INDIAN INSTITUTE
OF TECHNOLOGY
PALAKKAD

Unsupervised Beamforming for Ultrasound Imaging

Submitted by
Joel Jeffrey

Mentor
Dr. Mahesh R Panicker

Contents

I INTRODUCTION

Ultrasonography, widely referred to as Ultrasound (US), is one of the most popular medical imaging tools, predominantly due to its non-invasive nature and real-time performance. Although US finds its applications principally in medical imaging (such as gynecology, cardiology, and emergency medicine), it is also widely used in leak detection and fingerprinting.

In a typical US imaging system, the US transducer, which consists of an array of elements, transmits an acoustic pulse into the human body. These US waves encounter an acoustic impedance mismatch. This causes the waves to be reflected back and is then received by the transducer elements (which doubles up as a receiver). The signals obtained in each of the elements are combined to form a high-resolution image through the process of beamforming. The raw data can be visualized as an input of dimension $N_t \times N_{ch}$, where N_t is the number of time samples and N_{ch}, is the number of transducer elements or channels. Beamforming converts this raw data into an image of dimension $N_r \times N_c$, where N_r is the number of rows and N_c, is the number of columns in the final image.

Two of the most common transmission techniques in any US imaging system are "Plane Wave Transmission" and "Focused Transmission". While beamforming in the plane wave transmission aims at generating images from parallel waves, beamforming in the focused wave transmission focuses US waves to a specific point or region of interest. The wavefront curvature in focused wave beamforming in most cases, often adds to the challenge of providing a noiseless high-resolution image, making focused wave beamforming computationally more complex.

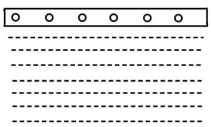

Figure 1: Plane Wave Transmission

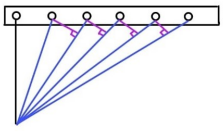

Figure 2: Focused Transmission

To accelerate the receive beamforming process, the usage of deep learning algorithms have been employed. This study aims to develop a couple of neural

networks with different architectures and compare the beamformed images generated by them. The Convolutional Neural Network (CNN) based beamformer and the Fully Connected Neural Network (FCNN) based beamformer have been developed and studied.

II WORK

II.1 LITERATURE REVIEW

Improving the image quality of US imaging while ensuring minimal complexity is one of the most researched fields in US imaging. Beamforming is a well-researched field that processes the ultrasound waves emitted by the transducers to form a high-resolution image. The increased interest in beamforming at high frame rates and image quality improvement has led to the development of the Plane-Wave Imaging Challenge in Medical Ultrasound (PICMUS) which aims to provide tools to compare the methods developed.[1].

Phenomena such as interference, noise, multi-path reverberation, and off-axis scattering (collectively called acoustic clutter) often degrades the image quality during the process of beamforming. Several approaches such as Delay and Sum (DAS) [2], Delay Multiply and Sum (DMAS) [3], Filtered Delay Multiply and Sum (FDMAS) [4], Generalised Coherence Factor (GCF) [5], and Minimum Variance Distortionless Response (MVDR) have been tried and tested, aiming to solve this issue.

DAS focuses on utilizing a predetermined delay and geometry-driven weights to interpolate individual transducer signals thereby forming a high-resolution image. The predetermined delay can be estimated from the speed of sound in the medium and the range (distance of the point of reconstruction from the probe). Though the DAS beamformer is the most popular, the presence of side lobes and limited resolution limits its application.

Further studies proved that DMAS, MVDR, and the other approaches to provide better results albeit using much more computation power. Although the DMAS and DAS techniques are very similar, the inclusion of a multiplication step in DMAS facilitates better pattern extraction and imaging. MVDR also known as Capon's method estimates the channel weights (also known as

apodization) by minimizing the variance of signal power (beam pattern) by maximizing the power in the desired direction.

The FDMAS is a variation of the DAS approach, where additional processing steps such as filtering and non-linear operations are incorporated to enhance the image quality. The filtering (generally low pass or band pass) helps in denoising. Although each of the above techniques provides better results, they are computationally more intense and require significant amounts of time for image generation. In the recent past, several attempts have been made to employ neural networks to accelerate the beamforming process. This project aims to develop an accelerator that can speed up the beamforming process while ensuring image quality does not deteriorate significantly.

The ultrasound image reconstruction methods can be broadly divided into classical techniques and novel deep learning-based reconstruction techniques. Analytical methods [6] and iterative methods [7] are a few of the most popular classical ultrasound image reconstruction methods. While, the analytical methods require proper and precise sampling while imaging, discrepancies between models and factors impede the image reconstruction in the iterative methods[8].

Harmonic imaging and Time reversal are popular classical techniques employed in US imaging. However, harmonic imaging is less sensitive than the normal brightness mode (B-Mode) imaging[9]. Time reversal was found to fail when attenuating media such as bones were encountered.[10]. These challenges limit the scope of classical techniques and encourage the shift to deep neural network-based techniques.

In recent years, the development of Neural Networks (NN) has led to a shift towards NN-based approaches to the challenge. NN-based architectures such as Generative Adversarial Networks (GANs) [11] and recurrent NN (RNN) have paved the way for a more efficient image reconstruction.

Byram et al [12] developed a beamforming method called the aperture domain model image reconstruction (ADMIRE). ADMIRE focused on suppressing the denoising and degeneration of the image. ADMIRE, however, has a huge runtime requiring hours per frame. This encouraged looking at this problem as a Deep Neural Network (DNN) problem.

Another study [13] proposes an MVDR-based approach, with 4 hidden layers following the input layer. The CNN employed here outputs apodization weights corresponding to each pixel in the image. The dataset used to train was the Time of Flight Corrected (ToFC) dataset and the loss function was Mean Squared Error (MSE).

Nair et al [14] proposed a U-Net architecture, with a VGG-13 encoder and Adam optimizer, to detect cysts. However, the Dice Similarity Coefficient (DSC), representing the overlap between the true cyst and the one detected, was found to decrease with increasing radius signifying a degradation in the network performance. In another study undertaken by the same authors [15], the U-Net architecture with Convolutional and Max Pooling layers have been employed for contracting and UpConvolution for expansion. As in their previous work, the authors encounter reduced efficiency when dealing with smaller cysts encouraging the use of end-to-end DNNs.

Perdios et al [16] employs a Convolutional Neural Network (CNN) based approach utilizing a network architecture similar to the U-Net to map low-quality images obtained to high-quality images from synthetic aperture (SA) measurements. The architecture makes use of skip operations to forward information directly preventing loss of information while downsizing.

Luchies et al [17] proposed a DNN with a Fully Connected Neural Network (FCNN) based architecture employing Short-Time Fourier Transform (STFT) and Inverse Short-Time Fourier Transform (ISTFT), with Adam as the optimizer and MSE as the loss function. While the model was found to work better with large neural networks, the performance (in terms of validation loss) was found to decrease with smaller neural networks.

These methods chiefly focused on developing static models that are trained to learn a single approach such as MVDR. Katare et al [18], came up with an active learning-based NN. Active Learning-based NN takes user inputs for the best image and trains the NN based on the image which is considered to be the ground truth. The apodization weights are then updated using backpropagation.

The Minimum Variance (MV) based beamforming was found to be computationally complex and Luijten et al [19] proposed a learning-based method called the Adaptive Beamforming by deep Learning (ABLE). ABLE utilizes

the structure of adaptive beamforming to compute the apodization weights. Further, ABLE uses signed-mean-squared-logarithmic error (SMSLE) as a loss function.

A study into the effectiveness of using GANs for Ultrasound beamforming was also conducted [20]. While CNN was used for image segmentation from raw data, GAN was implemented to learn from the FIELD-II training dataset. The GAN generated a B-Mode image and a DNN segmentation matching that of nearby cysts.

In another study utilizing MobileNetV2 as the architecture [21], the proposed method was found to provide better imaging when compared to the DAS beamforming and similar to Minimum Variance Beamforming (MVB). The proposed architecture was also found to have reduced complexity due to its architecture.

II.2 DATASET

The data fed into the beamformer is the delay-compensated data of dimension $N_r \times N_c \times N_{ch}$ for the CNN beamformer and $(N_r N_c) \times N_{ch}$ for the FCNN beamformer where N_r is the number of rows, N_c is the number of columns, and N_{ch} is the number of channels. A total of 1000 scanned images are taken in with a split of 800 for training and 200 for testing. The batch size was set to 40.

II.3 NEURAL NETWORKS

The study focuses on performing the beamforming algorithm using two architectures - the Convolutional Neural Network architecture and the Fully Connected Neural Network architecture. Some commonalities between both architectures include the usage of the same optimizer, the Adam optimizer a Stochastic Gradient Descent (SGD) algorithm-based optimizer, which builds upon SGD for a more robust and faster converging optimizer. Both architectures also employ a learning rate of 1e-4.

II.3.1 CNN Architecture

The architecture of the CNN beamformer consists of an input layer followed by 4 hidden layers and a softmax layer.

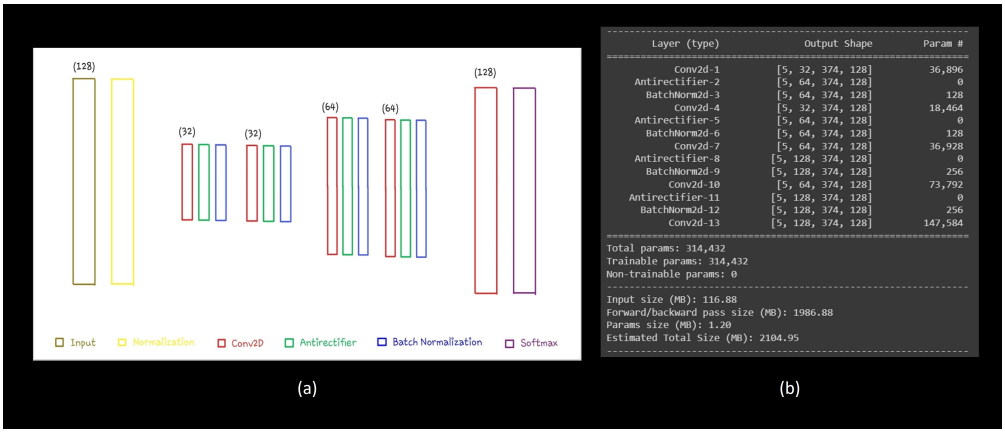

Figure 3: (a) CNN Architecture, (b) The model summary as obtained in PyTorch

The input to the model is L_2 normalized and is of dimension (x, y, c) where x,y, and c are the rows, columns, and channels of the delay-compensated data. The data used here is of dimension $(374, 128, 128)$. The 4 hidden layers are convolutional layers and increase the dimension to 128 from 32. The kernel size used throughout is $(3, 3)$ and same padding is performed to ensure an output with the same dimension as the input. An antirectifier is used as an activation function with the rectified linear unit (ReLU) preventing data from being mapped to zero. Batch normalization has also been used to ensure stability while training by normalizing the activations in each layer. The final softmax layer converts the apodization weights to values between 0 and 1 such that their sum is 1. Once the model is trained, the apodization weights received are dot multiplied to the input image, and the output is summed along the channel axis to obtain the beamformed data. Hilbert transform is also performed on the beamformed data helping retrieve the envelope of the beamformed signal, signifying the magnitude of the signal as a function of time.

II.3.2 FCNN Architecture

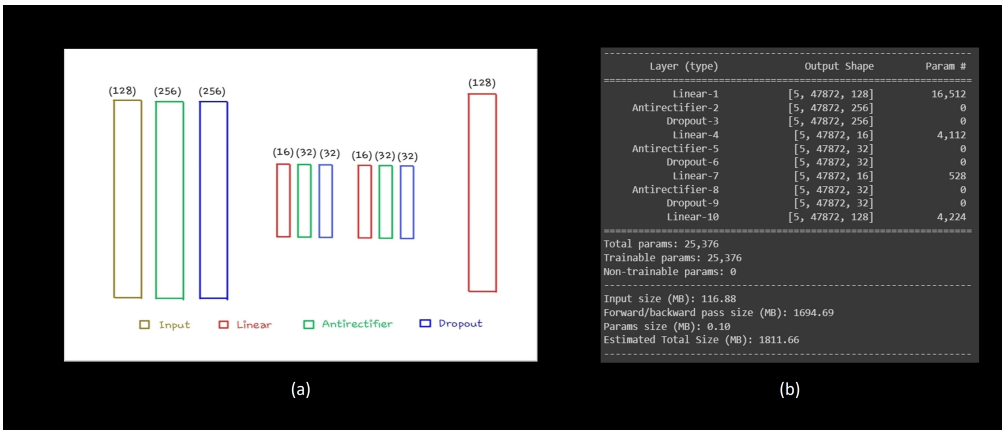

Figure 4: (a) FCNN Architecture, (b) The model summary as obtained in PyTorch

The FCNN beamformer consists of an input layer followed by 3 hidden layers. The input to the model is flattened to 1 dimension and is fed into the network. The input and the 3 hidden layers are linear layers that take in different data of different dimensions. Similar to the CNN architecture, Antirectifier is also used here. Dropout layers are also implemented in the beamformer to prevent overfitting of data. The remaining architecture including the multiplication of apodization weights, summing and taking the Hilbert transform remains the same.

II.4 IMAGE QUALITY METRICS

Besides the qualitative estimation of the beamformed image, we implement a quantitative analysis to get a better idea of the image quality. Parameters such as contrast ratio, and contrast to noise ratio are found out as part of this quantitative analysis. The Contrast Ratio (CR) compares how bright the brightest spot of an image is with respect to the darkest part. Similarly, Contrast to Noise Ratio (CNR) quantifies how much an image is visible in the presence of background noise. The Generalized Contrast-to-Noise Ratio (GCNR) is an improvement on the CNR which is robust to dynamic changes. We also looked at the Full Width Half Maximum (FWHM) parameter which gives the sharpness of a peak.

III RESULTS

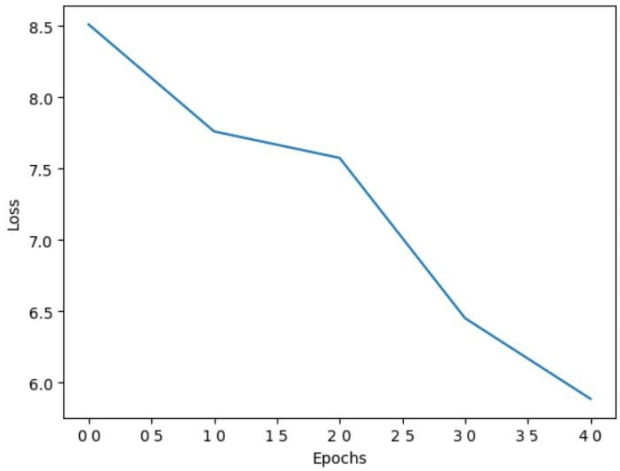

Figure 5: Training Loss of the CNN Model

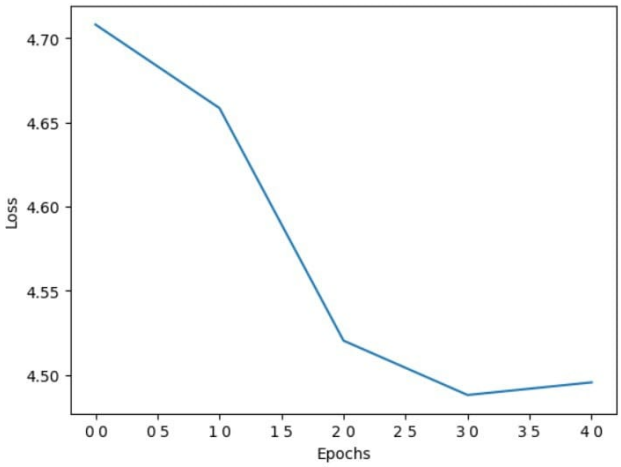

Figure 6: Training Loss of the FCNN Model

Beamforming using CNN and FCNN were compared and studied. Here, we compare the beamformed image from both beamformers with the ground truth as well as a comparison between the images beamformed.

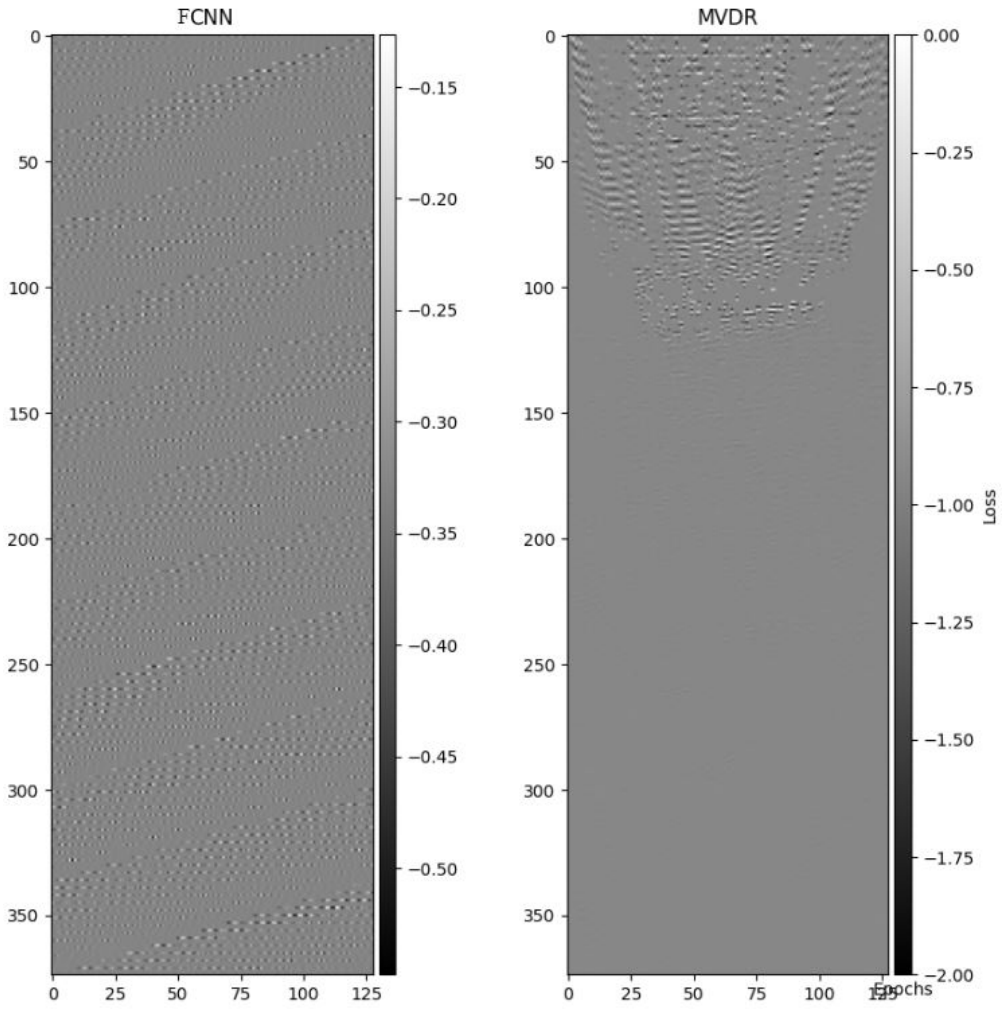

Figure 7: An example of an image obtained via the FCNN beamformer. The MVDR image is the ground truth while the FCNN image is the beamformed image

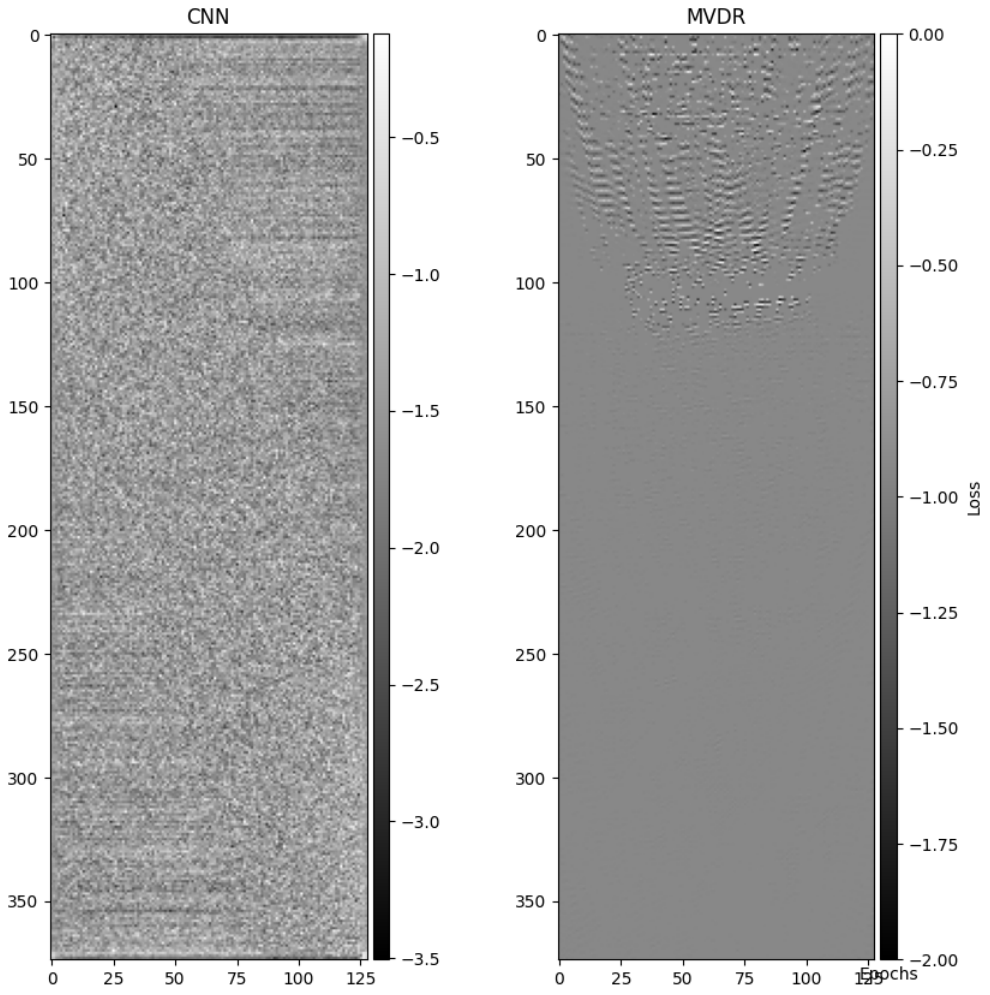

Figure 8: An example of an image obtained via the CNN beamformer. The MVDR image is the ground truth while the CNN image is the beamformed image

IV FUTURE IMPLEMENTATIONS

The image quality metrics have been developed. However, implementation of the quantitative analysis of the model is yet to be done. The image quality metrics have to be estimated for both the CNN and FCNN beamformer. The training and testing data can be more varied. The image quality metrics were obtained from an image in the 200 scanned data set aside for testing. Ideally, an image from the PICMUS dataset should be used for testing. Work also has to be done in modifying the training data to obtain a better model. Comparing the beamforming process using other neural network architectures such as U-Net, MobileNet can be done. In addition, we intend to investigate Graph Neural Networks (GNNs) and Generative Adversarial

Networks (GAN), both of which are capable of implementing Vision Transformers and Diffusion Models, and apply them to Computed Tomography (CT) and Magnetic Resonance Imaging (MRI). Finally, a study into the idea of deleting the loss function and training the neural network based on picture quality metrics such as contrast ratio (CR) and contrast to noise ratio (CNR) can also be implemented.

IV. REFERENCES

[1] H. Liebgott et al. "Plane-Wave Imaging Challenge in Medical Ultrasound". In: *2016 IEEE International Ultrasonics Symposium (IUS)*. 2016, pp. 1–4. DOI: 10.1109/ULTSYM.2016.7728908.

[2] Giulia Matrone et al. "The Delay Multiply and Sum Beamforming Algorithm in Ultrasound B-Mode Medical Imaging". In: *IEEE Transactions on Medical Imaging* 34.4 (2015), pp. 940–949. DOI: 10.1109/TMI.2014.2371235.

[3] Adrien Besson et al. "Compressed delay-and-sum beamforming for ultrafast ultrasound imaging". In: *2016 IEEE International Conference on Image Processing (ICIP)*. 2016, pp. 2509–2513. DOI: 10.1109/ICIP.2016.7532811.

[4] Asraf Mohamed Moubark et al. "Enhancement of contrast and resolution of B-mode plane wave imaging (PWI) with non-linear filtered delay multiply and sum (FDMAS) beamforming". In: *2016 IEEE International Ultrasonics Symposium (IUS)*. 2016, pp. 1–4. DOI: 10.1109/ULTSYM.2016.7728678.

[5] Yuanguo Wang et al. "A dynamic generalized coherence factor for side lobe suppression in ultrasound imaging". In: *Computers in Biology and Medicine* 116 (2020), p. 103522. ISSN: 0010-4825. DOI: https://doi.org/10.1016/j.compbiomed.2019.103522. URL: https://www.sciencedirect.com/science/article/pii/S0010482519303828.

[6] D.G Vince et al. "Comparison of texture analysis methods for the characterization of coronary plaques in intravascular ultrasound images". In: *Computerized Medical Imaging and Graphics* 24.4 (2000), pp. 221–229. ISSN: 0895-6111. DOI: https://doi.org/10.1016/S0895-6111(00)00011-2. URL: https://www.sciencedirect.com/science/article/pii/S0895611100000112.

[7] John Nuyts et al. "Iterative reconstruction for helical CT: a simulation study". In: *Physics in Medicine Biology* 43.4 (Apr. 1998), p. 729. DOI: 10.1088/0031-9155/43/4/003. URL: https://dx.doi.org/10.1088/0031-9155/43/4/003.

[8] Samuel Cahyawijaya. *Biomedical Image Reconstruction: A Survey*. 2023. arXiv: 2301.11813 [eess.IV].

[9] Michalakis A Averkiou. "Tissue harmonic imaging". In: *2000 IEEE Ultrasonics Symposium. Proceedings. An International Symposium (Cat. No. 00CH37121)*. Vol. 2. IEEE. 2000, pp. 1563–1572.

[10] Mathias Fink, Gabriel Montaldo, and Mickael Tanter. "Time-reversal acoustics in biomedical engineering". In: *Annual review of biomedical engineering* 5.1 (2003), pp. 465–497.

[11] Ian Goodfellow et al. "Generative Adversarial Networks". In: *Commun. ACM* 63.11 (Oct. 2020), 139â144. ISSN: 0001-0782. DOI: 10.1145/3422622. URL: https://doi.org/10.1145/3422622.

[12] Brett Byram et al. "A model and regularization scheme for ultrasonic beamforming clutter reduction". In: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 62.11 (2015), pp. 1913–1927. DOI: 10.1109/TUFFC.2015.007004.

[13] Roshan P Mathews and Mahesh Raveendranatha Panicker. "Towards Fast Region Adaptive Ultrasound Beamformer for Plane Wave Imaging Using Convolutional Neural Networks". In: *2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*. 2021, pp. 2910–2913. DOI: 10.1109/EMBC46164.2021.9630930.

[14] Arun Asokan Nair et al. “A Fully Convolutional Neural Network for Beamforming Ultrasound Images”. In: *2018 IEEE International Ultrasonics Symposium (IUS)*. 2018, pp. 1–4. DOI: 10.1109/ULTSYM.2018.8579960.

[15] Arun Asokan Nair et al. “A Deep Learning Based Alternative to Beamforming Ultrasound Images”. English (US). In: *2018 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2018 - Proceedings*. ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings. Institute of Electrical and Electronics Engineers Inc., Sept. 2018, pp. 3359–3363. ISBN: 9781538646588. DOI: 10.1109/ICASSP.2018.8461575.

[16] Dimitris Perdios et al. “Deep Convolutional Neural Network for Ultrasound Image Enhancement”. In: *2018 IEEE International Ultrasonics Symposium (IUS)*. 2018, pp. 1–4. DOI: 10.1109/ULTSYM.2018.8580183.

[17] Adam C Luchies and Brett C Byram. “Deep neural networks for ultrasound beamforming”. In: *IEEE transactions on medical imaging* 37.9 (2018), pp. 2010–2021.

[18] Mayank Katare et al. *Learning while Acquisition: Towards Active Learning Framework for Beamforming in Ultrasound Imaging*. 2022. arXiv: 2208.00464 [eess.SP].

[19] Ben Luijten et al. “Adaptive Ultrasound Beamforming Using Deep Learning”. In: *IEEE Transactions on Medical Imaging* 39.12 (2020), pp. 3967–3978. DOI: 10.1109/TMI.2020.3008537.

[20] Arun Nair et al. “A Generative Adversarial Neural Network for Beamforming Ultrasound Images Invited Presentation”. In: Mar. 2019, pp. 1–6. DOI: 10.1109/CISS.2019.8692835.

[21] Sobhan Goudarzi, Amir Asif, and Hassan Rivaz. “Ultrasound Beamforming using MobileNetV2”. In: *2020 IEEE International Ultrasonics Symposium (IUS)*. 2020, pp. 1–4. DOI: 10.1109/IUS46767.2020.9251565.