

Deep Learning Framework for B-Mode Ultrasound Image Reconstruction

Mahsa Mikaeili¹

mahsa.mikaeili@okan.edu.tr

Istanbul Okan University, Mechatronic Engineering, Istanbul,
Turkey¹Hasan Şakir Bilge^{2,3}

bilge@gazi.edu.tr

Gazi University, Biomedical Calibration and Research Center,
Ankara, Turkey²Gazi University, Electrical and Electronics Engineering, Ankara,
Turkey³

Abstract— We present a U-Net-based pipeline for B-mode ultrasound image reconstruction that ingests post-processed RF data, predicts a log-compressed image, and then performs an interpolation for display. To identify effective design choices, we compare six configurations formed by two loss functions: mean-squared error and compound MMUAE+TV+gradient loss, and three output activation functions: linear, ReLU, and tanh. Evaluation with PSNR, SSIM, and visual inspection of scan line profiles and B-mode images indicates that the activation function is the dominant factor. Tanh function consistently preserves lesion boundaries, maintains realistic speckle, and avoids dynamic-range saturation; linear function is acceptable but yields softer edges; ReLU degrades contrast owing to negative-value clipping. Switching from MSE to the compound loss function produces only modest changes, suggesting that regularization is secondary to activation choice in this setting. Overall, a tanh-activated U-Net with a simple MSE objective offers a strong accuracy-complexity trade-off for reconstruction.

Keywords—Ultrasound Image; raw RF data; reconstruction; deep neural network

I. INTRODUCTION

Ultrasound imaging (US) is among the most widely used medical imaging modalities due to its noninvasiveness, lack of ionizing radiation, affordability, and real-time imaging capability. Because ultrasound offers key advantages, it has received sustained research interest, and significant work has targeted better image quality, including higher resolution, improved contrast, and greater signal-to-noise ratio (SNR).

One of the key approaches to enhancing US image quality is beamforming techniques. Conventional beamforming methods include delay-and-sum (DAS), minimum variance beamforming (MVB), and short-lag spatial coherence (SLSC). However, these methods are often limited by reduced image quality and high computational complexity [1].

With the emergence of deep neural networks (DNNs) and their demonstrated high performance in various tasks, including classification, segmentation, and detection, some researchers have adopted this powerful approach for beamforming applications and enhancing US resolution. In [2], the authors classified these approaches into three categories based on the data type utilized: image-based, pre-beamforming, and post-beamforming methods. Among these approaches, image-based methods apply deep neural networks (DNNs) at the final stage of the beamforming process. Pre-beamformed RF-based methods operate directly on raw RF channel data before beamforming. In contrast, post-beamformed methods utilize RF data after beamforming but before envelope detection and logarithmic compression. The block diagram in Fig. 1 illustrates the US image reconstruction process and data categories between stages. In [3], the authors proposed a convolutional neural network (CNN)-based two-step reconstruction method to address the diffraction issue in ultrafast ultrasound imaging. In [4], the authors employ a U-Net architecture to learn the nonlinear mapping between low-quality and high-quality images, focusing on reconstructing high-quality RF images and preserving speckle patterns. Also, in [5], the authors propose an

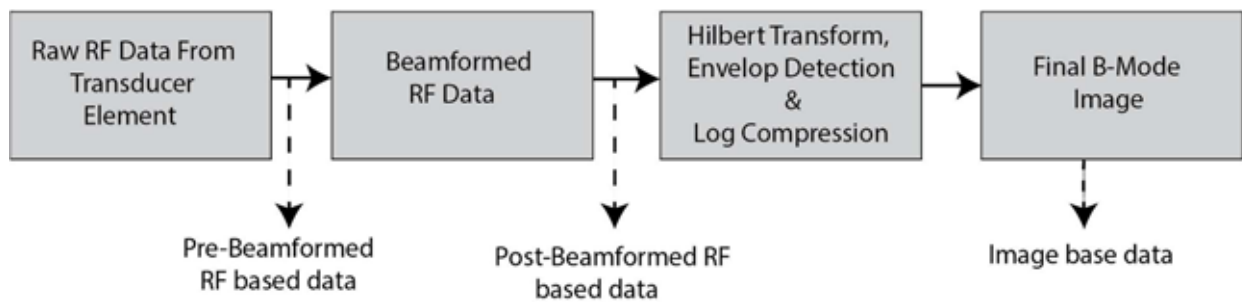


Fig.1. Block diagram of the US image reconstruction process, including data categories between stages.

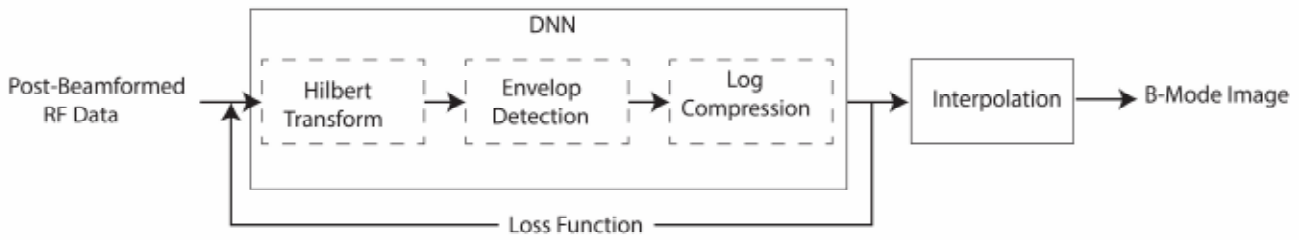


Fig.2. Block diagram of the proposed method. The last processing part of the US image reconstruction process is replaced with a DNN.

end-to-end deep learning framework for plane wave imaging (PWI) that integrates Spatial Matched Filter (SMF)-based beamforming with a residual U-Net for speckle reduction, followed by envelope detection and log compression to enhance image quality. In another study, authors propose a deep learning-based beamforming method for PWI that uses a Fully Convolutional Neural Network with multi-resolution wavelet decomposition and step-by-step training to map RF channel data to high-quality images defined by a sharp Gaussian Point Spread Function (PSF), enabling sharper reconstructions without reducing frame rate [6]. In [7], the authors apply a Generative Adversarial Network (GAN) for multi-focus ultrasound image recovery that maintains frame rate while achieving constant lateral resolution. Their approach learns a mapping from single-focus to multi-focus images using adversarial and Mean Squared Error (MSE) losses, combined with a boundary-seeking strategy to improve training stability.

Evaluated another study is [8]. This work introduces two deep learning-based methods: Inception for DW (Diverging Wave) Network (IDNet), an inception CNN for high-quality diverging wave image reconstruction from few transmissions, and a GAN-based approach for multi-focus ultrasound recovery, enhancing resolution and contrast while maintaining frame rate without requiring raw channel data. Also, in [9], a self-supervised deep neural network that works with an encoder and decoder strategy with sparse regularization-based loss is proposed for ultrafast plane-wave ultrasound image reconstruction from RF data, achieving high-quality results with substantially reduced computational time compared to traditional methods. In [10], the authors propose a neural network-based beamforming framework for ultrasound that uses normalization-independent loss functions to reconstruct speckle-reduced B-mode images from sub-aperture RF signals, demonstrating effective generalization from simulated training to real-world imaging while preserving resolution and detail. Another investigated work is [11]. This work proposes a frequency-domain DNN approach for suppressing off-axis scattering in ultrasound channel data, using short-time Fourier transform processing and per-frequency network training to achieve enhanced contrast and side-lobe suppression compared to conventional beamforming methods. Also, in [12], authors introduce a deep learning-based reconstruction framework that uses an encoder-decoder CNN to generate high-quality B-mode ultrasound images from temporally and spatially sub-sampled channel data. They claim their method can achieve resolution and contrast comparable to advanced beamforming methods while reducing acquisition requirements. In [13], the authors

propose a two-stage deep learning framework combining a fully convolutional PixelNet and a conditional GAN to reconstruct high-quality ultrasound images from single plane-wave RF data by reducing angle dependence through learned data transformation.

This study aims to evaluate the performance of a U-Net architecture for reconstructing ultrasound B-mode images while omitting last stages of conventional pipelines, and to systematically assess the impact of the loss function and output activation on reconstruction quality. We benchmark two loss formulations (mean-squared error and a compound MMUAE+TV+gradient loss) and three activations: linear, ReLU, and tanh. The remainder of the paper is organized as follows: Section II details the methodology, network configuration, datasets, and the evaluation protocol for losses and activations; Section III presents and discusses the results; and Section IV concludes the study.

II. MATERIALS AND METHODS

This study employs a U-Net architecture to reconstruct a US B-mode image. Post-beamformed data undergoes multiple processing stages (Fig.1) in the conventional US image reconstruction process before producing a B-mode image. These stages typically include the Hilbert transform for quadrature signal generation, envelope detection to extract the signal magnitude, logarithmic compression to adjust the dynamic range, and finally, interpolation to improve image resolution. All stages from the Hilbert transform through logarithmic compression are omitted in the proposed approach. Instead, the U-Net directly reconstructs log-compressed data from the post-beamformed data. Finally, an interpolation step is applied to generate the B-mode image. This approach simplifies the US image reconstruction process, replacing the last processing part with a DNN. A schematic representation of the proposed method is shown in Fig. 2.

A. Applied Network

A U-Net architecture was employed to map post-beamformed ultrasound data directly to log-compressed data. The network adopts a standard encoder-decoder configuration, comprising four down-sampling stages in the encoder and four corresponding up-sampling stages in the decoder. Each encoder stage consists of two 3×3 convolutional layers followed by 2×2 max-pooling, with the number of feature channels doubled at each level to progressively capture higher-level representations. The decoder mirrors this process, utilizing up-sampling and

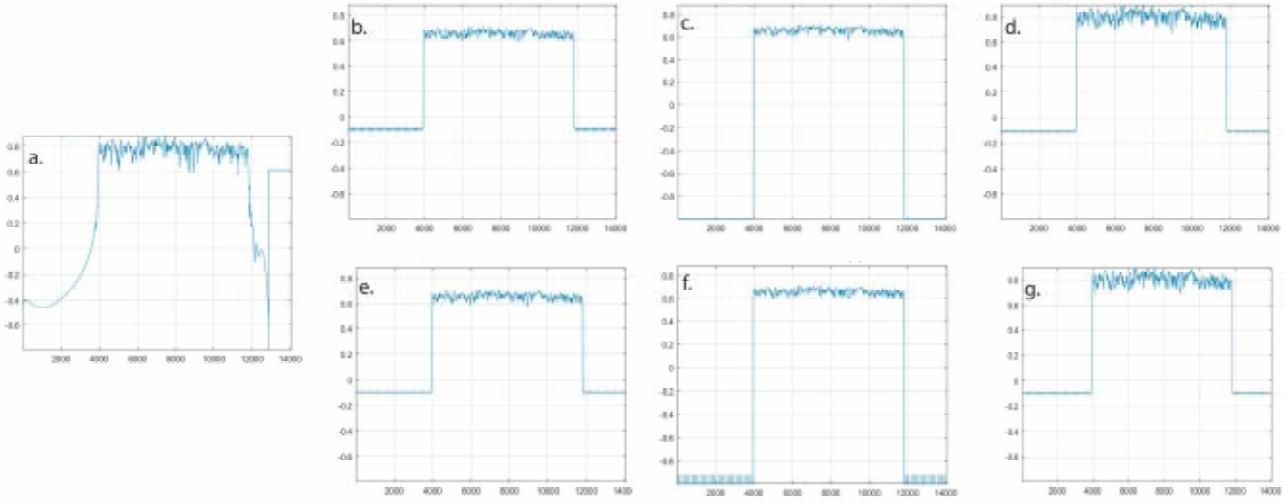


Fig.3. (a) Ground truth signal. A typical scan line reconstructed with (b) MSE loss and Linear activation function, (c) MSE loss and ReLU activation function, (d) MSE loss and Tanh activation function, (e) MMUAE+TV+Grad loss and Linear activation function, (f) MMUAE+TV+Grad loss and ReLU activation function, (g) MMUAE+TV+Grad loss and Tanh activation function.

skip connections to integrate high-resolution spatial information from the encoder, followed by convolutional refinement. To investigate the effect of nonlinearity on the reconstruction task, the model was trained under different activation function configurations, including linear, ReLU, and tanh, applied within the convolutional blocks and at the output layer. The final layer incorporates a 1×1 convolution, where the activation choice determines the output range and intensity scaling. Furthermore, two different loss functions were systematically evaluated. One of these is the conventional mean squared error (MSE), which penalizes voxel-wise intensity discrepancies between predictions and targets, Eq.1

$$L_{MSE}(y_{true}, y_{pred}) = \frac{1}{N} \sum_{i=1}^N (y_{true}^{(i)} - y_{pred}^i)^2 \quad (1)$$

The second loss function is a composite formulation that combines masked mean absolute error (MMUAE), total variation (TV), and gradient similarity (MMUAE+TV+Grad), Eq.2, designed to enforce smoothness while preserving structural edges. This comparative analysis enables a rigorous assessment of the combined impact of loss function design and activation strategy on ultrasound image reconstruction performance.

$$L_{total} = L_{MMUAE} + \alpha L_{TV} + \beta L_{grad} \quad (2)$$

In this formulation, $\alpha=0.01$ regulates the degree of smoothness enforced by the TV loss, while $\beta=0.01$ balances the preservation of structural boundaries.

B. Data Acquisition and Experimental Setup

To generate the data required for training and testing the proposed model, the Field II simulation program [14] was employed. In total, 542 US images were simulated, representing phantoms containing cysts and regions of high scatter density. The diameters of these regions were set to 6, 5, 4, 3, or 2 mm, covering a range of clinically relevant lesion sizes to assess the model's ability to reconstruct structures of varying scales. Each

simulated phantom contained one, two, or three structures, which could be reflectors, cysts, or a combination of both, thereby introducing structural diversity and improving the model's generalization capability. This simulation setup was designed to mimic realistic ultrasound imaging conditions by incorporating variations in target size, scatterer distribution, and structural composition, enabling robust evaluation of the proposed reconstruction method.

The simulation parameters were a linear array transducer with a center frequency of 3.5 MHz, 64 active elements, and a scan line of 128. Both input and output datasets were normalized to the range $[-1, 1]$ before training to ensure numerical stability. Of the simulated dataset, 100 images were reserved for testing, and the remainder was used for training. Model training was performed with a learning rate of 0.0001 over 25 epochs. The experiments were conducted on a workstation with an Intel Core i7 8th-generation CPU (4 GHz, 32 GB RAM) and an NVIDIA GeForce GTX GPU.

III. RESULTS AND DISCUSSION

In this study, by employing the U-Net structure, we evaluate the performance of multiple combinations of loss functions (MSE, MMUAE+TV+Grad) and activation functions (Linear, ReLU, Tanh) in reconstructing a US B-mode image (Fig.3). Based on Fig.3.b, by employing MSE+Linear, the overall shape is recovered, but mild ringing appears near the edges, and the plateau is slightly uneven. However, MSE+ReLU clips negative values, biasing the baseline and flattening transitions, which explains its poor fidelity. Also, in Fig.3.d best matches the reference, preserving edge sharpness and plateau level while keeping outputs bounded. Using the compound MMUAE+TV+Grad loss with Linear (e) yields a profile similar to (b) but a bit smoother on the plateau due to TV regularization, at the cost of slight edge softening. The ReLU variant with the compound loss (f) again suffers from clipping and distorted transitions. Finally,

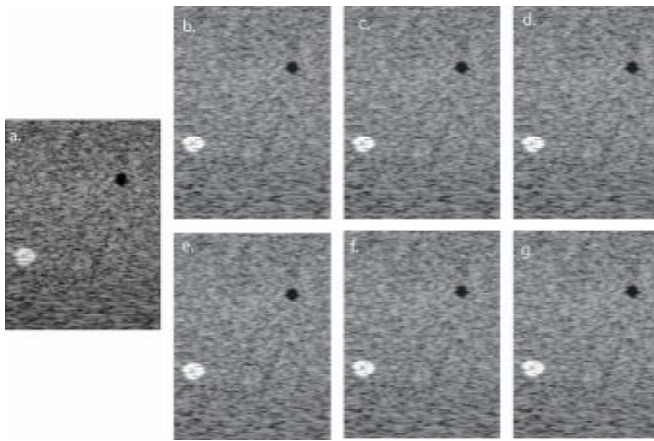


Fig.4. (a) Ground truth Image. Images reconstructed with (b) MSE loss and Linear activation function, (c) MSE loss and ReLU activation function, (d) MSE loss and Tanh activation function, (e) MMUAE+TV+Grad loss and Linear activation function, (f) MMUAE+TV+Grad loss and ReLU activation function, (g) MMUAE+TV+Grad loss and Tanh activation function.

Table. 1. Acquired results with combinations of loss and activation functions.

	MSE			MMUAE+TV+Grad		
	Linear	ReLU	Tanh	Linear	ReLU	Tanh
PSNR	18.281	9.321	19.119	18.251	9.597	19.160
SSIM	0.601	0.552	0.631	0.605	0.551	0.623

MMUAE+TV+Grad+Tanh (g) closely tracks (d), suppressing small ripples with only minimal loss of edge crispness.

Also, Fig.4 and Table 1 demonstrate qualitative and quantitative results of the applied method. The tanh outputs MSE+tanh (d) and MMUAE+TV+Grad+tanh (g) most closely resemble the ground truth: the anechoic cyst interior is uniformly dark, its boundary is crisp with minimal haloing, and the bright point target is compact without blooming. Background speckle looks statistically consistent, which matches their top metrics (PSNR $\approx$ 19.1–19.2 dB, SSIM $\approx$ 0.623–0.631). The linear variants (b) and (e) recover structure but show slightly elevated background variance and mild edge ringing; cyst edges are a touch softer, and the point target is a bit more spread, consistent with the modestly lower scores. The ReLU cases (c) and (f) clip low-intensity regions, compressing dynamic range: speckle appears washed, the cyst shows noticeable fill-in, and edges and point targets lose contrast. This visual degradation aligns with the significant metric drop. Comparing losses within the same activation, MMUAE+TV+Grad mainly smooths small ripples in the speckle (e.g., g vs d, e vs b) but also slightly softens edges, which is why metric changes are minor. Overall, the tanh activation yields the best compromise of edge sharpness, speckle fidelity, and lesion contrast; linear is acceptable; ReLU is unsuitable for B-mode reconstruction because of dynamic-range clipping.

IV. CONCLUSION

Using a U-Net architecture, we reconstruct B-mode images from post-beamformed RF inputs, predict the log-compressed output, and then apply interpolation to obtain the final B-mode image. Across the examined settings, the activation function was the primary determinant of image quality: the tanh function

produced the most faithful reconstructions with crisp boundaries and realistic speckle; the linear function was acceptable but slightly softer; and the ReLU function consistently degraded contrast due to dynamic-range clipping. In contrast, changing the loss from MSE to the compound MMUAE+TV+Grad yielded only marginal differences, reducing minor ripple artifacts while mildly softening edges. For this post-beamformed RF to log-compressed data pipeline followed by interpolation, a tanh-activated U-Net with a simple MSE loss (optionally with light TV regularization) offers the best accuracy-complexity trade-off.

ACKNOWLEDGMENT

This work was supported by the Scientific and Technical Research Council of Turkey (TÜBİTAK) under Grant Number 122E140.

REFERENCES

- [1] Li, Z., Wiacek, A., & Bell, M. A. L. (2020, September). Beamforming with deep learning from single plane wave RF data. In *2020 IEEE International Ultrasonics Symposium (IUS)* (pp. 1-4). IEEE.
- [2] Nguon, L. S., Seo, J., Seo, K., Han, Y., & Park, S. (2022). Reconstruction for plane-wave ultrasound imaging using a modified U-Net-based beamformer. *Computerized Medical Imaging and Graphics*, 98, 102073.
- [3] Perdios, D., Vonlanthen, M., Martinez, F., Arditi, M., & Thiran, J. P. (2021). CNN-based image reconstruction method for ultrafast ultrasound imaging. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 69(4), 1154-1168.
- [4] Perdios, D., Vonlanthen, M., Besson, A., Martinez, F., Arditi, M., & Thiran, J. P. (2018, October). Deep convolutional neural network for ultrasound image enhancement. In *2018 IEEE International Ultrasonics Symposium (IUS)* (pp. 1-4). IEEE.
- [5] Mor, E., & Bar-Hillel, A. (2020). A unified deep network for beamforming and speckle reduction in plane wave imaging: A simulation study. *Ultrasonics*, 103, 106069.
- [6] Goudarzi, S., & Rivaz, H. (2022). Deep reconstruction of high-quality ultrasound images from raw plane-wave data: A simulation and in vivo study. *Ultrasonics*, 125, 106778.
- [7] Goudarzi, S., Asif, A., & Rivaz, H. (2020). Fast multi-focus ultrasound image recovery using generative adversarial networks. *IEEE Transactions on Computational Imaging*, 6, 1272-1284.
- [8] Lu, J., Millioz, F., Garcia, D., Salles, S., Liu, W., & Friboulet, D. (2020). Reconstruction for diverging-wave imaging using deep convolutional neural networks. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 67(12), 2481-2492.
- [9] Zhang, J., He, Q., Xiao, Y., Zheng, H., Wang, C., & Luo, J. (2021). Ultrasound image reconstruction from plane wave radio-frequency data by a self-supervised deep neural network. *Medical Image Analysis*, 70, 102018.
- [10] Hyun, D., Brickson, L. L., Looby, K. T., & Dahl, J. J. (2019). Beamforming and speckle reduction using neural networks. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 66(5), 898-910.
- [11] Luchies, A. C., & Byram, B. C. (2018). Deep neural networks for ultrasound beamforming. *IEEE transactions on medical imaging*, 37(9), 2010-2021.
- [12] Mamistvalov, A., Amar, A., Kessler, N., & Eldar, Y. C. (2022). Deep-learning-based adaptive ultrasound imaging from sub-nyquist channel data. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 69(5), 1638-1648.
- [13] Wasih, M., Ahmad, S., & Almekkawy, M. (2023). A robust cascaded deep neural network for image reconstruction of single plane wave ultrasound RF data. *Ultrasonics*, 132, 106981.
- [14] <https://field-ii.dk/>