sum_{n=1}^{N} \left\| p(x_n, t_n) - \mathcal{F}_\phi(x_n, t_n) \right\|_2^2, \tag{6}$$

and that the representation complies with the partial differential equation (PDE), which models the wave propagation, as expressed by the loss term,

$$\mathcal{L}_{\text{PDE}} = \frac{1}{M} \sum_{m=1}^{M} \left\| \left(\frac{\partial^2}{\partial x^2} - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) \mathcal{F}_\phi(x_m, t_m) \right\|. \tag{7}$$

The optimal network parameters ϕ^* minimize the total loss on N collocation points for the data-fitting term and on M collocation points for the PDE-fitting term,

$$\phi^* = \arg\min_{\phi} \ \lambda \mathcal{L}_{\text{data}} + \mathcal{L}_{\text{PDE}}, \tag{8}$$

where λ is a regularization parameter, which controls the relative importance between the corresponding terms. Note that the gradients for the calculation of the PDE-loss term (7) are evaluated with the automatic differentiation method[17], which is based on the chain rule to backpropagate gradients through the network. We employ the sinusoidal representation network (SIREN) architecture[10], which uses a sinusoidal activation function. Such smooth and periodic non-linearities that are infinitely differentiable result in smooth representations of complex signals and their spatial and temporal derivatives. The associated parameter ω of the sinusoidal activation function controls the frequency mapping of the network parameters[18] and facilitates training[10].

4 RESULTS

For this study, we have used a 5-layer perceptron architecture for the neural network with parameters as listed in Table 1 implemented in Python with PyTorch[19]. The Adam optimizer[20] with a learning rate of 10^{-4} is used for the network training. Backscattering from an air-filled spherical shell of 1 m diameter and 0.025 m thickness in free field, insonified with a linear frequency modulated pulse with a central frequency of 20 kHz and 30 kHz bandwidth, is computed through the analytical solution[21]. The backscattered signals are recorded with a sampling frequency of 120 kHz and spatially sampled every 1 cm along a 6 m linear trajectory. The resulting matched-filtered signals constitute the baseline dataset.

To demonstrate the effect of spatial aliasing, an undersampled dataset is obtained from the full baseline dataset by retaining only the matched-filtered response at every 10 pings. Using the undersampled dataset to evaluate the data-loss term in Eq. (6) and evaluating the PDE-loss term in Eq. (7) on the same number of collocation points sampled uniformly at random within the spatiotemporal domain, we train the PINN described in Table 1 over 10000 iterations. We have used a regularization parameter $\lambda = 1e3$, to prioritize the fidelity on the available data. Figure 2 shows the evolution of the data-loss (6), the PDE-loss (7), as well as the total loss over the training iterations. The optimization converges, i.e., the rate of change for the training loss becomes infinitesimal, roughly after 5000 iterations.

After training, the PINN can be used for predicting the matched-filtered backscattered soundfield at any point within the continuous spatiotemporal domain. Hence, we use the trained PINN model to interpolate the soundfield over the missing recordings. The reconstructed backscattered soundfield is, then, compared with the baseline dataset as well as a soundfield reconstruction obtained with cubic spline interpolation. The results are shown in Fig. 3, where the along-track axis represents the spatial coordinate

Table 1: Network parameters

| Layer	$\dim(\mathbf{u}_{\ell-1})$	Parameters ϕ			σ_ℓ	ω_ℓ
		$\dim(\mathbf{u}_\ell)$	$\dim(\mathbf{W}_\ell)$	$\dim(\mathbf{b}_\ell)$		
Input	2	512	2×512	512	$\sin$	3
Hidden	512	512	512×512	512	$\sin$	30
Hidden	512	512	512×512	512	$\sin$	30
Hidden	512	512	512×512	512	$\sin$	30
Output	512	1	512×1	1	$\mathbb{I}$	1
		Total: 790017				
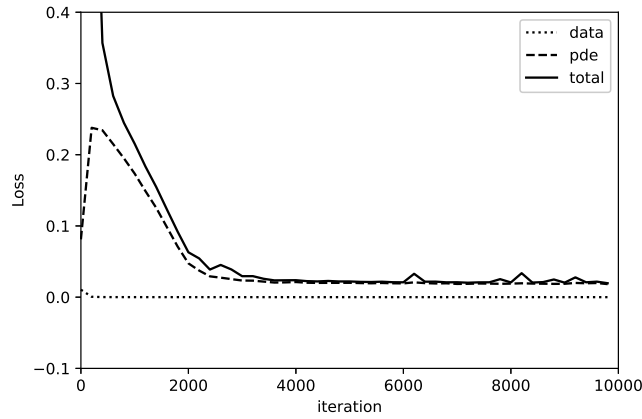
Figure 2: Loss terms towards convergence.

x and the range axis represents the temporal coordinate t transformed into slant range y as $y = ct/2$ for two-way propagation and sound speed c. A detailed view of the initial part of the trajectory indicates that polynomial interpolation fails to capture accurately the wavefront curvature at larger incident angles. Note that equidistant along-track spatial sampling of a linear trajectory results in a nonlinear incident angle spacing, $\theta = \sin(\arctan(x/y_0))$ where y_0 is the range at the closest point of approach. However, the PINN prediction interpolates the wavefronts smoothly; see Fig. 3(f). Additionally, the accuracy of the reconstruction is evaluated through the normalized mean square error (MSE)[13,14] between the actual $p(x,t)$ and the reconstructed soundfield $\hat{p}(x,t)$ in each case defined as,

$$\text{MSE} = 10 \log_{10} \frac{\mathbb{E}\left[(p(x,t) - \hat{p}(x,t))^2\right]}{\mathbb{E}\left[p(x,t)^2\right]}. \tag{9}$$

The MSE for the reconstructed soundfield with spline interpolation is -10.23 dB, whereas with PINN interpolation is -14.65 dB.

The effect of spatial undersampling manifests as spatial aliasing in the reconstructed SAS images. Figure 4 shows the SAS images obtained by backprojecting the corresponding matched-filtered soundfields in Fig. 3. Specifically, spatial aliasing artifacts are present in Fig. 4(b), which correspond to the SAS
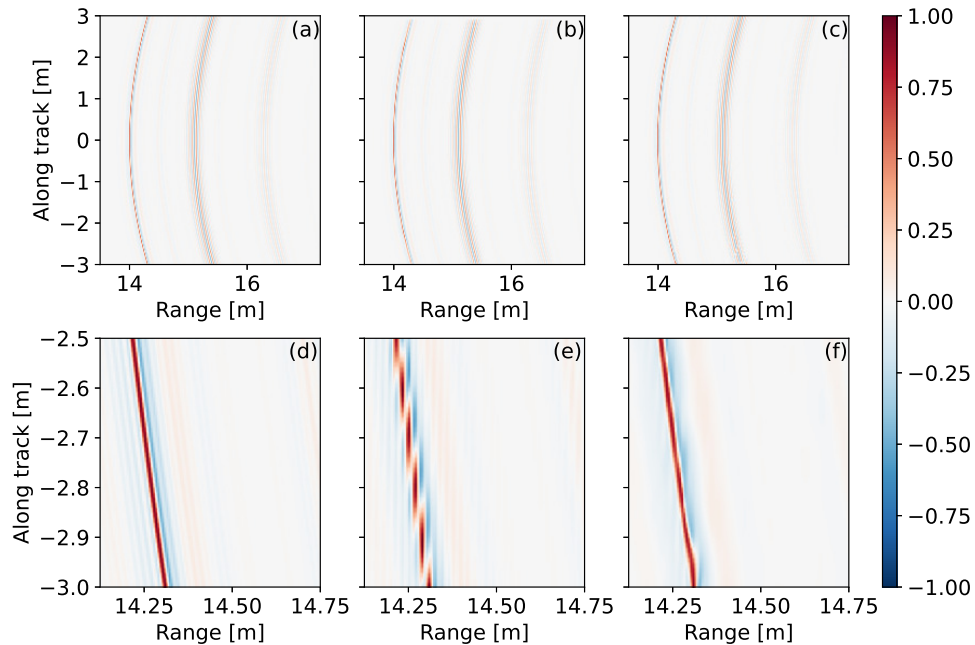
Figure 3: (a) Simulated matched-filtered backscattered signals from an air-filled spherical shell in free field, spatially sampled every 1 cm along a 6 m linear trajectory. Reconstructed field with (b) spline interpolation and (c) PINN. (d)-(f) Detailed view over the initial 0.5 m trajectory of the corresponding reconstructed fields in (a)-(c).
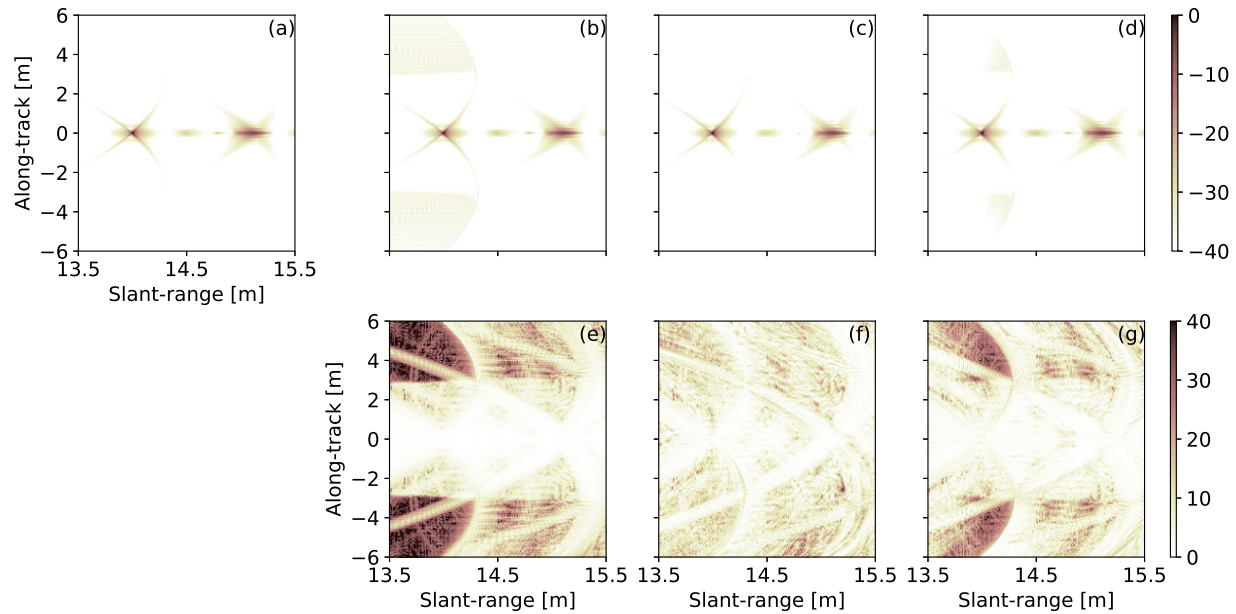
image from the undersampled dataset. Interpolating the matched-filtered soundfield using the trained PINN before backprojection eliminates the spatial aliasing as shown in Fig. 4(c), which is not the case with spline interpolation, Fig. 4(d). Comparing Figs. 4(e)–(g) indicates that PINN interpolation reduces significantly the absolute pixel-wise reconstruction error.

5   CONCLUSION

Inadequate spatial sampling in aperture synthesis induces aliasing artifacts in SAS imaging. This study proposes a PINN as a spatiotemporal function approximator to reconstruct the backscattered soundfield in the continuous spatiotemporal domain. Constraints based on wave propagation models supplement the information content from limited data during training resulting in physically viable network generalization performance. Simulation results show that PINNs can interpolate accurately the soundfield from sub-Nyquist spatial samples of matched-filtered recordings, resulting in SAS images free of aliasing artifacts.

ACKNOWLEDGMENTS

Figure 4: SAS reconstruction with backprojection of the (a) baseline dataset, (b) undersampled dataset, (c) interpolated dataset with PINN, (d) interpolated dataset with splines. Absolute error between the baseline dataset and (e) the undersampled dataset, (f) the interpolated dataset with PINN, (g) the interpolated dataset with splines

REFERENCES

1.　P. Vouras, K. V. Mishra, A. Artusio-Glimpse, S. Pinilla, A. Xenaki, D. Griffith, and K. Egiazarian, "An overview of advances in signal processing techniques for classical and quantum wideband synthetic apertures," *IEEE J. Sel. Top. Sig. Proces.*, vol. 17, no. 2, pp. 317–369, 2023.

2.　L. J. Cutrona, "Comparison of sonar system performance achievable using synthetic aperture techniques with the performance achievable by more conventional means," *J. Acoust. Soc. Am.*, vol. 58, no. 2, pp. 336–348, 1975.

3.　P. T. Gough and D. W. Hawkins, "Unified framework for modern synthetic aperture imaging algorithms," *Int. J. Imag. Syst. Tech.*, vol. 8, no. 4, pp. 343–358, 1997.

4.　K. D. Rolt and H. Schmidt, "Azimuthal ambiguities in synthetic aperture sonar and synthetic aperture radar imagery," *IEEE J. Oceanic Eng.*, vol. 17, no. 1, pp. 73–79, 1992.

5.　R. E. Hansen, *Sonar Systems*. Croatia: InTechOpen, 2011, ch. 1, pp. 3–28.

6.　P. Gough and D. Hawkins, "Imaging algorithms for a strip-map synthetic aperture sonar: Minimizing the effects of aperture errors and aperture undersampling," *IEEE J. Oceanic Eng.*, vol. 22, no. 1, pp. 27–39, 1997.

7.　S. Synnes, A. Hunter, R. Hansen, T. Sæbø, H. Callow, R. Van Vossen, and A. Austeng, "Wideband synthetic aperture sonar backprojection with maximization of wave number domain support," *IEEE J. Oceanic Eng.*, vol. 42, no. 4, pp. 880–891, 2016.

8.　M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *J. Comput. Phys.*, vol. 378, pp. 686–707, 2019.

9.　G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nature Reviews Physics*, vol. 3, no. 6, pp. 422–440, 2021.

10. V. Sitzmann, A. Martel, J. Bergman, D. Lindell, and G. Wetzstein, "Implicit neural representations with periodic activation functions," *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 33, pp. 7462–7473, 2020.

11. M. Qadri, M. Kaess, and I. Gkioulekas, "Neural implicit surface reconstruction using imaging sonar," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 1040–1047.

12. A. Reed, J. Kim, T. Blanford, A. Pediredla, D. Brown, and S. Jayasuriya, "Neural volumetric reconstruction for coherent synthetic aperture sonar," *ACM Transactions on Graphics (TOG)*, vol. 42, no. 4, pp. 1–20, 2023.

13. M. Pezzoli, F. Antonacci, and A. Sarti, "Implicit neural representation with physics-informed neural networks for the reconstruction of the early part of room impulse responses," *arXiv preprint*, no. arXiv:2306.11509, 2023.

14. X. Karakonstantis, D. Caviedes-Nozal, and E. Richard, A.and Fernandez-Grande, "Room impulse response reconstruction with physics-informed deep learning," *J. Acoust. Soc. Am.*, vol. 155, no. 2, pp. 1048–1059, 2024.

15. A. Bellettini and M. A. Pinto, "Theoretical accuracy of synthetic aperture sonar micronavigation using a displaced phase-center antenna," *IEEE J. Oceanic Eng.*, vol. 27, no. 4, pp. 780–789, 2002.

16. P. M. Morse and K. U. Ingaard, *Theoretical Acoustics*. Princeton, New Jersey: Princeton University Press, 1986, ch. 6-7.

17. B. A. Baydin, A. G.and Pearlmutter, A. Radul, and J. M. Siskind, "Automatic differentiation in machine learning: a survey," *J. Mach. Learn. Res.*, vol. 18, pp. 1–43, 2018.

18. N. Benbarka, T. Höfer, and A. Zell, "Seeing implicit neural representations as Fourier series," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 2041–2050.

19. A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, and e. a. Antiga, L., "Pytorch: An imperative style, high-performance deep learning library," *Adv. Neural Inf. Process. Syst.*, 2019.

20. D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint*, no. arXiv:1412.6980, 2014.

21. A. B. Baynes, "Scattering of low-frequency sound by compact objects in underwater waveguides," tech. rep., Naval Postgraduate School, 2018.