BME 590L: Machine Learning for Ultrasound Lesion Mapping with Apodization Optimization

Courtney A. Trutna
Department of Biomedical Engineering
Duke University
Durham, NC 27707
courtney.trutna@duke.edu

Abstract

Machine learning techniques have much to offer to the field of ultrasound imaging. This work uses a U-Net architecture to perform lesion (echogenicity) mapping. Additionally, B-mode ultrasound image simulations are performed directly in the machine learning architecture, allowing for an optimization of the "physical" apodization applied to the transducer. Results show that the U-Net architecture is able to reliably identify the general location of the lesion, though the shapes are somewhat distorted. Physical layer optimization appeared to generally construct a point spread function similar to the full aperture case for all configurations tested, though the actual apodization functions differ. This suggests that traditional focusing and apodization methods are necessary for (or at minimum assist) machine learning echogenicity mapping.

1 Introduction

Machine learning is a growing field with many potential applications to ultrasound imaging. Two such utilities are explored within this project.

First, ultrasound imaging is subject to deterministic speckle noise. Speckle is present in all B-mode ultrasound images, and is considered to have negative effects on the diagnostic ability of ultrasound imaging. This project focuses on the typical test task of anechoic lesion identification. In this task, speckle noise leads to 'write-in' over the lesion, such that a lesion that should be devoid of signal instead has speckle present. This leads to a lowering of lesion contrast and contrast-to-noise-ratio. This project proposes to "despeckle" an image through machine learning, to make an echogenicity map unaffected by speckle noise. This will allow for easy visualization of lesions.

Second, the field of ultrasound imaging typically applies an apodization to the transducer transmit and receive profiles to modify the point spread function (PSF) of the imaging system. In the extreme case, entire elements can be removed, which is sometimes beneficial in matrix array imaging, where it can be difficult to record data from the entire array simultaneously. This project explored using the apodization as a trainable series of weights to determine the ideal apodization for image input into a neural network. The project specifically used a 64x16 element array to explore the apodization desired when using a matrix transducer for 2D imaging.

2 Related work

The lesion identification task used in this project is regarded as a segmentation problem, and is approached with the U-Net architecture [1]. This architecture is popular and has many applications in segmentation tasks. Additionally, several recent papers in the field of ultrasound imaging have investigated the use of machine learning. One such paper of interest is [2], which uses a neural net

to beamform data to reduce speckle noise. That paper is similar in goals to this project, though this project assumes delay and sum beamforming, and only alters the apodization, while that work investigated both apodization and more complicated (learned) beamforming. Another work of interest is [3], which is very similar to the work herein: both use U-Nets for a lesion segmentation task. However, that work uses a linear transducer and does not attempt to optimize any physical parameters.

3 Methods

3.1 Ultrasound image modeling

The underlying data used for this project consisted of modelled echogenicity fields. To model the typical results seen in ultrasound, a random scatterer field was generated, with enough scatterers to create speckle that is "fully developed". The random scatterers outside of the lesion all have equal echogenicity. Scatterers within random lesions were set to 1/100th the value of the background, to model anaechoic lesions. Lesions were random in number (between 0 and 3), size (between 0.5mm and 3 mm) and position (anywhere in the scatterer field), though were all modeled as anechoic.

The physical layer was modeled using the concept of a "complete data set" from ultrasound imaging. A 64x16 element matrix transducer was focused at 4 cm, and the resulting sensitivity pattern was modeled for each of the transducer's 1024 elements. These simulations were performed using Field-II, a standard ultrasound simulation package [4, 5].

The signal corresponding to each element was multiplied by an individual trainable value, representing the weight of that "element" on the transducer. Following these weightings, the channels are summed together coherently to give a two dimensional PSF. In general, the lateral resolution of the PSF is represented by the Fourier transform of the aperture, so using all elements equally (rect) results in a tight sinc function, and applying any apodization to the elements results in a wider PSF. However, apodizations are frequently used to reduce the side lobes present in the sinc function.

The PSF is then convolved with a field of scatterers to produce an radiofrequency image. This image must then be demodulated using a Hilbert transform to become the familiar B-mode ultrasound image. Note that the U-Net in this project received an non-log-compressed version of the B-mode image to avoid zero-division errors, though B-mode images are typically displayed with log compression. Both are displayed throughout this document for reference.

Two key assumptions about the imaging system were made in this work. First, the PSF was assumed to be constant in the region of interest. This is a commonly made assumption within the transducer depth of field. As only a small region of interest is simulated, this is believed to be a valid assumption for this work.

Second, only the two dimensional PSF was modeled. This was necessary due to the extreme computational cost of modelling the three dimensional PSF, and the cost of the three dimensional convolution with a scatterer field. However, out-of-plane scatterering is a known contributor to ultrasonic noise. Importantly, an advantage of matrix probes is their superior elevational focusing, which can decrease the effects of "out-of-plane" focusing. This advantage of matrix transducers is not accounted for or modelled in this project, which is a limitation of this study. However, many first-order ultrasound analyses use this 2D PSF assumption.

3.2 Machine learning architecture

The B-mode images were fed into a U-Net design with the goal of recovering the original echogenicity map. Because of the limited difficulty of this problem, the number of filters at every level of the UNET was divided by 8 relative to the original paper. This increased training speed and discouraged overtraining. An Adam optimization function with learning rate 1e-5 was used for fitting, with a binary cross entropy loss function between the neural net output and the true echogenicity map.

When training the physical layer, various initialization, constraints and regularization techniques were employed to explore the connection between the physical layer and the neural net, in addition to exploring a further reduction of the number of filters in the UNET architecture.

				Training		Test	
Intializer	Constraints	Reg.	UNet Factor	Acc.	C. Entropy	Acc.	C. Entropy
Ones	None	None	8	0.9712	0.1737	0.7323	3.8436
Ones	[0,1)	None	8	0.9553	0.1611	0.5584	0.5438
Ones	[0,1)	L1	8	0.9857	0.2265	0.5973	1.3324
Rand. uniform	[0,1)	L1	8	0.9755	0.5987	0.6312	4.1503
Ones	[0,1)	L1	16	0.9783	0.3463	0.5818	6.6999

4 Results

First, the U-Net was tested on the task of echogenicity mapping assuming no apodization (all matrix elements contributing equally). This task achieved an training and test accuracy of 0.9583 and 0.7722, respectively, and a training and test binary cross entropy (loss) of 0.6409 and 2.8070, respectively. These results suggest that the model may be overfitting, though the output images of the test set look qualitatively correct. An example set of (validation) images is shown figure 1. For reference, the PSF and weights of each element (uniformly one) are also shown in figure 1.

The machine learning architecture was then changed to allow the physical layer weights to train. Various configurations were tested to explore the interaction of the network with the physical layer, with results reported in Table 1. In the order reported, the various combinations are as follows. Unless otherwise noted, initialization was uniformly one for the matrix weights.

First, the physical layer was allowed to train with no constraints or additional loss penalties. An example set of images is shown in figure 2. Second, the physical layer was given a constraint to be within the values of 0 and 1, with a progressive 0.9 enforcement rate applied (results in figure 3). It can be seen in some later configurations that this constraint was not fully obeyed, however, it does force the model to train downwards, rather than randomly, when initialized at one. Third, an L1 regularization was added to suppress the matrix values (results in figure 4). Fourth, the initialization was changed to a random uniform, and the regularization changed back to L1 (results in figure 5). Fifth, a case where the UNet was given half as many filters at each layer compared to previous cases, or 1/16 the number of filters in the original paper (results in figure 6). An L2 regularization was also explored and found to give similar results (albiet slightly smoother appearing weights) as the L1 regularization.

Qualitatively, in all cases, the U-Net is able to reconstruct an echogenecity map with a lesion in the correct location, but the shape of the lesion is distorted relative to the truth.

It should be noted that the random case includes negative weights despite constraints to keep within 0 and 1. The constraints were likely not reached because of the limited number of training iterations. Negative weights in practice would resemble phase shifts, which are usually not considered past of ultrasonic apodization, but are physically realizable.

5 Discussion

Overall, every case tested has similar quantitative and qualitative results, with an overall trend of high training accuracy, lower validation accuracy, and good qualitative results on lesion placement (though not necessarily structure). Some of this may be due to selection bias: during testing some attempts to run the training would not converge, in which case the training would be restarted. This general trend of lower validation accuracy also suggests this task is fairly simple for the number of trainable parameters used to achieve it, and the model may be over-fitting to the training data. The validation results still look qualitatively good, possibly due to the encoding/decoding structure of the U-Net architecture used. The final case trained, with a lower number of parameters, still exhibits low validation accuracy, but it was also observed that training failed frequently when using this smaller number of parameters, suggesting perhaps a different alteration of the architecture would be more effective.

It should also be noted that particularly for the constant initialization, the physical layer is only slightly altered. It seems possible that the gradient contribution of the the physical layer is relatively small due to the multiple complex processes (PSF convolution, envelope detection) between the

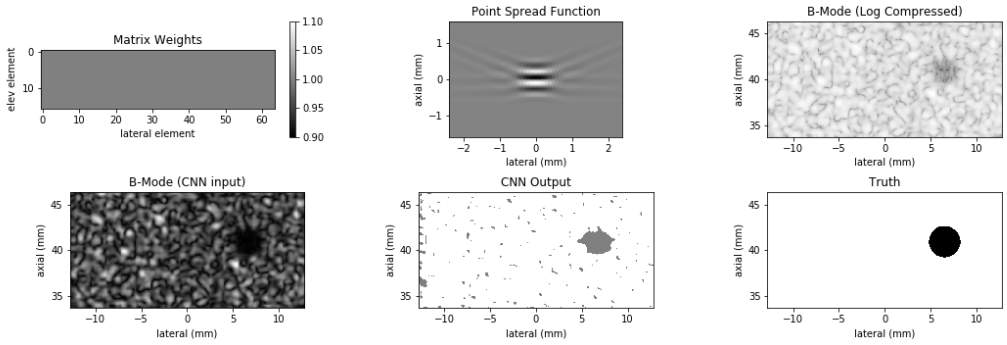

Figure 1: Results for no physical layer training. Subplots are as labeled. Note that the CNN output approximates the truth, but at a lower value (truth and CNN are on the same color axis), and with noise present. Still, the lesion is easily identifiable.

weights and the input to the U-Net. These observed mild changes persisted even when changing the scale of the weights or loss function to encourage a more sparse, 0's and 1's weight output.

However, it can be seen that the PSF for the optimized physical layer is the same in almost every case, including the randomly initialized case. This suggests that the full aperture is beneficial to the CNN's processing, unsurprisingly, as it is the theoretical best resolution. Particularly the convergence of the random initialization to this PSF suggests that contrary to the issues mentioned above, the neural network is actively updating the physical layer, and that this "generally focused PSF" is the optimal solution. Note that here "focused" does not necessarily mean that all the matrix values are equal, but rather as in the randomly initialized case, enough elements are added to create a generally focused PSF. This observation aligns with observations in the field of ultrasound that in low-or-zero noise simulations less elements are needed to successfully create an image. (This is generally found not to be true in vivo, where the additional noise sources present appear to be suppressed by the use of many spatially unique elements.)

Additionally, it can be seen that When initialized to a constant value, the matrix weights develop vertical stripes. Because only the 2D imaging plane is modeled, this is unsurprising, because each element in an elevational column contributes approximately the same signal as the others, and thus would be effected in similar ways. Future work should address the effects of the three dimensional point spread function and scatterer space, which would likely have an effect on the elevational direction of apodization.

6 Conclusion

A U-Net machine learning architecture is capable of successfully reconstructing an echogenicity map of lesions from ultrasound B-mode images. Results from physical layer optimization suggest that the traditional focusing methods or apodization that approximate the full-aperture PSF lead to the most accurate mappings. These results suggest that the U-Net architecture can be used in segmentation and de-speckling ultrasound applications, as long as the imaging system is reasonably focused.

Acknowledgments

Thank you to Dr. Roarke Horstmeyer and the 590 teaching assistants Kevin Zhou and Ouwen Huang.

References

[1] O. Ronneberger, P. Fischer, T. Brox "U-Net: Convolutional Networks for Biomedical Image Segmentation". (2015). arXiv:1505.04597

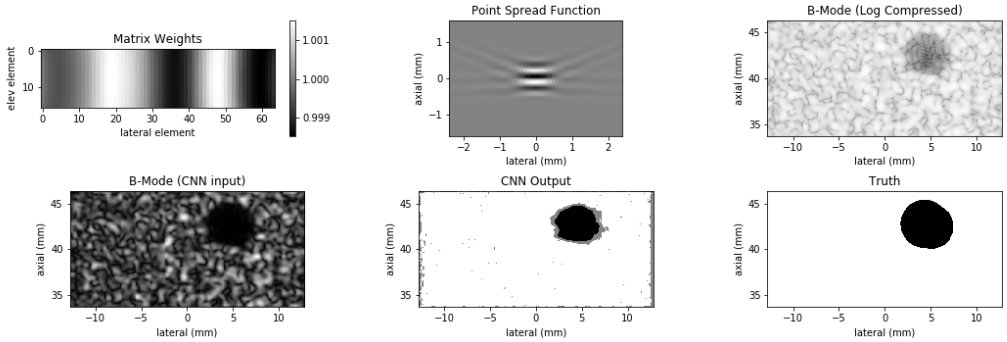

Figure 2: Results for physical layer training with no constraints or regularizations. Note that while the matrix weights train to equal values across the elevational dimension. Also, note that while a pattern is visible in the weights, the values are all very similar, and the PSF is not substantially changed.

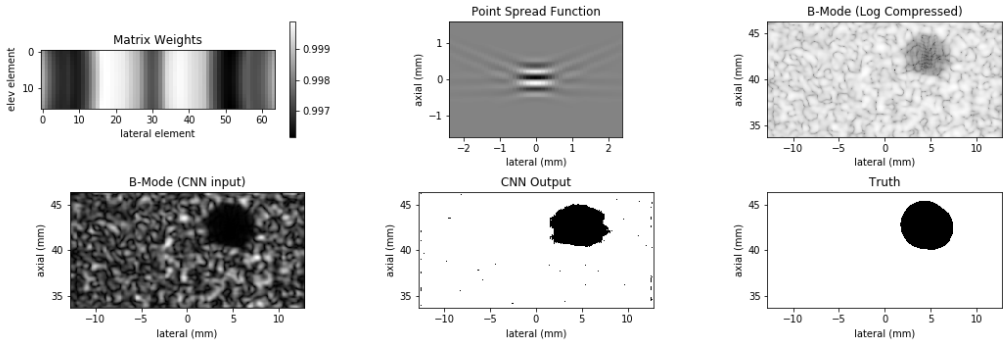

Figure 3: Results for physical layer training with constraints but no regularizations. Again, note the matrix weights train to equal values across the elevational dimension, but the values are very similar, and the PSF is not substantially changed. The CNN is able to recover the lesion location.

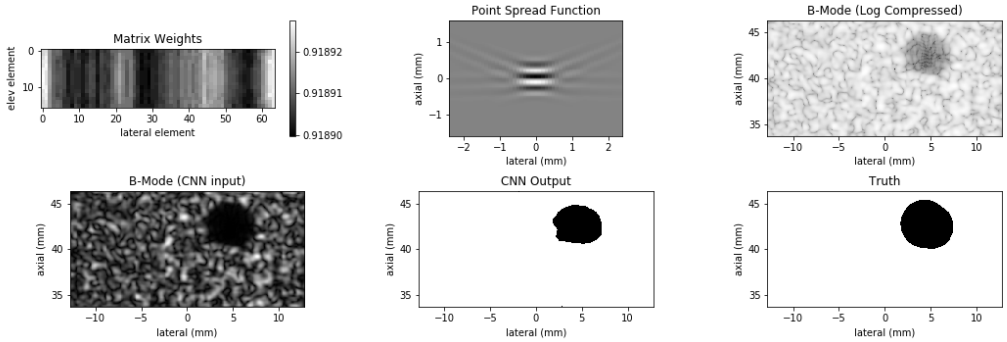

Figure 4: Results for physical layer training with constraints and a L1 regularization on the weights. Again, note the matrix weights train to equal values across the elevational dimension, but the values are very similar, and the PSF is not substantially changed. The CNN is able to recover the lesion location.

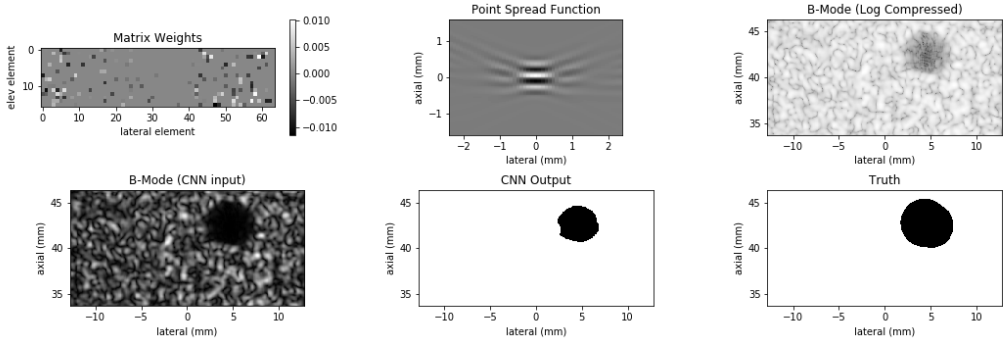

Figure 5: Results for physical layer training with constraints, a L1 regularization on the weights, and a random uniform initialization. Notably, here the vertical banding is not present, but the PSF remains similar to the other cases. The CNN is able to recover the lesion location.

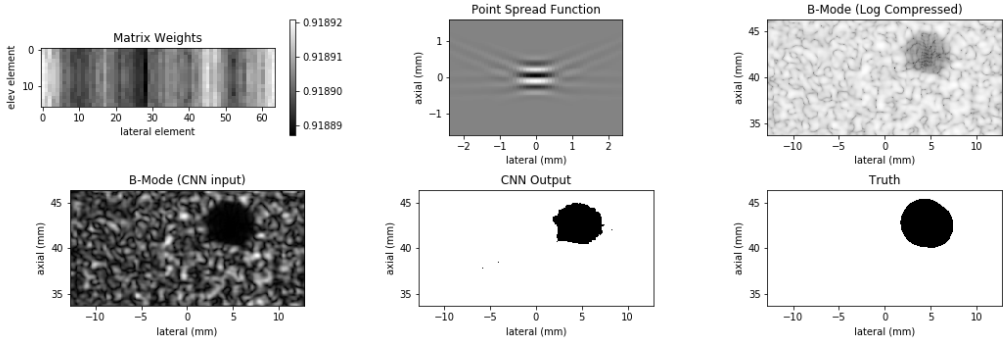

Figure 6: Results for physical layer training with constraints, a L1 regularization on the weights, and a U-Net with half as many filters per layer as previous cases. Again, note the matrix weights train to equal values across the elevational dimension, but the values are very similar, and the PSF is not substantially changed. The CNN is able to recover the lesion location.

[2] D. Hyun, L. L. Brickson, K. T. Looby, and J. J. Dahl, "Beamforming and Speckle Reduction Using Neural Networks," IEEE Trans. Ultrason. Ferroelectr. Freq. Control, vol. 66, no. 5, pp. 1–1, (2019).

[3] A. A. Nair, M. R. Gubbi, T. Duy Tran, A. Reiter, and M. A. Lediju Bell, "A Fully Convolutional Neural Network for Beamforming Ultrasound Images," IEEE Int. Ultrason. Symp. IUS, vol. 2018–October, pp. 1–4, 2018.

[4] J.A. Jensen: Field: A Program for Simulating Ultrasound Systems, Paper presented at the 10th Nordic-Baltic Conference on Biomedical Imaging Published in Medical and Biological Engineering and Computing, pp. 351-353, Volume 34, Supplement 1, Part 1, (1996).

[5] J.A. Jensen and N. B. Svendsen: Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers, IEEE Trans. Ultrason., Ferroelec., Freq. Contr., 39, pp. 262-267, (1992).