# Full Title of Article

**Author name(s) withheld** EMAIL(S) WITHHELD
*Address withheld*

**Editors:** Under Review for MIDL 2023

## Abstract

We present a physics-inspired NeRF for ultrasound that learns tissue properties from overlapping ultrasound sweeps and leverages a ray-tracing-based neural rendering for ultrasound novel view synthesis. Recent publications demonstrated that NeRF encodes the representation of a three-dimensional scene from a set of two-dimensional ultrasound frames in the weights of a neural network. However, the rendering scheme presented in the existing work does not consider view-dependent changes in appearance and geometry observed in ultrasound imaging. In this paper, we discuss direction-dependent changes in the scene and show that a physics-inspired rendering improves the fidelity of ultrasound image synthesis. In particular, we demonstrate experimentally that the proposed method generates accurate b-mode images for regions ambiguously represented in the dataset owing to view-dependent differences of the ultrasound images. We conduct our experiments using generated b-mode images of simulated liver sweeps and spine phantom sweeps tracked with a robotic arm. The experiments corroborate that our method generates ultrasound frames that enable consistent volume compounding from previously unseen views. To the best of our knowledge, the presented work is the first to address view-dependent ultrasound image synthesis using NeRF.

**Keywords:** ultrasound, neural radiance fields, image reconstruction.

## 1. Introduction

Building a three-dimensional mental model of an organ from a set of two-dimensional ultrasound frames is a demanding task that requires extensive training and clinical expertise. The view-dependent nature of ultrasound imaging contributes to the increased difficulty, since an image of a specific region in the patient's body depends on the probe position. Therefore, the mental model is constantly updated based on images that may carry different information for the same region. Nevertheless, a trained operator approaches this problem effortlessly owing to the consciousness of the underlying anatomical structure and the effect of the probe position on its representation. However, manual visual analysis remains expensive and error-prone. A system that can reconstruct an ultrasound volume could reduce the cost of ultrasound acquisition and error rate. Much research is recently being developed for utilizing 3D ultrasound in diagnostic applications, as well as interventional radiology. 3D Ultrasound volumes are conventionally generated using special wobbler probes, 2D transducers, or tracked probes to compound a 3D volume from 2D slices. In the last decade, new approaches such as computational sonography (Hennersperger et al., 2015), sensorless 3D ultrasound (Gee et al., 2006), and deep learning-based image formation techniques (Prevost et al., 2017) have aimed at improving the 3D compounding quality of this portable and affordable modality. The anisotropic nature of the ultrasound modality

© 2023 Anonymous.

generates various view-dependent artifacts which have negative effects on multi-modal medical imaging applications. Our proposed approach focuses on learning the 3D structure of an organ using volumes scanned from different view directions. This method enables us to generate isotropic 3D ultrasound volumes, which introduces a new intermediate modality for the medical image processing community to explore.

Although viewing-direction dependency is a prominent characteristic of ultrasound imaging, it is not a unique property of ultrasound. To some extent, a similar phenomenon characterizes natural images. For instance, since the non-Lambertian assumption does not hold for most real world objects, appearance due to reflections might be inconsistent between views. Scene reconstruction from a set of two-dimensional view-dependent observations has been hence extensively studied (Seitz and Dyer, 1999) (Niemeyer et al., 2020). Recently, neural radiance fields (NeRF) emerged as a new, potent method for generating photorealistic, view-dependent images of a static 3D scene from a collection of pose-annotated images (Mildenhall et al., 2022).

In computer vision, NeRF became a baseline for a prolific research addressing various aspects such as dynamic scenes (Park et al., 2021), large scale scenes (Rematas et al., 2022), or scene generalization (Yu et al., 2021). Moreover, as presented in iNeRF (Yen-Chen et al., 2021) representing a 3D model as a neural network provides a reference for 6DoF pose estimation, which potentially can find an application in ultrasound tracking. The idea behind NeRF, however, was primarily developed for the purpose of natural image synthesis and takes advantage of established methods from computer graphics. In particular, it builds on a ray-based rendering formulation (Kajiya and Von Herzen, 1984) which underlying physics is not directly applicable to an ultrasound image formation model. In this paper, we propose a physics-inspired NeRF for ultrasound that facilitates the synthesis of b-mode images from novel views. Our contributions are as follows:

- a method that synthesises accurate b-mode images by learning the view-dependent appearance and geometry of a scene from multiple ultrasound sweeps;
- a rendering formulation based on a ray-tracing model, which considers the physical isotropic characteristics of a tissue w.r.t. ultrasound;
- open source datasets[1] comprising multiple pose-annotated 2D ultrasound sweeps acquired with different probe directions.

In our experiments, we use synthetic liver and spine phantom data. We evaluate our method quantitatively and qualitatively. Particularly, we reason about the shortcomings of learning geometry without taking into account the rendering based on the physics behind ultrasound. To the best of our knowledge, this paper presents the first application of NeRF to ultrasound that considers view-dependent characteristics of ultrasound imaging.

## 2. Related Work

**Implicit representation.** Recently, *implicit representation* (Sitzmann et al., 2020), a novel method of representing signals as a neural network proved to be a convenient way for

1. The link to the data and implementation will be provided upon acceptance.

storing information about a 3D scene. Here, the neural network can be seen as a universal function approximator, which represents a scene as a continuous function parameterised by its weights. As a consequence, it allows for a mapping from a 3D continuous coordinate space to intensity. As discussed in ImplicitVol (Yeung et al., 2021), implicit volume representation enables the extraction of any cross-sectional plane; therefore can be used as a training signal for sensorless volume reconstruction.

In IREM, the authors propose an image reconstruction framework based on implicit representation, which reproduces high-resolution magnetic resonance (MR) images using low-resolution training data (Wu et al., 2021). In ultrasound imaging, an implicit representation network is used to store a volume learnt from a set of ultrasound frames (Gu et al., 2022). However, the observed geometry and appearance of a specific location might be inconsistent when the location is present in more than one image, since ultrasound is a view-dependent modality. Therefore, an implicit representation in ultrasound might be inadequate to represent view-dependent changes in the scene's geometry and appearance.

**Neural Radiance Field.** Recent progress in neural continuous shape representation sparked interest in methods reproducing realistic view-dependent appearance. In particular, in a seminal work introducing NeRF (Mildenhall et al., 2022), the authors propose combining representation of a scene, namely position and viewing direction, and fully differentiable volumetric rendering method for a novel view synthesis. In NeRF a fully-connected neural network is mapping between a 5D vector, a 3D location $\mathbf{x} = (x, y, z)$, 2D viewing direction $\mathbf{d} = (\theta, \phi)$, volume density $\sigma$, and radiance $\mathbf{c} = (r, g, b)$. For each pixel in the image, a camera ray is defined as $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ with the camera origin $\mathbf{o}$ in the center of the pixel defining the near plane. The final colour value of each pixel is calculated using a rendering method that considers contribution of each spatial location to a pixel in the generated image. Specifically, it renders a 2D image using the following formula:

$$C(\mathbf{r}) = \int_{t_n}^{t_f} T(t)\sigma(t)c(t)dt \tag{1}$$

$$\text{where } T(t) = \exp - \int_{t_n}^{t_f} \sigma(\mathbf{r}(s))ds \tag{2}$$

The volume rendering integral in Equation (1) accumulates a radiance field along a ray $\mathbf{r}$, therefore each sampled position contributes to the final colour of a pixel. The input of each sample is controlled by the transmittance factor $T(t)$ (Equation (2)). The transmittance can be further interpreted as a probability that the ray does not terminate between $t_n$ and $t_i$. Next, the render pixel value is compared with a value in an image using a photometric total squared error. Due to its flexibility and high fidelity novel view synthesis NeRF has been been recently applied in medical imaging. For instance, MedNeRF propose a NeRF framework to reconstruct CT-projections from X-ray (Corona-Figueroa et al., 2022), and EndoNeRF adopts NeRF for surgical scene 3D reconstruction (Wang et al., 2022).

Yet, researchers have not thoroughly investigated the potential of NeRF in the context of ultrasound imaging. One of a few studies focuses on its application to reconstruct the 3D volume of a spine (Li et al., 2021). In this paper, the authors applied the original formulation of volume rendering to ultrasound image synthesis and achieved promising reconstruction results. However, the accumulation of information along the ray (Equation (1)) is not

justified by the ultrasound image formation. To address this shortcoming, we instead focus on reformulating rendering such that it is based on the physics of ultrasound.

## 3. Method

### 3.1. Background: Ultrasound Physics

Ultrasound images are generated by mapping reflected sounds from the tissue within a thin transversal slice of the body. Intrinsic acoustic parameters such as travelling speed of sound, acoustic impedance, attenuation coefficient, and spatial distribution of sound scattering micro-structures are the main contributing factors affecting the sound reflection within the tissue. By knowing the mapping of these parameters in space one would be able to simulate renderings of 3D ultrasound in arbitrary views (Salehi et al., 2015).

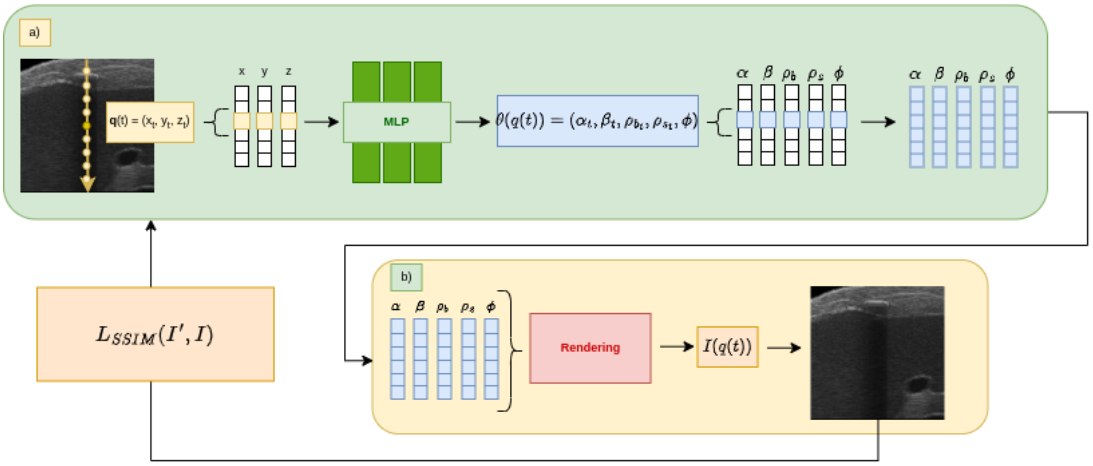


Figure 1: **Proposed framework in the single-frame case.** The method consists of two modules: (a) a multi-layer perceptron (MLP) and (b) volume rendering function. For a query point $\mathbf{q} \in \mathbb{R}^{\nVDash}$ sampled along a probe ray, the network extracts a parameter vector from an internal volume representation. Once all the queries along the ray are computed, the resulting vectors are passed to the rendering function. The rendering computes a per-query intensity that depends on parameters at queried and preceding positions along the ray. To render the complete image, we query rays covering a complete image space. Finally, we composite the pixel intensities into an image. The output frame is compared with the target image using SSIM.

### 3.2. Ultrasound NeRF

As shown in Figure 1, our method follows the original NeRF structure described in Section 2, namely, it consists of two components: a neural network (Figure Figure 1a) and volumetric rendering (Figure 1b). Our network $F_\Theta$ represents a volume as a 3D vector-valued

continuous function that maps a position $\mathbf{q} = (x, y, z)$ in a Cartesian coordinate space into a parameter vector $\theta \in \mathbb{R}^{4}$ which elements correspond to attenuation $\alpha$, reflectance $\beta$, border probability $\rho_b$, scattering density $\rho_s$, and scattering intensity $\phi$. These parameters serve as input to the rendering function. We do not provide explicit viewing directions to the network because the parameter vector consists of isotropic physical tissue properties. We ensure that the regressed physical properties remain consistent between views, whereas the view-dependent changes are encouraged by the rendering which implicitly receives viewing directions from a set of queries along a ray. Figure 2 illustrates definition of a ray and query points.

### 3.3. Ultrasound Volume Rendering

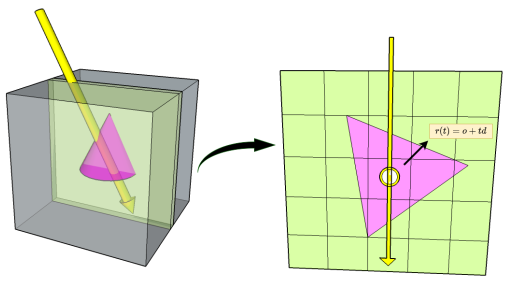


Figure 2: **A ray and query points.** Each ray $r(t)$ corresponds to a single scan-line with origin $o$ at top of the image plane and traverses through the medium perpendicularly to the acquisition direction $d$. Along each ray we sample a set of query points defined by $r(t) = o + td$ that correspond to the spatial location within a 3D volume.

Our ultrasound volume rendering model builds upon a formulation presented by Salehi et al. (Salehi et al., 2015) that proposes a simulation model combining ray-tracing and convolution. The advantage of this model is its flexibility in representing ultrasound artifacts coming from backscattering effects. For each scan-line $r$, we define a function for the recorded ultrasound echo $E(r, t)$, measured at distance $t$ from the transducer, as a sum of reflected $R(r, t)$ and backscattered $B(r, t)$ energy:

$$E(r, t) = R(r, d) + B(r, t) \tag{3}$$

The reflected energy is defined by:

$$R(r, t) = |I(r, t) \cdot \beta(r, t)| \cdot PSF(r) \otimes G(r', t') \tag{4}$$

The term $I(r, t)$ corresponds to the remaining energy at the distance $t$, $\beta(r, t)$ represents the reflection coefficient, and $PSF(r)$ is a predefined 2D point-spread function. G(r, t) is an indicator function that admits 1 for points at the boundary and 0 otherwise. It is sampled from a Bernoulli distribution parameterized by the border probability $\rho_b$. We use

a probabilistic approach because the underlying anatomical structure is unknown therefore we do not know the position of the tissue borders. The energy loss is traced along each scan-line, and the remaining energy $I(r,t)$ is modelled using the loss of the energy due to reflection $\beta$ at the boundaries and attenuation $\alpha$ compensated by applying an unknown time-gain compensation function (TGC). The final formulation for $I(r,t)$ assumes an initial unit intensity $I_0(r,0)$ and loss of energy at each step $\delta t$.

$$I(r,t) = I_0 \cdot \prod_{n=0}^{t-1} [(1-\beta(r,n)) \cdot G(r,n)] \cdot \exp^{(-\int_{n=0}^{t-1} (\alpha - TGC_n) \cdot f \cdot dt)} \tag{5}$$

We can further simplify the above equation by modeling the compensated attenuation by a single parameter since TGC is a scaling factor:

$$I(r,t) = I_0 \cdot \prod_{n=0}^{t-1} [(1-\beta(r,n)) \cdot G(r,n)] \cdot \exp^{(-\int_{n=0}^{t-1} (\alpha \cdot f \cdot dt))} \tag{6}$$

Consequently, $\alpha's$ correspond to the physical attenuation only up to an unknown scaling factor. The backscattered energy $B(r,t)$ from the scattering medium is a function of remaining energy $I(r,t)$ and a 2D map of scattering points $T(r,t)$:

$$B(r,t) = I(r,t) \cdot PSF(r) \otimes T(r',t') \tag{7}$$

The map $T(r,t)$ is learnt using a generative model similar to (Zhang et al., 2020):

$$T(r,t) = H(r,t) \cdot \phi(r,t) \tag{8}$$

In this model, $H(r,t)$ is an indicator function that admits 1 for a query point being a scattering point and 0 otherwise. This function is sampled from the Bernoulli distribution parameterized by scattering density $\rho_s$. The intensity of a scattering point is controlled by its amplitude $\phi$ which models sampling from a normal distribution with the mean $\phi$ and unit variance.

### 3.4. Network Structure

Our network is a fully-connected network, known as a multi-layer perceptron (MLP). It consists of 8 layers with 128 neurons. As discussed in NeRF, we are using positional encoding on an input vector, which maps it to a higher dimensional space to help the network learn high-frequency details. We are optimizing the weights of our network using the Adam optimizer (Kingma and Ba, 2014). Our loss is a Structural Similarity Index Measure (SSIM) (Wang et al., 2004) between the rendered and true b-mode images.

## 4. Experiments & Results

### 4.1. Data

We acquired two types of data: synthetic and phantom b-mode images. For both datasets straight sweeps were recorded with different constant tilt angles (Figure 3). **Synthetic data**

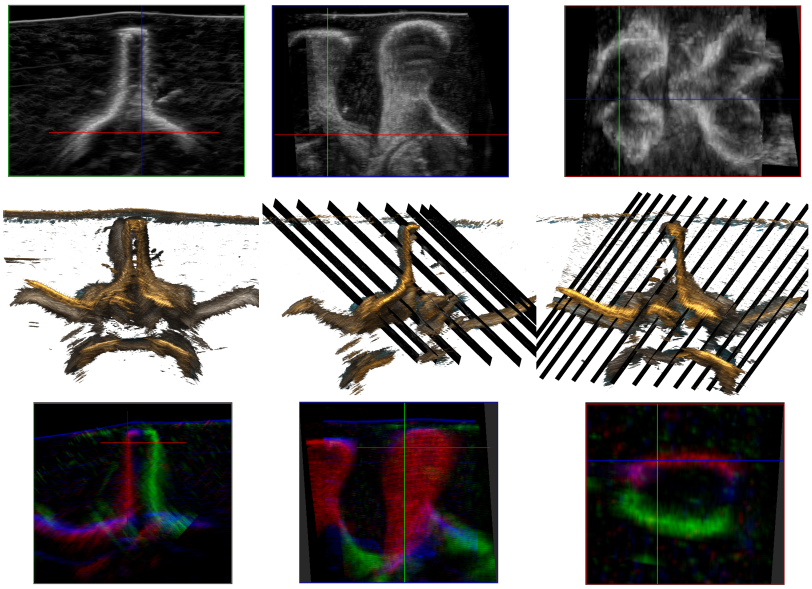

Figure 3: **Spine phantom data.** The first row shows fused sweeps from all views that serve as input to the model. As shown, throughout the training a network receives complete information about the underlying scene. The second row is a compounding from all training data with examples of tilted views illustrated in black in the two last images. The last row illustrates overlaid sweeps and presents how different views contribute to the total volumetric information; red, green, and blue intensities are captured from different view directions.

We simulated b-mode images of a liver from a label map using an ultrasound simulation algorithm of the ImFusion software version 2.42 (ImFusion GmbH, Munich, Germany). Each sweep comprised 2D ultrasound images and respective tracking information. Our synthetic dataset consisted of seven sweeps: six with an acquisition angle tilted and one with an acquisition angle perpendicular w.r.t the acquisition direction (Figure 2). Each sweep consisted of 200 2D US images with respective tracking information. The tilted sweeps differed in the slope's degree and direction. Therefore, an organ was observed from different viewing angles and directions. Our frames contained occlusions caused by scanning between ribs in different directions respective to the probe direction. We used four tilted sweeps for training, totalling 800 frames, and we tested on three sweeps: one perpendicular and two tilted, totalling 600 images. **Phantom data** We acquired phantom data of a lumbar spine, gelatine-based phantom. We used a robotic manipulator (KUKA LBR iiwa 7 R800) and linear probe to obtain ultrasound sweeps. The position of the probe was tracked using robotic tracking. We accessed real-time images and tracking information using ImFusion software version 2.42 (ImFusion GmbH, Munich, Germany). We scanned our phantom with a probe in paramedian sagittal orientation. The collected data comprised 13 sweeps with 150 frames each: six pairs of tilted sweeps and one perpendicular sweep. The trajectory of each pair of tilted sweeps was defined such that the data acquired for training completely covered tissue visible in the test data. To keep the constant spacing of images in each sweep and cover the whole testing region through training data, we reduced the number of testing frames per sweep to 100 frames. The scans occupied an area of two lumbar vertebrae. Analogously to synthetic data, the slope and direction of a probe differed between sweep pairs. As a consequence, the spinous process occluded different regions depending on the viewing direction (Figure 3). We used four tilted sweep pairs for training, totaling 1200 frames, and three sweeps for testing, totaling 300 images.

### 4.2. Results

We evaluated our Ultra-NeRF method both quantitatively and qualitatively. **Quantitative Results.** As shown in Table 1., we evaluated the quality of a novel view synthesis by computing SSIM (Wang et al., 2004) between synthetic and reference testing data. To analyze the effect of rendering, we present the metric for both versions, with and without rendering. For the synthetic data with rendering we achieve $SSMI_{median} = 0.43$ for both tilted and perpendicular views, whereas for the spine phantom the $SSIM_{median}$ is higher for perpendicular than tilted views ($SSIM_{median} = 0.54, SSIM_{median} = 0.51$). The method without rendering attains higher $SSIM$ for synthetic dataset ($SSIM_{median} =$ 0.49 - tilted, $SSIM_{median} = 0.5$ - perpendicular), and lower, or similar for spine dataset ($SSIM_{median} = 0.50$ - tilted, $SSIM_{median} = 0.58$ - perpendicular).

**Qualitative Results.** Figure 4 illustrates examples of synthetic b-mode images generated with our method. Further, we compare the same views generated with and without the rendering function. In Figure 6 we display volumes compounded from perpendicular liver ultrasound sweeps. We cut through the sagittal plane to illustrate that the rendering improved accuracy of image synthesis. We show that the resulting compounding (b) features occlusions visible in the reference data (a). For the spine dataset we chose three representative views of a compounded tilted volume (Figure 5). The volumes demonstrate

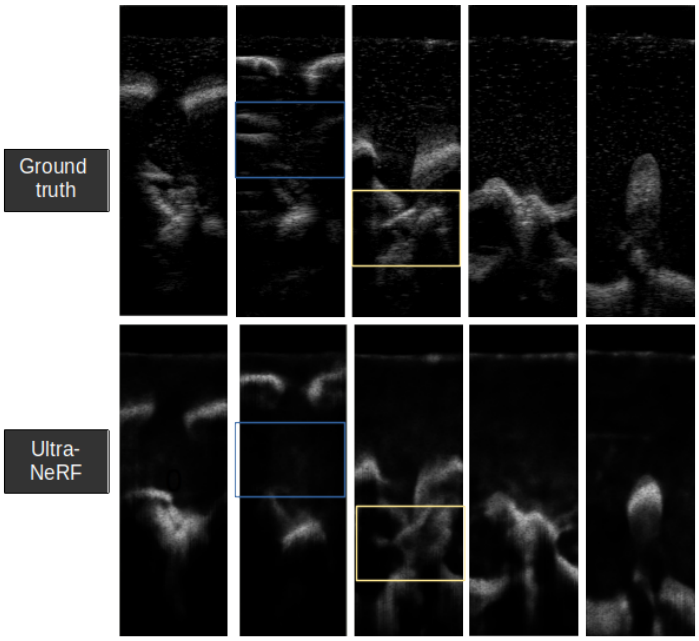


Figure 4: **B-mode image synthesis.** Our method can infer novel views, as shown in the examples. However, it does not produce artifacts inconsistent with a ray-based model, such as reverberations (blue rectangle). The method suffers from low rendering quality w.r.t appearance. In particular, it does not reproduce scattering patterns and over-smooths renderings. As marked by the yellow rectangle, it has failure modes when representing complex structures.

Table 1: Structural Similarity Index Measure (SSIM) measured between synthetic and reference b-mode images.

| dataset type | | with rendering | | | | w/o rendering | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | median | mean | min | max | median | mean | min | max |
| liver synthetic | tilted | 0.41 | 0.43 | 0.37 | 0.58 | **0.49** | **0.50** | 0.45 | 0.64 |
| | perpendicular | 0.41 | 0.43 | 0.38 | 0.56 | **0.50** | **0.48** | 0.46 | 0.61 |
| spine phantom | tilted | **0.54** | **0.51** | 0.36 | 0.60 | 0.50 | 0.48 | 0.36 | 0.59 |
| | perpendicular | 0.57 | **0.54** | 0.42 | 0.65 | **0.58** | **0.54** | 0.41 | 0.64 |

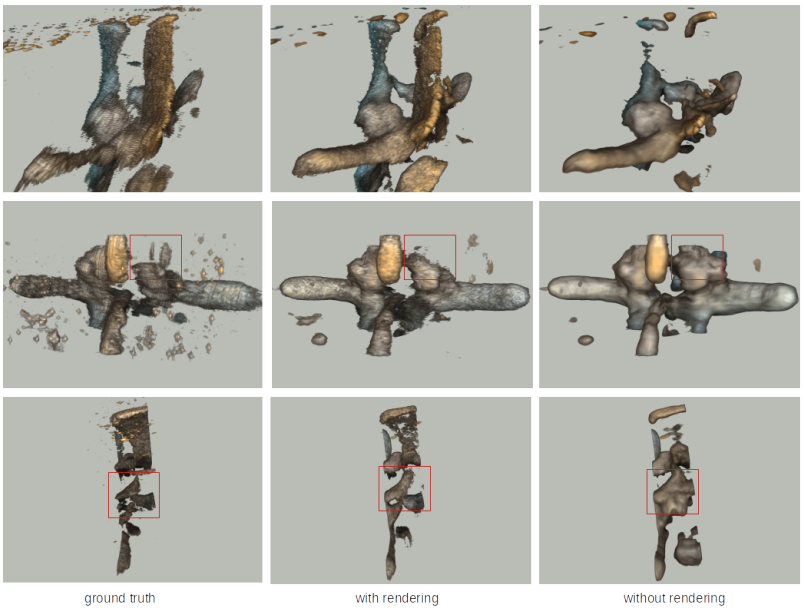


Figure 5: **Compounded volumes of a spine.** The first row shows that without rendering the model is not aware of the viewing direction. In the second and last rows, we observe that occluded parts of the lamina are visible (red rectangle) when rendering is not used.

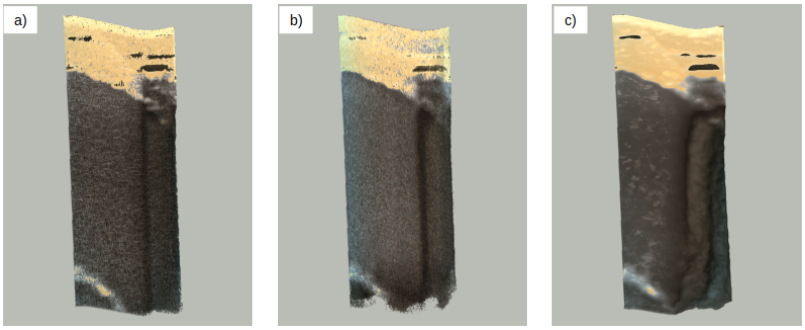


Figure 6: **Sagittal plane of compounded volumes.** Volume (a) is the reference compounded from a perpendicular test view of a liver. Volume (b) is compounded from data generated with Ultra-NeRF and volume (c) with the network without rendering. Our method is able to infer view-dependent changes in appearance resulting in the shadow adjusted to the viewing-direction.

the significance of rendering when a vertebra is visible from a tilted view. We compounded volumes using compounding algorithm of the ImFusion software.

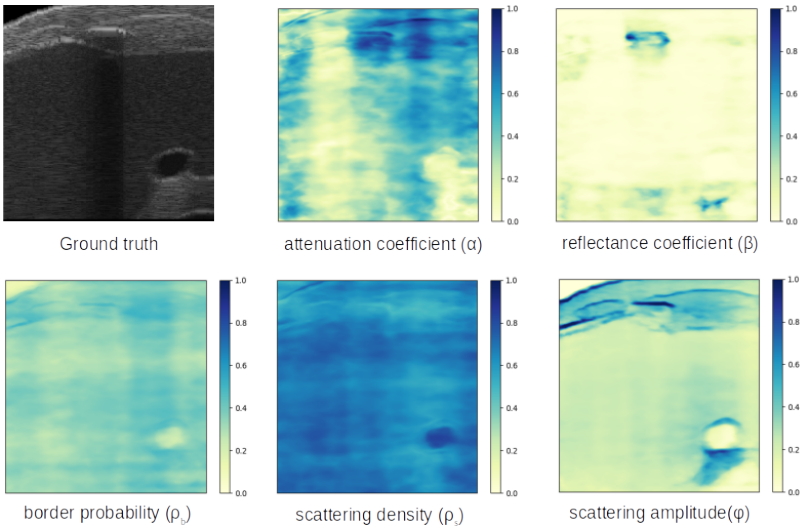


Figure 7: Intermediate maps illustrating each element of rendering parameter vector $\theta$. The rendering component receives these maps from the MLP and computes final intensity values. Each element corresponds to a tissue's physical property. The attenuation map presents weaker attenuation for brighter regions in the reference image. On the reflectance coefficient map, we observe that the method is aware of a highly reflective structure (a rib). The border probability admits higher values for larger intensities suggesting stronger reflections. Scattering density and scattering amplitude define scattering distribution in the image and are the maps.

## 5. Discussion

In this paper, we presented Ultra-NeRF, a framework to learn the appearance and geometry of a scene from ultrasound images. The experiments corroborate that our method integrates knowledge about viewing direction into integral scene representation, which enables the view-dependent synthesis of ultrasound frames and produces plausible b-mode images.

We achieve this task by learning view-invariant physical properties of the underlying tissue. As presented in Figure 7, the network learns to differentiate between a highly reflective structure, such as a rib, and soft tissue with different levels of attenuation. Our rendering method receives isotropic physical properties and incorporates the rendering direction implicitly which enables the synthesis of view-dependent appearance and geometry. As illustrated in Figure 5 and Figure 6, we showed that Ultra-NeRF can generate acceptable results that allow us to create consistent compounded volumes from unseen views. Despite grounding our rendering function in the physics of ultrasound, the underlying model approx-

imates physical phenomena related to ultrasound wave propagation in tissue. Therefore, the network does not reproduce ultrasound artifacts that are not explained by the model. For instance, reverberations caused by bone tissue produce numerous reflections that appear in a b-mode image (Figure 4). However, our model assumes that if a high-intensity signal is produced, the remaining energy is lower than the total intensity observed in a b-mode image along one scan-line. The network, therefore, has to decide whether to generate strong reflections or artifacts caused by an echoing sound. In the current network configuration, the network prioritizes creating strong reflections over artifacts.

Moreover, the rendering procedure samples a scene with a single ray per scan-line, which does not express more complex ray interactions, reflections, or refractions. This is a potential bottleneck for real-world applications of high-quality ultrasound synthesis and volume representation. An additional shortcoming of the model is an engineered PSF, which assumes two axes in the image plane and a discrete grid approximation, and does not produce scattering patterns visible in the b-mode images. As shown in Figure 4, although the method learns accurate geometry, appearance diverges considerably, which might explain modest quantitative results (maximum $SSIM_{median} = 0.57$) (Table 1).

Intermediate maps showing input vectors of the rendering function (Figure 7) illustrate another shortcoming of Ultra-NeRF, namely, large number of independent parameters representing degrees of freedom in output intensity. Large variation in possible outputs leading to the same intensity indicates that the unconstrained network might not provide a unique solution to the same queries between multiple training instances. For example, as shown in Figure 7, if the model's task is to reproduce a low scattering region, the network can compensate a high scattering density by a small amplitude. However, the same result would be produced by a high amplitude and a low scattering density.

The learnt parameters correlate with physical quantities, however, the values do not reflect actual characteristics of underlying tissue, therefore cannot be interpreted as such. For instance, the unknown TCG factor does not allow us to learn a true distribution of attenuation values, which are compensated and subject to device-specific settings. This could be one of the reasons for the observed discrepancy in attenuation and scattering density (Figure 7) within a tissue that cannot be explained by its heterogeneity. These facts, along with the high degree of freedom and settings-dependence, might result in the same tissue being represented by different intermediate maps. In future works, these shortcomings could be addressed by providing an anatomical prior information, uncertainty measures, and regularization. Like the original NeRF, our model:

- is scene-specific, meaning that it cannot generalize beyond one patient and scanning setup;
- is randomly initialized and we expect the model to train only from available images, which results in long training time and unstable training.

Anatomical priors, and more complex network architectures might help to address these challenges.

## 6. Conclusion

We presented a view-dependent ultrasound synthesis method based on NeRF, which learns patient-specific geometry and the appearance of anatomical structures. We demonstrated that viewing-direction awareness and 3D representation of a scene enable the generation of b-mode images from previously unseen views. Results show the feasibility of this method. Since, to our knowledge, this is the first work directly addressing volumetric rendering for ultrasound in NeRF, it facilitates further extensions and application directions. In future works, the method can be adapted even more to fit ultrasound imaging characteristics. First, an improved method could incorporate prior information about an anatomical structure and anatomy-based regularization. Moreover, the assumption about a static scene does not support many real-world medical scenarios in which anatomical structures are subject to deformations due to external and internal forces. Although we achieved good results on synthetic and phantom data, we have not evaluated our method on patient data due to the lack of adequate datasets, which should be addressed in subsequent works. We believe, that solving these challenges will enable many novel applications in medical imaging beyond b-mode image synthesis.

## References

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.

Andrew H Gee, R James Housden, Peter Hassenpflug, Graham M Treece, and Richard W Prager. Sensorless freehand 3d ultrasound in real tissue: speckle decorrelation without fully developed speckle. *Medical image analysis*, 10(2):137–149, 2006.

Ang Nan Gu, Purang Abolmaesumi, Christina Luong, and Kwang Moo Yi. Representing 3d ultrasound with neural fields. In *Medical Imaging with Deep Learning*, 2022.

Christoph Hennersperger, Maximilian Baust, Diana Mateus, and Nassir Navab. Computational sonography. In *International conference on medical image computing and computer-assisted intervention*, pages 459–466. Springer, 2015.

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

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Honggen Li, Hongbo Chen, Wenke Jing, Yuwei Li, and Rui Zheng. 3d ultrasound spine imaging with application of neural radiance field method. In *2021 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4. IEEE, 2021.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. *NeRF*, volume 65. Springer International Publishing, 2022. ISBN 9783030584528.

Michael Niemeyer, Lars Mescheder, Michael Oechsle, and Andreas Geiger. Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3504–3515, 2020.

Keunhong Park, Utkarsh Sinha, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Steven M Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5865–5874, 2021.

Raphael Prevost, Mehrdad Salehi, Julian Sprung, Alexander Ladikos, Robert Bauer, and Wolfgang Wein. Deep learning for sensorless 3d freehand ultrasound imaging. In *International conference on medical image computing and computer-assisted intervention*, pages 628–636. Springer, 2017.

Konstantinos Rematas, Andrew Liu, Pratul P Srinivasan, Jonathan T Barron, Andrea Tagliasacchi, Thomas Funkhouser, and Vittorio Ferrari. Urban radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12932–12942, 2022.

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.

Steven M Seitz and Charles R Dyer. Photorealistic scene reconstruction by voxel coloring. *International Journal of Computer Vision*, 35(2):151–173, 1999.

Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. *Advances in Neural Information Processing Systems*, 33:7462–7473, 2020.

Yuehao Wang, Yonghao Long, Siu Hin Fan, and Qi Dou. Neural rendering for stereo 3d reconstruction of deformable tissues in robotic surgery. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 431–441. Springer, 2022.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

Qing Wu, Yuwei Li, Lan Xu, Ruiming Feng, Hongjiang Wei, Qing Yang, Boliang Yu, Xiaozhao Liu, Jingyi Yu, and Yuyao Zhang. Irem: High-resolution magnetic resonance image reconstruction via implicit neural representation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 65–74. Springer, 2021.

Lin Yen-Chen, Pete Florence, Jonathan T Barron, Alberto Rodriguez, Phillip Isola, and Tsung-Yi Lin. inerf: Inverting neural radiance fields for pose estimation. In *2021*

*IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1323–1330. IEEE, 2021.

Pak-Hei Yeung, Linde Hesse, Moska Aliasi, Monique Haak, Weidi Xie, Ana IL Namburete, et al. Implicitvol: Sensorless 3d ultrasound reconstruction with deep implicit representation. *arXiv preprint arXiv:2109.12108*, 2021.

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.

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.