2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) | 979-8-3315-8618-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/EMBC58623.2025.11252942

Synthesizing Ultrasound B-mode Images from Subsampled RF Data: A Data-Driven Deep Learning Approach*

Nasrin Sheibani-Asl[1], Sogand Zamanikhah[1], Gregory Czarnota[2,3], & Ali Sadeghi-Naini[1,2,3], *Senior Member, IEEE*

[1]Department of Electrical Engineering and Computer Science, York University, Toronto, ON, Canada
[2]Department of Radiation Oncology, Sunnybrook Health Sciences Centre, Toronto, ON, Canada
[3]Physical Sciences Platform, Sunnybrook Research Institute, Toronto, ON, Canada

***Abstract*— Ultrasound, a widely used and safe imaging modality, utilizes radio frequency (RF) signals obtained from pulse-echo imaging to generate B-mode images, offering a visual representation of internal body organs and tissue microstructures. The processing details and the specific parameters applied for converting RF data to B-mode images in ultrasound devices are typically not made available by manufacturers. In this study, we investigated a convolutional network architecture for conversion of RF data into two different types of B-mode images generated by the Ultrasonix scanners. Additionally, we assessed the network's efficacy in translating subsampled RF data with fewer scan lines into the B-mode images, aiming to expedite the data acquisition and transmission process. Our results on an unseen test set demonstrate the feasibility of reconstructing B-mode images from full and subsampled RF frames with a quality similar to the scanner-generated B-mode images, even when these images undergo multiple nonlinear processing steps. The proposed approach for B-mode image reconstruction from subsampled RF frames permits scanning wider lateral field of view with the same frame rate, offering increased imaging efficiency while maintaining comparable image quality. In addition, it provides a practical solution for reconstructing high-quality B-mode images from engineered RF data in research settings, where scanner processing details are unavailable.**

***Keywords*— Ultrasound imaging, RF signal, Fast B-mode Image Generation, Deep Learning**

I. Introduction

Ultrasound is a non-invasive medical imaging modality that utilizes high-frequency sound waves to create real-time images of internal organs and tissues. Unlike the X-ray or computed tomography (CT), ultrasound does not use ionizing radiation for imaging, making it a safe and accessible modality [1]. The brightness mode (B-mode) is the most commonly used ultrasound imaging mode that produces 2D or 3D grayscale images of the scanned tissues. In the B-mode imaging, ultrasound pulses are transmitted by a transducer and the generated echoes from the pulse-tissue interactions are received and processed to image the underlying tissue structures. The grayscale intensity in the resulting images are associated with the magnitude of the received echoes, with brighter and darker areas indicating the higher and lower magnitudes, respectively [1].

A challenge associated with utilizing ultrasound imaging in research applications lies in the conversion of RF data to B-mode images while preserving a quality comparable to the images produced by the scanner. This is important, for example, in applications where the utility and efficacy of altered or engineered RF data should be assessed through a comparison between the associated B-mode images and those originally generated by the scanner [2]. Broadly, the RF to B-mode conversion process involves envelope detection, log compression, filtering, and nonlinear intensity mapping [3]. However, some ultrasound images may undergo additional refinement steps to achieve enhanced visualization and a smoother appearance with reduced granularity.

The ultrasound RF frames store the scan lines acquired during the pulse-echo imaging. Given that these scan lines are not necessarily parallel to each other in the physical space, their geometrical position in the image space needs to be arranged based on the probe's geometry, before interpolation and further image processing. This process is usually called scan conversion. For example, the Ultrasonix scanners output two types of B-mode images, namely pre-scan-converted images (bpr files) and post-scan converted images (b8 and b32 files). Specifically, the bpr files are pre-scan-converted, envelope detected, log compressed and scaled to 8-bit b-mode image data, with no further processing. These images are sent directly from the cine buffer of the device. The b8 files, in contrast, are post-scan-converted with further processing applied, i.e., filtering, and adjustments to the greyscale map parameters (brightness, contrast, and gamma parameters) [4]. The exact procedures involved in generating the final B-mode images are device-specific and their details and specific parameters are typically not made available by the manufacturers, introducing an additional layer of complexity to the B-mode image reconstruction task.

In addition, a significant stride in ultrasound imaging could be achieved by proposing a method that not only converts RF data to B-mode images but also reduces the minimum required scan lines for maintaining a comparable image quality. While modern ultrasound devices perform well in real-time B-mode imaging, potential still exists for further improvement. Reducing the number of scan lines necessary to cover a fixed image sector can hold promise in optimizing imaging efficiency. This approach enables the

*This study was supported by research grants from the Natural Sciences and Engineering Research Council (NSERC) of Canada, Terry Fox Foundation, and the Lotte and John Hecht Memorial Foundation. A. Sadeghi-Naini holds the York Research Chair in Quantitative Imaging and Smart Biomarkers, and an Early Researcher Award from the Ontario Ministry of Colleges and Universities.

scanning of a larger image sector within an equivalent time frame, or expediting the data acquisition and transmission process for a fixed sector size [5].

Deep learning has emerged as a powerful method with significant potential in various image processing and generation tasks [6, 7]. A few studies have utilized deep learning approaches to transform RF data into B-mode images. For instance, Ren utilized RF data in conjunction with 11 additional parameters as input to a deep learning framework to generate B-mode images [8]. In another study, researchers employed deep learning to interpolate subsampled RF data, and subsequently applied the conventional delay-and-sum (DAS) reconstruction method to the interpolated RF data to generate the B-mode images [5]. The deep learning model in [5], however, is not end-to-end. Also, the proposed method cannot reconstruct comparative B-mode images to those generated by a scanner when the exact processing steps and the applied parameters are not available.

In this paper, we have investigated a modified U-Net framework for fast generation of high-quality B-mode images using intact or subsampled RF data. This network has been applied as the generative model to reconstruct two types of B-mode images, including the pre-scan converted (bpr) and post-scan converted (b8). The model has been trained on a clinical dataset acquired from patients with suspicious breast lesions. The reconstructed B-mode images in an independent test set have been compared with ground-truth counterparts generated by the scanner using both pixel-wise evaluation metrics and performance metrics from a benign/malignant classification task. The results show high potential for the proposed method in processing the intact and subsampled RF data and reconstructing high quality B-mode images that are very comparable to the scanner-generated images.

II. Materials and Methods

A. Dataset

The dataset applied in this study for the image reconstruction task includes the ultrasound RF data and B-mode images (bpr and b8) obtained from 151 patients with breast lesions, following the institutional research ethics board approval. The ultrasound scans were performed using an RF-enabled Sounix Touch (Ultrasonix, Vancouver, Canada) system utilizing an L14-5/60 transducer, operating at a nominal frequency of 10 MHz, with an imaging depth and lateral field of view of 4 cm and 6 cm, respectively. Each patient was scanned at multiple imaging planes spanning the entire lesion volume (with multiple RF/B-mode pairs acquired at each plane), as well as through a single continuous sweep over the lesion volume. The dataset was randomly split at patient level into a training set (113 patients), and an independent test set (38 patients). From the large pool of RF frames and B-mode images in the training set, 1,000 pairs were chosen randomly for training the reconstruction models, with another 200 pairs selected as the validation set for monitoring and optimizing the models during the training process. For independent evaluation of the models, 300 RF/B-mode pairs were randomly selected from the unseen patients in the test set. The same training, validation, and test data were used for all the reconstruction models in this study.

For the classification task, the same 113 and 38 patients from the image reconstruction experiments were included in the training and test sets, respectively. Additionally, the publicly available breast ultrasound images (BUSI) dataset [9], containing 647 (benign: 437; malignant: 210) B-mode images, was incorporated. The training set for the classification model comprised 1,128 B-mode images, with 581 (benign: 393; malignant: 188) and 547 (benign: 249; malignant: 298) images from the BUSI dataset and our previous training set of 113 patients, respectively. From this training set, 10% was randomly selected as the validation set for monitoring the classification model during training, while maintaining the balance between the selected samples from the two datasets and a similar proportion of benign/malignant images in the training and validation sets. The test set included five groups of 38 (benign: 20; malignant: 18) images (one image per test-cohort patient). The five groups were associated with the scanner-generated B-mode images and the corresponding images generated using the trained reconstruction models. A separate set of 66 (benign: 44; malignant: 22) B-mode images from the BUSI dataset was also used for further evaluation of the classification model.

B. Data Preprocessing

The acquired RF frames comprised of 2080 samples × 510 lines, while the scanner-generated B-mode images were of size 520×510 and 373×541 pixels for the bpr and b8 files, respectively. For processing by the U-net model, the B-mode images and RF frames were resampled to 2048×512 pixels using linear interpolation. For the experiments with subsampled RF frames, the RF frames were subsequently down-sampled to 2048×256, 2048×128, and 2048×64 by retaining a scan line every 2, 4, and 8 lines in the frame, respectively. The pixel values in the B-mode images and RF frames were normalized to the range [-1,1] using the maximum values of their dynamic range.

For the classification task, the B-mode images were resized to 224×224 pixels and normalized to the range [-1,1] to serve as input to the pretrained DenseNet model (described in *Section D*).

C. Image Reconstruction Model

Fig. 1 demonstrates the scheme of the proposed network architecture for converting RF frames to high-quality B-mode images. The proposed architecture is based on the U-Net model [9] and composed of down-sampling and up-sampling paths connected by skip connections. The input undergoes the down-sampling path, where its size decreases step-by-step by a factor of 2, either solely in the first dimension or in both dimensions. This down-sampling process continues until reaching the bottleneck layer, after which the operation is reversed for up-sampling.

The configuration of the last layer in the network's up-sampling path is determined based on the size of the input RF

frame. The role of this layer is to interpolate the data to match the output dimensions with those of the original B-mode image for subsequent comparison through the loss function. In models where the input RF frame size is (2048, 512), the final layer employs a transpose convolution with a kernel size of (4, 1) and a stride of (2, 1). For the RF size of (2048, 256) and (2048, 128), a kernel size of (4, 4) and a stride of (2, 2) and (2, 4) are applied, respectively. Finally, for the input size of (2048, 64) a kernel size of (8,8) and a stride of (2,8) is used.

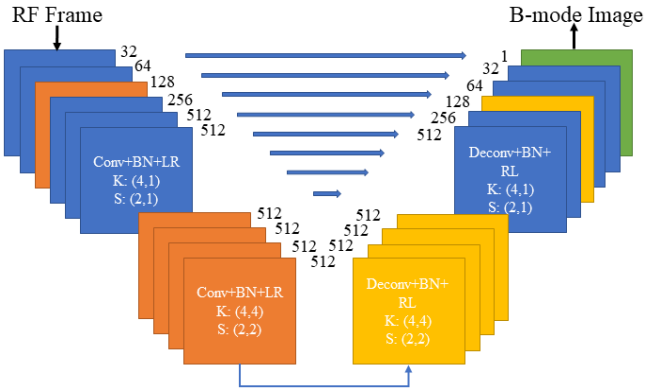

Figure 1. The proposed network architecture for RF to B-mode conversion. The number of the kernels, their sizes (K) and the strides (S) are given for each block. Same colors indicate same configuration, kernel size and stride. Conv: convolution, Deconv: deconvolution, BN: batch normalization, LR: Leaky ReLU, RL: ReLU.

The slope parameter for the Leaky ReLU activation function was set to 0.2. Prior to training, the network weights were initialized using a normal distribution with a mean of zero and standard deviation 0f 0.02. The training was performed using an Adam optimizer [10] with an L1 loss function, and an initial learning rate of 0.001. The learning rate was decayed after each epoch according to the following equation:

$$\mathrm{lr} = 0.001 \times \left(1 - \frac{\mathrm{n}}{150}\right) \quad (1)$$

where n represents the epoch number during the training process. The network was trained for 100 epochs with a batch size of 1. The performance of the models on the training and validation sets was monitored during the training process. The model with the lowest validation loss was preserved and utilized for subsequent evaluations.

D. Evaluation Metrics

The performance of the reconstruction models in generating high-quality B-mode images compared to those generated by the scanner was initially assessed through pixel-level evaluation metrics, including the mean absolute error (MAE), normalized root mean squared error (NRMSE) with min-max normalization, NRMSE with Euclidean normalization, and peak signal to noise ratio (PSNR) [11]. These metrics were applied to the output images after denormalization.

Subsequently, the efficacy of the generated B-mode images was evaluated in classifying benign versus malignant breast lesions and compared to the original images generated by the scanner. A DenseNet-121 model [12] pre-trained on the ImageNet dataset was utilized for image classification. The backbone of the pre-trained model was applied as a feature extractor, while fine-tuning the fully connected layers with a sigmoid activation function for benign/malignant classification on the dataset described in *Section A*. The model was fine-tuned for 30 epochs using the Adam optimizer with a binary cross-entropy (BCE) loss function, a learning rate of 2e-3 and a weight decay of 1e-3 to help prevent overfitting. Additionally, early stopping with a patience of 10 epochs was applied to terminate model training if no improvement in validation performance was observed. The accuracy and area under the curve (AUC) were used as the classification performance metrics in this study for comparing the efficacy of the B-mode images.

III. Results

Fig. 2 presents representative B-mode images generated by the network using test RF frames with 512 scan lines. The quality of the synthesized bpr and b8 images are quite comparable with their counterparts originally generated by the scanner. As shown in Fig. 2, the bpr images exhibit a granular appearance in contrast to the smoothed b8 images. This difference arises from the additional processing steps applied to the b8 images by the ultrasound device.

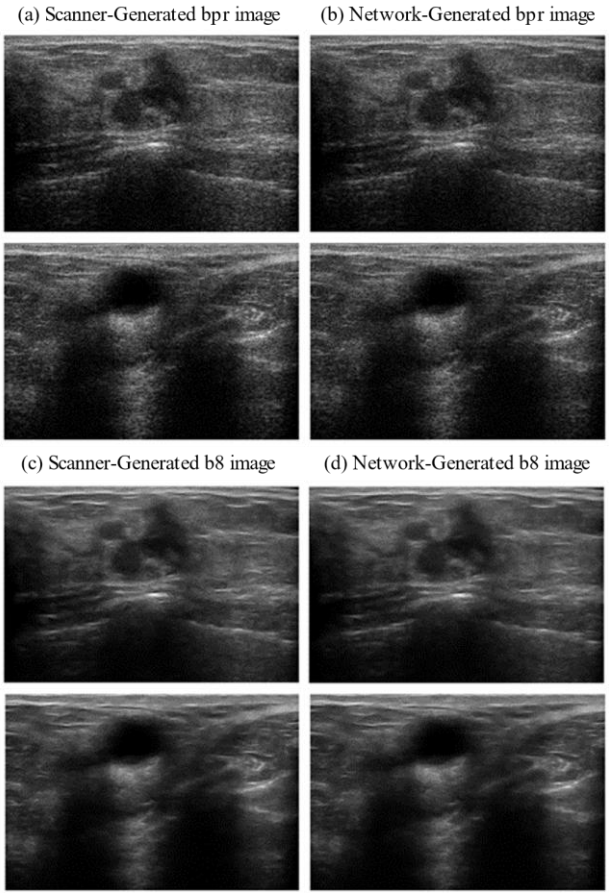

Figure 2. B-mode images generated by the scanner and the proposed network for two representative RF frames with 512 scan lines, in the independent test set.

Table I provides a summary of the average pixel-level evaluation metrics obtained on the independent test set,

comparing the network-generated and the original scanner-generated B-mode images. The metrics have been computed on denormalized images, with pixel values ranging from 0 to 255. Based on the metrics presented in Table I, the proposed architecture has a decent performance in converting RF data to B-mode images, for both the pre-and post-scan-converted B-mode images.

Table I. Results of pixel-level evaluation on the test set (average ± standard deviation), comparing the B-mode images synthesized using full RF frames (512 scan lines) and the original B-mode images generated by the scanner.

Metric	Pre-scan-converted (bpr) images	Post-scan-converted (b8) images
MAE	0.852±0.121	1.36±0.28
$NRMSE_{MinMax}$	0.007±0.001	0.011±0.002
$NRMSE_{Euclidean}$	0.026±0.005	0.042±0.007
PSNR	46.48±1.551	42.66±1.70

For comparison, the bpr images were also reconstructed using the conventional method based on envelope detection of RF signals. The process included Hilbert transform-based envelope detection applied along each RF line, logarithmic compression ($20\log_{10}$), dynamic range clipping (−50 dB to 0 dB), and linear scaling to the range [0, 255]. Fig. 3 presents a representative image generated using this method. Quantitatively, the average MAE on the test set was 38.77 ± 6.48, $NRMSE_{MinMax}$ was 0.29 ± 0.05, $NRMSE_{Euclidean}$ was 1.01 ± 0.22, and the PSNR was 14.68 ± 1.7. These metrics show a notable discrepancy between the images generated through the conventional method and the scanner-generated bpr images. The contrast is different between these images due to the unknown specifications of the scanner's dynamic range adjustments. In addition, the speckle pattern appears sharp and grainy, and vertical line artifacts are visible near the top and bottom of the generated image. The quality of the output in this method is influenced by the choice of compression parameters and dynamic range. However, even with various tunings, it still falls short of replicating the image quality achieved by the scanner. It is also worth noting that similar b8 images cannot be produced using this methos due to unknown specifications of the additional processing steps applied by the scanner, e.g., resampling, filtering, and nonlinear intensity mapping, to generate these images.

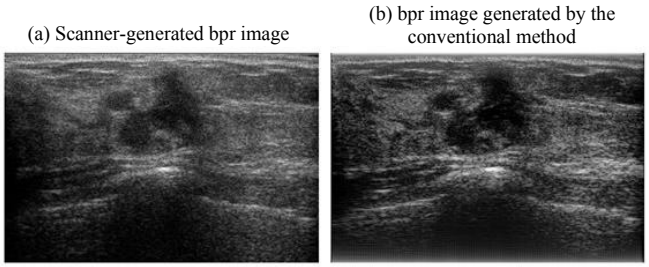

Figure 3. Pre-scan-converted (bpr) images generated by the scanner and the conventional method.

Given that converting RF data to the b8 images is a more challenging task, we focused the subsequent experiments on the conversion of subsampled RF data exclusively to the post-scan-converted B-mode images. Table II reports the average pixel-wise evaluation metrics for the B-mode image generation task using the subsampled RF data in the test set. As the number of scan lines decreases, there is an inevitable impact on the quality of the synthesized B-mode image; nevertheless, such quality impact is minimal for the synthesized images based on the results presented in Table II. The impact becomes more pronounced with 64-scan-line RF frames, where the column count has been reduced to one-eighth of its original value.

Fig. 4 demonstrates the synthesized images for representative subsampled RF frames with different number of scan lines. While the figure shows a slight blurring as the number of scan lines decreases in the subsampled RF frame, the quality of the synthesized images is still comparable with those generated by the scanner.

Table II. Comparison of the B-mode images generated by the network using subsampled RF data in the test set with the original B-modes (b8) in terms of different pixel-level evaluation metrics (average ± standard deviation).

Metric	RF Lines: 256	RF Lines: 128	RF Lines: 64
MAE	1.73±0.28	2.44±0.35	3.29±0.54
$NRMSE_{MinMax}$	0.014±0.001	0.020±0.001	0.027±0.003
$NRMSE_{Euclidean}$	0.053±0.005	0.075±0.004	0.101±0.007
PSNR	40.54±1.13	37.49±1.02	34.91±1.20

Table III presents the results of benign/malignant classification experiments. The B-mode images generated by the network from the full and 256-scan-line RF frames demonstrate on-par accuracy and AUC with the scanner-generated B-mode images. The classification performance drops slightly with the images generated from 128-scan-line RF frames, and further with those generated from 64-scan-line frames. The breast lesions included in the test set of this study are mostly complicated cases from BI-RADS 4 and 5 categories that are challenging to classify. This can be evident by comparing the performance of the classification model on the BUSI test set, where an accuracy of 93.9% and an AUC of 0.93 was achieved. Nevertheless, the aim of the classification experiments in this study is to compare the efficacy of the network-generated B-mode images with scanner-generated counterparts in differentiating between benign and malignant lesions, with results showing strong agreement in classification performance.

Table III. Comparison of the accuracy and AUC values for benign/malignant classification using the original B-mode images generated by the scanner and the images generated by the network from the full and subsampled RF data.

Test Images	Train. Acc (%)	Val. Acc (%)	Test Acc (%)	Test AUC
Original B-mode images	98.2	88.6	76.3	0.76
B-mode from 512-line RF	---	---	76.3	0.76
B-mode from 256-line RF	---	---	76.3	0.76
B-mode from 128-line RF	---	---	73.7	0.73
B-mode from 64-line RF	---	---	71.1	0.71

IV. Conclusion

In this study, a modified U-Net framework was developed and explored to translate ultrasound RF data into two distinct types of B-mode images: pre-scan-converted and post-scan-converted B-mode images. We also investigated the impact of reducing the number of scan lines in the RF data on the

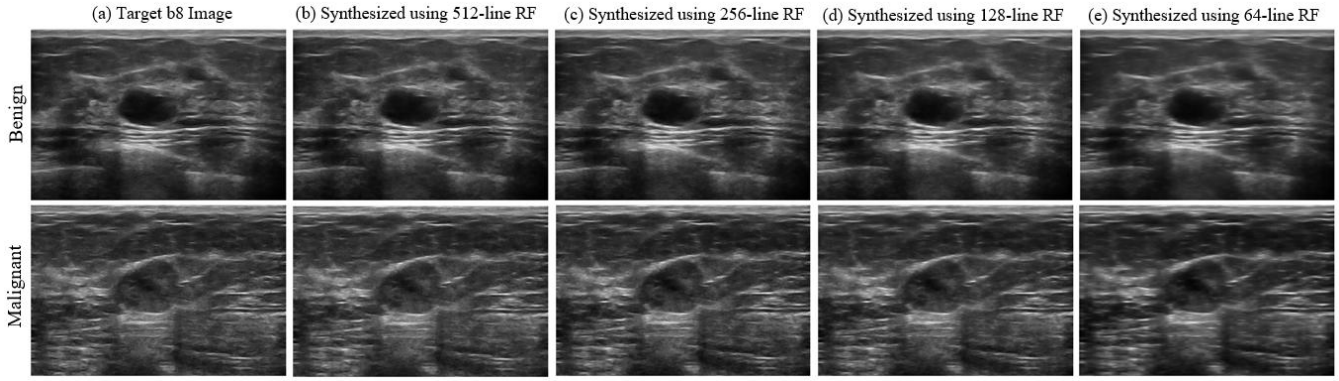

Figure 4. Representative B-mode images generated using subsampled RF data in the test set: (a) target images produced by the scanner; (b) images generated using the full RF frame; (c-e) images generated using the half, one-fourth, and one-eighths of the original scan-line count.

quality of generated B-mode images. Our results demonstrate the feasibility of reconstructing B-mode images with a similar quality to those produced by the scanner, even when the scanner-generated images have gone through multiple nonlinear processing steps, as observed in the post-scan-converted B-mode images. Additionally, the proposed method can potentially accelerate the data acquisition and transmission process in B-mode imaging by reducing the number of required scan lines while maintaining comparable image quality. This approach enables scanning wider lateral fields of view at the same frame rate, while also offering a practical method for reconstructing high-quality B-mode images from engineered RF data in research settings, where scanner processing details are unavailable.

Compliance with Ethical Standards

This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of York University (Date: 2023/09/18).

References

[1] J. R. Costello, H. Arif, B. Kalb, and D. R. Martin, "Ultrasound," in *An Introduction to Medical Physics*, M. Maqbool Ed. Cham: Springer International Publishing, 2017, pp. 329-370.

[2] M. Dai, S. Li, Y. Wang, Q. Zhang, and J. Yu, "Post-processing radio-frequency signal based on deep learning method for ultrasonic microbubble imaging," *Biomedical engineering online,* vol. 18, pp. 1-19, 2019.

[3] S. Azizi *et al.*, "Transfer learning from RF to B-mode temporal enhanced ultrasound features for prostate cancer detection," (in eng), *Int J Comput Assist Radiol Surg,* vol. 12, no. 7, pp. 1111-1121, Jul 2017, doi: 10.1007/s11548-017-1573-x.

[4] A. H. L. Dau, "RF2B - RF to B-Mode Ultrasound," Bachelor's degree, Faculty of Informatics, Technical University of Munich, Munich, Germany, 2010.

[5] Y. H. Yoon, S. Khan, J. Huh, and J. C. Ye, "Efficient B-mode ultrasound image reconstruction from sub-sampled RF data using deep learning," *IEEE transactions on medical imaging,* vol. 38, no. 2, pp. 325-336, 2018.

[6] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review," *Medical image analysis,* vol. 58, p. 101552, 2019.

[7] H. Taleghamar, S. A. Jalalifar, G. J. Czarnota, and A. Sadeghi-Naini, "Deep learning of quantitative ultrasound multi-parametric images at pre-treatment to predict breast cancer response to chemotherapy," *Scientific reports,* vol. 12, no. 1, p. 2244, 2022.

[8] J. Ren, "From RF signals to B-mode images using deep learning," Master's degree, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, Stockholm, Sweden, 2018.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, 2015: Springer, pp. 234-241.

[10] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, (ICLR 2015)*, San Diego, CA, USA, May 7-9, 2015.

[11] S. Van der Walt *et al.*, "scikit-image: image processing in Python," *PeerJ,* vol. 2, p. e453, 2014.

[12] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700-4708.