Caltech Summer Undergraduate Research Fellowship (SURF)

Lung Ultrasound Imaging with Neural Operators and Neural Networks for Practical Use

Armeet Singh Jatyani[1], Prof. Animashree Anandkumar[2], Dr. Jiayun Wang[2]

[1] *Applicant at California Institute of Technology, Pasadena, CA*
E–mail: *armeet@caltech.edu*

[2] *Mentors at Anima AI + Science Lab*
E–mail: *anima@caltech.edu, peterw@caltech.edu*

1. INTRODUCTION

1.1. Lung Ultrasound Imaging

Renowned for its lower cost, portability, and efficiency, lung ultrasound is a radiation-free practical alternative to chest radiography, particularly in critical care and emergency settings. Ultrasound tests can be administered bedside in under 3 minutes and diagnose asthma, chronic obstructive pulmonary disease, pneumonia, pneumothorax, pulmonary edema, and pulmonary embolism, often with better diagnostic accuracy than chest radiography methods Marini et al. (2021).

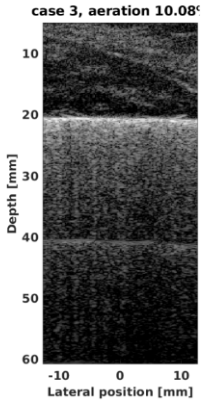

Fig. 1: Baseline: A chest ultrasound B-Mode image. Note the limited spatial resolution. Medical professionals rely on experience, pattern recognition, and comparisons with reference images to extract meaning from these images. The chest wall (0-20 mm depth) can be roughly distinguished from deeper lung tissue.

1.2. Baseline: B-Mode Imaging

In this project, we seek to improve upon the current state-of-the-art for ultrasound-based lung imaging: B-Mode imaging. In this technique, transducers at the skin surface send focused ultrasound beams directly into the body. These beams reflect off of soft tissue, bones, and air pockets inside the region of interest and are received by transducers Matrone et al. (2019). The received signals are processed to yield the B-Mode image for a single line. This process is repeated to cover the entire region of interest.

The following are key imaging features we are concerned with when evaluating and improving ultrasound medical imaging techniques Demi (2018). Spatial resolution refers to the smallest spatial distance in which scatterers (bones, tissue, air) can be discerned from each other. Temporal resolution refers to the time between samples. Contrast is the ease by which different scatterers are differentiated from one another. Finally, we are concerned with the depth of imaging. While techniques that can image deeper into the patient's lung are preferred, ultrasound pulses that travel deeper are absorbed and reflected more frequently, resulting in received signals that have been confounded with higher noise. Thus at greater depths, we are less certain about the accuracy of our reconstructed images.

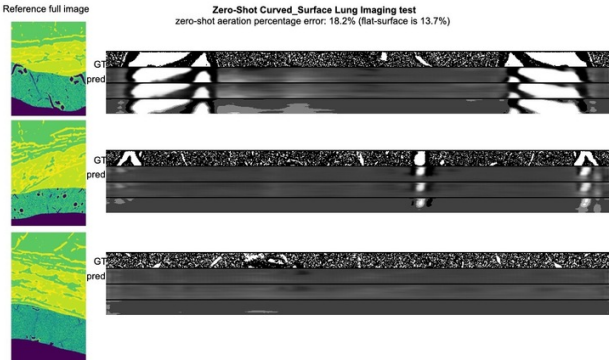

Fig. 2: Curved Chest Wall Inference. On the left, we see three distinct samples with curved lung chest walls. On the right, we see ground truth and model predictions for the 32 mm depth model.

2. OBJECTIVES

The objective of this research is to iterate and improve on the novel neural network-based approach (see 3) to ultrasound imaging, with the specific goal of adapting the method for practical use in clinical settings. In more detail...

1. Design, train, and fine-tune various neural network architectures to generate accurate lung density maps from raw ultrasound radio frequency data. Explore the effectiveness of Self-Supervised and Semi-Supervised Learning approaches, among other state-of-the-art machine-learning techniques to learn from limited unlabeled and labeled data.

2. Improve practical viability. Develop models that are invariant to curved lung chest walls of varying depths and ultrasound measurement configurations (see Figure 2). Explore domain adaptation techniques that would allow models to be deployed in practical settings Ben-David et al. (2006).

3. Implement certainty layers to reveal inference confidence at various chest depths. Evaluate inference certainty at varying depths to analyze the extent to which diffraction limits and instrument noise confound predictions.

4. Quantify performance and improvements in comparison to the current state-of-the-art: B-mode images.

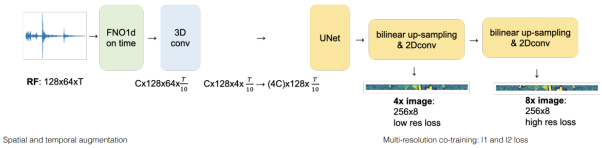

Fig. 3: Early Architecture. Raw ultrasound radio frequency data first passes through the 1D Fourier Neural Operator, is convolved to reduce dimension, and passes through the UNet to yield 4x and 8x resolution inferences. The model is trained for 60 epochs, with the AdamW optimizer and a composite loss function (L1 and L2 loss).

3. APPROACH

3.1. A Novel Approach: Neural Operator & Neural Network-Based Inverse Solver

3.1.1. Theoretical Basis

Equation 1 is derived from the Elastic Wave Equation and describes a model for the nonlinear propagation of waves in a heterogeneous medium. Three-dimensional solutions to the equation have been confirmed with experimental measurements Pinton et al. (2009). As such, the elastic wave equation is an effective model for ultrasound wave propagation in lung tissue.

$$\nabla^2\Psi - \frac{1}{c_0^2}\frac{\partial^2\Psi}{\partial t^2} + \frac{\delta}{c_0^4}\frac{\partial^3\Psi}{\partial t^3} + \frac{\beta}{\rho c_0^4}\frac{\partial^2\Psi}{\partial t^2} + \frac{1}{\rho}\nabla\Psi\cdot\nabla\Psi - \sum_{m=1}^{\nu}\xi_m = 0 \tag{1}$$

$$\xi_m + \omega_m\xi_m = \alpha_m(\omega_m)\frac{\Delta_m}{c_0^2}\nabla^2\Psi \tag{2}$$

In this equation, the raw ultrasound radio frequency data is denoted by Ψ and the image (density map) is given by ρ. As such, the problem of imaging the lung can be reduced into an inverse problem in which the objective is to reconstruct the lung density map ρ. We train a machine learning model to learn the mapping from observed radio frequency data Ψ to a density map (image) ρ.

3.1.2. Early Architecture

Dr. Jiayun (Peter) Wang has implemented an early architecture incorporating neural operators and a UNet-inspired network to learn the mapping from ultrasound radio frequency data to lung aeration images 3. Neural operators learn mappings between infinite-dimensional function spaces; it has been shown that neural operators are well-suited for partial differential equation solution approximation Kovachki et al. (2023).

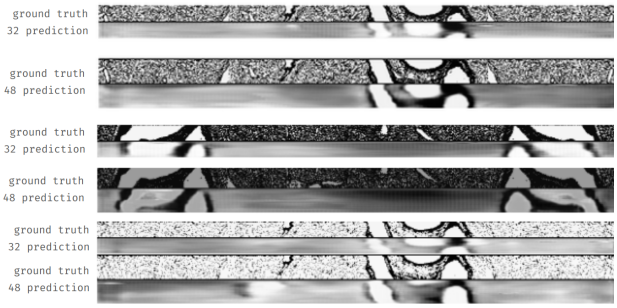

Fig. 4: Lung aeration ground truth and predictions images for 32-depth and 48-depth variant models. Pixels take on values continuously from 0 (air) to 1 (liquid). These predictions were made by the UNet model architecture discussed in 3.1.2

.

3.1.3. Advantages and Novelties

Our models output a precise density map ρ, allowing medical professionals to distinguish tissue types clearly. Additionally, the resulting density maps are well-suited for segmentation Ronneberger et al. (2015). Finally, neural networks can be altered in design to output pixel-wise inference confidence maps. This allows medical practitioners to take into account the model's certainty when making predictions, a feature that is invaluable in practical use cases. In comparison, the current baseline (B-Mode images) vaguely distinguishes tissue types by brightness. Softer tissues show up as darker spots, while rigid structures show up as bright spots.

3.2. Dataset

3.2.1. Overview

Each sample in the dataset is a tuple (Ψ_i, ρ_i). Ψ_i denotes raw radio frequency data recorded by ultrasound transducers for sample i; ρ_i denotes the corresponding two-dimensional lung density map. We want to train a model to learn the mapping from $\Psi \to \rho$; as such, we treat each Ψ_i as the input data and each ρ_i as the target.

The dimensions of each RF sample are given by $\Psi_i \in \mathbb{R}^{T \times M \times N} \mid T = 1509, M = 64, N = 128$. The first dimension T represents time while the second and third dimensions M, N represent spatial information, specifically transducer id's. The dimensions of each density map are given by $\rho_i \in \mathbb{R}^D$. Both dimensions represent spatial information; the first dimension D represents depth (under the skin) while the second dimension L represents lateral position.

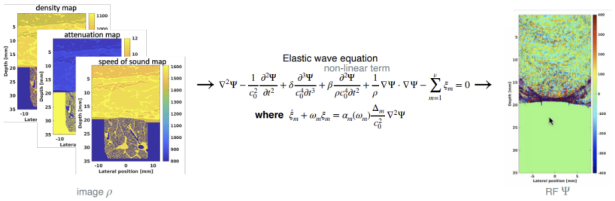

Fig. 5: Data Generation Process. Solving the elastic wave equation from (Equation 1), we generate a dataset of radio frequency data to density map image mappings.

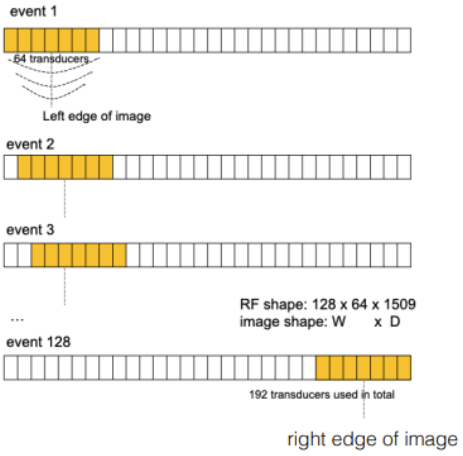

Fig. 6: Ultrasound Transducer Configuration.

3.2.2. Generation

We generate a dataset of 4,500 mappings as described in 3.2.1. For each sample's density map, we use the elastic wave equation outlined in (Equation 1) to generate the corresponding radio frequency data. This process is repeated for each density map resulting in a set of 4,500 tuples (Ψ_i, ρ_i) composing the entire dataset. The process is illustrated in 5.

3.2.3. Ultrasound Transducer Configuration

The labeled dataset described in 3.2.1 and 3.2.2 is collected using the following configuration. 256 transducers are placed in direct contact with the skin. In a single event, pulses are sent out from 64 adjacent transducers and received by 128 transducers. This setup is illustrated in Figure 6.

4. WORK PLAN

Weeks 1-2: Run baseline tests to quantify 32 depth early model performance.
Weeks 2-4: Develop confidence layers and evaluate inference certainty at depths of 48 and 64 mm.
Weeks 4-6: Apply domain adaptation techniques to improve model performance on real clinical datasets.
Weeks 6-8: Implement SSL approaches and varied neural network architectures; evaluate performance and report on findings.
Weeks 8-10: Evaluate improvements, assemble insights & conclusions, and prepare materials for final presentation.

REFERENCES

Ben-David, S., Blitzer, J., Crammer, K., and Pereira, F. 2006, in Advances in Neural Information Processing Systems, ed. B. Schölkopf, J. Platt, and T. Hoffman, Vol. 19 (MIT Press)

Demi, L. 2018, Applied Sciences, 8

Kovachki, N., Li, Z., Liu, B., et al. 2023, Journal of Machine Learning Research, 24, 1

Marini, T. J., Rubens, D. J., Zhao, Y. T., et al. 2021, Radiology: Cardiothoracic Imaging, 3, e200564, pMID: 33969313

Matrone, G., Ramalli, A., and Tortoli, P. 2019, Applied Sciences, 9

Pinton, G., Dahl, J., Rosenzweig, S., and Trahey, G. 2009, IEEE Trans. Ultrason. Ferroelectr. Freq. Control, 56, 474

Ronneberger, O., Fischer, P., and Brox, T. 2015, CoRR, abs/1505.04597