ORIGINAL ARTICLE

UCGNet: Capsule-Guided GAN for Ultrasound Image Reconstruction from Single-Plane Wave RF Data

Maryam Asad Samani [1], Ali Gharekhani [1], Parastoo Farnia [2,3], Bahador Makki Abadi [2,3*]

[1] *Department of Electrical Engineering, Iran University of Science and Technology, Tehran, Iran*

[2] *Department of Medical Physics and Biomedical Engineering, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran*

[3] *Advanced Medical Technologies and Equipment Institute, Tehran University of Medical Sciences, Tehran, Iran*

*Corresponding Author: Bahador Makki Abadi

Email: b-makkiabadi@sina.tums.ac.ir

Received: 27 May 2025 / Accepted: 15 October 2025

Abstract

Purpose: This study aims to improve ultrasound image reconstruction from single-plane wave RF data using Capsule Neural Networks, which can produce comparable image quality to Convolutional Neural Networks while requiring significantly fewer parameters. In addition to reducing model size, the proposed approach preserves clinically important image features and is better suited for real-time implementation in embedded systems with constrained computational resources.

Materials and Methods: We propose a novel ultrasound image reconstruction architecture, UCGNet (U-Caps-GAN Network), which combines Capsule Networks with a Generative Adversarial Network framework. UCGNet reconstructs high-quality B-mode ultrasound images directly from single-plane wave RF data and is evaluated on the Plane-wave Imaging Challenge in Medical Ultrasound (PICMUS) dataset. Capsule Networks play a key role in achieving parameter efficiency by encoding spatial hierarchies through vectorized feature representations. Their dynamic routing mechanism captures part–whole relationships and pose variations, enabling the network to preserve fine structural details essential for diagnostic imaging, without relying on deep, redundant convolutional layers. This makes the proposed architecture particularly well-suited for real-time applications in embedded systems with limited computational resources.

Results: The reconstructed images achieved a mean Signal-to-Noise Ratio (SNR) of 18.45 and a Peak Signal-to-Noise Ratio (PSNR) of 40.92, outperforming the baseline UNet model in terms of image quality. Additionally, UCGNet required about 23% of the training parameters compared to UNet, demonstrating its suitability for real-time applications on resource-constrained devices.

Conclusion: UCGNet provides an efficient and accurate solution for ultrasound image reconstruction from raw RF data. Its improved image fidelity and reduced computational complexity make it a strong candidate for practical use in portable and embedded medical imaging systems.

Keywords: Capsule Network; Generative Adversarial Network; Ultrasound Image Reconstruction; Lightweight Deep Learning; Single-Plane Wave.

DOI:

1. Introduction

Ultrasound imaging is a widely used diagnostic tool in various medical applications, ranging from cancer detection [1, 2] to real-time control in robotic systems guided through ultrasound-based visual feedback [3]. Compared to other imaging modalities like CT and MRI, ultrasound offers several advantages, including portability, affordability, real-time imaging capability, and the absence of ionizing radiation [4]. These benefits have made it valuable for point-of-care and bedside applications. However, the modality is often limited by image quality issues arising from its underlying physical and hardware constraints. Key challenges include speckle noise, low contrast, lateral resolution variability, and acoustic artifacts—all of which stem from the physics of wave propagation and transducer limitations. Raw Radio-Frequency (RF) signals carry these modality-specific issues, complicating the formation of high-fidelity B-mode images. Furthermore, system-level issues such as suboptimal Point Spread Functions (PSFs) and the need for computationally expensive post-processing limit real-time and embedded applications. These challenges are particularly critical in advanced scenarios like real-time 3D imaging and portable Point-Of-Care Ultrasound (POCUS) [5,6].

To address these limitations, advancements in image reconstruction techniques are essential. One of the fundamental approaches to improving ultrasound image quality is beamforming, which processes raw radio-frequency data to generate interpretable images. [7]. Conventional beamforming methods, such as Delay-and-Sum (DAS) and Short-Lag Spatial Coherence, have been widely used to enhance image resolution and contrast by optimizing the spatial coherence of received signals [8]. However, while DAS offers simplicity, it often suffers from wide main lobes and high side lobe levels [9]. Plane-wave imaging techniques, especially single-plane wave acquisitions, provide ultra-high frame rates but often at the cost of image quality. Multi-angle acquisitions can mitigate this issue, but typically compromise speed. This trade-off highlights a central challenge: developing reconstruction methods that maintain both high frame rates and image quality, especially in resource-limited or real-time environments.

Recent efforts have investigated hybrid beamforming approaches that combine analog and digital processing stages to reduce system complexity and support emerging applications such as wearable and 3D ultrasound systems [10]. In parallel, Deep Learning has emerged as a powerful framework for medical image reconstruction, offering significant improvements in segmentation, classification, and image processing tasks [9–20]. In ultrasound, Convolutional Neural Network (CNN) models have been successfully applied to beamforming and RF-to-image reconstruction, including direct reconstruction from single-plane wave data (Table 1). However, their

Table 1. An overview of different approaches to beamforming

Ref.	Dataset	Methods	Objective
[7]	Mixed (Simulation, Phantom, in vivo)	Autoencoder UNet	Beamforming from RF data after single-plane wave transmission
[9]	Mixed	CNN	PSF-optimized beamforming for TRF reconstruction in single-plane wave imaging
[24]	Mixed	Multi-task UNet	Simultaneous image and segmentation outputs from raw ultrasound channel data using deep learning
[25]	Mixed	CNN	Beamforming of ultrasound channel signals with speckle reduction
[26]	Simulation, Phantom	Deep Neural Networks	End-to-end ultrasound beamforming and image reconstruction directly from RF channel data, improving resolution and CNR, and enabling multitask learning.
[27]	Simulation, Phantom	Generative Adversarial Network (GAN)	An alternative to traditional beamforming, extracting information from raw RF data to produce B-mode images and segmentation maps.
[29]	Mixed	Bayesian Posterior Inference with Sylvester Normalizing Flow	Adaptive compressed sensing and ultrasound image reconstruction for real-time imaging

deep architectures often involve high computational loads, making them unsuitable for real-time or embedded scenarios. Besides, CNNs are often sensitive to geometric transformations, lack explicit spatial relationships due to pooling layers, and demand substantial data and computation [21, 22, 23].

To further enhance deep models, attention mechanisms have been integrated into CNNs to enable more focused feature extraction in complex, noisy images [30]. Additionally, Vision Transformers (ViTs) have emerged as compelling alternatives to CNNs in medical imaging, using self-attention to capture global context and long-range dependencies [31,32]. Transformer-based models, particularly when equipped with attention gates like those in Attention U-Net, offer improved noise suppression and feature localization. Techniques such as semi-supervised and self-supervised learning are also gaining traction, allowing effective training with minimal labeled data—highly beneficial in the often data-scarce field of medical imaging [33, 34].

An alternative to traditional CNNs is the Capsule Network (CapsNet), which represents features as vectors rather than scalars, preserving pose and part–whole relationships via dynamic routing. CapsNets maintain spatial hierarchies, offering robustness to geometric transformations without relying on deep architectures or extensive data augmentation. Their ability to encode orientation and structural detail makes them particularly attractive for ultrasound image reconstruction. Hybrid architectures like UCapsNet have combined capsule layers with UNet-like structures for improved performance in tasks such as breast lesion segmentation [21]. Recent studies (Table 2) demonstrate CapsNet's strong performance in segmentation and classification tasks with fewer parameters and less training data [35-41]. Nevertheless, CapsNets have not yet been applied to ultrasound beamforming or RF-to-image reconstruction, representing a promising direction for exploration.

In this work, we introduce UCGNet (U-Caps-GAN Network), a novel Deep Learning architecture that integrates Capsule Networks into a Generative Adversarial Network framework for high-quality ultrasound image reconstruction. Unlike prior CapsNet applications limited to segmentation or classification, UCGNet reconstructs B-mode images directly from single-plane wave RF data. Capsule layers provide efficient, spatially-aware feature encoding with fewer parameters, while adversarial learning sharpens image realism and enhances perceptual quality. This design directly addresses the dual challenges of computational efficiency and image fidelity, making it particularly suitable for real-time and embedded ultrasound systems. Our architecture bridges a critical gap in the literature by combining the spatial awareness of CapsNets with the generative power of GANs. The result is a Deep Learning solution that delivers high-quality, high-frame-rate reconstructions with reduced model complexity, tailored for deployment in real-world, resource-constrained ultrasound imaging environments.

Table 2. An overview of different approaches to CapsNets

Ref.	Dataset	Methods	Objective
[22]	various modalities	CapsNet	Apply CapsNet for accurate medical image segmentation
[35]	MICCAI 2013	CapsNet with reconstruction extension	Object segmentation using locally-connected routing
[36]	MSD	Multi-SegCaps, EM-routing	Segmentation using context from adjacent 2D slices
[37]	JSRT	CapsNet with CHT & Fourier preprocessing	Automatic LV segmentation in cardiac MRI
[40]	Breast Ultrasound (Private Dataset)	UCapsNet (U-Net + CapsNet)	Breast tumor segmentation and classification with high accuracy

2. Materials and Methods

This section presents the network architecture and its training procedure, followed by a discussion on the utilized dataset, the implementation environment, and the evaluation metrics employed.

2.1. Network Design

UCGNet is motivated by UNet [42] and CapsNet, and is trained within a GAN framework [43] (generator-discriminator). U-CapsNet was selected as the generator due to its lightweight yet expressive architecture. Compared to SegCaps and EM-routing-

based models [36], U-CapsNet maintains high spatial fidelity while significantly reducing computational overhead—an essential consideration for real-time ultrasound applications. Additionally, U-CapsNet avoids the routing convergence issues sometimes observed in deeper capsule structures. Given our objective of reconstructing detailed ultrasound images from RF data, this architecture offered a suitable balance between performance and simplicity. Alternative GAN variants, such as StyleGAN [44] or CycleGAN [45], were not considered appropriate as they are typically used for style transfer or cross-domain mappings, which are not directly aligned with the goals of this work.

Our capsule-based generator (Figure 1) adopts an encoder–decoder architecture, similar to UNet, where max-pooling operations are replaced with capsule layers. The encoder captures hierarchical features from the RF data using a series of convolutional and capsule layers, while the decoder reconstructs the image from these features using transposed convolutions and skip connections. Capsule layers output vector representations instead of scalars, capturing both the presence and spatial relationships of features [21]. The reconstructed image is produced through progressive upsampling layers and integrated with Batch Normalization [46] to retain fine spatial details.

The discriminator, designed to differentiate between real and reconstructed images, is a CNN consisting of four sequential blocks (Figure 1). Each block comprises a 2D convolutional layer, followed by Batch Normalization and a LeakyReLU activation function with a negative slope of 0.2. These layers progressively reduce spatial resolution while expanding the feature depth, enabling the discriminator to focus on increasingly abstract features and better detect subtle differences between real and generated images. The final layer is a convolutional layer, producing a single-channel output, which serves as a patch-level real/fake prediction map. This architecture is inspired by the PatchGAN discriminator [47], which has demonstrated effectiveness in capturing high-frequency details and preserving fine textures in image generation tasks. The design choices aim to balance expressive power and training stability, which are critical for adversarial learning settings.

To train and develop the designed network, the Mean Squared Error (MSE) loss function is used to compare each pixel of the reconstructed image with the ground truth image (Equation 1):

$$\mathrm{MSE} = \frac{1}{n}\sum_{t=1}^{n}(p_t - \hat{p}_t)^2 \tag{1}$$

where p_t, $\hat{p}_t$, and n are the ground truth, reconstructed images, and number of pixels values, respectively.

In addition to MSE, adversarial loss [43] guides the generator G to produce images indistinguishable from real images, by competing against the discriminator D. The adversarial loss $\mathcal{L}_{GAN}(G)$ for the generator is (Equation 2):

$$\mathcal{L}_{GAN}(G) = \mathbb{E}_x\left[\log\left(1 - D\left(G(x)\right)\right)\right] \tag{2}$$

The discriminator loss $\mathcal{L}_{GAN}(D)$ is defined as (Equation 3):

$$\mathcal{L}_{GAN}(D) = -\mathbb{E}_y[\log D(y)] - \mathbb{E}_x\left[\log\left(1 - D\left(G(x)\right)\right)\right] \tag{3}$$

where x is the input RF data, y represents the real ground truth images, $\mathbb{E}_x$ is the expectation over noise samples used to generate reconstructed images, and $\mathbb{E}_y$ is the expectation over real data samples. The total generator loss $\mathcal{L}$ is then a weighted combination of the reconstruction loss and the adversarial loss (Equation 4).

$$\mathcal{L} = \mathrm{MSE} + \mathcal{L}_{GAN}(G) \tag{4}$$

The 7 main stages taking place in reconstructing ultrasound images from single-plane wave RF data are summarized in the following algorithm.

UCGNet was ultimately trained over a span of 160 epochs with a total of 3,615,073 parameters. The training loss function is depicted in Figure 2. Despite some incidental noise, the training loss had already reduced to less than 0.005 after approximately 75 epochs. Upon conclusion of the training process, with regard to the final network weights, the training loss of 0.0008 was achieved.

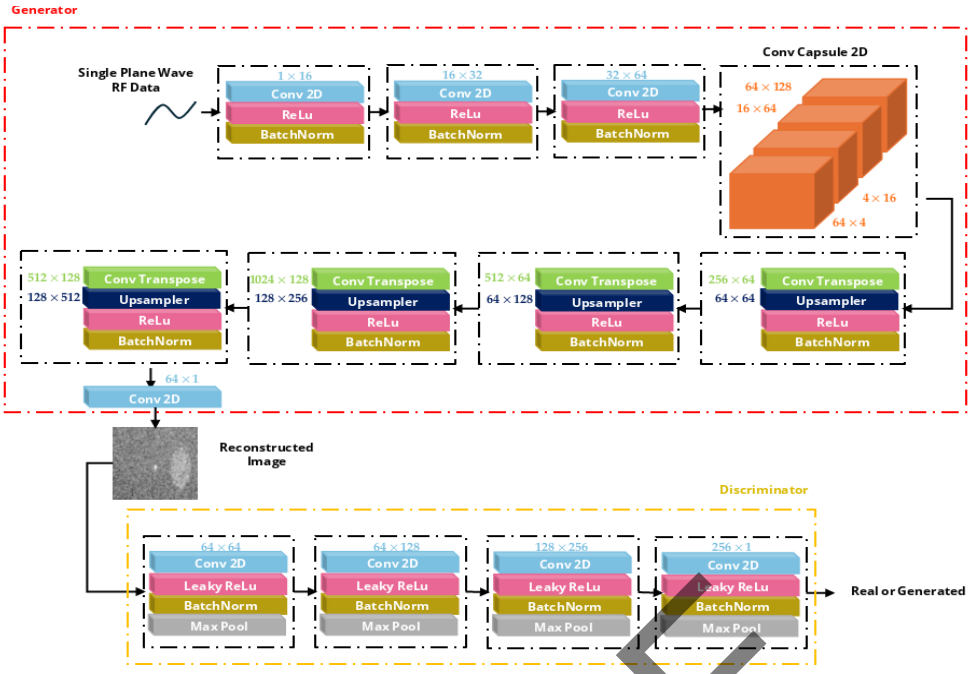

Figure 1. GAN framework of UCGNet. The proposed network is composed of a CapsNet generator and CNN discriminator. The number of parameters for each layer is indicated near the arrows, and the color of the number corresponds to the color of the associated layer for clarity

UCGNet Training Algorithm

for ***n*** iterations **do**

Step1: Extract features from the RF data throughout 2Dconvolution layers and Leaky ReLU activation function

Step2: Encode the extracted features using 2D Capsule layers

Step3: Decode encoded features with 2D deconvolution layers and ReLU activation function

Step4: Upsample features with transpose convolutional layers

Step5: Convert reconstructed features to a 2D grayscale image by convolutional layers and passing through a tanh activation function

Step6: Compute the loss of the reconstructed and the ground truth image

Step7: Update the network weights

return loss value

Despite dropout and standard weight decay, some fluctuations in the training and were observed (Figure 2), which are expected given the limited dataset size and variability of phantom and simulated images. These fluctuations highlight that additional regularization or advanced techniques could further

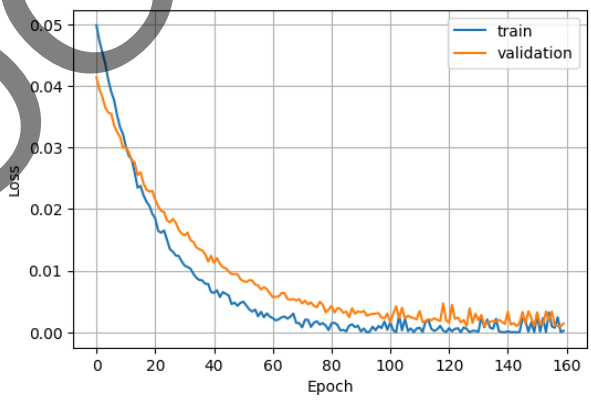

Figure 2. Mean Squared Error (MSE) comparison between reconstructed and ground-truth images. This figure demonstrates the training and evaluation loss of the proposed UCGNet model in relation to the original images

stabilize training, but the current configuration provides a reasonable balance between convergence stability and reconstruction performance.

UCGNet is compared to a typical GAN, consisting of the same discriminator as before and a conventional UNet generator. This specific network comprises encoders and decoders composed of 2D convolutional layers, along with several pooling layers within these layers (Figure 3). The model is trained for approximately 160 epochs and achieves a final loss of 0.0012, indicating its high level of success.

Training Details

Number of epochs: 160

Learning Rate: 0.0002

Dropout: 0.1

Optimizer: Adam

Batch Size: 32

Loss Function: Weighted combination of reconstructic loss and adversarial loss

Learning Rate Scheduler: Step decay every 20 epoch

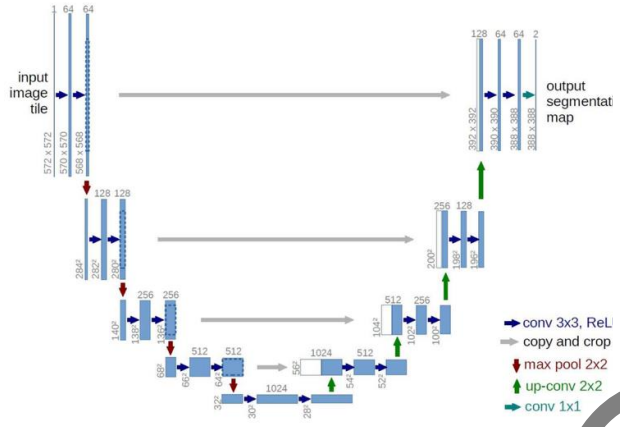

Figure 3. Schematic of the UNet architecture [42]. This network is adopted as the generator module within UCGNet

2.2. Dataset

The training and evaluation of UCGNet were conducted using the publicly available PICMUS dataset [48], which includes two simulated and two phantom datasets. Each dataset contains RF data from 75 steered plane wave transmissions, with angles uniformly distributed between -16° and +16°. For ground truth generation, all 75 plane wave transmissions were compounded using Delay-And-Sum (DAS) beamforming to produce high-quality reference images.

For training, 422 frames were used across all four datasets, randomly selected while ensuring a balanced representation of phantom and simulated data. In addition, 62 frames were reserved for validation during training to monitor convergence and compute the evaluation loss (as shown in Figure 2). Finally, 124 frames were assigned exclusively for testing, and all reported results (Figures 3–5) are based on this test dataset. We ensured subject-level separation between the splits, meaning that frames from a given subject appear in only one of the training, validation, or test sets. This guarantees that the model is evaluated on completely unseen subjects, preventing data leakage and providing a fair assessment of generalization performance. To increase data variability and mitigate overfitting, the following data augmentation techniques were applied during training: small-angle rotations (±5°), horizontal flipping with a 50% probability, random cropping to 256×256 pixels, and Gaussian noise injection with σ = 0.01 applied with a 30% probability.

2.3. Evaluation Metrics

Signal-to-Noise Ratio (SNR) is indeed a commonly used image quality metric to evaluate and compare the output of the UCGNet network with ground truth images. SNR measures the ratio of the signal power to the noise power in an image, indicating the level of background noise and the smoothness of the tissue. A higher SNR value indicates a more desirable aspect of the image, with a stronger signal relative to the noise. SNR is a quantitative measure that can provide valuable insights into the quality of the reconstructed ultrasound images. Equation 5 is the formula for calculating SNR for a background Region Of Interest (ROI), where S_o and σ_o are the mean and standard deviation, respectively:

$$\text{SNR} = \frac{S_o}{\sigma_o} \tag{5}$$

The ratio between the maximum signal power and the power of affecting noise, known as the Peak Signal-to-Noise Ratio (PSNR), is calculated as a representation of the similarity between the ground truth image and the reconstructed images (Equation 6):

$$PSNR = 10\log_{10}\left(\frac{\text{MAX}^2}{\text{MSE}}\right) \tag{6}$$

where MAX indicates the maximum possible pixel value of the image, and MSE is the mean square error between two images.

The Structural Similarity Index Measure (SSIM) is a perceptual metric that quantifies the similarity between the ground truth image and the reconstructed

images by considering structural information, luminance, and contrast. It is defined as (Equation 7):

$$\mathrm{SSIM}(x,y)2 = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x{}^2 + \mu_y{}^2 +)(\sigma_x{}^2 + \sigma_y{}^2 + C_1)} \quad (7)$$

where μ_x and μ_y are the mean intensities of images x and y, $\sigma_x{}^2$ and $\sigma_y{}^2$ are the corresponding variances, σ_{xy} is the covariance between the two images, and $and\ C_2$ are constants introduced to stabilize the division.

3. Results

3.1. Small World

This section compares the reconstructed outputs of UCGNet and a conventional UNet-GAN, providing quantitative insights. The UNet-GAN baseline was re-implemented by us and trained under the same preprocessing and training setup as UCGNet. All experiments were conducted using Google Colaboratory with Python 3.8, utilizing a T4 GPU for training. While the hardware differs from that reported in the original UNet-GAN paper, this setup ensures a fair and consistent comparison between the two models.

Table 3 summarizes the parameter count for UNet and UCGNet. UCGNet contains 3,615,073 parameters, which is about 23% of the 14,787,777 parameters of UNet-GAN. This substantial reduction in parameter count highlights the improved model efficiency of UCGNet, leading to lower memory usage and faster inference time. All text and table entries have been updated to reflect this exact value.

Table 3. Number of parameters of each network

Network	No. of parameters
UNet-GAN	14787777
UCGNet	3615073

For inference, the method proposed in [27] is also utilized. Table 4 and Figure 5 compare the average SNR, PSNR, SSIM, and execution time among UNet-GAN, UCGNet, and the approach in [27]. UCGNet achieves the highest performance in all metrics. Importantly, the execution time of UCGNet is substantially lower, suggesting a trade-off between complexity and performance.

Table 4. SNR, PSNR, SSIM, and Execution time of UNet and UCGNet

Network	SNR (dB)	PSNR (dB)	SSIM	Execution Time (seconds)
UNet-GAN	17.04 ± 0.45	39.37 ± 0.40	0.89 ± 0.02	0.0280 ± 0.0012
[27]	15.90 ± 0.50	36.65 ± 0.45	0.87 ± 0.03	0.0189 ± 0.0009
UCGNet	18.45 ± 0.42	40.92 ± 0.38	0.91 ± 0.02	0.0068 ± 0.0005

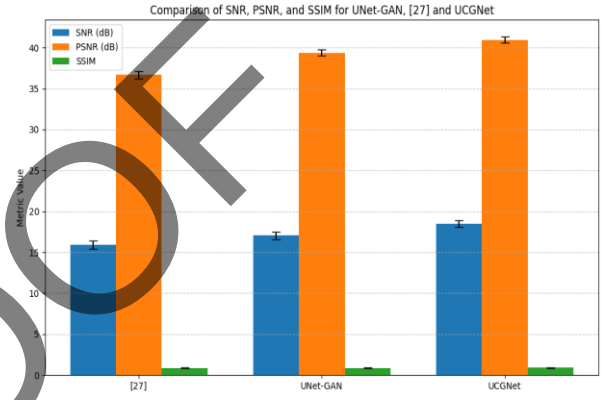

Figure 4. Comparison of image quality metrics - SNR, PSNR, and SSIM for UNet, the proposed method in [27], and UCGNet

Figures 5 and 6 show representative examples of reconstructed ultrasound images. Visually, UCGNet demonstrates slightly improved structural coherence and edge definition compared to UNet and UNet-GAN. However, it is important to note that these improvements come with certain trade-offs. While UCGNet significantly reduces parameter count and execution time, it requires a longer training period, likely due to the iterative nature of capsule routing and the complexity of the GAN framework.

Although UCGNet shows enhanced quantitative metrics, the gains over UNet-GAN are modest, suggesting that the combination of GAN and CapsNet offers incremental, rather than transformative, benefits for this particular application. This observation highlights an important consideration: the improved image quality of UCGNet must be weighed against the increased training complexity and potential difficulty in tuning capsule-based architectures. Future work could explore more efficient capsule routing

algorithms or hybrid models that selectively apply capsules to reduce computational burden.

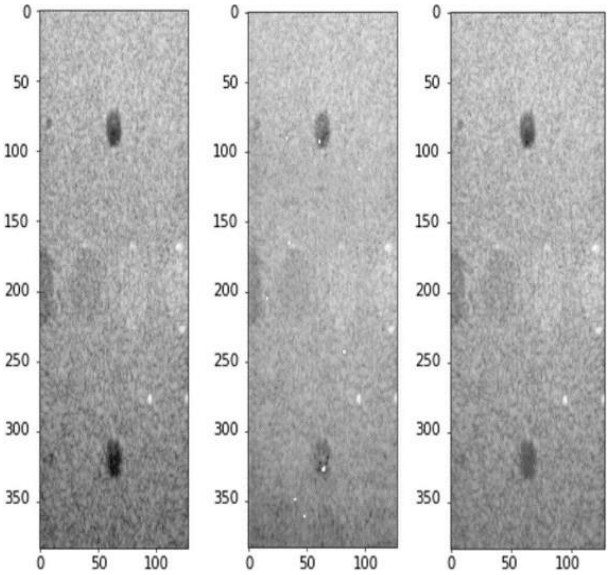

Figure 5. Qualitative comparison of reconstructed images. (a) Ground truth image, (b) output of UCGNet, and (c) output of UNet. Both networks achieve similar reconstruction accuracy; however, UCGNet requires considerably fewer parameters to reach this level of performance, demonstrating its parameter efficiency

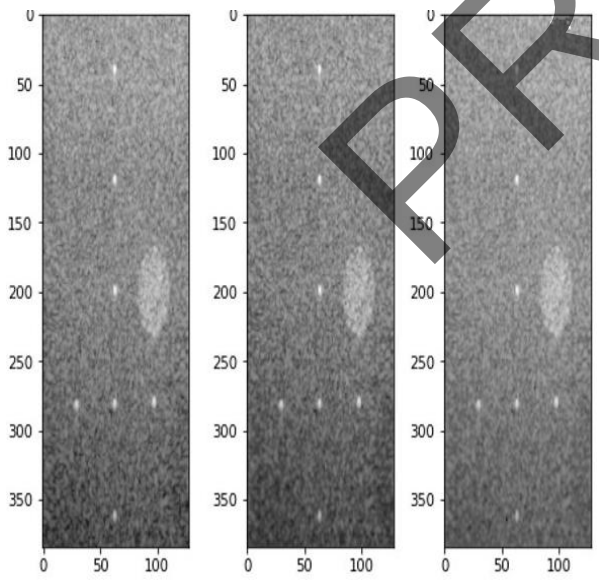

Figure 6. Additional qualitative comparison on diverse samples. (a) Ground truth image, (b) UCGNet reconstruction, and (c) UNet reconstruction. The results show that UCGNet maintains comparable reconstruction quality to UNet while significantly reducing the number of parameters, thereby providing a more efficient solution

4. Conclusion

This manuscript introduces UCGNet—a capsule-enhanced UNet architecture trained within a GAN framework—for reconstructing ultrasound images. The proposed method demonstrates promising results on the PICMUS dataset, achieving SNR=18.45 dB, PSNR=40.92 dB, and SSIM=0.91, representing noticeable improvements over the delay-and-sum beamforming algorithm. Furthermore, UCGNet achieves these improvements using only 23% of the parameters required by a standard UNet-GAN, enhancing computational efficiency and making it more suitable for real-time or embedded deployment in clinical applications. For future development and future work, the impact of alternative routing algorithms (e.g., Expectation-Maximization routing or dynamic routing with additional iterations) should be investigated. Deeper capsule hierarchies need to be analyzed to better preserve spatial hierarchies in complex anatomical structures. Furthermore, perceptual loss functions (e.g., VGG-based or feature-matching losses) could be incorporated in combination with MSE and adversarial losses to improve textural realism. In addition, progressive training strategies or curriculum learning can be applied to stabilize GAN convergence and reduce training time while preserving high-quality outputs. Lastly, evaluation metrics could be extended by incorporating clinical metrics such as Contrast-to-Noise Ratio (CNR) or lesion detectability. These directions could pave the way toward more robust and clinically applicable ultrasound reconstruction.

Acknowledgments

The National Elite Foundation of Iran has provided support for the present investigation.

References

1- V. Kumar *et al.*, "Automated and real-time segmentation of suspicious breast masses using convolutional neural network." *PLOS One,* Vol. 13 (No. 5), p. e0195816, (2018).

2- G. Pons, J. Martí, R. Martí, S. Ganau, and J. A. Noble, "Breast lesion segmentation combining b-mode and elastography ultrasound." *Ultrasonic Imaging,* Vol. 38 (No. 3), pp. 209-24, (2016).

3- R. Mebarki, A. Krupa, and F. Chaumette, "2-D ultrasound probe complete guidance by visual servoing using image moments." *IEEE Transactions on Robotics,* Vol. 26 (No. 2), pp. 296-306, (2010).

4- R. Wang, "High-resolution image reconstruction for portable ultrasound imaging devices." *EURASIP Journal on Advances in Signal Processing,* (2019).

5- Y. Lyu, Y. Shen, M. Zhang, and J. Wang, "Real-time 3D ultrasound imaging system based on a hybrid reconstruction algorithm." *Chinese Journal of Electronics,* Vol. 33 (No. 1), pp. 245-55, (2024).

6- H. Yoon, J. Kim, K. Lee, and T. K. Song, "Design and Implementation of Analog-Digital Hybrid Beamformers for Low-Complexity Ultrasound Systems: A Feasibility Study." *Bioengineering,* Vol. 11 (No. 1), p. 8, (2023).

7- Z. Li, A. Wiacek, and M. L. Bell, "Beamforming with deep learning from single plane wave RF data." Presented at the *Proc. IEEE International Ultrasonics Symposium (IUS)*, (2020).

8- M. A. Lediju, G. E. Trahey, B. C. Byram, and J. J. Dahl, "Shortlag spatial coherence of backscattered echoes: Imaging characteristics." *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control,* Vol. 58 (No. 7), pp. 1377-88, (2011).

9- S. Goudarzi and H. Rivaz, "Plane-Wave Ultrasound Beamforming: A Deep Learning Approach." *arXiv,* (2021).

10- A. K. Tehrani and H. Rivaz, "Displacement estimation in ultrasound elastography using pyramidal convolutional neural network." *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control,* (2020).

11- M. Amiri, "Fine-tuning UNet for ultrasound image segmentation: different layers, different outcomes." *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control,* (2020).

12- M. Azimi, A. Gharekhani, and S. Ebadollahi, "An Integration of a Neural-Network-Based Computer Vision Model and a 2-DOF Object Tracker Robot." Presented at the *2023 11th RSI International Conference on Robotics and Mechatronics (ICRoM)*, (2023).

13- R. Delaunay, "An unsupervised learning approach to ultrasound strain elastography with spatio-temporal consistency." *Physics in Medicine & Biology,* (2021).

14- S. Jeon and C. Kim, "Deep learning-based speed of sound aberration correction in photoacoustic images." Presented at the *Photons Plus Ultrasound: Imaging and Sensing 2020*, (2020).

15- D. Karimi, "Accurate and robust deep learning-based segmentation of the prostate clinical target volume in ultrasound images." *Medical Image Analysis,* Vol. 57pp. 186-96, (2019).

16- M. A. Samani, K. Hooshanfar, H. S. Jey, and S. M. Esmailzadeh, "Eye-Tracking Based Control of a Robotic Arm and Wheelchair for People with Severe Speech and Motor Impairment (SSMI)." Presented at the *2023 11th RSI International Conference on Robotics and Mechatronics (ICRoM)*, (2023).

17- M. Sharifzadeh, "Phase aberration correction: A convolutional neural network approach." *IEEE Access,* Vol. 8p. 112400, (2020).

18- K. Song, J. Feng, and D. Chen, "A survey on deep learning in medical ultrasound imaging." *Frontiers in Physics,* Vol. 12p. 1398393, (2024).

19- A. K. Tehrani, "Semisupervised training of optical flow convolutional neural networks in ultrasound elastography." Presented at the *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, (2020).

20- A. K. Tehrani, "Mpwc-net++: evolution of optical flow pyramidal convolutional neural network for ultrasound elastography." Presented at the *Medical Imaging 2021: Ultrasonic Imaging and Tomography*, (2021).

21- G. Madhu *et al.*, "UCapsNet: A two-stage deep learning model using U-Net and Capsule Network for breast cancer segmentation and classification in ultrasound imaging." *Cancers,* Vol. 16 (No. 22), p. 3777, (2024).

22- S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules." Presented at the *Advances in Neural Information Processing Systems (NeurIPS)*, (2017).

23- S. Survarachakan, "Capsule Nets for Complex Medical Image Segmentation Tasks." *figshare,* (2020).

24- M. Tran, "CapsNet for Medical Image Segmentation." *arXiv,* (2022).

25- A. A. Nair, "Deep learning to obtain simultaneous image and segmentation outputs from a single input of raw ultrasound channel data." *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control,* (2020).

26- D. Hyun, "Beamforming and speckle reduction using neural networks." *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control,* Vol. 66 (No. 5), pp. 898-910, (2019).

27- E. Dahan and I. Cohen, "Deep-learning-based multitask ultrasound beamforming." *Information,* Vol. 14 (No. 10), p. 582, (2023).

28- A. A. Nair, T. D. Tran, A. Reiter, and M. A. Bell, "A generative adversarial neural network for beamforming ultrasound images: Invited presentation." Presented at the *2019 53rd Annual Conference on Information Sciences and Systems (CISS)*, (2019).

29- S. W. Penninga, H. van Gorp, and R. J. van Sloun, "Deep Sylvester posterior inference for adaptive compressed sensing in ultrasound imaging." Presented at the *ICASSP 2025 - IEEE International Conference on Acoustics, Speech and Signal Processing*, (2025).

30- H. Liu, "Attention residual network for medical ultrasound image segmentation." *Scientific Reports,* Vol. 15 (No. 1), p. 22155, (2025).

31- M. Vafaeezadeh, H. Behnam, and P. Gifani, "Ultrasound image analysis with vision transformers." *Diagnostics,* Vol. 14 (No. 5), p. 542, (2024).

32- K. Xia and J. Wang, "Recent advances of transformers in medical image analysis: a comprehensive review." *MedComm–Future Medicine,* Vol. 2 (No. 1), p. e38, (2023).

33- B. Behboodi and H. Rivaz, "Ultrasound segmentation using U-Net: learning from simulated data and testing on real data." Presented at the *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (2019).

34- H. Rasaee, M. Samuel, B. Behboodi, J. Afilalo, and H. Rivaz, "Ultrasound segmentation using semi-supervised learning: Application in point-of-care sarcopenia assessment." *IEEE Open Journal of Engineering in Medicine and Biology,* (2025).

35- V. Akoto-Adjepong, O. Appiah, P. K. Mensah, and P. Appiahene, "TTDCapsNet: Tri Texton-Dense Capsule Network for complex and medical image recognition." *PLOS One,* Vol. 19 (No. 3), p. e0300133, (2024).

36- S. Bonheur, D. Štern, C. Payer, M. Pienn, H. Olschewski, and M. Urschler, "Matwo-capsnet: a multi-label semantic segmentation capsules network." Presented at the *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, (2019).

37- J. Deepika, C. Rajan, and T. Senthil, "Improved CAPSNET model with modified loss function for medical image classification." *Signal, Image and Video Processing,* Vol. 16 (No. 8), pp. 2269-77, (2022).

38- R. LaLonde and U. Bagci, "Capsules for object segmentation." *arXiv,* (2018).

39- R. LaLonde, Z. Xu, I. Irmakci, S. Jain, and U. Bagci, "Capsules for biomedical image segmentation." *Medical Image Analysis,* Vol. 68p. 101889, (2021).

40- F. Long, J. J. Peng, W. Song, X. Xia, and J. Sang, "BloodCaps: A capsule network based model for the multiclassification of human peripheral blood cells." *Computer Methods and Programs in Biomedicine,* Vol. 202p. 105972, (2021).

41- A. F. M. Saif, T. Imtiaz, S. Rifat, C. Shahnaz, W. P. Zhu, and M. O. Ahmad, "CapsCovNet: A modified capsule network to diagnose Covid-19 from multimodal medical imaging." *IEEE Transactions on Artificial Intelligence,* Vol. 2 (No. 6), pp. 608-17, (2021).

42- O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation." Presented at the *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, (2015).

43- I. Goodfellow *et al.*, "Generative adversarial nets." Presented at the *Advances in Neural Information Processing Systems (NeurIPS)*, (2014).

44- T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks." Presented at the *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2019).

45- J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks." Presented at the *IEEE International Conference on Computer Vision (ICCV)*, (2017).

46- S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift." Presented at the *International Conference on Machine Learning (ICML)*, (2015).

47- C. Fan, H. Lin, and Y. Qiu, "U-Patch GAN: A medical image fusion method based on GAN." *Journal of Digital Imaging,* Vol. 36 (No. 1), pp. 339-55, (2023).

48- H. Liebgott, "Plane-Wave Imaging Challenge in Medical Ultrasound." Presented at the *IEEE International Ultrasonics Symposium (IUS)*, (2016).