Beyond Ultra-NeRF: Explainable Neural Fields for Ultrasound

Magdalena Wysocki[1,2] Mohammad Farid Azampour[1,2,3] Felix Tristram[1,2]
Benjamin Busam[1,2,4] Nassir Navab[1,2]
[1] Technical University Munich [2] MCML [3] Sharif University of Technology [4] 3Dwe.ai

Abstract

Current ultrasound image synthesis techniques often fall short in semantic accuracy and physical realism or produce images with a significant domain gap. Ultra-NeRF addresses these issues by creating a Neural Field from reconstructed acoustic properties via pose-annotated B-mode images and shows that it can be used for novel view synthesis of B-mode images. While Ultra-NeRF generates plausible results, it lacks explainability in the acoustic parameter space. In this paper, we revisit neural fields for ultrasound and introduce the Sonographic Neural Reflection Field (SuRF), which adheres to the physical properties of acoustic ultrasound. By redesigning Ultra-NeRF's differentiable forward synthesis model and incorporating physics-inspired regularizations, we ensure the interpretability of learned acoustic parameters. Tested on the Ultra-NeRF in-silico dataset and a new multi-view ex-vivo 3D ultrasound dataset, our method demonstrates enhanced reconstruction and interpretation across various tissue types, including fat, muscle, and bone.

1. Introduction

Inspired by Neural Radiance Fields (NeRF)[15], initial advancements in Neural Fields [26] for image-based novel view synthesis[1] concentrated on generating photorealistic RGB images[2, 14, 17, 19, 27] through a differentiable forward model grounded in classical ray-based volume rendering principles [12]. Over time, Neural Fields for novel view synthesis have evolved to include more intricate interactions between electromagnetic waves and the scene being imaged. These developments have introduced elements such as reflections [10], polarization [7], and scattering in complex environments, including underwater imaging [4, 13, 22], underscoring the significance of detailed physics-based differentiable forward models for capturing a wider range of visual features in RGB images. Furthermore, recent research has shown that sensor-specific forward models are essential for accurately depicting the image formation process and defining Neural Fields for data obtained from various sensors across the electromagnetic spectrum [6, 8, 28]. Ultra-NeRF (Neural Reflection Field for Ultrasound) [25] has proven that the NeRF framework can be successfully adapted to acoustic waves physics in medical ultrasound imaging. It employs a specialized, differentiable, ray-based forward synthesis model, common in ultrasound synthesis [3, 18], to accurately capture tissue acoustic properties from pose-annotated B-mode (brightness mode) images. However, further analysis revealed that using Ultra-NeRF's forward model to regress acoustic parameters can lead to degenerate solutions [25], where the model fails to accurately represent the true acoustic characteristics of tissues. This shortcoming affects the interpretability of Ultra-NeRF's volumetric ultrasound representation and limits its application in areas like tissue classification [20].

In this paper, we tackle the issue of acoustic parameter interpretability by revisiting the forward synthesis model of Ultra-NeRF. We propose a new Neural Field for Ultrasound, named Sonographic Neural Reflection Field (SuRF).

To summarize, our contributions are:

- We propose Sonographic Neural Reflection Field, a refined, ultrasound-specific Neural Field that captures a 3D scene based on the fundamental acoustic properties: attenuation, scattering, and reflection;
- We incorporate physics-informed regularization into the optimization process to enhance interpretability, aligning optimized tissue characteristics with their physical counterparts;
- We introduce an ex-vivo spine dataset with muscle and fat tissue and a second phantom dataset with known attenuation properties, enabling the evaluation of the correspondence between learned and actual attenuation characteristics.

2. Related Work

The adoption of Neural Fields is growing in medical imaging, which is evident in recent publications highlighting

Corresponding author: Magdalena Wysocki (*magdalena.wysocki@tum.de*).

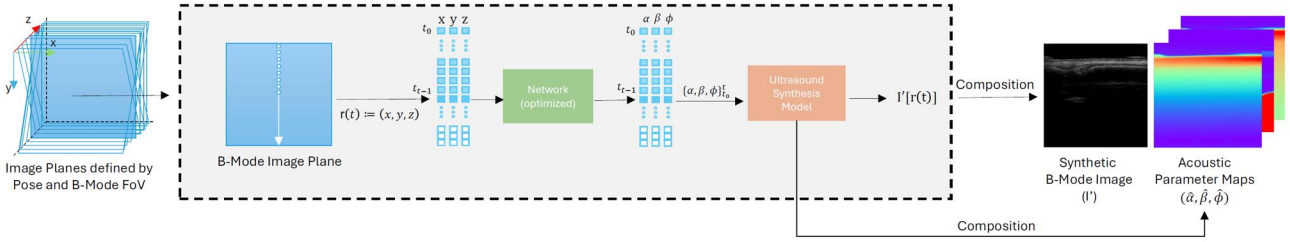

Figure 1. **Overview of SuRF pipeline** The method consists of two modules: a neural network and a ultrasound synthesis model. At each sampled point, the neural network generates a vector of rendering parameters, which are subsequently mapped by the rendering function to pixel intensities. The synthetic B-mode image is created by synthesizing all the pixels for the B-mode image plane.

its relevance when considering the distinctive properties of medical imaging modalities. Notably, Corona-Figueroa et al. [6] propose a method to render CT reconstruction from multi-view X-rays. Iddrisu et al. [11] use NeRF to reconstruct 3D MRI volumes from 2D MRI slices while considering potential motion artifacts. Zha et al. [28] and Fang et al. [9] propose a Neural Attenuation Field based on NeRF for sparse view CBCT (Cone Beam Computed Tomography) reconstruction by learning the attenuation coefficient field. The physics and mechanism behind an image formation using ultrasound differs from the imaging with electromagnetic wave [23]. To model ultrasound propagation in human tissue for for novel B-mode synthesis, Ultra-NeRF introduces introduces a specialized volumetric forward synthesis model for ultrasound imaging, demonstrating NeRF's adaptability to B-mode imaging. However, it has been shown that Ultra-NeRF's forward model can produce degenerate solutions, failing to accurately represent tissue acoustic properties [25]. In this paper, we improve parameter interpretability by revising Ultra-NeRF's forward synthesis model and adding physics-inspired regularization.

3. Method

3.1. Overview

We aim to reconstruct tissue acoustic parameters from B-mode images for a forward synthesis model that generates B-mode images for any image plane within the scanned region. We refine the ultrasound-specific forward synthesis model of Ultra-NeRF by proposing synthesis based on a minimal set of parameters (attenuation, reflection, and scattering) and incorporating log-compression into the model, reducing complexity while enabling plausible B-mode synthesis. As shown in Fig 1, our method uses a neural network that maps coordinates sampled along the ray to acoustic parameter at these coordinates. The network is trained with B-mode observations and similarity losses between synthetic and ground truth images. To avoid degenerate solutions, we apply physics-inspired regularization by integrating regularization terms into the loss function.

3.2. Forward Synthesis Model for Ultrasound

We model the rendering of a B-mode image for Field of View (FOV) of a linear probe. A B-mode image is composed of scanlines, which in our model correspond to ultrasound rays. We refer the reader to the supplementary material for the definition of the ray. We model the recorded echo E along a ray as a function of distance t from the transducer by considering two types of reflections: major reflections (specular reflections) R and backscattering (diffuse reflections) B:

$$E(r,t) = R(r,t) + B(r,t) \tag{1}$$

Having formulated the echo E, we define the design choices for R and B.

In ultrasound systems, received reflections, denoted as R', are enhanced through Dynamic Range adjustment on the ultrasound machine. To consider the amplification of weak reflections, we calculate the amplified reflection $R(r,t)$ using a log-compression function with a hyperparameter γ [24], resulting in the reflection at a point p defined as:

$$R(r,t) = \ln(1+\gamma \cdot R'(r,t)) \cdot \ln(1+\gamma) \tag{2}$$

$$\text{where } R'(r,t) = T(r,t) \cdot \beta(r,t) \tag{3}$$

with $T(r,t)$ as the transmitted ultrasound signal at the distance t and $\beta(r,t)$ the learned reflection coefficient. To compute $T(r,t)$, the propagation of sound is tracked along r and$T(r,t)$ along r is modeled by considering the intensity loss attributed to reflection and attenuation from t_n to t_f as follows:

$$T(r,t) = T(r,0) \cdot \exp^{-\int_{t_n}^{t_f} \beta(r,n)dn} \cdot \exp^{-\int_{t_n}^{t_f} \alpha(r,n)dn} \tag{4}$$

where t_n is the origin of the ray, t_f is a location on the ray infinitesimally close to t. We assume that the transmitted ultrasound signal at the ray origin defined by $T(r,0)$ is 1. In Equation 4, the α represents the learned attenuation coefficient and refers to the gradual weakening or reduction in the intensity of an ultrasound wave. In this context, we do not distinguish between various attenuation sources and do not model the effect of Time Gain Compensation (TGC).

The backscattered energy $B(r,t)$ originating from the scattering medium is influenced by the transmitted sound $T(r,t)$ and a 2D map of scattering S and convolved with the 2D point-spread function (PSF):

$$B(r,t) = T(r,t) \cdot \iint S(u,v) \cdot \text{PSF}(r-u, t-v)\, du\, dv \quad (5)$$

The map S is learned using a generative model inspired by [29]:

$$S(u,v) = 1(u,v) \cdot \phi(u,v) \quad (6)$$

where ϕ represents the scattering amplitude, and $1(u,v)$ is an indicator function that equals 1 for a scattering point and 0 otherwise. The value for a point p is sampled from the Bernoulli distribution parameterized by scattering density ρ_s. As demonstrated in [29], when dealing with a fully developed speckle, it is assumed that the scattering density remains constant. Therefore, only the scattering amplitude is learned.

Evaluating Equations (1) to (6) requires knowledge of the three acoustic properties of tissue at point p, namely attenuation coefficient α, reflection coefficient β, and scattering amplitude ϕ. Together, these properties form the tissue characteristics vector $\theta(p) = [\alpha(p), \beta(p), \phi(p)]$.

3.2.1 B-mode Synthesis and Acoustic Parameter Maps

Equations (1) to (6) allow us to compute echo values for each point $p \in \Psi$. However, in practice, we approximate Ψ by its discrete version, denoted as $\hat{\Psi}$. Since the computation of the echo for each pixel of $\hat{\Psi}$ depends on the tissue characteristics vector θ at that point, we similarly create tissue characteristics maps for the same image plane as $\hat{\Psi}$. We denote these maps as the attenuation map $\hat{\alpha}$, reflection map $\hat{\beta}$, and scattering map $\hat{\phi}$. We refer the reader to the supplementary material for the exact discrete formulation.

3.3. Physics Inspired Regularization

3.3.1 Interrelation of acoustic properties

We exploit the physical relationship between scattering and attenuation in ultrasound imaging, where attenuation is influenced by absorption, reflection, and scattering within the tissue. To leverage this relationship, we propose a regularization term that enforces correlation between the inferred scattering and attenuation maps. This is achieved by using the Local Normalized Cross-Correlation (LNCC) to guide the network in maximizing the similarity between $\hat{\alpha}$ and $\hat{\phi}$.

The local correlation between $\hat{\alpha}$ and $\hat{\phi}$ for each $k = \hat{\alpha}(\hat{p}), l = \hat{\phi}(\hat{p})$ over a window Ω is defined as follows:

$$\text{LNCC}(k,l) = \frac{\sum_{\xi \in \Omega} \frac{1}{w(\hat{p})} \hat{\phi}'(\xi) \hat{\alpha}'(\xi)}{\sqrt{\sum_{\xi \in \Omega} [\hat{\phi}'(\xi)]^2 \sum_{\xi \in \Omega} [\hat{\alpha}'(\xi)]^2}} \quad (7)$$

with the weighted, mean-shifted values $\hat{\phi}'(\xi)$ and $\hat{\alpha}'(\xi)$ defined as:

$$\hat{\phi}'(\xi) = w(\hat{p})(\hat{\phi}(\hat{p}+\xi) - \mu(\hat{p})) \quad (8)$$

$$\hat{\alpha}'(\xi) = w(\hat{p})(\hat{\alpha}(\hat{p}+\xi) - \nu(\hat{p})) \quad (9)$$

Finally, the regularization loss is computed over all points $\hat{p} \in \hat{\Psi}$ to maximize LNCC between scattering and attenuation as follows:

$$\mathcal{L}_{\text{LNCC}}(\hat{\alpha}, \hat{\phi}) = -\sum_{\hat{p} \in \hat{\Psi}} \text{LNCC}(\hat{\alpha}(\hat{p}), \hat{\phi}(\hat{p})) \quad (10)$$

3.3.2 Local homogeneity

In ultrasound imaging, it is common to assume that tissue is locally homogeneous, implying minimal variation in tissue characteristics. To leverage this assumption, we apply total variation (TV) regularization to the scattering map based on the gradient in this map. However, traditional TV regularization can penalize legitimate abrupt changes in tissue properties between different types. To address this, we introduce a reflection-weighted version of TV regularization,

$$\mathcal{L}_{\text{TV}}(\hat{\phi}, \hat{\beta}) = \sum_{\hat{p} \in \hat{\Psi}} \sigma(\hat{p}) \phi'(\hat{p}) \quad (11)$$

$$\text{where } \sigma(\hat{p}) = \beta_{max} - \hat{\beta}(\hat{p})$$

In this equation, the penalty function uses a weight factor, $\sigma(\hat{p})$, to constrain the reflection coefficient based on variability in the scattering attenuation map. The correlation between attenuation and scattering means that penalizing the total variation of the scattering map also affects the attenuation map, reflecting their bidirectional relationship.

3.3.3 Loss Function

The aim is to find an implicit neural representation of a vector field parametrized by a neural network f_Θ. To find the trainable parameters Θ we optimize the network to minimize the following loss function:

$$\begin{aligned} \mathcal{L} = \lambda_1 (1 - \mathcal{L}_{\text{MSSSIM}}(I, I_r)) + \lambda_2 \mathcal{L}_2(I, I_r) \\ + \lambda_3 \mathcal{L}_{\text{TV}}(\hat{\phi}, \hat{\beta}) + \lambda_4 \mathcal{L}_{\text{LNCC}}(\hat{\alpha}, \hat{\phi}), \end{aligned} \quad (12)$$

consisting of primary optimization terms, namely MSSSIM and L_2, and auxiliary regularization, each associated with respective weights λ_i.

4. Experiments

4.1. Data

In addition to datasets presented in [25], we introduce two additional datasets: an ex-vivo phantom dataset to assess

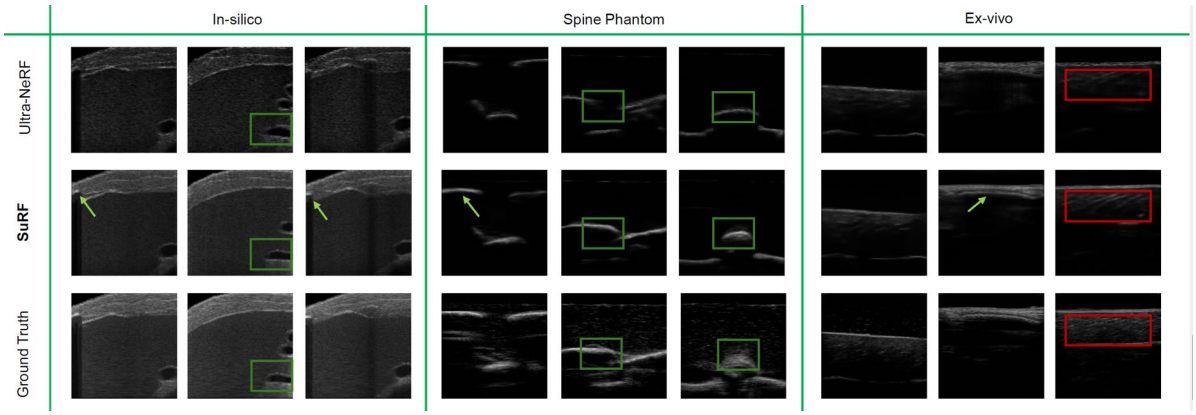

Figure 2. **Qualitative evaluation of rendering results.** Synthetic B-mode images created using poses from test datasets for Ultra-NeRF (upper row) and SuRF (middle row) in comparison to the original B-mode images (lower row) captured at these poses; examples for (from the left) synthetic dataset, phantom dataset, ex-vivo dataset. The examples show that our methods synthesizes plausible B-mode images that preserve view-dependent features like acoustic shadows underneath the bone pointed in the images by the arrows. We observe that the shape of structures like vessels (bounding boxes, in-silico dataset) or bones (bounding boxes, spine phantom dataset) are better preserved in SuRF. Similarly to Ultra-NeRF, SuRF falls short in reproducing complex tisse texture details such as muscle fibres (bounding boxes, ex-vivo dataset).

rendering performance on real tissue, and a dataset collected on a general purpose ultrasound phantom with known attenuation (CIRS Phantom). Each dataset comprises B-mode images with corresponding pose annotations. We refer the reader to the suplementary material for the details regarding the dataset.

4.2. Qualitative Evaluation

4.2.1 Novel View Synthesis

We qualitatively evaluate novel view synthesis by comparing the synthetic B-mode images with corresponding real B-mode images from the test set. As illustrated in Figure 2, the renderings closely match real B-mode images from various datasets, effectively capturing key view-dependent phenomena like acoustic shadowing. For example, when a bone obscures the underlying tissue, the renderings reflect this interaction accurately. SuRF improves Ultra-NeRF by better preserving the shape of the structures such as vessels and bones. However, like its predecessor, SuRF still struggles with accurately reproducing complex tissue textures, such as muscle fibers.

4.2.2 Explainability of The Acoustic Parameter Space

We qualitatively assessed interpretability by comparing the decomposition of synthetic B-mode images into reflection, attenuation, and scattering between Ultra-NeRF and SuRF (Fig 3). Without regularization, the maps show a degenerate solution where scattering encodes most of the information, making it difficult to distinguish tissue types. In these maps, scattering fully defines the B-mode image, with no correlation between scattering and attenuation. However, with regularization, scattering and attenuation are correlated, indicating that stronger scattering is associated with greater attenuation.

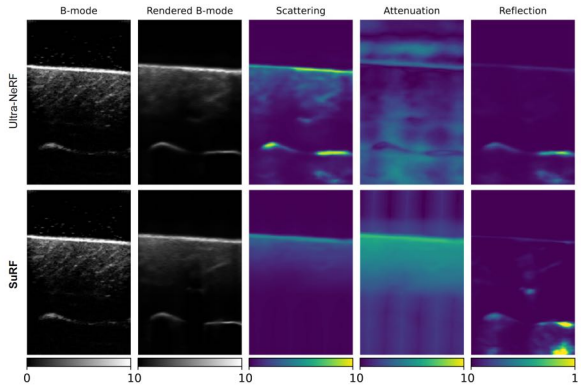

Figure 3. **Decomposition into tissue acoustic characteristics maps** SuRF helps in learning tissue parameters that produce plausible B-mode images and enhance tissue acoustic characteristics. Top row shows the decomposition of a rendered B-mode for Ultra-NeRF [25] and bottom row shows it for SuRF.

5. Conclusion

In this paper, we presented SuRF, a novel, interpretable, ultrasound-tailored neural field designed for physically plausible B-mode synthesis. SuRF highlights the potential of neural fields beyond imaging with electromagnetic waves. Our experiments on various ultrasound datasets, including the newly acquired, multi-view, pose-annotated ex-vivo data, demonstrate the method's utility for novel-view B-mode synthesis, robustness across different tissue types, and its ability to capture tissue geometry. Morover, SuRF significantly improves interpretability within the acoustic parameter space by incorporating physics-based regularizations. This improvement enhances our ability to understand and interpret B-mode data beyond their pixel intensities.

References

[1] S. Avidan and A. Shashua. Novel view synthesis by cascading trilinear tensors. *IEEE Transactions on Visualization and Computer Graphics*, 4(4):293–306, 1998. 1

[2] Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5855–5864, 2021. 1

[3] Benny Burger, Sascha Bettinghausen, Matthias Radle, and Jürgen Hesser. Real-time gpu-based ultrasound simulation using deformable mesh models. *IEEE transactions on medical imaging*, 32(3):609–618, 2012. 1

[4] Lifang Chen, Yuchen Xiong, Yanjie Zhang, Ruiyin Yu, Lian Fang, and Defeng Liu. Sp-seanerf: Underwater neural radiance fields with strong scattering perception. *Computers & Graphics*, page 104025, 2024. 1

[5] CIRS. General Purpose Ultrasound Phantom, 2014. Accessed on 05.03.2024. 1

[6] Abril Corona-Figueroa, Jonathan Frawley, Sam Bond-Taylor, Sarath Bethapudi, Hubert PH Shum, and Chris G Willcocks. Mednerf: Medical neural radiance fields for reconstructing 3d-aware ct-projections from a single x-ray. *arXiv preprint arXiv:2202.01020*, 2022. 1, 2

[7] Akshat Dave, Yongyi Zhao, and Ashok Veeraraghavan. Pandora: Polarization-aided neural decomposition of radiance. In *European Conference on Computer Vision*, pages 538–556. Springer, 2022. 1

[8] Thibaud Ehret, Roger Marí, Dawa Derksen, Nicolas Gasnier, and Gabriele Facciolo. Radar fields: An extension of radiance fields to sar. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 564–574, 2024. 1

[9] Yu Fang, Lanzhuju Mei, Changjian Li, Yuan Liu, Wenping Wang, Zhiming Cui, and Dinggang Shen. Snaf: Sparse-view cbct reconstruction with neural attenuation fields. *arXiv preprint arXiv:2211.17048*, 2022. 2

[10] Yuan-Chen Guo, Di Kang, Linchao Bao, Yu He, and Song-Hai Zhang. Nerfren: Neural radiance fields with reflections. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18409–18418, 2022. 1

[11] Khadija Idrissu, Sylwia Malec, and Alessandro Crimi. 3d reconstructions of brain from mri scans using neural radiance fields. *bioRxiv*, pages 2023–04, 2023. 2

[12] James T Kajiya and Brian P Von Herzen. Ray tracing volume densities. *ACM SIGGRAPH computer graphics*, 18(3):165–174, 1984. 1

[13] Deborah Levy, Amit Peleg, Naama Pearl, Dan Rosenbaum, Derya Akkaynak, Simon Korman, and Tali Treibitz. Seathru-nerf: Neural radiance fields in scattering media. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 56–65, 2023. 1

[14] Ricardo Martin-Brualla, Noha Radwan, Mehdi SM Sajjadi, Jonathan T Barron, Alexey Dosovitskiy, and Daniel Duckworth. Nerf in the wild: Neural radiance fields for unconstrained photo collections. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7210–7219, 2021. 1

[15] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020. 1

[16] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 165–174, 2019. 1

[17] Keunhong Park, Utkarsh Sinha, Jonathan T. Barron, Sofien Bouaziz, Dan B Goldman, Steven M. Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. *ICCV*, 2021. 1

[18] Mehrdad Salehi, Seyed-Ahmad Ahmadi, Raphael Prevost, Nassir Navab, and Wolfgang Wein. Patient-specific 3d ultrasound simulation based on convolutional ray-tracing and appearance optimization. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 510–518. Springer, 2015. 1

[19] Katja Schwarz, Yiyi Liao, Michael Niemeyer, and Andreas Geiger. Graf: Generative radiance fields for 3d-aware image synthesis. *Advances in Neural Information Processing Systems*, 33:20154–20166, 2020. 1

[20] Caifeng Shan, Tao Tan, Jungong Han, and Di Huang. Ultrasound tissue classification: a review. *Artificial Intelligence Review*, 54(4):3055–3088, 2021. 1

[21] Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in neural information processing systems*, 33:7537–7547, 2020. 1

[22] Yunkai Tang, Chengxuan Zhu, Renjie Wan, Chao Xu, and Boxin Shi. Neural underwater scene representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11780–11789, 2024. 1

[23] Maria Tirindelli, Christine Eilers, Walter Simson, Magdalini Paschali, Mohammad Farid Azampour, and Nassir Navab. Rethinking ultrasound augmentation: A physics-inspired approach. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24*, pages 690–700. Springer, 2021. 2

[24] Wolfgang Wein, Ali Khamene, Dirk-André Clevert, Oliver Kutter, and Nassir Navab. Simulation and fully automatic multimodal registration of medical ultrasound. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007: 10th International Conference, Brisbane, Australia, October 29-November 2, 2007, Proceedings, Part I 10*, pages 136–143. Springer, 2007. 2

[25] Magdalena Wysocki, Mohammad Farid Azampour, Christine Eilers, Benjamin Busam, Mehrdad Salehi, and Nassir

Navab. Ultra-nerf: Neural radiance fields for ultrasound imaging. In *Medical Imaging with Deep Learning*, pages 382–401. PMLR, 2024. 1, 2, 3, 4

[26] Yiheng Xie, Towaki Takikawa, Shunsuke Saito, Or Litany, Shiqin Yan, Numair Khan, Federico Tombari, James Tompkin, Vincent Sitzmann, and Srinath Sridhar. Neural fields in visual computing and beyond. In *Computer Graphics Forum*, pages 641–676. Wiley Online Library, 2022. 1

[27] Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa. pixelnerf: Neural radiance fields from one or few images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4578–4587, 2021. 1

[28] Ruyi Zha, Yanhao Zhang, and Hongdong Li. Naf: neural attenuation fields for sparse-view cbct reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 442–452. Springer, 2022. 1, 2

[29] Lin Zhang, Valery Vishnevskiy, and Orcun Goksel. Deep network for scatterer distribution estimation for ultrasound image simulation. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 67(12):2553–2564, 2020. 3