Neural Implicit Representations for Medical Ultrasound Volumes and 3D Anatomy-specific Reconstructions

Ananya Bal

CMU-RI-TR-23-82
Dec 7, 2023

The Robotics Institute
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA

Thesis Committee:
Dr. Howie Choset *(chair)*
Dr. John Galeotti *(co-chair)*
Dr. Shubham Tulsiani
Dr. Yehonathan Litman

Submitted in partial fulfillment of the requirements for the degree of Master of Science in Robotics.

To my family, friends, and mentors.

Abstract

Most Robotic Ultrasound Systems (RUSs) equipped with ultrasound-interpreting algorithms rely on building 3D reconstructions of the entire scanned region or specific anatomies. These 3D reconstructions are typically created via methods that compound or stack 2D tomographic ultrasound images using known poses of the ultrasound transducer with the latter requiring 2D or 3D segmentation. While fast, this class of methods has many shortcomings. It requires interpolation-based gap-filling or extensive compounding and still yields volumes that generate implausible novel views. Additionally, storing these volumes can be memory-intensive.

These challenges can be overcome with neural implicit learning which provides interpolation in unobserved gaps through a smooth learned function as well as a lighter representation for the volume in terms of memory. In this thesis, a neural implicit representation (NIR) based on the physics of ultrasound image formation is presented. With this NIR, a physically-grounded version of tissue reflectivity function (TRF) is learned by regression using observed intensities in ultrasound images. Additionally, this NIR also learns a spatially-varying point spread function (PSF) of the ultrasound imaging system to improve the photorealism of rendered images. The TRF learned through this method can handle contrasting observations from different viewing-directions due to a differentiable rendering function that incorporates the angle of incidence between ultrasound rays and the tissue interfaces in the scanned volume. It is a stable representation of the tissue volume that when combined with the viewing-direction, can produce true-to-orientation ultrasound images.

Given that many diagnostic and surgical applications, robotic or otherwise, require anatomy-specific 3D reconstructions, it is not sufficient to learn entire ultrasound volumes without discerning the required anatomies. To circumvent the use of traditional 3D segmentation methods that are computationally-heavy, I demonstrate that the obtained TRF can be used to learn a neural implicit shape representation for anatomies that are largely homogeneous. This is formulated as a weakly-supervised binary voxel occupancy function that is learned in parallel with the NIR. All these contributions are substantiated on simulated, phantom-acquired and live subject-acquired ultrasound images capturing blood vessels. Finally, an application for the anatomy-specific reconstruction is discussed in the context of physical simulations for deformation modeling of soft tissue.

Acknowledgments

I would like to express my heartfelt gratitude towards my advisors Professors Howie Choset and John Galeotti for their continued support and guidance throughout my masters and during this work. They have provided me with invaluable opportunities and sage advice. I would also like to thank FNU Abhimanyu, Ashutosh Gupta, Andrew Orekhov, Nicolas Zevallos and Yizhu Gu for their constant valuable support, discussion and feedback. They have always been there to help me with experiments and other technical problems in the lab. I would like to thank Magdalena Wysocki and Mohammad Farid Azampour for insightful technical discussions and for being very enthusiastic collaborators. I would also like to thank my committee member Professor Shubham Tulsiani for his perspective and feedback. The courses I have taken under him at CMU have contributed a lot to my knowledge and work. I am also very grateful to our collaborators at UPMC.

I want to thank all my friends and peers at CMU that made my journey very memorable and fun, but especially FNU Abhimanyu, Akshaya Kesarimangalam, Prasanna Sriganesh and Sarvesh Patil. I want to thank Pranav Srinivasan for constantly supporting and motivating me. Lastly, I want to thank my parents Debasis Bal and Dr. Madhusmita Jena, for raising me to be a strong and independent woman in science.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Ultrasound (US) imaging or Sonography is a contact-based tomographic imaging modality that uses high-frequency pressure waves to image soft tissues such as muscles and internal organs. Being non-ionizing, real-time, portable, and low cost, ultrasound imaging is a popular diagnostic first step for many medical interventions - from emergency procedures such as obtaining vascular access for Extracorporeal Membrane Oxygenation (ECMO) and Resuscitative Endovascular Balloon Occlusion (REBOA) [79], [51] to diagnostic procedures such as guiding needles for biopsies, monitoring stones in the gallbladder and kidneys [62], [21], and tumours in the liver and breast tissue [53], [87]. A 3D volumetric representation of the internal scene from a subject's body is valuable to clinicians for diagnostics as well as surgery. However, most commercial ultrasound systems yield a single 2D cross-sectional view of the anatomy at a time. Trained sonographers are capable of fusing multiple 2D observations into a 3D model through their knowledge of the human anatomy. But this abstract understanding can be error-prone and impossible to achieve in the absence of skilled sonographers.

More recently, robotic ultrasound systems coupled with ultrasound-interpreting algorithms are being developed to function in the absence of sonographers. Robotic ultrasound scanning has many benefits such as known probe poses, constant pressure, consistent imaging quality and repeatability [47]. Robot poses in specific, have been used extensively for stacking-based 3D volume reconstruction from 2D images [50], [61], [60]. However this approach has many shortcomings. Firstly, unless an adaptive acquisition method is used that tracks the unfilled gaps in the volume, many regions will remain unobserved, making interpolation necessary in some form [84]. Secondly, interpolation methods often either require hand-tuning parameters or are specific to the imaged anatomy and produce non-existent artifacts [20]. Thirdly, the generated volumes are a non-dynamic representation with no physical grounding, meaning that any new image produced by slicing this volume will not show true intensities that would be observed in reality by the transducer at that pose. Fourthly, contrasting intensities observed for the same locations from different observation poses need to be handled specifically. Finally, storing and manipulating these volumes can demand considerable amounts of memory.

The use of segmentation in anatomy-specific reconstruction is widespread [73], [5], [18], [44]. More often, segmentation in 2D image planes (2D U-Net) is preferred rather than in 3D voxel grids (3D U-Net) as the quality yield is low given high computational costs [94]. Traditional U-Net and CNN-models require a lot of anatomy-specific expert-annotated labels and can perform poorly for out-of-distribution data where discontinuities or deformations are observed in ultrasound [3]. Issues like discontinuities can affect mesh generation which is a key step in ultrasound simulations [12], deformation modeling [72] and planning in robot-assisted surgeries [13].

These fundamental challenges with typical volume and anatomy-specific reconstruction methods for ultrasound motivate the emergence of a novel approach rather than improvements to existing methods.

1.2 Contributions

I present neural implicit representations for both 3D ultrasound volume reconstruction and 3D anatomy-specific reconstruction. The NIR for volume reconstruction (volume NIR) is an advanced version of Ultra-NeRF [90] addressing substantial lacunae. The NIR for anatomy-specific reconstruction is an original breakthrough contribution.

For the volume NIR, I derive a simple physical model that emulates the physics of ultrasound wave propagation for image formation. This model learns two essential components - the tissue reflectivity function (TRF) and the point spread function

(PSF). The tissue reflectivity function quantifies the spatially-varying energy reflected back to the ultrasound transducer after encountering tissues. This function is influenced by tissue properties such as attenuation and acoustic impedance. The point spread function (PSF) is the local spatial impulse response of an ultrasound imaging system, or namely, the resulting image when the medium is solely composed of a single scatterer [67]. Convolving the PSF with the TRF produces the final observed intensities in ultrasound images [14], and by having both these parameters in a differentiable renderer, a multi-layer perceptron (MLP) can regress over multi-view 2D images to learn a volume. This is the basis of the volume NIR.

The PSF affects the general specularity and blurring observed in the rendered image. Ultra-NeRF takes in a pre-defined Gaussian kernel for the PSF and operates with the assumption that the PSF remains constant throughout the imaging-plane, when in reality, it varies with multiple factors, like the transducer geometry, plane-wave steering angle, and scatterer position to name a few. To mimic a realistic PSF with room for spatial variation, I have two additional parameters in our volume NIR that model per-point Gaussian PSFs based on the resolutions of the ultrasound system. Learning these additional parameters in a depth-constrained manner helps improve the photorealism of the rendered images, especially for more complex scenes observed in ultrasound data collected from real-world living subjects.

The wave-propagation model used in the differentiable rendering formulation in Ultra-NeRF produces correct viewing-direction dependent silhouettes and shadows, but it lacks the understanding of the variation of reflectance of tissue with respect to the viewing-direction. This results in correct geometry but incorrect intensities in rendered images for novel-views. The tissue reflectance signal received at the ultrasound transducer from a point depends on the angle of the incidence of the ultrasound ray at the tissue surface at that point. Finding this angle of incidence is reliant on determining surface normals of internal tissue interfaces throughout the learned volume. In order to render fully correct view-dependent images, I propose a new approach to learn a normal field in addition to the TRF with the volume NIR. This contribution is particularly useful in determining the best viewing pose for a particular tissue so that it appears with highest visual contrast to the observer.

For obtaining anatomy-specific reconstructions, the volume NIR is modified to also learn an implicit voxel occupancy function for a specific tissue. Provided the anatomy is largely homogeneous, the tissue properties for the anatomy are characterized by minimal variations. The physical tissue parameters learnt by the NIR adhere to this property. Additionally, the correlation between the parameters differentiate different types of tissue. These properties are sufficient to demarcate tissue such as blood vessels and in this work have been implicitly learned by a shape representation coupled with the volume NIR. This implicit shape representation, which is a larger dual-purpose NIR by itself, is formulated as a weakly-supervised binary occupancy function. Like the volume NIR, it is parameterized by the 3D coordinate space but I add the learned tissue properties as parameters as well. In my work, I analyse vessels and observe that a handful of 2D masks are sufficient for generating high-fidelity 3D reconstructions.

Shortly before these NIR methods were developed, I was focused on developing subject-specific tissue deformation simulations. Most methods in this area depend on detailed computed tomography (CT) scans for surface mesh generation but CT data is not always readily available. Ultrasound volumes can be a substitute but the issues discussed in section 1.1 make mesh generation very noisy. The implicit shape representation enables more accurate mesh generation from ultrasound data. The impact of good 3D anatomy-specific reconstruction is discussed in the context of FEM simulations for soft tissue and so is a novel subject-specific stiffness calibration method specific to 2D ultrasound.

Chapter 2 discusses all relevant prior work, chapter 3 dicusses the experimental setup and data used for experiments, chapter 4, chapter 5, chapter 6 and chapter 7 discuss methods, and their respective experiments and results, and finally, chapter 8 discusses conclusions and future work.

Chapter 2

Background and Related Works

2.1 Reconstruction in Robotic Ultrasound

2.1.1 Building Volumes

In stacking-based reconstruction, there exist two major categories of methods: pixel-based and voxel-based methods. The pixel nearest neighbor (PNN) method is the most common pixel-based method [60]. Each pixel in 2D images is mapped to the nearest voxel in the 3D volume. It is simple and computationally inexpensive but it results in the loss of important details and produces blocky or non-smooth artifacts due to low resolution [15]. In [75], the pixel method is employed to reconstruct blood vessels from ultrasound images by positional calibration of the ultrasound probe. Voxel-based methods use 2D ultrasound frames to reconstruct a detailed 3D volume by interpolating or averaging pixel values [60]. This method considers the spatial relationship between voxels and pixels, resulting in more accurate images. However, it requires more computation and may be sensitive to the quality and resolution of 2D images. Voxel-nearest neighbor (VNN) and distance-weighted (DW) are common methods of computing the voxel values. VNN can preserve the most original texture from ultrasound images, but it tends to generate large reconstruction artifacts when the distance of the voxel to the image plane is large. DW suppresses speckle noise but smooths out the volume, losing information. The authors of [34] presented the squared-distance-weighted (SDW) reconstruction algorithm to reduce the smoothing effect and preserve details in the reconstructed volume.

Radial Basis Function (RBF) interpolation was proposed in [66] as an approximation with splines that try to use the underlying shape of the data in the volume reconstruction. Overfitting is typically unavoidable for the spline method. In [69], the authors present the Rayleigh reconstruction/interpolation with a Bayesian framework that estimates a function for the tissue by statistical methods. The Rayleigh method tends to suppress speckle noise but over-smoothes the boundaries. To summarize, there are still problems to be solved in this area.

2.1.2 Segmenting Anatomies

Quite some works exist that try to reduce supervision for ultrasound image segmentation. In [9], the authors proposed iterative pseudo label-based learning for semi-supervised cardiac structure segmentation. The authors of [48] argued that pseudo labels could be generated via uncertainty estimation during the semi-supervised segmentation. Uncertainty-based confidence-aware refinement at the pixel level was proposed in [83] to ensure the accuracy of pseudo-labels. Although these methods are commonly used, they do not guarantee the accuracy of the generated pseudo labels when applied to US images with typical shadow artifacts. Mean Teacher proposed in [77], is a method that averages model weights instead of label predictions for promoting the consistency between model predictions and the target. Nevertheless, these methods do not consider the negative impact of random artifacts on the prediction of labels in the face of insufficient annotations, which may limit their applicability in accurately segmenting US images with missing or uncertain boundaries.

2.2 Ultrasound Image Synthesis

2.2.1 Interpolative approaches

The interpolative approaches to ultrasound image synthesis use prerecorded 3D ultrasound volumes and reslicing techniques combined with post-processing like adding deformations and artificial shadows in the slice. This technique is very fast, but because the artifacts in the ultrasound images are inserted ad-hoc, they suffer from method-immanent disadvantages such as restricted orientations and the absence of several artifacts like mirroring, refractions, and reverberations [4], [27]. An evaluation of a simulation training system for gynecological sonography was presented in [32]. Re-slicing pre-acquired 3D freehand ultrasound data is directly used for simulation of 2D ultrasound images. Given all the issues that persist, reslicing-based methods are far more suitable to data augmentation for segmentation [61] to present more variations in organs/anatomy to networks in training rather than generate novel-views for any diagnostics.

2.2.2 Generative approaches

Advances in deep learning have enabled various learning-based approaches for ultrasound image synthesis. Some of them incorporate wave-propagation based parameters. Conditional GANs are widely used in generating ultrasound images conditioned on physical input, such as calibrated coordinate values [33] and echogenicity maps [78]. A 2020 paper [49] employed a generative autoencoder model, trained on a large amount of tracked ultrasound data, to perform patient-specific image generation from transducer position and orientation. In [80], a cycleGAN model is used to improve the realism of simulated ultrasound images produced from a ray-casting approach. A GAN approach for image translation is proposed in [95] that bypasses rendering during inference. To simulate realistic ultrasound speckle patterns, the authors of [10] introduced a speckle layer to incorporate the physical model of speckle generation into a GAN-based data augmentation network. However, this was based on Fourier optics and not the physics of ultrasound wave propagation.

2.2.3 Physics-based Simulation

The physics of ultrasound image formation has been touched upon in many works that investigate ultrasound simulation and multiple models have been proposed. Fully synthetic ultrasound simulation has been proposed by [39], [36], [38] based on an acoustic wave-propagation model and the concept of spatial impulse response which is implemented in a program called Field II [37]. Field II can be used to simulate any linear ultrasound system given apodization, focus, pulse excitation scheme and aperture geometry. The program requires location and strength of scatterers as input and produces best results with carefully designed and synthetically generated scattering patterns. As such, the program is mostly used to determine the effects of various parameters on transducer design. Additionally, the simulations for even a single B-mode image takes an extremely long amount of time and needs to be parallelized, making it impractical for real-time simulation.

Some works like [42] and [6] used the Westervelt equation, solved with the finite difference method, to create ultrasound wave-propagation models. These models fell out of favour due to high computation complexities. Simple ray-based models were introduced in [46] and [71] that discussed reflection and lambertian scattering. Later works expanded on these models and perhaps the foundational paper in the context of current standard wave-propagation models is [12]. It describes the convolution-based model for ultrasound. The simulation uses deformable mesh models of tissue for which acoustic parameters can be defined. This work was expanded by [68] where the authors showed optimization for tissue parameters using MRI data. Almost all these simulators are heavily reliant on user-provided material properties which is a huge disadvantage.

2.3 Implicit Learning

2.3.1 RGB

Implicit neural representations refer to a class of methods that use neural networks to encode complex, high-dimensional data without explicitly defining the underlying structure or function. The authors of [25] showed that implicit functions could represent a template for a detailed surface geometry. Integration an implicit surface representation into a neural network was suggested in [57], demonstrating that this approach enabled the inference of shapes with greater geometric intricacies compared to voxel-based representations. A neural scene representation that integrated rendering for a scene through ray-casting with deep learning was introduced in [74]. It did not require shape supervision. This bleeds into neural radiance fields (NeRF) [58], where a fusion of volume rendering and neural networks is employed. However, in this approach, an analytical function is utilized for differentiable rendering, distinguishing it from methods solely relying on neural networks for scene representation.

2.3.2 Medical

MedNeRF [19] proposed the incorporation of GRAF (Generative Radiance Field) [70] to render CT projections from single or multi-view X-ray images. GRAF is a combination of NeRF and GAN. This is proposed as NeRF struggles to handle complex scenes with large amounts of geometric complexity. To handle this limitation, NeRF is trained to minimize the differences between the rendered and ground truth images, while the GAN is trained to distinguish between the generated image and a ground truth image, and utilized to refine the NeRF outputs and improve image quality.

Neural implicit representations have been successfully employed for tasks such as 3D reconstruction from ultrasound images, e.g. in ImplicitVol [92], segmentation with continuous functions [43], deformable image registration [88], and high-resolution MR reconstruction [89]. Another category of methods combines implicit representations with complex imaging models that address intricacies of medical imaging modalities. For instance, [65] incorporates implicit neural representations into a 4D-CT reconstruction pipeline, utilizing them as a fixed prior alongside a motion field. Multiple neural networks are proposed in [91] to represent bias field, noise variance, and volume intensities implicitly, enhancing 3D reconstruction from motion-corrupted 2D slices. The use of NeRF in the medical domain is an emerging research area since adapting volume-rendering methods for medical imaging requires addressing the unique characteristics of image formation models unique to medical imaging. Existing literature, largely centered on CT or MR, includes examples like MedNeRF [19], specializing in CT reconstruction from X-ray data, and [35], extending the approach to brain MR scans by incorporating a radiation attenuation response. However, ultrasound imaging stands in stark contrast to MR and CT due to its inherent anisotropic nature.

2.3.3 Ultra-NeRF

Ultra-NeRF [90] presented the first differentiable rendering-based NIR for ultrasound with a convolution-based ray-tracing model. The method is able to reconstruct a continuous volume from multiple angled ultrasound sweeps. It produces partially view-dependent rendering, rendering shadows precisely but lacking consideration of view-dependent reflection. The formulation of the rendering presented in [90] is discussed in chapter 4. Ultra-NeRF serves as the baseline for the contributions presented in this thesis.

2.3.4 Ultrasound vs RGB imaging

Being tomographic in nature, ultrasound allows us to image cross-sections of a scene or object by using penetrating waves. This is fundamentally different from RGB imaging which is projection-based. This difference leads to a key distinction in the way neural representations for both these modalities differ. In RGB imaging, knowing the poses of the camera for multiple images does not amount to a 3D representation of the scene. Either feature-matching based triangulation is required for an explicit point-cloud like representation, or ray propagation is required for a NeRF-like implicit representation. However, with tomographic images, known poses can help stack observed pixels in a 3D voxel grid, providing us with a direct explicit 3D representation of the scene. Given the discussed challenges with this compounding-based representation, a NeRF-like NIR for ultrasound is justified for better reconstructions.

It is also worthwhile to note that there exist methods coupled with NeRF that attempt label-propagation in 3D from weak 2D supervision for segmenting out objects in RGB scenes. But these methods are more complex, often combining additional knowledge [59], [81] from other larger pre-trained models to understand 3D geometry. In contrast, a handful of annotated 2D masks are sufficient for our equivalent task of anatomy-specific reconstruction in the ultrasound domain. This is possible because firstly, we learn physical parameters which convey more meaningful information about tissues instead of using appearance parameters like intensities [28]. Secondly, most tissues (fat, nerves, sheaths, vessels etc.) are largely homogeneous in appearance, eliminating the requirement for using other sources for shape understanding.

2.4 Deformation Modeling

Force data collected from a multi-axial force sensor mounted on the robotic manipulator, and tissue deformation data collected from a stereo camera system are used for estimating mechanical parameters of soft tissue in [11]. The authors of [24] use RGB-D sensing to learn force values in an ex-vivo set up with a da Vinci

Surgical System for brain tissue. In [41], a novel approach to simulate the soft-body deformation of an observed object is introduced. The approach tracks an object's movement using an RGB-D sensor and simulates its deformation iteratively. The method could be applied to track skin deformation but since the scope of this thesis is vessels, exterior RGB-D sensing is not applicable.

Quite a few studies have explored simulating soft-tissue deformations but few optimize simulations using medical images. In [30], the authors propose a comprehensive pipeline to create patient-specific biomechanical models and optimize deformation predictions in FEM through iteratively updating model parameters by maximizing image similarity between FEM-predicted MR images and the experimentally acquired MR images of a breast. To predict deformations in real-time, in [96] a liver model with biomechanical properties similar to a real one is created using FEM and a data set of deformations with different forces is generated. The mechanical behaviour is simulated in real time by a LightGBM (Light Gradient Bossting Machine) regression model trained with the generated data set. Vessels are modeled and deformed in real-time using a tensor-mass method in [29] and the authors perform experiments for determining realism but do not use medical imaging to quantify it and rely on qualitative results. Other papers [40], [56] simulate vessel deformations due to blood flow.

In [54], the authors propose using deep neural networks to learn large deformations occurring in ultrasound-guided breast biopsy as FEM is not real-time. They train a U-Net architecture on a relatively small amount of synthetic data generated in an offline phase from FEM simulations of probe-induced deformations to provide accurate prediction of lesion displacement. 3D-PhysNet, proposed in [85], can predict three-dimensional deformations in solids under applied forces by encoding the physical properties of materials and applied forces in the network, essentially learning the FEM simulation.

Chapter 3

Experimental Setup and Data

Two categories of data are used in this work - 1) Synthetic ultrasound data and 2) Real-world ultrasound data captured through a robotic system.

3.1 Synthetic Ultrasound Data

The synthetic ultrasound data is generated from liver CT scans using ImFusion [1] which has a proprietary ray-tracing based simulator. The subject has hepatic vessels visible. Six angled sweeps with 200 images each and pose tracking, are generated from the CT data. The tilted sweeps differ in slope compared to the outer surface. Therefore, the scene is observed from different viewing directions. The frames contain occlusions caused by bone structures (ribs).

3.2 Real-world Ultrasound Data

The real world ultrasound data is acquired with a robotic ultrasound system as shown in Figure 3.1. The system consists of a PaoLus UF-760AG Portable Diagnostic Ultrasound Imaging Equipment (FUKUDA DENSHI,UK) using a 5-12 MHz 2D linear transducer mounted on the 6-DoF Universal Robot UR3e robot. The robot provides poses for the ultrasound transducer at 150 Hz and is well-calibrated, having sub-millimeter accuracy.

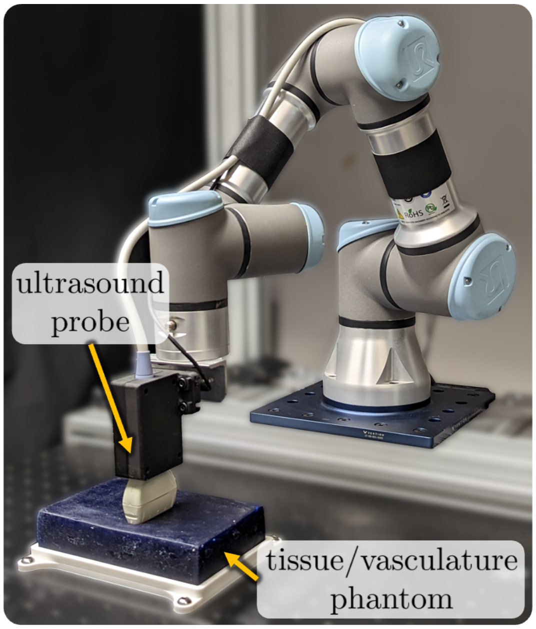

Figure 3.1: The Robotic Ultrasound System used for data acquisition.

This system is used to acquire data from two types of subjects - 1) Blue-gel Phantom and 2) Live-pig subjects.

3.2.1 Blue-gel Dataset

The medical phantom is a Blue-gel phantom with a pair of bifurcating vessels. It is a homogeneous subject without any features corresponding to muscle, fat, nerves etc. The blue-gel dataset is collected by imaging the phantom placed in a water-bath. This is done to eliminate direct contact with the phantom which will cause deformations. 3 sweeps, 2 transverse to the vessels (with 360 images each) and 1 raster scan longitudinal to the vessels (500 images), are captured perpendicular to the surface of the the phantom (0° from the surface normal). 8 transverse sweeps with an average of 350 images each, tilted from angles $-20°$ to $+20°$ at intervals of 5° are collected along the same trajectory as the first transverse sweep at 0°. The tilting rotation occurs along the longitudinal plane, which implies that all tilted sweeps image in transverse planes.

3.2.2 Live-Pig Dataset

The live animal subjects are 2 living pigs. For both these live-pig datasets, we use our robotic system to scan the femoral vessels in the transverse orientation. For live-pig1, 5 angled sweeps are collected, tilted from angles $0°$ to $-20°$ at intervals of $5°$. Each of these sweeps have 245 images. The tilting rotation occurs along the longitudinal plane, which implies that all tilted sweeps image along the transverse plane. For live-pig2, 7 angled sweeps are collected, tilted from angles $-15°$ to $+15°$ at intervals of $5°$. Each of these sweeps have 200 images on an average.

3.3 Compute

All the learning-based methods are implemented in PyTorch [64] and trained on an NVIDIA TITAN RTX GPU with 24 GB memory.

Chapter 4

Learning point-wise Point Spread Functions and Architecture Changes to Ultra-NeRF

4.1 Ultrasound Volume Rendering Revisited

Ultra-NeRF used a $\mathbb{R}^3 \rightarrow \mathbb{R}^5$ NIR where attenuation α, reflection coefficient β, border probability ρ_b, scattering density ρ_s and scattering amplitude ϕ are predicted for each sampled (x, y, z) location in the volume. In experiments conducted by the authors later, it was found that the border probability term is redundant. The following ray-propagation model encapsulates the modified volume rendering formulation from Ultra-NeRF.

For each scan-line or ultrasound ray r, eq. (4.1) defines a recorded US echo $E(r, t)$, measured at distance t from the transducer, as a sum of reflected energy $R(r, t)$ and backscattered $B(r, t)$ energy:

$$E(r, t) = R(r, t) + B(r, t) \tag{4.1}$$

As shown in eq. (4.2), the reflected energy at a point (r, t) is a product of the remaining energy in the ray $I(r,t)$ and the reflection coefficient at that point $\beta(r,t)$.

$$R(r,t) = |I(r,t) \cdot \beta(r,t)| \tag{4.2}$$

$I(r,t)$ is dependent on both tissue attenuation and energy lost due to reflections up till point t in the ray. Given that the initial energy of the ultrasound ray at the transducer I_0 is of unit intensity, $I(r,t)$ is given by eq. (4.3)

$$I(r,t) = I_0 \cdot \prod_{n=0}^{t-1} (1 - \beta(r,n)) \cdot e^{(-\int_{n=0}^{t-1} (\alpha \cdot f \cdot dt))} \tag{4.3}$$

where α values correspond to the physical attenuation of tissue only up to an unknown scaling factor, f corresponds to the frequency of the ray, and loss of energy happens at an infinitesimal step dt along the ray. The final α learned by the network is representative of the product of the attenuation and frequency.

To obtain the backscattered energy $B(r,t)$ term, the product of the remaining energy $I(r,t)$ and the PSF is convolved with a 2D map of scattering points, as shown in eq. (4.4). This map is obtained by multiplying an admittance function H with the scattering amplitude ϕ. H is sampled from a Relaxed Bernoulli distribution with parameter ρ_s showing the inherent uncertainty in the observation of the scattering effect of each point scatterer.

$$B(r,t) = I(r,t) \cdot PSF(r) \otimes (H(r,t) \cdot \phi(r,t)) \tag{4.4}$$

Ultra-NeRF uses a constant 7x7 Gaussian kernel with mean 0 and variance 1 as the PSF. This is an informed guess that worked well for the datasets in consideration in [90].

4.2 Drawbacks of Using a User-specified Global PSF

As described in section 1.1, the PSF observed at a location depends on multiple factors such as transducer geometry, beam shape, steering angle, location of scatterer etc.

Therefore, a single PSF, constant throughout the volume is an incorrect assumption. Additionally, relying on a user-input necessitates fine-tuning.

While testing on real-world ultrasound data (blue-gel and live-pig), the PSF used by Ultra-NeRF results in rendered images which have a very different speckle-pattern from the ground truth images. Figure 4.1 captures this phenomenon.

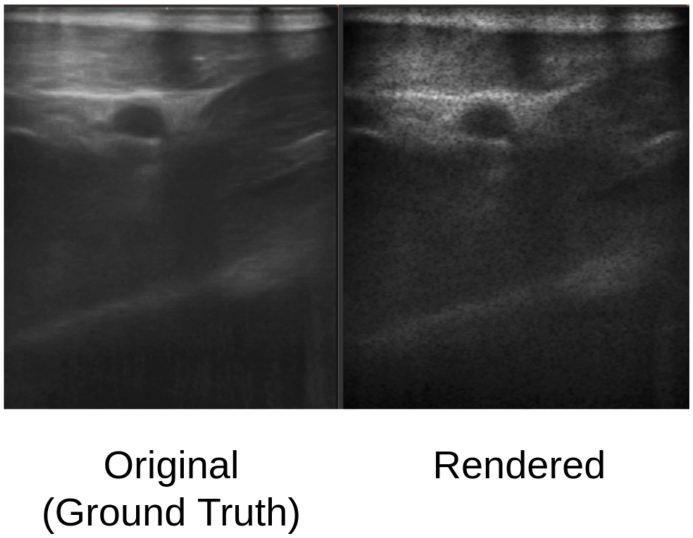

Figure 4.1: The original (ground truth) ultrasound image (left) and the corresponding rendered image with a constant PSF (formulation from Ultra-NeRF) (right). It shows the mismatched speckle pattern.

4.3 Literature

Homomorphic filtering in the cepstrum domain is used in [52] for extracting PSFs from raw RF (Radio-Frequency) data. Although the paper mentions the use of ultrasound images for the method, raw signals from the RF image are used as the input and not B-mode ultrasound images. RF images are not accessible on most commercial ultrasound systems. Therefore, to the best of my knowledge, no accessible method using B-mode ultrasound images exists for estimating the PSFs of imaging systems.

4.4 Method: Learning a per-point PSF

Modeling transducer-specific properties is very parameter-intensive and in most cases, infeasible due to proprietary ultrasound systems. However, the effect of the scatterer position on the PSF can be modeled using regression over observed intensities through the volume rendering equations while considering beam resolutions.

We know from prior work that the PSF [52] is a cosine-modulated 2D Gaussian kernel with the parameters being the axial (σ_x) and spatial (σ_y) resolutions in pixels eq. (4.5).

$$PSF(r,t) = \exp\left(-\frac{1}{2}\left(\frac{x^2}{{\sigma_x}^2} + \frac{y^2}{{\sigma_y}^2}\right)\right) \cdot \cos\left(2\pi f\right) \tag{4.5}$$

Here, x and y refer to the coordinates in the PSF kernel.

To learn the per-point PSFs, the network predicts two further parameters per sampled point - σ_x and σ_y, making the NIR a $\mathbb{R}^3 \to \mathbb{R}^6$ function. The axial resolution refers to the ability of the system to distinguish between two structures that are aligned along the axis of the ultrasound beam. The spatial resolution is a measure of how well the system can depict fine details and differentiate between structures that are adjacent to each other.

We allow the network to predict σ_x and σ_y values within the actual axial and spatial resolution ranges of most commercially-available ultrasound systems. This is done with scaled sigmoid activation functions. σ_x and σ_y are between 0.145 - 1.45 mm and 0.4 - 3 mm respectively [12]. The learned resolutions are multiplied by the system-specific mm-to-pixel scaling factor before being applied to eq. (4.5) in the rendering. We use a 7x7 kernel for the Gaussian PSF.

We expect the PSFs to be consistent along the width of the imaging plane at a particular depth, implying that the primary variation should be along the height of the image. In order to enforce this, I regularize the learned values for σ_x and σ_y in a depth-wise manner using a loss. To achieve this, the following loss eq. (4.6) is used.

$$\mathcal{L}_{\text{depth\_reg}} = \frac{\left(\sum_{h=1}^{H} L2(\text{Var}_h(\sigma_x), 0) + \sum_{h=1}^{H} L2(\text{Var}_h(\sigma_y), 0)\right)}{2} \tag{4.6}$$

Here, H is the height of the image in pixels and Var refers to the computed variance.

The rendering loss applied to the NIR is the same as [90]. It is given by eq. (4.7).

$$\mathcal{L}_{\text{render}} = \lambda \, \text{SSIM} \left(Im', Im \right) + (1 - \lambda) L2 \left(Im', Im \right) \tag{4.7}$$

Here, Im is the original image, Im' is the rendered image and $\lambda = 0.7$. The final loss used for the training is:

$$\mathcal{L}_{\text{total}} = 0.8 \mathcal{L}_{\text{render}} + 0.2 \mathcal{L}_{\text{depth\_reg}} \tag{4.8}$$

The network is trained with the Adam Optimizer [45] and a learning rate of 0.001.

4.5 Method: Neural Network Architecture Changes

While applying the Ultra-NeRF architecture for learning ultrasound volumes from live-pig data, some aberrations are observed in the rendered images, especially near high-frequency regions. Some typical artifacts like pixelation, non-smooth interpolation and failed interpolation were observed, indicating that the network was struggling to fit to the data. Examples can be seen in fig. 4.2.

Drawing inspiration from other works in the domain of medical imaging addressing inverse learning problems [89], [76], I expanded the MLP network to contain 14 hidden layers in addition to the prediction layer. I also added 3 skip connections to layers 4, 8 and 12. The modified network architecture which learns the PSF is shown in fig. 4.3.

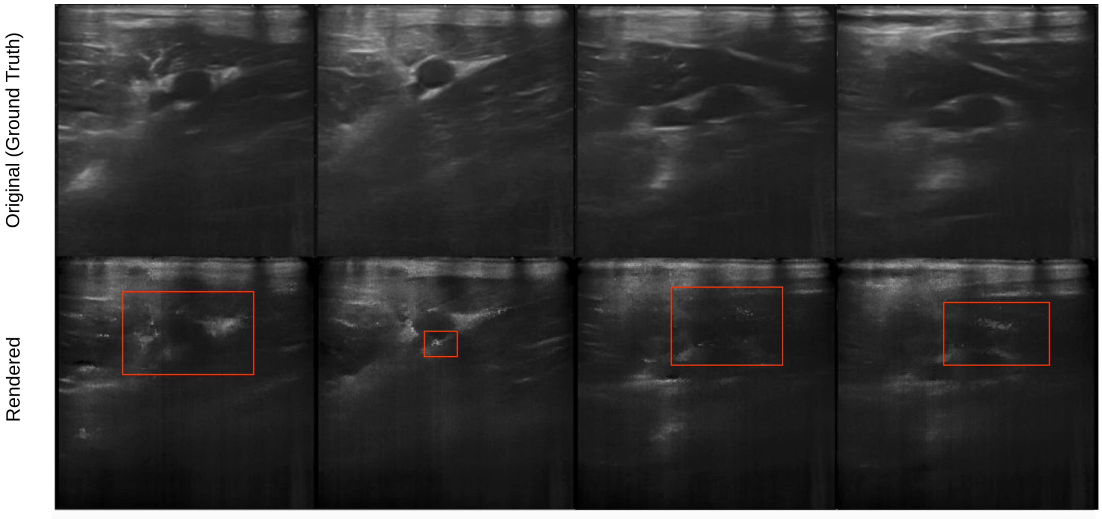

Figure 4.2: The comparison of rendered images from the learned volume to the ground truth images of live-pig1 dataset when using the 8-layer MLP architecture from Ultra-NeRF. The red boxes highlight the high-frequency details missed.

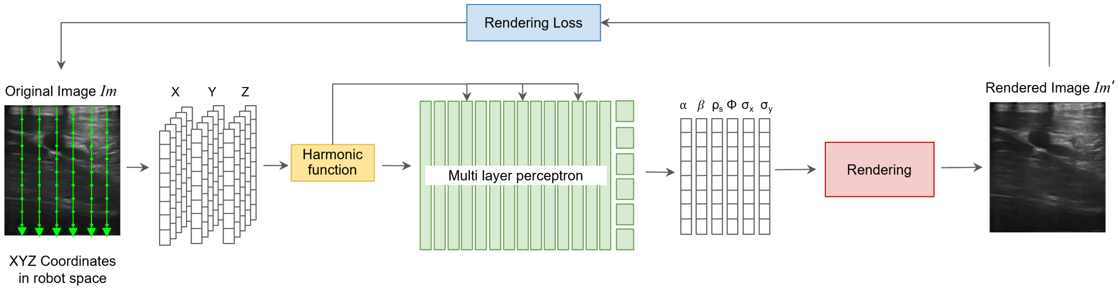

Figure 4.3: The new network architecture used for learning the TRF as well as the PSFs.

4.6 Results

4.6.1 New Network Architecture

The new architecture is better-suited to learning more complex scenes that are present in data from real-world subjects. The improved rendered images demonstrate that the new network architecture helps learn high-frequency artifacts well, being able to interpolate seamlessly. Observe improved rendered images on the live-pig1 datatset in fig. 4.4.

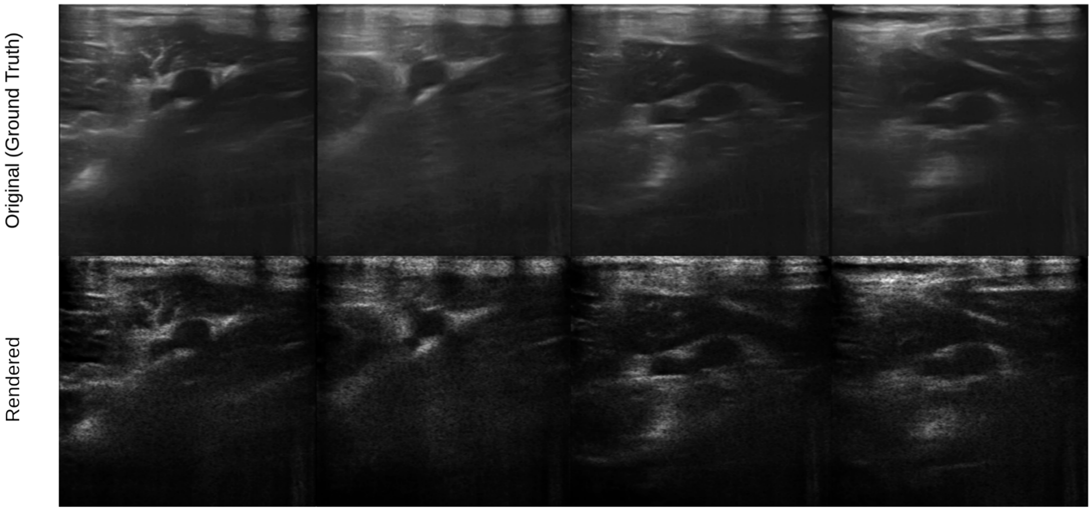

Figure 4.4: The comparison of rendered images from the learned volume to the ground truth images of live-pig1 dataset when using the new 15-layer MLP architecture. Most of the high frequency artifacts are captured.

While the network performs better in capturing these high frequency artifacts, this is not directly reflected in the SSIM scores. The general smoothing and higher average intensities obtained in the rendered images from the 8 layer architecture produce SSIM scores **(0.486)** which are equivalent to the SSIM scores for rendered images from the 15 layer architecture **(0.487)**.

Dataset	Ultra-NeRF (pre-defined, constant PSF)	Proposed NIR (learned, varying PSFs)
Synthetic	0.369	**0.455**
Phantom: Blue-Gel	0.394	**0.553**
Live-pig1	0.486	**0.632**
Live-pig2	0.487	**0.643**

Table 4.1: SSIM scores of rendered images from Ultra-NeRF and the proposed NIR

4.6.2 Learned PSFs

It is observed that learning the local spatial and axial resolutions for PSFs help improve the quality of the rendered images and provide higher mean SSIM scores. The comparisons with SSIM scores of rendered images from models with a single, constant PSF are provided in table 4.1. For all the datasets, we observe an increase in the mean SSIM scores.

Qualitative results are provided in fig. 4.5, fig. 4.6, fig. 4.7. The general observation is that the rendered images are smoother which particularly helps discern smaller structures. With a grainier pattern from the constant PSF, it is entirely possible that speckle makes smaller features discontinuous, making them harder to observe. This is a very important distinction for smaller anatomies such as nerves.

The observed variation in the axial and spatial resolutions learned by the method is minimal, observed only in the fourth decimal place while measured in millimeters. While this does not indicate what the beam shape looks like, the changes in the kernel values affect the final intensities after multiple convolutions.

Discussion

As stated earlier, to the best of my knowledge, no open-sourced method exists to determine the Point Spread Function for an ultrasound imaging system. Field II provides a library to compute the PSF but this requires a special phantom which was not accessible to me. However, one good check for validating that the network is learning meaningful resolution values, is to compare the learned resolutions on two different datasets collected by the same system. In table 4.2, we can see that the

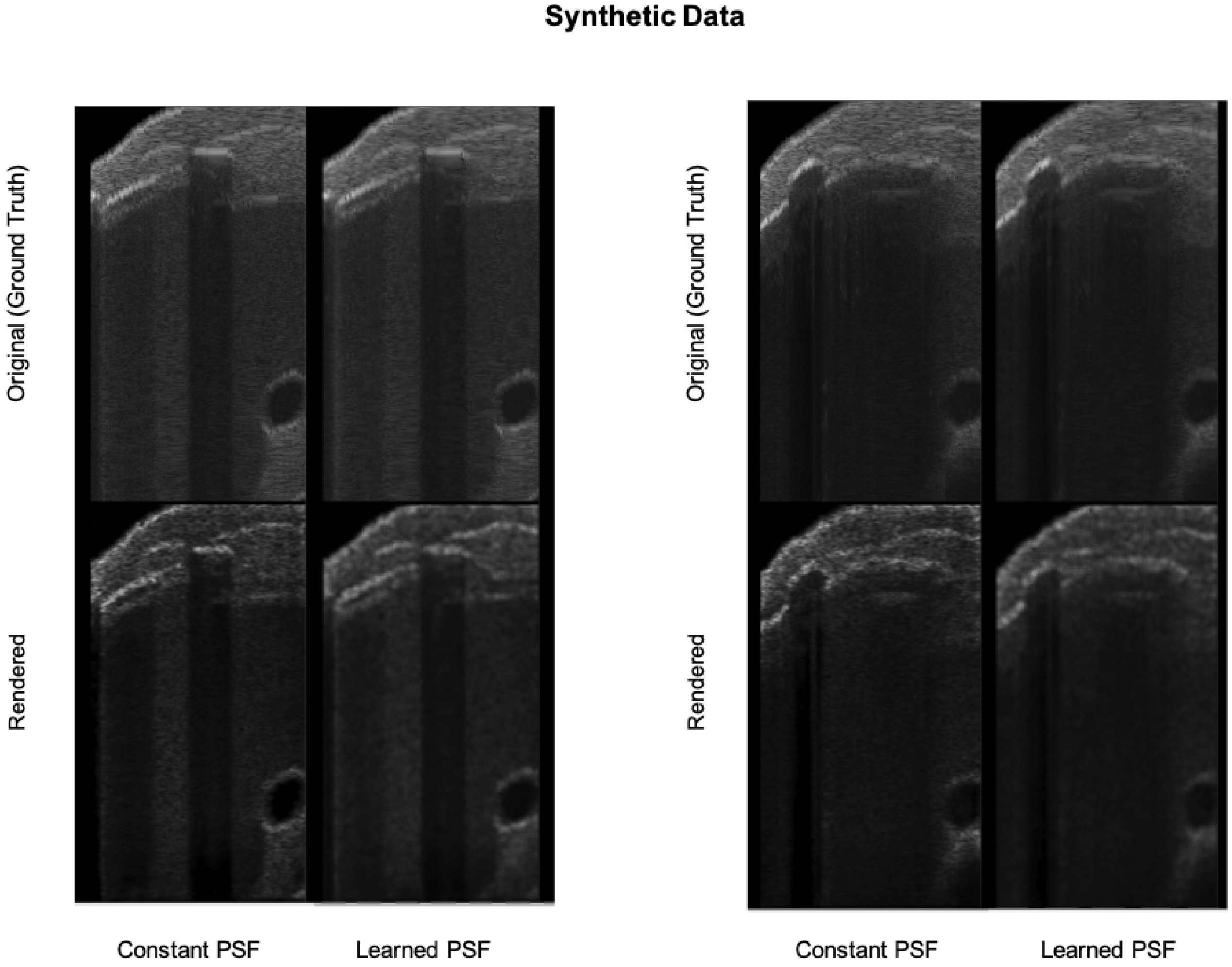

Figure 4.5: The comparison of rendered images from the synthetic dataset while 1) using constant PSF estimate and 2) learning per-point PSFs. Images from two poses are provided.

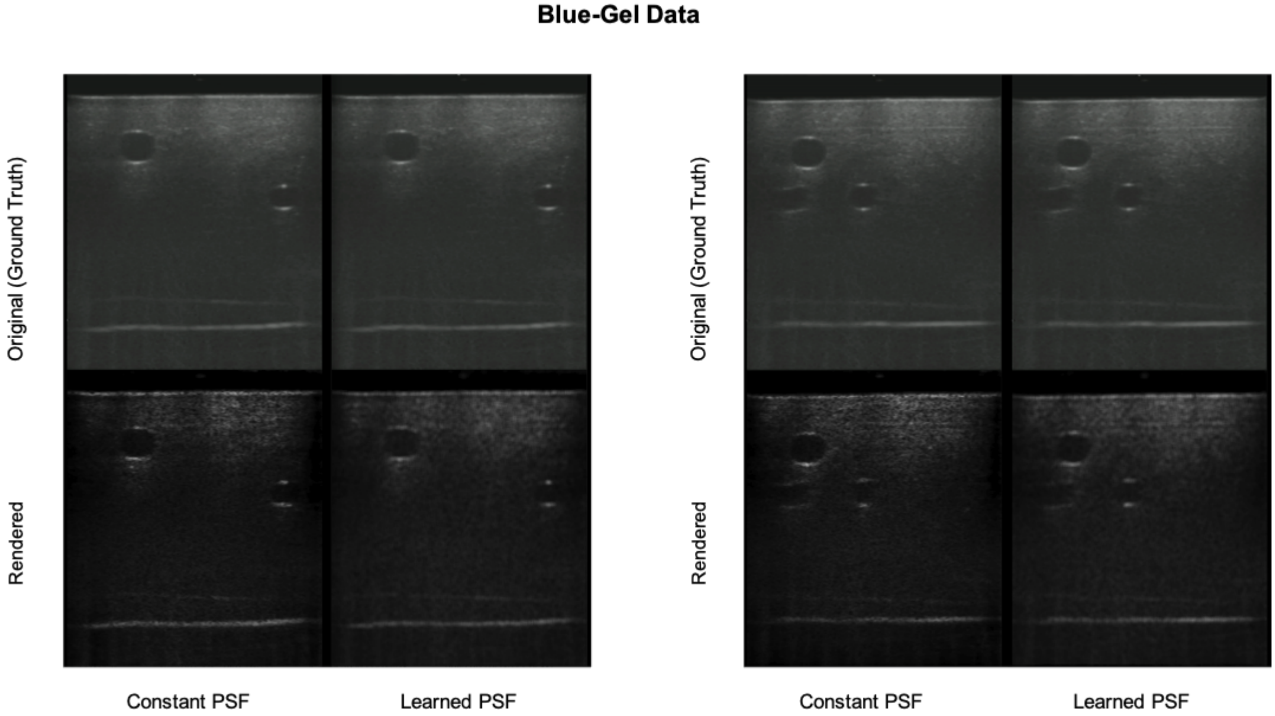

Figure 4.6: The comparison of rendered images from the blue-gel dataset while 1) using constant PSF estimate and 2) learning per-point PSFs. Images from two poses are provided.

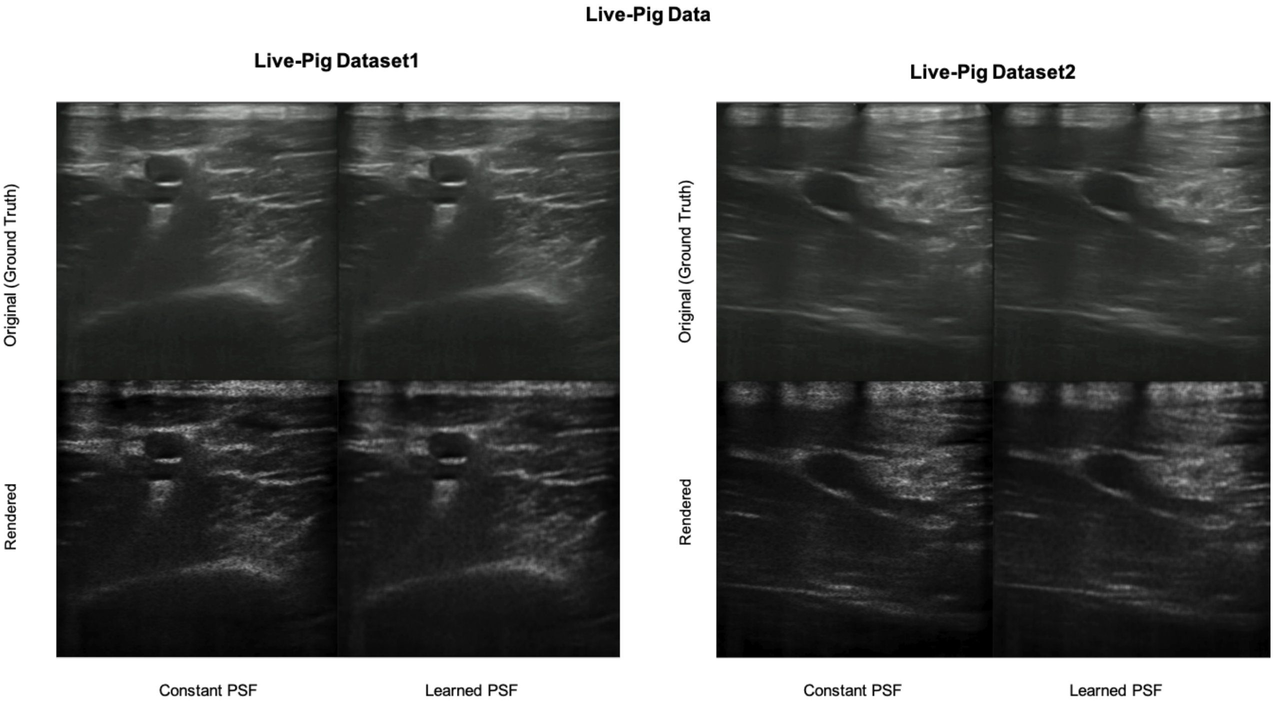

Figure 4.7: The comparison of rendered images from the live-pig datasets while 1) using constant PSF estimate and 2) learning per-point PSFs. Images from one pose are provided per dataset.

Dataset	Mean σ_x (Axial Resolution) in mm	Mean σ_y (Spatial Resolution) in mm
Blue-Gel	0.723	2.126
Live-pig1	0.743	2.098

Table 4.2: Mean estimations for PSF parameters for a single ultrasound imaging system on two different subjects.

network estimates nearly the same mean axial and spatial resolutions for datasets from two different subjects acquired by the same imaging system showing that meaningful values are learned for the PSF. However, to ensure that the PSFs follow the expected profiles according to bean shape, further constraints and regularization needs to be added to the method.

Chapter 5

Viewing Direction-Dependence for Ultrasound Volume Rendering

5.1 Literature

Reflection occurs at large scale tissue boundaries and is incidence-angle dependent. A common physical model for the reflection, which is described in textbooks [31] and [93] and further used [86] and [82], is to calculate the reflected signal by multiplying a $\cos(\theta)$ term with the reflectance coefficient where θ refers to the local incidence angle. This formulation is derived from Lambertian reflectance with the key distinction being that the $\cos(\theta)$ term has an exponent when it comes to modeling tissue reflectance in the context of ultrasound.

5.2 Method

The literature presents a physics model which incorporates the angle of incidence of ultrasound rays with tissue interfaces. This can be utilized in the differentiable rendering module that my NIR uses. However, it is evident that the incidence angle must be known in order to use the incidence-angle dependent rendering.

Let us first examine the changes to the rendering formulation for incorporating the incidence angles. The reflected energy, computed by eq. (4.2), is now computed by the modified eq. (5.1).

$$R(r,t) = |I(r,t) \cdot \beta(r,t) \cdot \cos(\theta)^m| \tag{5.1}$$

And the remaining transmitted energy $I(r,t)$ as seen in eq. (4.3) is computed by:

$$I(r,t) = I_0 \cdot \prod_{n=0}^{t-1} \{1 - \beta(r,n) \cdot (cos(\theta(r,n))^m)\} \cdot e^{(-\int_{n=0}^{t-1} (\alpha \cdot f \cdot dt))} \tag{5.2}$$

As θ features in the differentiable rendering module, one could assume that training the network with this module on images from different viewing angles will suffice and that the regression will converge to the correct incidence angles. The network would directly predict the normal field (N) parameters N_x, N_y, and N_z for each (r,t). However, given that the data is collected in sweeps with unknown gaps between frames, few pixels in the entire observed volume have observations from different imaging planes corresponding to different viewing angles. This means that unless many more angled sweeps are acquired and used in training, the network will not be able to converge on meaningful values for θ. Additionally, abrupt geometries in tissue interfaces could also mean that the smooth function that the network learns would have degenerate solutions for θ at many locations.

Another method to estimate N, is to compute surface normals from the volumetric gradients. To obtain high-fidelity volumetric gradients, the learned volume must be sampled at multiple planes, preferably at infinitesimal increments, in order to compute differences. Computing the gradient in the ultrasound domain will result in wrongly computed surface normals because the scattering artifacts and shadows lead to incorrect gradients. Therefore, gradients from ultrasound images would be noisy and incorrect in magnitudes by a large margin. Another complexity is the added computation of rendering new ultrasound images from the learned volume at multiple poses.

In the NIR, reflectance fields learned by the network are cleaner (no scattering) and view-independent (no shadow effects). This makes reflectance maps a better choice, albeight not flawless, to learn a close approximation of the true N. However, reflectance learned by the network is not guaranteed to indicate all tissue interfaces correctly, especially in regions where the network has difficulty attributing pixel intensities in the right ratios between reflectivity and backscattering. For this reason, we do not rely on recomputing N from gradients along the reflectance field at every epoch, even though this can be performed differentiably. However, the normals computed from the gradients of the reflectance field can serve as a rough prior for N that the network can refine as it trains with varying intensities from data at different viewing-angles. Therefore, in the proposed method, we combine the prior computed from volumetric reflectance gradients with the view-dependent rendering formulation.

Initially, the network is trained for a few epochs on a single sweep, preferably with the ultrasound probe at the same orientation throughout the sweep. The network predicts the TRF and PSF only, with the N prediction initially frozen. Once a rudimentary TRF is learnt, the training pauses and the normal field prior P is computed from sampled reflectance maps. The training then resumes and processes all the training sweeps with per-point N_x, N_y, and N_z values being predicted.

The predicted N is used to compute incidence angles between known rays and tissue interfaces which are used in eq. (5.2) and eq. (5.1). As for the supervision through the rough prior, we apply a soft loss as shown in eq. (5.3), allowing the network to deviate from and correct over P where the signal from the rendering dictates.

$$\mathcal{L}_{\text{normal}} = 0.9 - \frac{\sum_{n=1}^{H\text{x}W} N_n \cdot P_n}{H\mathbf{x}W} \tag{5.3}$$

Where H and W are the height and width of the image in pixels. The value 0.9 was chosen empirically and provides the best results in the experiments conducted.

When combined with learning the PSFs, the loss used in training the network is given by eq. (5.4).

$$\mathcal{L}_{\text{total}} = 0.75\mathcal{L}_{\text{render}} + 0.1\mathcal{L}_{\text{depth\_reg}} + 0.15\mathcal{L}_{\text{normal}} \tag{5.4}$$

From multiple experiments, the value of m in eq. (5.1) and eq. (5.2) is determined to be 1. The network is trained with the Adam Optimizer [45] and a learning rate of 0.001. Observe fig. 5.1 for the architecture.

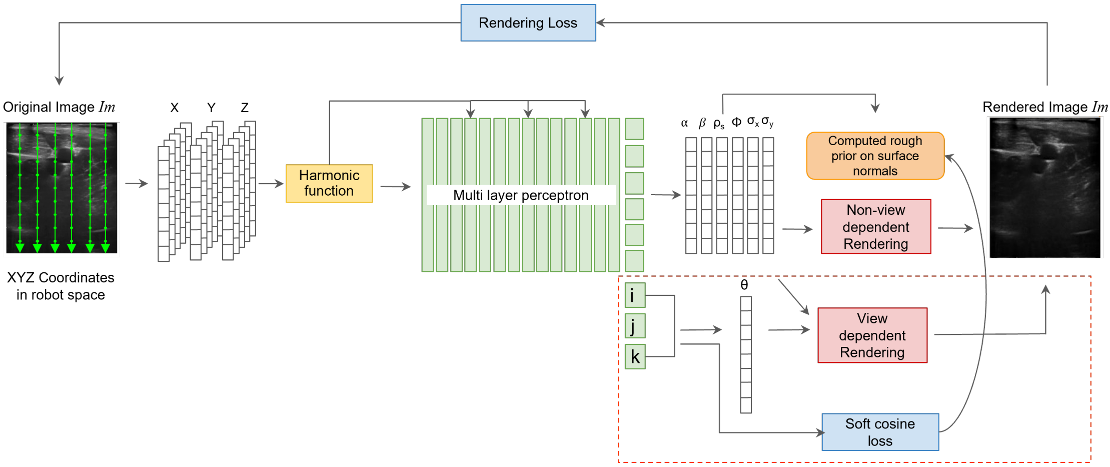

Figure 5.1: Network architecture for learning surface normal field for view-dependent rendering. The modules in the red box are activated once the prior P is computed.i, j, k correspond to N_x, N_y and N_z.

5.3 Experiments and Results

We test this method on the blue-gel and live-pig2 datasets. In the actual ultrasound images in fig. 5.3, we can observe the changes in observed intensities at the top and bottom edges of the phantom as well as the vessel walls when we view it from varying orientations. Our goal is to duplicate these findings in the NIR renderings.

For training on the blue-gel dataset, sweeps at 0°, 10°and -20°were used. The validation is performed on unseen sweeps at orientations 20°and -15°. As observed in the regions corresponding to the vessel walls and the phantom edges in fig. 5.4 and fig. 5.5, the network is able to render realistic images at different orientations, suggesting that the network is learning a reasonable N. The learnt surface normals for a given slice are seen in fig. 5.2. One interesting thing to note is that the overall intensities in the rendered images when the NIR is trained for different viewing directions, is lower in comparison to the ground truth images. It is possible that the NIR is unable to fully model the complexity of the scene and is averaging intensities across views to some extent. More tests need to be conducted to ascertain the reason behind this.

For training on the live-pig2 dataset, sweeps at 0°, -5°, 5°, -15°and 15°were used. The validation is performed on unseen sweeps at orientations +10°and -10°. With

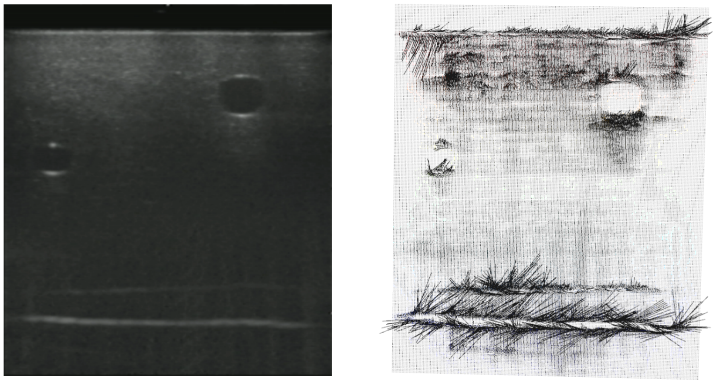

Figure 5.2: Left: Ground truth ultrasound image at a given pose; Right: Surface normals pertaining to the slice at the same pose, extracted from learned N.

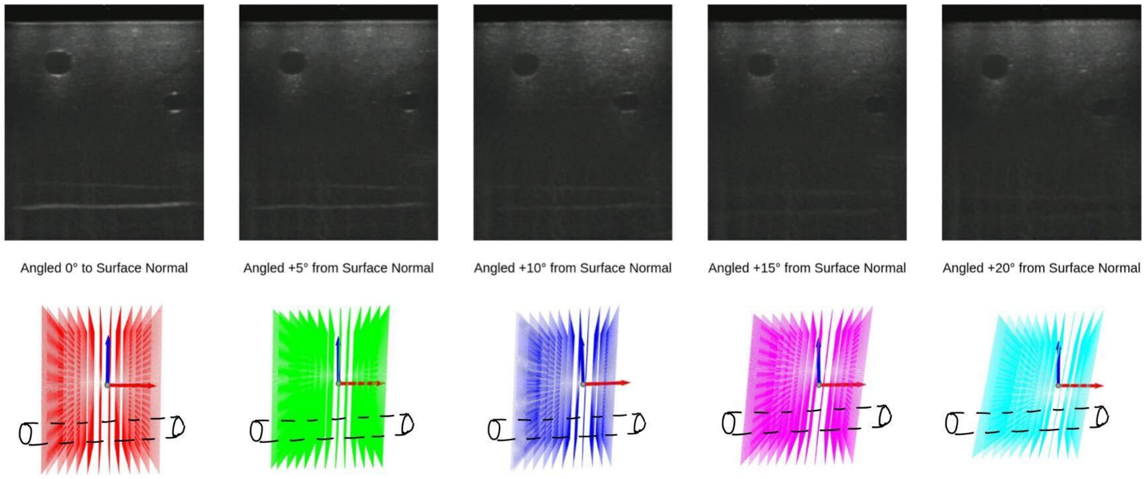
Angled 0° to Surface Normal
Angled +5° from Surface Normal
Angled +10° from Surface Normal
Angled +15° from Surface Normal
Angled +20° from Surface Normal

Figure 5.3: Illustration depicting ground truth images of the blue-gel phantom at varying orientations in a fanning motion of the probe along with the corresponding slices with a visualization of vessel for understanding relative pose.

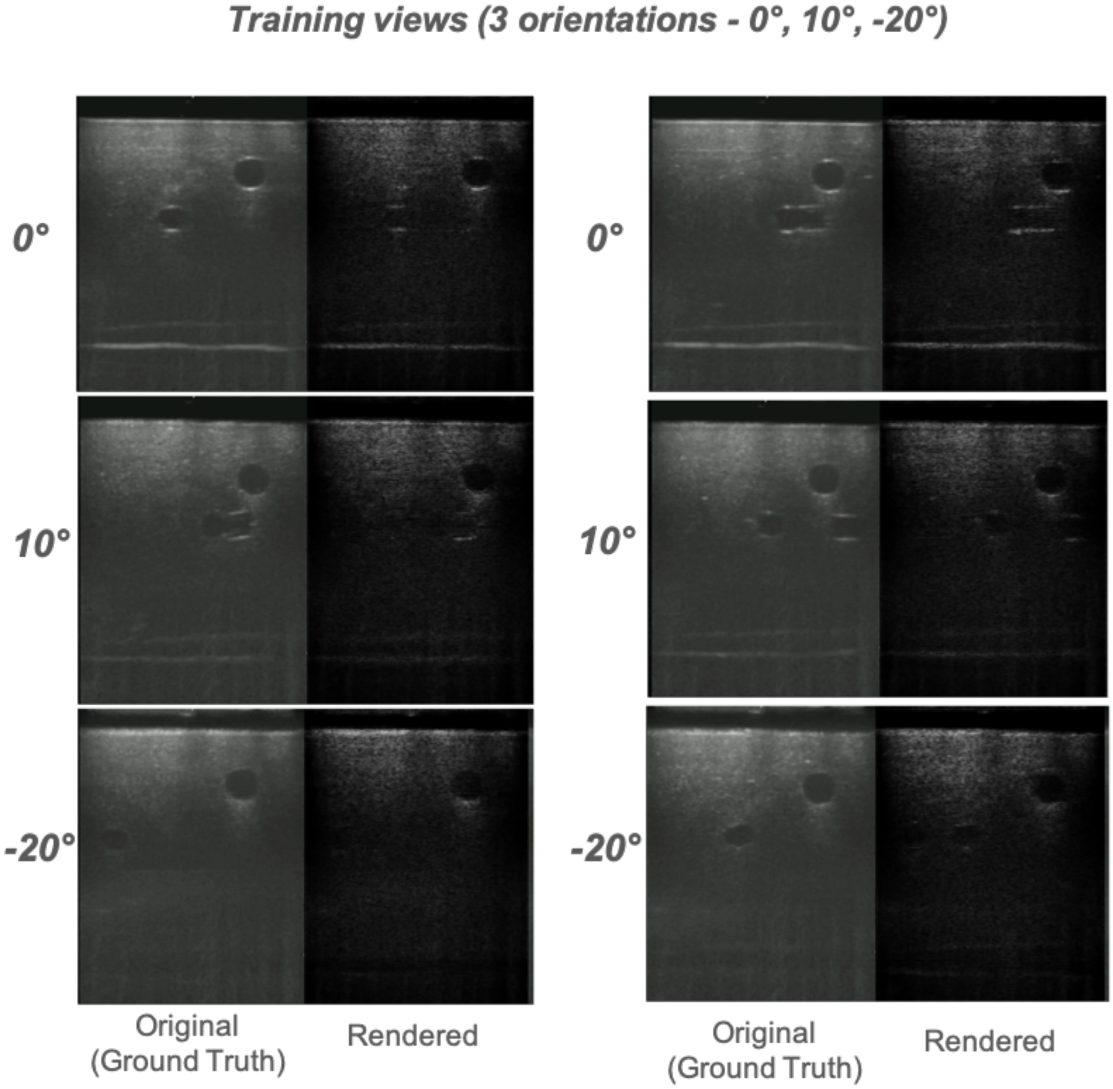

Figure 5.4: The ground truth images at training orientations and the corresponding rendered images as seen in training from the blue-gel dataset.

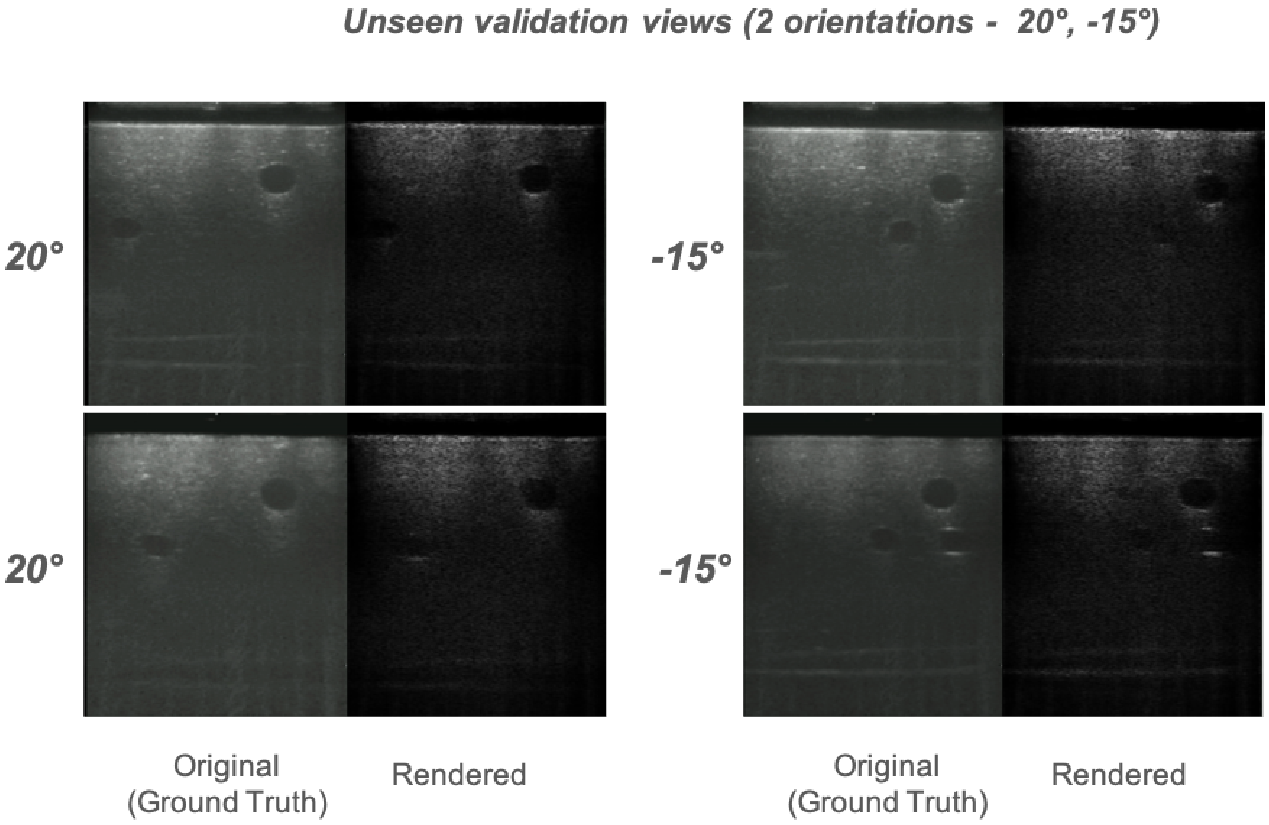

Figure 5.5: The ground truth images at validation orientations and the corresponding rendered images as seen during inference from the blue-gel dataset.

Dataset	SSIM Non view-dependent rendering	SSIM View-dependent rendering
Phantom: Blue-Gel	0.38	**0.42**

Table 5.1: Mean SSIM scores for blue-gel dataset.

the pig data, there are many mixed results. A few sections of the sweep are captured well by the model and in these patches, view dependence is observed but not for all the tissue interfaces. While the training views show some amount of success, the validation results are not good fig. 5.6. Before attributing the subpar results to scene complexity, one should also consider that the pig data is not collected in a water-bath, and so there exist varying compression-based deformations in all the angled sweeps. This means that a non-coherent set of observations is fed to the network. Unsurprisingly, it fails to render good images.

The mean SSIM scores of rendered images from view dependent rendering vs non view-dependent rendering are presented in table 5.1 for blue-gel data. As the results on the live-pig data are not structurally sound (do not represent tissue boundaries well) due to deformations, no SSIM is provided.

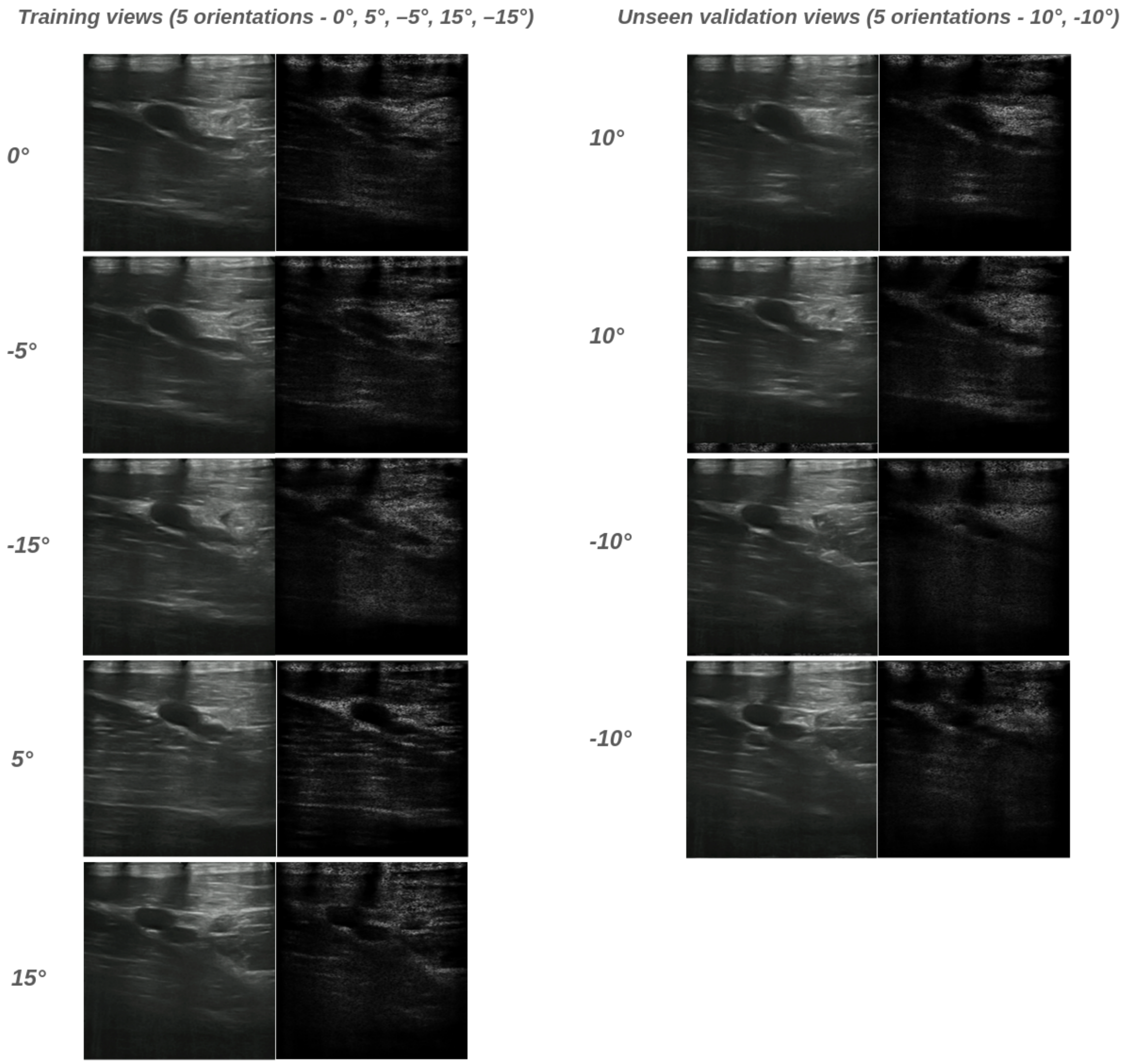
Training views (5 orientations - 0°, 5°, –5°, 15°, –15°)
Unseen validation views (5 orientations - 10°, -10°)
0°
-5°
-15°
5°
15°
10°
10°
-10°
-10°

Figure 5.6: The ground truth images and their corresponding rendered images during training and validation.

Chapter 6

Implicit Shape Representation for Anatomy-specific 3D Reconstruction

6.1 Literature

There is rich history of using implicit voxel-based representations for reconstructing 3D objects [55], [17], [8]. In non-medical applications, surface-based shape representations are being preferred [63] over volume-based shape representations like voxel occupancy grids. This is primarily due to the graphical requirement for high resolution representations, especially for objects with complex geometries that appear more choppy with voxel grids. However, most medical applications and anatomical reconstructions commonly use voxel-based representations. The closest example to our work [8], shows how implicit representations can help with anatomical reconstructions when only sparse labels are available. Most implicit shape representation methods, even [8], use shape embeddings to reconstruct multiple 3D shapes during inference.

6.2 Method

The key difference of my approach from prior voxel-based shape representation networks is that I apply the reconstruction to a single subject/scene/geometry at a time. This means that my method does not require shape embeddings or any prior understanding of the geometry of the anatomy that is to be reconstructed. The 3D shape is conditioned on the TRF parameters.

Three layers added to the Ultra-NeRF network represent the implicit shape representation. These layers are not a part of the core network learning the TRF. For this chapter, all training has been done without including contributions from chapter 5 and chapter 4. As seen in fig. 6.1, this module takes as input the (x, y, z) coordinates of (r, t) locations, appends the parameters of the TRF, namely attenuation, reflectance, scattering density and scattering amplitude and predicts a 0/1 occupancy per (r, t) location. The idea is to leverage the homogeneous acoustic properties of tissue learned by our INR.

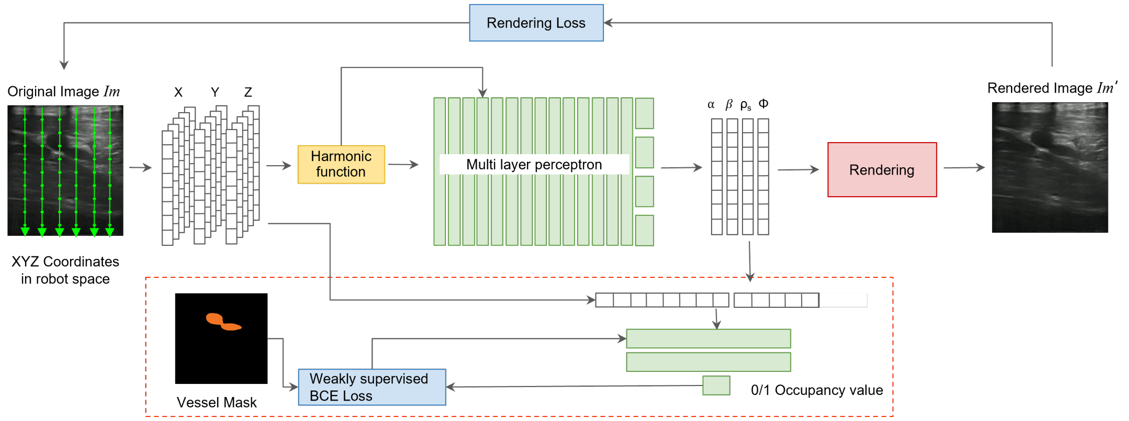

Figure 6.1: Network architecture for learning anatomy-specific implicit shape representation while learning the NIR.

The network is trained jointly for rendering as well as the occupancy prediction. As this contribution uses the original Ultra-NeRF implementation, it is extensible to any Ultra-NeRF-like implementation that learns a TRF. The shape representation is trained in a weakly-supervised manner. A Binary Cross Entropy loss (eq. (6.1)) computed with predicted 2D occupancy maps and sparse ground truth segmentation masks of vessels, is applied to the network.

$$\mathcal{L}_{\text{Occ}} = \mathcal{L}_{\text{BCEwithLogits}}(\text{Occ}_{Pr}, \text{Occ}_{GT}) \tag{6.1}$$

The method is a dual optimization on two objectives - 1) learn the TRF that produces the best rendered images and 2) learn the occupancy function that best matches the ground truth segmentation masks. As the occupancy prediction is for a region within the observed volume, the optimization for the occupancy acts like an attention mechanism in an indirect way, forcing the NIR to learn to render the region corresponding the masks first. This could result in some incorrectly learned TRF parameters as it breaks from the depth-wise learning in ray propagation model. Additionally, if the network focuses too much on the occupancy prediction, it is likely that the TRF learned is one that facilitates the best occupancy function, and is not representative of the actual underlying tissue properties. More than one unique set of TRF parameters exist that can render the same image. There is currently no existing mechanism to ensure that a non-degenerate solution to the TRF isn't learned.

However, to ensure that the occupancy prediction does not corrupt the TRF function learned by the NIR, we use a warm up mechanism within a single optimizer training framework instead of using a memory-intensive and complex two optimizer framework. The individual losses for rendering and occupancy prediction are weighted and the weight for rendering loss gradually decreases from 1 to 0.5 while the weight for the occpuancy loss increases from 0 to 0.5. Therefore, the loss for the shape representation training is given by eq. (6.2).

$$\mathcal{L}_{\text{Total}} = \gamma\mathcal{L}_{\text{Render}} + (1 - \gamma)\mathcal{L}_{\text{Occ}} \tag{6.2}$$

6.3 Experiments and Results

We evaluate the reconstructed vessels for a 3D Dice score which is a volumetric measure of reconstruction accuracy. The results for synthetic, blue-gel and live-pig2 data are provided at three levels of supervision in table 6.1 and the visualizations are provided in fig. 6.2. The visualizations for the synthetic data is created in a 500x250x500 grid, whereas the visualizations for the remaining two datasets are created in 500x500x500 grids. A high overlap in the predicted shape in blue with

Dataset	5% Supervision	10% Supervision	20% Supervision
Synthetic	0.741	0.857	0.867
Blue-gel	0.852	0.904	0.915
Live-pig2	0.708	0.801	0.830

Table 6.1: 3D Dice scores for vessel reconstructions obtained with the proposed method against ground truth voxel occupancy maps.

the ground truth is red is observed. As expected, as the supervision reduces, the performance decreases. In the case of synthetic and blue-gel data, we can recover the ground truth voxel occupancy map from the CT scan. However, for pig data, as CT scans were unavailable, voxel occupancy maps are created through dense segmentation and compounding, on originally acquired frames as well new intermediate frames queried from the NIR.

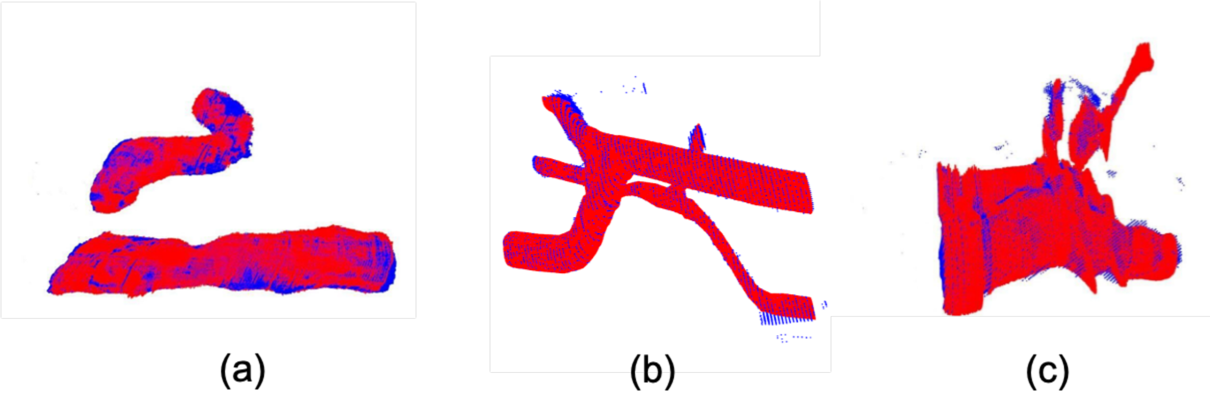

Figure 6.2: (a) Reconstructed vessel from synthetic dataset, trained with 10% supervision; (b) Reconstructed vessel from blue-gel dataset, trained with 10% supervision; (c) Reconstructed vessel from live-pig2 dataset, trained with 10% supervision. Here red represents the ground truth and blue represents the learned implicit shape representation.

The implicit shape representation is unique in the sense that it provides a continuous 3D volume of the vessel which cannot be acquired directly by any standard ultrasound method. 2D U-Net/similar methods yield 2D segmentation masks (no 3D information). 3D U-Net/similar methods are computationally expensive and yield segmented volumes but these inherently contain gaps from ultrasound voxel compounding (discontinuous). Both 2D U-Net and 3D U-Net are heavily supervised

Dataset	IoU from U-Net	IoU from implicit shape representation
Live-pig2 (10% supervision)	0.8011	0.8868
Live-pig2 (20% supervision)	0.7934	0.8914

Table 6.2: 2D Dice Scores for U-Net and the proposed method.

methods (100s of images at the very least) whereas this new implicit method is weakly supervised on a handful of frames. There does not exist a truly apples-to-apples comparison against standard baseline methods. The current best way to validate the proposed method is in 2D against a pretrained 2D U-Net (on pig ultrasound images), fine-tuned on the dataset in consideration, but only on the same frames that the NIR used. I performed this comparison for the live-pig2 dataset. The 2D Dice scores from the implicit shape representation outperformed the weakly-supervised fine-tuned U-Net (see table 6.2), proving that the proposed method does better than a traditional 2D U-Net when using so little subject-specific labeled data.

Chapter 7

3D Soft Tissue Deformation in Simulation

7.1 Motivation

Ultrasound imaging requires the application of a significant force to maintain contact between the probe and the subject, and in some cases, improve the quality of imaging. An undesirable effect is that the applied force deforms the elastic tissue and can cause vessel collapses or lateral displacement termed rolling. While an in-plane needle insertion (needle fully visible in the 2D ultrasound plane) into a vessel under a constant force is feasible with ultrasound, insertions out of the the ultrasound plane (needle partially or fully out of the 2D ultrasound plane) run the risk of missing the vessels as the localized vessel centers might shift due to deformations. In such a scenario, having an estimate of the 3D vessel deformation can turn a blind insertion into an informed insertion.

Traditionally, simulators using the Finite Element Method (FEM) have been used to estimate deformations of 3D structures under applied forces. Porting this method to concealed anatomy such as vessels comes with its own set of challenges : 1. The material properties of the tissue are not known exactly; 2. Tissue is typically non-homogeneous material; 3. The 3D shape of the entire vessel cannot be obtained easily; 4. How do we ascertain that the simulation is realistic? Additionally, FEM

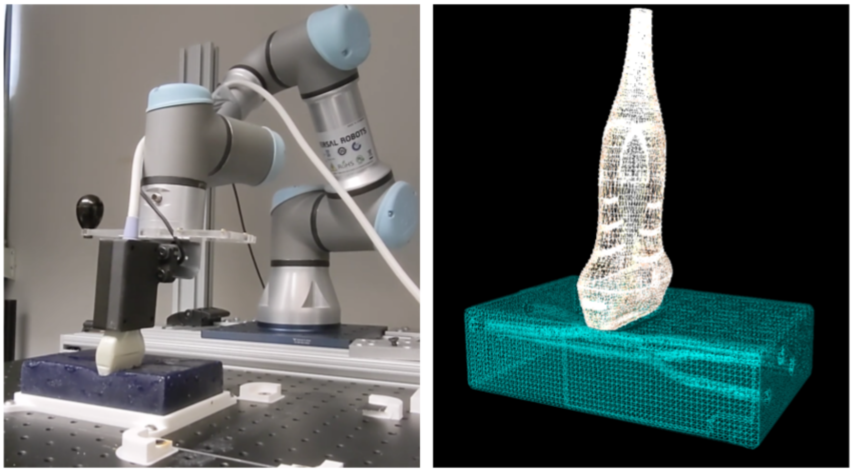

Figure 7.1: Left: Robotic system; Right: The registered simulation setup.

simulation, even with fast simulators utilizing GPU compute, is time-consuming and cannot be applied in real time for high resolutions.

Inference with neural networks can be much faster than FEM simulation [54]. Networks that process 3D representations such as point clouds and meshes can be conditioned on forces to predict deformed shapes [85], effectively mimicking FEM simulation. While such networks may not learn the underlying physics of deformation with the highest precision, they have shown to produce realistic deformations. Few such networks exist in the domain of medical data, perhaps due to the difficulties in obtaining training data from concealed anatomy and using 2D imaging to observe 3D deformations. Therefore realistic deformation data generation is a key challenge.

7.2 Method

I present a method to utilize 2D tomographic medical imaging for building and calibrating a simplified FEM simulation of 3D vessel structures. This simulation mimics vessel compression due to forces from an ultrasound probe. The material properties in calibration are estimated by using an optimization for the Young's modulus and Poisson's ratio. This is achieved by maximizing the IoU area between the deformed vessels observed in real-world ultrasound images captured by a robot and vessel cross-section from the simulated deformed model. This method can be used to generate 3D training data for neural networks being developed to predict deformations.

7.2.1 3D Model Generation

Slices of the CT volume of the phantom are segmented through pixel thresholding in 3DSlicer [23] for labeling vessels. The vessel labels are manually rectified and propagated through the length of the CT volume to obtain a hollow triangular mesh. This mesh is uniformly downsampled with Blender [2] to ensure vertices are uniformly distributed with low distances between vertices. FEM simulations could fail with meshes that have sparse vertices placed far apart. This downsampled mesh is then converted to a volumetric tetrahedral mesh using Gmsh [26], with the vessel region not having any connectivity. The tetrahedral connectivity is only for the region of the phantom that mimics tissue.

For the porcine subjects, I previously (before Ultra-NeRF) employed a U-Net-based segmentation model to obtain masks for vessels from ultrasound images. These masks are then stacked by robot pose to create a solid volume of the vessel. This volume is processed through marching cubes to obtain a hollow mesh which is then fused with an artificially-added outer mesh that represents tissue. The resulting mesh is tetrahedralized for results similar to the phantom model. Figure 7.2 shows the medical image data and the generated 3D models of the phantom and the porcine vessels.

This approach to building 3D models is very tedious, requires manual corrections, and often results in non-smooth meshes with concavities and discontinuities due to the presence of gaps in the ultrasound volume. Both triangular and tetrahedral mesh fitting algorithms behave weirdly for such data. The models from chapter 6 are much smoother and seldom have gaps. Moving ahead, the implicit shape presentation can be a very effective means of 3D model generation, especially when CT data is unavilable or costly.

7.2.2 Simulation

I used the SOFA library [7] for simulations and parameterize the simulation in a simplistic manner. It is assumed that material density, Young's modulus E and

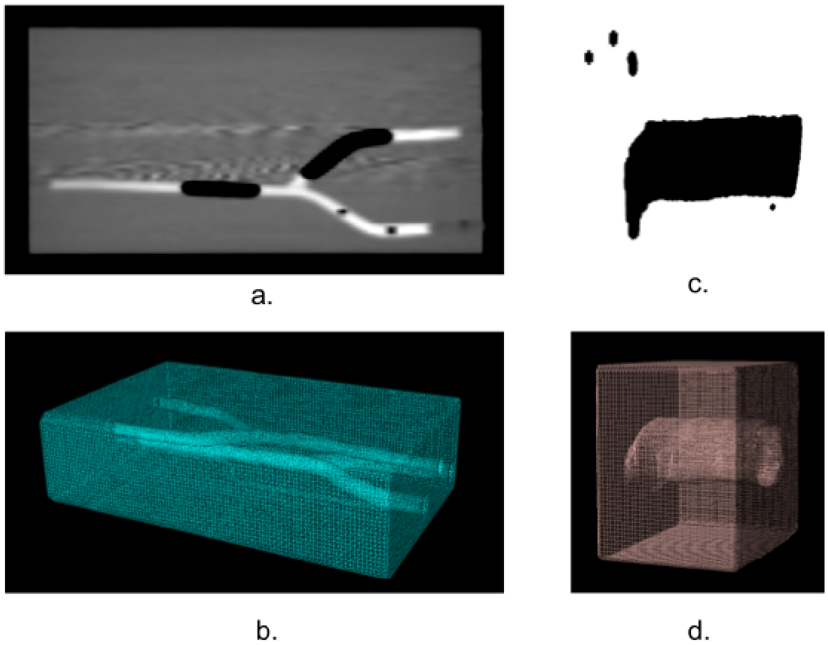

Figure 7.2: a. The CT volume of the blue-gel phantom; b. The tetrahedral mesh of the blue-gel phantom; c. The stacked ultrasound volume of a pig vessel; d. The tetrahedral mesh of the pig vessel

Poisson's ratio ν are the primary factors affecting the response of a model to applied forces. After defining the correct transformations that register the simulation frame to the robot frame, the 3D tetrahedral model is loaded for simulation. Downward gravity is defined and all the points at the base of the model are fixed. By defining simulation parameters such as material properties, force and application direction, I simulate the probe-phantom interaction resulting in vessel deformations.

7.2.3 Calibration with IoU

In this setup, the assumption is that the exact material properties of the subject are unknown. While the material properties for the phantom are known as it is standard equipment, this information has to be estimated for the tissue of every new subject that we would want to perform deformation simulation for. It then becomes necessary to apply some form of per-subject estimation. In this case, the estimation is achieved through an optimization for E and ν of the material by maximising the overlap of the vessel area between the ultrasound images and the corresponding cross-sections from generated simulations. This is performed for the model of the phantom for which the material density was known.

The idea is that since the simulation frame is registered to the robot, the cross section from the deformed model at a given pose p and force f, along the image plane, should resemble the vessel anatomy seen in the ultrasound image collected by the robot at p and force f. The IoU between vessel regions is used to determine the overlap and is used as a reward in a Cross-Entropy Maximization method.

7.3 Experiments and Results

An iterative Cross Entropy Maximization method [22] for optimization is applied with 8 agents, sampling values for E and ν from known ranges for the gel material of the phantom. 2 agents are purely exploratory and the ranges are 60-850 kPa [16] and 0.47-0.49 for E and ν. The optimization was considered converged when the standard deviation was lower than 5 and 0.005 for E and ν respectively. With the addition of scattering agents to the polymers used for manufacturing the blue-gel phantom, we expect a slight deviation from the expected 600 KPa and 0.48 values for both the parameters. The optimization is performed at two poses with 3 different force values (6, 8 and 10 N) and averaged over them for obtaining the final calibrated material properties. All 6 optimizations converged within 10 epochs.

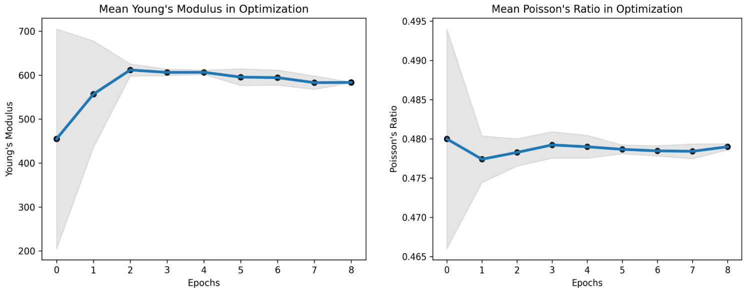

Figure 7.3: The mean E (left) and ν (right) values (with standard deviations in grey) after each epoch of the optimization for pose 1, force 8 N.

The graph showing the convergence of the optimization for the blue-gel phantom model is shown in Fig 7.3. The average converged values for Young's Modulus and Poisson's Ratio are 592.5433 ± 20.13 kPa and 0.482 ± 0.003. The highest IoU at the poses where the optimization was applied was 0.76. The simulation and optimization results for the blue-gel phantom showing the contours from the cross section of the deformed mesh progressively aligning better with the vessel masks from the ultrasound data are seen in Fig 7.4.

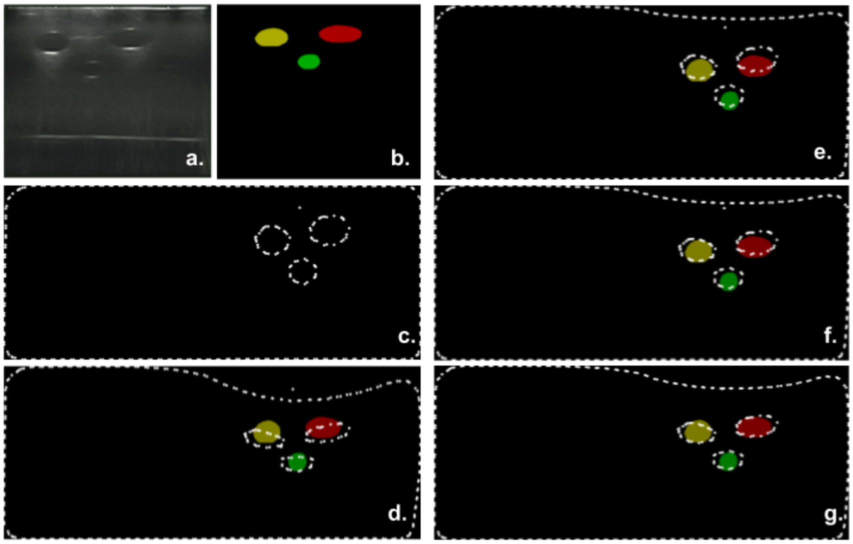

Figure 7.4: a. The ultrasound image captured by the robot at pose 1 with force 6 N; b. Vessel masks identified in the ultrasound image by our segmentation model; c. The cross-section contours from the undeformed mesh at pose 1. d-f. Cross-sections from the deformed phantom model with the registered vessel masks at pose 1. The cross-sections were generated with applied force 6 N, simulated with material properties estimated in epochs 0, 2, 6 and 8 of the optimization respectively.

7.4 Discussion

It is observed that after optimization, the highest IoU score in simulations across all the poses at three different forces is 0.79 with the average being 0.72. This shows that the parameter estimates work uniformly at all poses over the homogeneous phantom. The gap in the IoU from the perfect score of 1 can be attributed to some of the assumptions made in our simplified model and to discrepancies arising from obtaining vessel masks in both CT and ultrasound data. If ultrasound images are used to generate the 3D model using chapter 6, then the need for cross-model CT-ultrasound registration would be eliminated. No simulation is performed for fluid inside the vessels and the underlying assumption is that the Young's Modulus and Poisson's

Ratio are sufficient to model soft tissue properties. The vessel masks obtained either through thresholding or segmentation are prone to the usual noise in the data and to the robustness of the segmentation model.

Work presented in this chapter was accepted as a workshop paper for ICRA Workshop on Representing and Manipulating Deformable Objects.

Chapter 8

Conclusions

Prior methods for reconstructing ultrasound volumes (and the organs within) suffer from a number of challenges that I have addressed. My advanced method contributes significantly to the previous state of the art in neural implicit representations for ultrasound - Ultra-NeRF.

In chapter 4, I discuss modifying the architecture of the MLP network derived from NeRF. With the proposed changed, Ultra-NeRF is able to represent complex real-world tissue volumes from live subjects. However, the assumed PSF that Ultra-NeRF employs results in rendered images which have very different speckle from ground truth images. To remedy this and better define the image-formation parameters, I learn per-point Point Spread Functions, which is a property specific to an ultrasound imaging system. As a result, the rendered images are more photo-realistic. In the chapter 5, I propose a novel approach to learn viewing-direction dependence within the NIR, in order to render true-to-orientation images. In chapter 6, I present a method to learn an implicit shape representation for homogeneous tissue like vessels using the acoustic physics parameters the NIR learns for the observed tissue. Finally, linking the NIR work to my prior research in the direction of 3D deformation simulation for soft tissue, I propose the implicit shape representation as a better 3D model generation method for simulations.

8.1 Future Work

In chapter 4, I show that learning the PSF helps render better ultrasound images and mention that my learned resolution values show very little variation along the depth. According to the physics of ultrasound beam-formation, the variation in the PSF can help determine a general focus depth and beam shape. More research needs to be done into how to regularize the learned values to depict the beam shape. For incorporating view-dependence, in chapter 5, I show that the method does not work for deformable scenes. For this purpose, some form of deformation-aware non-rigid 3D registration is required for building one coherent volume. While the voxel occpuancy based implicit shape representation is useful, for many applications such as VR, deformation simulations and even robotic path planning for surgical interventions, having a surface representation is more valuable. Perhaps, SDFs can be learnt within this NIR framework. Finally, the 3D tissue model generation for deformation simulation needs to be tested with the models reconstructed from my proposed NIR method.

Bibliography

[1] URL https://www.imfusion.com/products/imfusion-suite. 3.1

[2] *Blender - a 3D modelling and rendering package.* 7.2.1

[3] FNU Abhimanyu, Andrew L. Orekhov, Ananya Bal, John Galeotti, and Howie Choset. Unsupervised deformable ultrasound image registration and its application for vessel segmentation, 2023. 1.1

[4] Dror Aiger and Daniel Cohen-Or. Real-time ultrasound imaging simulation. *Real-Time Imaging*, 4(4):263–274, 1998. 2.2.1

[5] O.A. Ajilisa, V.P. Jagathy Raj, and M.K. Sabu. Segmentation of thyroid nodules from ultrasound images using convolutional neural network architectures. *J. Intell. Fuzzy Syst.*, 43(1):687–705, jan 2022. ISSN 1064-1246. doi: 10.3233/JIFS-212398. 1.1

[6] Nathan Albin, Oscar P Bruno, Theresa Y Cheung, and Robin O Cleveland. Fourier continuation methods for high-fidelity simulation of nonlinear acoustic beams. *The Journal of the Acoustical Society of America*, 132(4):2371–2387, 2012. 2.2.3

[7] Jérémie Allard, Stéphane Cotin, François Faure, Pierre-Jean Bensoussan, François Poyer, Christian Duriez, Hervé Delingette, and Laurent Grisoni. Sofa-an open source framework for medical simulation. In *MMVR 15-Medicine Meets Virtual Reality*, volume 125, pages 13–18. IOP Press, 2007. 7.2.2

[8] Tamaz Amiranashvili, David Lüdke, Hongwei Bran Li, Bjoern Menze, and Stefan Zachow. Learning shape reconstruction from sparse measurements with neural implicit functions. In *International Conference on Medical Imaging with Deep Learning*, pages 22–34. PMLR, 2022. 6.1

[9] Wenjia Bai, Ozan Oktay, Matthew Sinclair, Hideaki Suzuki, Martin Rajchl, Giacomo Tarroni, Ben Glocker, Andrew King, Paul M Matthews, and Daniel Rueckert. Semi-supervised learning for network-based cardiac mr image segmentation. In *Medical Image Computing and Computer-Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part II 20*, pages 253–260. Springer, 2017. 2.1.2

[10] Lennart Bargsten and Alexander Schlaefer. Specklegan: a generative adversarial network with an adaptive speckle layer to augment limited training data for ultrasound image processing. *International journal of computer assisted radiology and surgery*, 15:1427–1436, 2020. 2.2.2

[11] Pasu Boonvisut and M Cenk Çavuşoğlu. Estimation of soft tissue mechanical parameters from robotic manipulation data. *IEEE/ASME Transactions on Mechatronics*, 18(5):1602–1611, 2012. 2.4

[12] 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. 1.1, 2.2.3, 4.4

[13] Enrico Checcucci, Angela Pecoraro, Daniele Amparore, Sabrina De Cillis, Stefano Granato, Gabriele Volpi, Michele Sica, Paolo Verri, Alberto Piana, Pietro Piazzolla, et al. The impact of 3d models on positive surgical margins after robot-assisted radical prostatectomy. *World Journal of Urology*, 40(9):2221–2229, 2022. 1.1

[14] Chuan Chen, Hendrik HG Hansen, Gijs AGM Hendriks, Jan Menssen, Jian-Yu Lu, and Chris L de Korte. Point spread function formation in plane-wave imaging: A theoretical approximation in fourier migration. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 67(2):296–307, 2019. 1.2

[15] Xiankang Chen, Tiexiang Wen, Xingmin Li, Wenjian Qin, Donglai Lan, Weizhou Pan, and Jia Gu. Reconstruction of freehand 3d ultrasound based on kernel regression. *Biomedical engineering online*, 13(1):1–15, 2014. 2.1.1

[16] Yao Chen, Jow-Lian Ding, Mahdieh Babaiasl, Fan Yang, and John P Swensen. Characterization and modeling of a thermoplastic elastomer tissue simulant under uniaxial compression loading for a wide range of strain rates. *Journal of the Mechanical Behavior of Biomedical Materials*, 131:105218, 2022. 7.3

[17] Zhiqin Chen and Hao Zhang. Learning implicit fields for generative shape modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5939–5948, 2019. 6.1

[18] Vahid Ashkani Chenarlogh, Mostafa Ghelich Oghli, Ali Shabanzadeh, Nasim Sirjani, Ardavan Akhavan, Isaac Shiri, Hossein Arabi, Morteza Sanei Taheri, and Mohammad Kazem Tarzamni. Fast and accurate u-net model for fetal ultrasound image segmentation. *Ultrasonic Imaging*, 44(1):25–38, 2022. doi: 10.1177/01617346211069882. PMID: 34986724. 1.1

[19] 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. 2.3.2

[20] Pierrick Coupé, Pierre Hellier, Xavier Morandi, and Christian Barillot. Probe trajectory interpolation for 3d reconstruction of freehand ultrasound. *Medical image analysis*, 11(6):604–615, 2007. 1.1

[21] Jessica C Dai, Michael R Bailey, Mathew D Sorensen, and Jonathan D Harper. Innovations in ultrasound technology in the management of kidney stones. *Urologic Clinics*, 46(2):273–285, 2019. 1.1

[22] Lih-Yuan Deng. The cross-entropy method: A unified approach to combinatorial optimization, monte-carlo simulation, and machine learning. *Technometrics*, 48 (1):147–148, 2006. doi: 10.1198/tech.2006.s353. 7.3

[23] Andriy Fedorov, Reinhard Beichel, Jayashree Kalpathy-Cramer, Julien Finet, Jean-Christophe Fillion-Robin, Sonia Pujol, Christian Bauer, Dominique Jennings, Fiona Fennessy, Milan Sonka, and et al. 3d slicer as an image computing platform for the quantitative imaging network. *Magnetic Resonance Imaging*, 30 (9):1323–1341, 2012. doi: 10.1016/j.mri.2012.05.001. 7.2.1

[24] Cong Gao, Xingtong Liu, Michael Peven, Mathias Unberath, and Austin Reiter. Learning to see forces: Surgical force prediction with rgb-point cloud temporal convolutional networks. In *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis: First International Workshop, OR 2.0 2018, 5th International Workshop, CARE 2018, 7th International Workshop, CLIP 2018, Third International Workshop, ISIC 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16 and 20, 2018, Proceedings 5*, pages 118–127. Springer, 2018. 2.4

[25] Kyle Genova, Forrester Cole, Daniel Vlasic, Aaron Sarna, William T Freeman, and Thomas Funkhouser. Learning shape templates with structured implicit functions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7154–7164, 2019. 2.3.1

[26] Geuzaine, Christophe and Remacle, Jean-Francois. Gmsh. URL `http://http://gmsh.info/`. 7.2.1

[27] Orcun Goksel and Septimiu E Salcudean. B-mode ultrasound image simulation in deformable 3-d medium. *IEEE transactions on medical imaging*, 28(11): 1657–1669, 2009. 2.2.1

[28] Ang Nan Gu, Purang Abolmaesumi, Christina Luong, and Kwang Moo Yi. Representing 3d ultrasound with neural fields. In *Medical Imaging with Deep Learning*, 2022. 2.3.4

[29] Shuxiang Guo, Xiaojuan Cai, and Baofeng Gao. A tensor-mass method-based vascular model and its performance evaluation for interventional surgery virtual reality simulator. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 14(6):e1946, 2018. 2.4

[30] Lianghao Han, John H Hipwell, Christine Tanner, Zeike Taylor, Thomy Mertzanidou, Jorge Cardoso, Sebastien Ourselin, and David J Hawkes. Development of patient-specific biomechanical models for predicting large breast deformation. *Physics in Medicine & Biology*, 57(2):455, 2011. 2.4

[31] W.R. Hedrick, D.L. Hykes, and D.E. Starchman. *Ultrasound Physics and Instrumentation*. Ultrasound Physics and Instrumentation. Elsevier Mosby, 2005. ISBN 9780323032124. 5.1

[32] IM Heer, K Middendorf, S Müller-Egloff, Martin Dugas, and A Strauss. Ultrasound training: the virtual patient. *Ultrasound in Obstetrics and Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 24(4):440–444, 2004. 2.2.1

[33] Yipeng Hu, Eli Gibson, Li-Lin Lee, Weidi Xie, Dean C Barratt, Tom Vercauteren, and J Alison Noble. Freehand ultrasound image simulation with spatially-conditioned generative adversarial networks. In *Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment: Fifth International Workshop, CMMI 2017, Second International Workshop, RAMBO 2017, and First International Workshop, SWITCH 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, 2017, Proceedings 5*, pages 105–115. Springer, 2017. 2.2.2

[34] QH Huang, YP Zheng, MH Lu, and ZR Chi. Development of a portable 3d ultrasound imaging system for musculoskeletal tissues. *Ultrasonics*, 43(3):153–163, 2005. 2.1.1

[35] Khadija Idrissu, Sylwia Malec, and Alessandro Crimi. 3d reconstructions of brain from mri scans using neural radiance fields. *bioRxiv*, pages 2023–04, 2023. 2.3.2

[36] Jørgen Arendt Jensen. Field: A program for simulating ultrasound systems. *Medical & Biological Engineering & Computing*, 34(sup. 1):351–353, 1997. 2.2.3

[37] Jørgen Arendt Jensen. Simulation of advanced ultrasound systems using field ii. In *2004 2nd IEEE International Symposium on Biomedical Imaging: Nano to Macro (IEEE Cat No. 04EX821)*, pages 636–639. IEEE, 2004. 2.2.3

[38] Jørgen Arendt Jensen and I Nikolov. Fast simulation of ultrasound images. In *2000 IEEE Ultrasonics Symposium. Proceedings. An International Symposium (Cat. No. 00CH37121)*, volume 2, pages 1721–1724. IEEE, 2000. 2.2.3

[39] Jørgen Arendt Jensen and Niels Bruun Svendsen. Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 39(2):262–267, 1992. 2.2.3

[40] Antônio Sousa Vieira De Carvalho Junior and Helton Hideraldo Bíscaro. Blood

flow sph simulation with elastic deformation of blood vessels. In *2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)*, pages 532–538. IEEE, 2019. 2.4

[41] Daeun Kang, Jaeseok Moon, Saeyoung Yang, Taesoo Kwon, and Yejin Kim. Physics-based simulation of soft-body deformation using rgb-d data. *Sensors*, 22 (19):7225, 2022. 2.4

[42] Athanasios Karamalis, Wolfgang Wein, and Nassir Navab. Fast ultrasound image simulation using the westervelt equation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 243–250. Springer, 2010. 2.2.3

[43] Muhammad Osama Khan and Yi Fang. Implicit neural representations for medical imaging segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part V*, pages 433–443. Springer, 2022. 2.3.2

[44] Kim Sunwoo Yoon Joon Shik Baek Seungjun Kim Beom Suk, Yu Minhyeong. Scale-attentional u-net for the segmentation of the median nerve in ultrasound images. *Ultrasonography*, 41(4):706–717, 2022. doi: 10.14366/usg.21214. 1.1

[45] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 4.4, 5.2

[46] Oliver Kutter, Ramtin Shams, and Nassir Navab. Visualization and gpu-accelerated simulation of medical ultrasound from ct images. *Computer methods and programs in biomedicine*, 94(3):250–266, 2009. 2.2.3

[47] Keyu Li, Yangxin Xu, and Max Q-H Meng. An overview of systems and techniques for autonomous robotic ultrasound acquisitions. *IEEE Transactions on Medical Robotics and Bionics*, 3(2):510–524, 2021. 1.1

[48] Yuexiang Li, Jiawei Chen, Xinpeng Xie, Kai Ma, and Yefeng Zheng. Self-loop uncertainty: A novel pseudo-label for semi-supervised medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part I 23*, pages 614–623. Springer, 2020. 2.1.2

[49] Cesare Magnetti, Veronika Zimmer, Nooshin Ghavami, Emily Skelton, Jacqueline Matthew, Karen Lloyd, Jo Hajnal, Julia A Schnabel, and Alberto Gomez. Deep generative models to simulate 2d patient-specific ultrasound images in real time. In *Annual Conference on Medical Image Understanding and Analysis*, pages 423–435. Springer, 2020. 2.2.2

[50] Nils Marahrens, Bruno Scaglioni, Dominic Jones, Raj Prasad, Chandra Shekhar Biyani, and Pietro Valdastri. Towards autonomous robotic minimally invasive

ultrasound scanning and vessel reconstruction on non-planar surfaces. *Frontiers in Robotics and AI*, page 178, 2022. 1.1

[51] Frederick A Masoudi, Angelo Ponirakis, Robert W Yeh, Thomas M Maddox, Jim Beachy, Paul N Casale, Jeptha P Curtis, James De Lemos, Gregg Fonarow, Paul Heidenreich, et al. Cardiovascular care facts: a report from the national cardiovascular data registry: 2011. *Journal of the American College of Cardiology*, 62(21):1931–1947, 2013. 1.1

[52] Oliver Mattausch and Orcun Goksel. Image-based psf estimation for ultrasound training simulation. In *Simulation and Synthesis in Medical Imaging: First International Workshop, SASHIMI 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 21, 2016, Proceedings 1*, pages 23–33. Springer, 2016. 4.3, 4.4

[53] Giovanni Mauri, Luigi Solbiati, Franco Orsi, and Lorenzo Monfardini. Thermal ablation of liver tumours: the crucial role of 3d imaging. *CardioVascular and Interventional Radiology*, 43:1416–1417, 2020. 1.1

[54] Andrea Mendizabal, Eleonora Tagliabue, Jean-Nicolas Brunet, Diego Dall'Alba, Paolo Fiorini, and Stéphane Cotin. Physics-based deep neural network for real-time lesion tracking in ultrasound-guided breast biopsy. In *Computational Biomechanics for Medicine: Solid and Fluid Mechanics for the Benefit of Patients 22*, pages 33–45. Springer, 2020. 2.4, 7.1

[55] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4460–4470, 2019. 6.1

[56] Monique Meuschke, Samuel Voss, Oliver Beuing, Bernhard Preim, and Kai Lawonn. Combined visualization of vessel deformation and hemodynamics in cerebral aneurysms. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):761–770, 2017. doi: 10.1109/TVCG.2016.2598795. 2.4

[57] Mateusz Michalkiewicz, Jhony K. Pontes, Dominic Jack, Mahsa Baktashmotlagh, and Anders Eriksson. Implicit surface representations as layers in neural networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019. 2.3.1

[58] 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. 2.3.1

[59] Ashkan Mirzaei, Yash Kant, Jonathan Kelly, and Igor Gilitschenski. Laterf: Label and text driven object radiance fields. In *European Conference on Computer*

Vision, pages 20–36. Springer, 2022. 2.3.4

[60] Farhan Mohamed and C Vei Siang. A survey on 3d ultrasound reconstruction techniques. *Artificial Intelligence—Applications in Medicine and Biology*, pages 73–92, 2019. 1.1, 2.1.1

[61] Cecilia G Morales, Jason Yao, Tejas Rane, Robert Edman, Howie Choset, and Artur Dubrawski. Reslicing ultrasound images for data augmentation and vessel reconstruction. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2710–2716. IEEE, 2023. 1.1, 2.2.1

[62] MC Murphy, B Gibney, C Gillespie, J Hynes, and F Bolster. Gallstones top to toe: what the radiologist needs to know. *Insights into Imaging*, 11(1):1–14, 2020. 1.1

[63] 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. 6.1

[64] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. 3.3

[65] Albert W Reed, Hyojin Kim, Rushil Anirudh, K Aditya Mohan, Kyle Champley, Jingu Kang, and Suren Jayasuriya. Dynamic ct reconstruction from limited views with implicit neural representations and parametric motion fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2258–2268, 2021. 2.3.2

[66] Robert Rohling, Andrew Gee, and Laurence Berman. A comparison of freehand three-dimensional ultrasound reconstruction techniques. *Medical image analysis*, 3(4):339–359, 1999. 2.1.1

[67] Lucien Roquette, Matthieu Simeoni, Paul Hurley, and Adrien Besson. On an analytical, spatially-varying, point-spread-function. In *2017 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4. Ieee, 2017. 1.2

[68] 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 *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference,*

Munich, Germany, October 5-9, 2015, Proceedings, Part II 18, pages 510–518. Springer, 2015. 2.2.3

[69] João M Sanches and Jorge S Marques. A rayleigh reconstruction/interpolation algorithm for 3d ultrasound. *Pattern recognition letters*, 21(10):917–926, 2000. 2.1.1

[70] 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. 2.3.2

[71] Ramtin Shams, Richard Hartley, and Nassir Navab. Real-time simulation of medical ultrasound from ct images. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008: 11th International Conference, New York, NY, USA, September 6-10, 2008, Proceedings, Part II 11*, pages 734–741. Springer, 2008. 2.2.3

[72] Hongjian Shi. *Finite element modeling of soft tissue deformation.* University of Louisville, 2007. 1.1

[73] Moh Nur Shodiq, Eko Mulyanto Yuniarno, Johanes Nugroho, and I Ketut Eddy Purnama. Ultrasound image segmentation for deep vein thrombosis using unet-cnn based on denoising filter. In *2022 IEEE International Conference on Imaging Systems and Techniques (IST)*, pages 1–6, 2022. doi: 10.1109/IST55454.2022.9827731. 1.1

[74] Vincent Sitzmann, Michael Zollhöfer, and Gordon Wetzstein. Scene representation networks: Continuous 3d-structure-aware neural scene representations. *Advances in Neural Information Processing Systems*, 32, 2019. 2.3.1

[75] Yuki Sugano, Shinya Onogi, Antoine Bossard, Takashi Mochizuki, and Kohji Masuda. Development of a 3d reconstruction of blood vessel by positional calibration of ultrasound probe. In *The 5th 2012 Biomedical Engineering International Conference*, pages 1–4. IEEE, 2012. 2.1.1

[76] Yu Sun, Jiaming Liu, Mingyang Xie, Brendt Wohlberg, and Ulugbek S Kamilov. Coil: Coordinate-based internal learning for imaging inverse problems. *arXiv preprint arXiv:2102.05181*, 2021. 4.5

[77] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems*, 30, 2017. 2.1.2

[78] Francis Tom and Debdoot Sheet. Simulating patho-realistic ultrasound images using deep generative networks with adversarial learning. In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, pages 1174–1177. IEEE, 2018. 2.2.2

[79] Janet Y Tsui, Adam B Collins, Douglas W White, Jasmine Lai, and Jeffrey A Tabas. Placement of a femoral venous catheter. *NEW ENGLAND JOURNAL OF MEDICINE*, 358(26):e30, 2008. 1.1

[80] Santiago Vitale, José Ignacio Orlando, Emmanuel Iarussi, and Ignacio Larrabide. Improving realism in patient-specific abdominal ultrasound simulation using cyclegans. *International journal of computer assisted radiology and surgery*, 15 (2):183–192, 2020. 2.2.2

[81] Suhani Vora, Noha Radwan, Klaus Greff, Henning Meyer, Kyle Genova, Mehdi SM Sajjadi, Etienne Pot, Andrea Tagliasacchi, and Daniel Duckworth. Nesf: Neural semantic fields for generalizable semantic segmentation of 3d scenes. *arXiv preprint arXiv:2111.13260*, 2021. 2.3.4

[82] Christian Wachinger and Nassir Navab. Alignment of viewing-angle dependent ultrasound images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 779–786. Springer, 2009. 5.1

[83] Guotai Wang, Shuwei Zhai, Giovanni Lasio, Baoshe Zhang, Byong Yi, Shifeng Chen, Thomas J Macvittie, Dimitris Metaxas, Jinghao Zhou, and Shaoting Zhang. Semi-supervised segmentation of radiation-induced pulmonary fibrosis from lung ct scans with multi-scale guided dense attention. *IEEE transactions on medical imaging*, 41(3):531–542, 2021. 2.1.2

[84] Yifan Wang, Tianyu Fu, Chan Wu, Jingfan Fan, Hong Song, Deqiang Xiao, Yucong Lin, Fangyi Liu, and Jian Yang. Adaptive tetrahedral interpolation for reconstruction of uneven freehand 3d ultrasound. *Physics in Medicine & Biology*, 68(5):055005, 2023. 1.1

[85] Zhihua Wang, Stefano Rosa, Bo Yang, Sen Wang, Niki Trigoni, and Andrew Markham. 3d-physnet: Learning the intuitive physics of non-rigid object deformations. *arXiv preprint arXiv:1805.00328*, 2018. 2.4, 7.1

[86] Wolfgang Wein, Shelby Brunke, Ali Khamene, Matthew R Callstrom, and Nassir Navab. Automatic ct-ultrasound registration for diagnostic imaging and image-guided intervention. *Medical image analysis*, 12(5):577–585, 2008. 5.1

[87] M.K. Welleweerd, A.G. de Groot, S.O.H. de Looijer, F.J. Siepel, and S. Stramigioli. Automated robotic breast ultrasound acquisition using ultrasound feedback. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9946–9952, 2020. doi: 10.1109/ICRA40945.2020.9196736. 1.1

[88] Jelmer M Wolterink, Jesse C Zwienenberg, and Christoph Brune. Implicit neural representations for deformable image registration. In *International Conference on Medical Imaging with Deep Learning*, pages 1349–1359. PMLR, 2022. 2.3.2

[89] 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. 2.3.2, 4.5

[90] Magdalena Wysocki, Mohammad Farid Azampour, Christine Eilers, Benjamin Busam, Mehrdad Salehi, and Nassir Navab. Ultra-nerf: Neural radiance fields for ultrasound imaging. In *MIDL*, 2023. 1.2, 2.3.3, 4.1, 4.4

[91] Junshen Xu, Daniel Moyer, Borjan Gagoski, Juan Eugenio Iglesias, P Ellen Grant, Polina Golland, and Elfar Adalsteinsson. Nesvor: Implicit neural representation for slice-to-volume reconstruction in mri. *IEEE Transactions on Medical Imaging*, 2023. 2.3.2

[92] 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. 2.3.2

[93] James A Zagzebski. Essentials of ultrasound physics. *(No Title)*, 1996. 5.1

[94] Nico Zettler and Andre Mastmeyer. Comparison of 2d vs. 3d u-net organ segmentation in abdominal 3d ct images. *arXiv preprint arXiv:2107.04062*, 2021. 1.1

[95] Lin Zhang, Tiziano Portenier, and Orcun Goksel. Learning ultrasound rendering from cross-sectional model slices for simulated training. *International Journal of Computer Assisted Radiology and Surgery*, 16:721–730, 2021. 2.2.2

[96] Jiahua Zhu, Yixian Su, Ziteng Liu, Bainan Liu, Yu Sun, Wenpeng Gao, and Yili Fu. Real-time biomechanical modelling of the liver using lightgbm model. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 18(6): e2433, 2022. 2.4