Deep Learning for Motion Tracking on the Micron Scale with Ultrasound

Derek Y. Chan
Department of Biomedical Engineering
Duke University
Durham, NC 27708
derek.chan@duke.edu

Abstract

Ultrasound elasticity imaging in soft tissue with acoustic radiation force requires extracting displacement information, typically on the order of several microns, from raw data. In this work, we implement a fully convolutional neural network for ultrasound displacement estimation. We present a novel method for generating ultrasound training data, in which virtual displacement volumes with a combination of randomly-seeded ellipsoids are created. Network performance was tested on the virtual displacement volumes as well as an experimental phantom dataset and human *in vivo* prostate data. In simulated and phantom data, the proposed network neural performed similarly well compared to a conventional phase-based displacement estimation algorithm. Application of the trained network to *in vivo* data enabled the visualization of the prostatic urethra and peripheral zone. Implications for data sampling based on a physical layer analysis are discussed.

1 Instructions

Several pathologies, including prostate cancer, breast cancer, and liver fibrosis, cause an increase in tissue stiffness with disease progression. Ultrasound techniques can be used to non-invasively measure stiffness. Specifically, acoustic radiation force impulse (ARFI) imaging uses focused ultrasound to displace tissue by several microns, with stiffer tissues having lower displacements. Raw 1-D in-phase and quadrature (I/Q) data are obtained before and after the induced displacements, and phase-shift autocorrelation techniques are typically used to extract the displacement information [1]. Repeated transmits across the field of view can build up a 2-D displacement image.

Displacement tracking in ARFI imaging can be degraded by factors such as thermal noise, data down-sampling, and finite window size. These lead to jitter errors, or variability in the estimated displacements. Furthermore, averaging of the heterogeneous displacement field within the point spread function of the ultrasound transducer leads to displacement underestimation [2]. Jitter can be reduced by using a larger smoothing kernel, at the expense of worse spatial resolution. Application of the Cramér-Rao Lower Bound (CRLB) to ultrasonic displacement estimation places a lower bound on the magnitude of jitter that results from an unbiased estimator [3]. Recent research has demonstrated the performance of biased estimators that surpass the CRLB using Bayesian methods, though these can be extremely computationally expensive [4].

In this work, a fully convolutional neural network is implemented for robust tracking of ARFI-induced displacements. The state-of-the-art performance of deep learning in many other imaging applications suggest that a neural network may be able to learn the complex patterns necessary to extract small displacements from noisy ultrasound data. Furthermore, once trained, the network should work in effectively real time, which is critical for immediate image visualization in a clinical setting. Obtaining more accurate displacement estimates will facilitate high-quality ARFI imaging of prostate cancer as well as other clinical applications of ultrasound elasticity imaging.

2 Related work

To our knowledge, there are no previous studies on using deep learning methods for the estimation of ARFI-induced displacements from ultrasound data. Two recent studies on using neural networks for strain elastography are described below, with both recently presented at the 21st International Conference on Medical Image Computing and Computer Assisted Intervention in September 2018. In strain elastography, the tissue is compressed by several millimeters and the corresponding strain is estimated. On the other hand, ARFI imaging is a more challenging scenario in which the displacements are three orders of magnitude smaller (several microns).

Kibria and Rivaz [5] describe a convolutional neural network for strain estimation with an architecture based on FlowNet, which was developed for optical flow imaging. They present results in simulation, phantoms, and *in vivo* liver data that suggest that their network is more robust to decorrelation noise, though the description of methods is limited with no discussion of the data that was used to train the network.

Wu et al. [6] present a 2-D strain field approach for strain imaging, also with a convolutional neural network. They demonstrate more compelling results with improved signal-to-noise ratio and contrast-to-noise ratio compared to a conventional strain estimation technique, but only used 40 samples in the training data.

3 Training data and preprocessing

All data that were used to train the network were simulated, since the ground-truth displacements are known in a simulation setting. In contrast, with experimental ARFI acquisitions, the actual displacement in the material is unknown, even in a controlled setting such as a tissue-mimicking phantom.

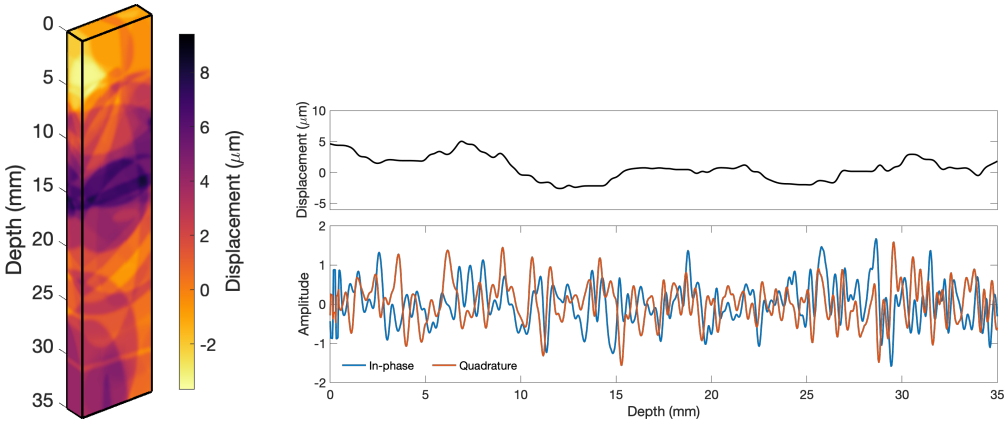
First, synthetic displacement fields were generated by using 3-D volumes consisting of 150 ellipsoids with random size, location, orientation, and amplitude inside a volume. A 3-D grid is necessary for realistic ultrasound tracking since the point spread function of the imaging system is three-dimensional. A 3-D Gaussian low pass filter with a standard deviation of 3.0 was applied in MATLAB to prevent excessively sharp edges in the volume. Figure 1(a) shows an example of a displacement field generated from a single random seed.

The ground truth displacement values were extracted from the center line of the virtual displacement volume, at lateral position 0 and elevation position 0. The top plot in Figure 1(b) shows the corresponding ground-truth displacement data for the volume shown in Figure 1(a).

Field II, an ultrasound simulation software, was used to place subresolution, sound-reflecting scatterers (corresponding to small structures such as organelles in tissue) in a field [7]. These scatterers were displaced by a given magnitude based on the 3-D displacement volumes that were generated. The software was then used to model the Siemens 12L4 ultrasound transducer to track the scatterers before and after the displacements were applied. In order to match the real data outputted by an ultrasound scanner resulting raw simulated radiofrequency data were demodulated and down-sampled from 1 GHz to 5 MHz sampling frequency. The down-sampled data was then re-upsampled to 25 MHz for more precise localization of the peaks (this step is performed in conventional displacement estimation as well).

Because the amplitude of Field II-simulated data is somewhat arbitrary (and on the order of 10^{-21} for the simulations in this study), the data was normalized across every feature in the data. First, the mean was subtracted across each dimension, and then each dimension was divided by its standard deviation. This type of pre-processing is commonly performed in deep learning and helps stabilize the learning. In fact, without this pre-processing step, the network returned the same constant value for any input because the weights were not able to be effectively adjusted.

The results of the ultrasonic tracking simulations are two sets of in-phase and quadrature (I/Q) data for each volume, corresponding to before and after the induced displacements. The displacement information is encoded between the two sets of I/Q data. The bottom plot in Figure 1(b) shows sample demodulated and normalized I/Q data after the displacements were applied.



(a) (b)

Figure 1: (a) Virtual 3-D displacement volume constructed by summing randomly-seeded ellipsoids. (b) Ground-truth displacements extracted from center depth line in the volume (top plot) and resulting simulated in-phase and quadrature (I/Q) demodulated data after scatterer displacement (bottom plot).

In total, 30,000 different virtual phantoms were generated and ultrasonically tracked. For each phantom, the two sets of I/Q data before and after the displacements were saved as the input to the network, and the ground truth displacements through depth were saved as the output to be reconstructed. The neural network was trained on 26,000 sets of these data, and 2,000 sets each were designated as validation and test datasets.

4 Fully convolutional neural network architecture

Figure 2 shows a diagram of the neural network architecture. The input to the network is a M$\times$1200$\times$2$\times$2 matrix, where M is the minibatch size (set to 75), 1200 is the number of depth samples, and the last two dimensions specify the time step (before or after the applied displacements) and I/Q channel, respectively. The output is a vector of length 1200, representing the displacement through depth.

A physical layer is first added that is multiplied by the input (Hadamard product). The weights of the physical layer were constrained to be either 0 or 1, to simulate downsampling of the input.

The data are then input into a series of convolutional and max pooling layers, with the number of features continually increasing due to the increasing complexity being represented. After four sets of double convolutional layers (3$\times$3 filter size) with 2$\times$1 max pooling, a series of transposed convolutional layers are used to build the image back up to a height of 1600. A final convolutional layer is used to collapse the last two dimensions to produce a 1-D output vector.

The L1 loss, or mean absolute error, was used to train the network and evaluate performance. The estimated displacements were subtracted element-wise from the ground truth displacements and the mean absolute difference across the entire minibatch was calculated. A minibatch size of 75 was used for training, with the ADAM (adaptive moment estimation) optimization algorithm and a learning rate of 0.001.

The network weights were initialized using the approach described by He et al., in which the variance of nodes in a layer is $2.0/n$ where n is the number of units in the previous layer [8]. This initialization was derived specifically for ReLU activations, which are used in this architecture, and prevents an exploding variance value as the number of input grows.

The fully convolutional architecture, without any fully connected layers, reduces the number of parameters in the model and is appropriate for this task since the displacement information is encoded locally (i.e., a fully connected layer is not needed to connect different spatial regions of the image together). The U-Net architecture described by Ronneberger et al. in 2015 is very similar to the one described here and was considered for this task, but the presence of skip connections did not improve the performance of the network based on the loss. This may be due to the fact that the displacement

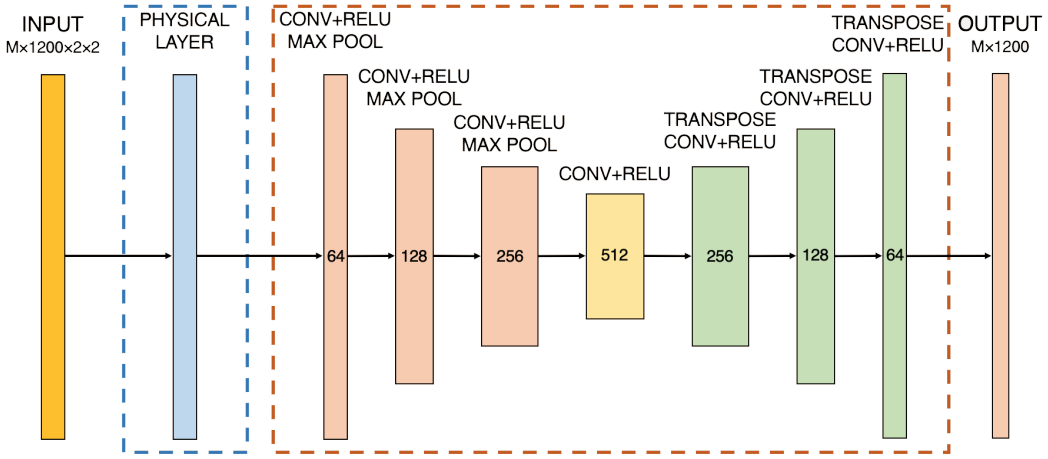


Figure 2: Diagram of the neural network architecture.

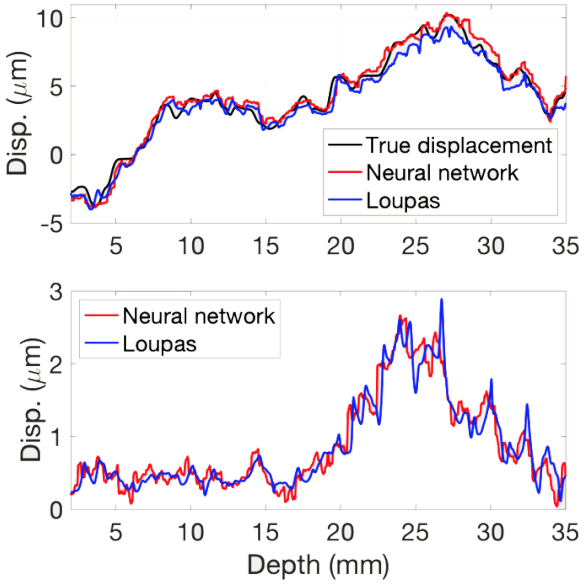


Figure 3: Comparison of neural network and Loupas' algorithm in simulated synthetic training data (top) and experimental phantom data (bottom).

data does not resemble the input I/Q data in any way, as clearly demonstrated in Figure 1, so the skip connections were less useful than they would be in a segmentation task.

5 Training results

On a Tesla V100 GPU, training took approximately 8 minutes for 10,000 iterations with a minibatch size of 75. Figure 3 demonstrates that both the neural network and a conventional phase-shift displacement estimator by Loupas et al. [9] are able to reasonably reconstruct the simulated ground-truth displacements from the virtual phantom data (top row). Deviations from the ground truth are due to an averaging effect of a finite imaging point spread function. In other words, scatterer displacement is inhomogeneous within the ultrasound track beam, and the different displacements are combined in the raw ultrasound data, making it difficult to determine the true displacement of the scatterers along the central axis. The neural network was able to generalize to experimental phantom data (bottom row of Figure 3).

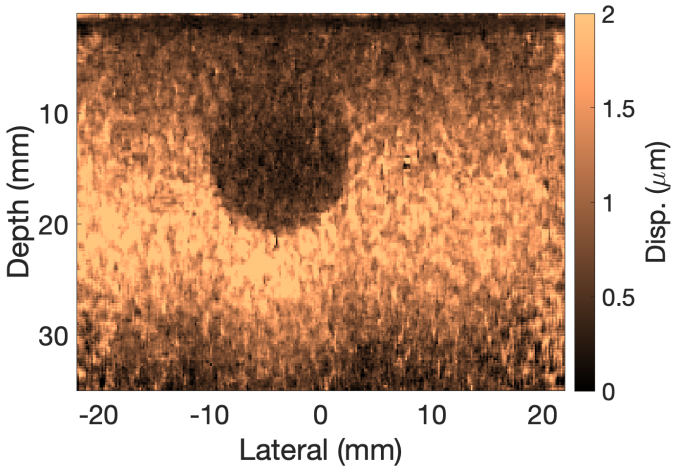


Figure 4: Reconstructed ARFI image for a phantom with a stiff inclusion.

Figure 4 shows a reconstructed two-dimensional displacement map for experimental data collected in a tissue-mimicking phantom with a stiff, spherical inclusion. The data were obtained using the same ultrasound transducer and transmit parameters that were simulated with Field II. Displacement estimation was performed along each 1-D lines through depth individually; with the neural network, the entire field of view was treated as a minibatch with size 328. The stiff inclusion can be visualized in the displacement image as a circular region of low displacement. This corresponded to the known location of the inclusion within the phantom.

Figure 5 shows a two-dimensional displacement map retroactively reconstructed from experimental *in vivo* human prostate data, obtained in an ongoing IRB-approved study at Duke University Medical Center. In Figure 5, the capsule of the prostate has been manually segmented and ARFI displacements are shown only within the prostate (B-mode ultrasound is displayed outside). The image reconstructed from the neural network enabled visualization of the prostate urethra (green arrow) as well as the peripheral zone of the prostate (yellow arrow).

6 Physical layer analysis

When the network was trained on just the L1 (mean absolute error) loss, all final weights of the physical layer were 1, with the mean absolute error equal to 0.393 microns after 10,000 training iterations. This indicated that if possible, full sampling is desirable and no down-sampling should be performed. However, assuming that we were interested in lower sampling rates to enable less expensive scanners at the potential expense of some accuracy, a term was added to the loss function to penalize the number of non-zero weights in the physical layer. A lambda parameter (similar to regularization parameters used in many machine learning applications) was multiplied by the weights to control the degree of penalization. For $\lambda = 2.0$, the mean absolute error after 10,000 iterations was 0.473 microns, with 83% of the weights equal to 1. In other words, there was some loss in accuracy (increase in mean absolute error) associated with fewer non-zero samples, but the fewer samples could potentially enable a decrease in hardware cost.

7 Discussion

In this study, a fully convolutional neural network was trained to extract small displacements from ultrasound data, using a novel method for generating synthetic displacement volumes. In simulated data, the network performed similarly well compared to a conventional phase-shift estimator, though it could not overcome fundamental errors due to the finite point spread function of the imaging system. In experimentally-acquired phantom data, the trained displacement estimator enabled visualization of a stiff inclusion, and in *in vivo* prostate data, the peripheral zone and urethra were well-visualized.

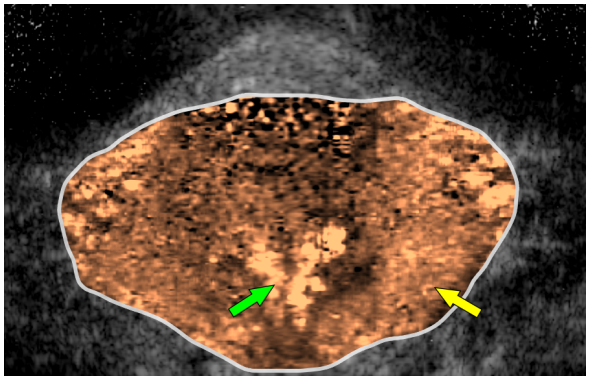


Figure 5: Reconstructed ARFI image for human *in vivo* prostate data showing the prostatic urethra (green arrow) and the peripheral zone (yellow arrow)..

One question that is currently unexplored with this project is the robustness of the network to different transducers or ultrasound transmit configurations, including transmit frequency, aperture width, and beamforming approach. Fundamentally, displacement estimation always involves extracting displacement data from local variations in the I/Q data, but the specific optimal weights may vary for different configurations. Transfer learning may potentially be a useful method for re-training a network for a different ultrasound setup, though at this stage it takes longer to generate the training data for a given transducer than to train the network from scratch.

Overall, the findings from this project suggest that with further network optimization, deep learning may become a viable technique for robust and fast stiffness imaging, and clinical translation (at least in a research setting) currently appears to be attainable. In future work, this network will be deployed in real-time in an ongoing prostate ARFI study at the Duke University Medical Center.

References

[1] G. F. Pinton, J. J. Dahl, and G. E. Trahey, "Rapid tracking of small displacements with ultrasound," IEEE Trans. Ultrason. Ferroelectr. Freq. Control, vol. 53, no. 6, pp. 1103–1117, June 2006.

[2] M. L. Palmeri, S. A. McAleavey, G. E. Trahey, and K. R. Nightingale, "Ultrasonic tracking of acoustic radiation force-induced displacements in homogeneous media," IEEE Trans. Ultrason. Ferroelectr. Freq. Control, vol. 53, no. 7, pp. 1300–1313, July 2006.

[3] W. F. Walker and G. E. Trahey, "A fundamental limit on delay estimation using partially correlated speckle signals," IEEE Trans. Ultrason. Ferroelectr. Freq. Control, vol. 42, no. 2, pp. 301–308, Mar. 1995.

[4] D. M. Dumont and B. C. Byram, "Robust tracking of small displacements with a Bayesian estimator," IEEE Trans. Ultrason. Ferroelectr. Freq. Control, vol. 63, no. 1, pp. 20–34, Jan. 2016.

[5] M. G. Kibria and H. Rivaz, "Global ultrasound elastography using convolutional neural network," 21st International Conference on Medical Image Computing & Computer Assisted Intervention, Granada, Spain, Sept. 2018.

[6] S. Wu et al., "Direct reconstruction of ultrasound elastography using an end-to-end network," 21st International Conference on Medical Image Computing & Computer Assisted Intervention, Granada, Spain, Sept. 2018.

[7] J. A. Jensen and N. B. Svendsen, "Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers," IEEE Trans. Ultrason. Ferroelectr. Freq. Control, vol. 39, no. 2, pp. 262-267, Mar. 1992.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: surpassing human-level performance on ImageNet classification," Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, Dec. 2015.

[9] T. Loupas, J. T. Powers, and R. W. Gill, "Experimental evaluation of velocity and power estimation for ultrasound blood flow imaging, by means of a two-dimensional autocorrelation approach." IEEE Trans. Ultrason. Ferroelectr. Freq. Control, vol. 42, no. 4, pp. 689–99, July 1995.