

# UCLA

## UCLA Previously Published Works

### Title

Enhancing Ultrasound Image Quality Across Disease Domains: Application of Cycle-Consistent Generative Adversarial Network and Perceptual Loss.

### Permalink

<https://escholarship.org/uc/item/82f951pp>

### Authors

Athreya, Shreeram

Radhachandran, Ashwath

Ivezić, Vedrana

et al.

### Publication Date

2024-12-17

### DOI

10.2196/58911

Peer reviewed

Original Paper

# Enhancing Ultrasound Image Quality Across Disease Domains: Application of Cycle-Consistent Generative Adversarial Network and Perceptual Loss

Shreeram Athreya<sup>1</sup>, MS; Ashwath Radhachandran<sup>2</sup>, BS; Vedrana Ivezić<sup>3</sup>, BS; Vivek R Sant<sup>4</sup>, MD; Corey W Arnold<sup>1,2,3,5,6</sup>, PhD; William Speier<sup>2,3,5</sup>, PhD

<sup>1</sup>Department of Electrical and Computer Engineering, University of California Los Angeles, Los Angeles, CA, United States

<sup>2</sup>Department of Bioengineering, University of California Los Angeles, Los Angeles, CA, United States

<sup>3</sup>Medical Informatics, University of California Los Angeles, Los Angeles, CA, United States

<sup>4</sup>Department of Surgery, The University of Texas Southwestern Medical Center, Dallas, TX, United States

<sup>5</sup>Department of Radiological Sciences, University of California Los Angeles, Los Angeles, CA, United States

<sup>6</sup>Department of Pathology and Laboratory Medicine, University of California Los Angeles, Los Angeles, CA, United States

**Corresponding Author:**

Shreeram Athreya, MS

Department of Electrical and Computer Engineering

University of California Los Angeles

924 Westwood Boulevard, Suite 600

Los Angeles, CA, 90024

United States

Phone: 1 4244206158

Email: [shreeram@ucla.edu](mailto:shreeram@ucla.edu)

## Abstract

**Background:** Numerous studies have explored image processing techniques aimed at enhancing ultrasound images to narrow the performance gap between low-quality portable devices and high-end ultrasound equipment. These investigations often use registered image pairs created by modifying the same image through methods like down sampling or adding noise, rather than using separate images from different machines. Additionally, they rely on organ-specific features, limiting the models' generalizability across various imaging conditions and devices. The challenge remains to develop a universal framework capable of improving image quality across different devices and conditions, independent of registration or specific organ characteristics.

**Objective:** This study aims to develop a robust framework that enhances the quality of ultrasound images, particularly those captured with compact, portable devices, which are often constrained by low quality due to hardware limitations. The framework is designed to effectively process nonregistered ultrasound image pairs, a common challenge in medical imaging, across various clinical settings and device types. By addressing these challenges, the research seeks to provide a more generalized and adaptable solution that can be widely applied across diverse medical scenarios, improving the accessibility and quality of diagnostic imaging.

**Methods:** A retrospective analysis was conducted by using a cycle-consistent generative adversarial network (CycleGAN) framework enhanced with perceptual loss to improve the quality of ultrasound images, focusing on nonregistered image pairs from various organ systems. The perceptual loss was integrated to preserve anatomical integrity by comparing deep features extracted from pretrained neural networks. The model's performance was evaluated against corresponding high-resolution images, ensuring that the enhanced outputs closely mimic those from high-end ultrasound devices. The model was trained and validated using a publicly available, diverse dataset to ensure robustness and generalizability across different imaging scenarios.

**Results:** The advanced CycleGAN framework, enhanced with perceptual loss, significantly outperformed the previous state-of-the-art, stable CycleGAN, in multiple evaluation metrics. Specifically, our method achieved a structural similarity index of 0.2889 versus 0.2502 ( $P < .001$ ), a peak signal-to-noise ratio of 15.8935 versus 14.9430 ( $P < .001$ ), and a learned perceptual image patch similarity score of 0.4490 versus 0.5005 ( $P < .001$ ). These results demonstrate the model's superior ability to enhance image quality while preserving critical anatomical details, thereby improving diagnostic usefulness.

**Conclusions:** This study presents a significant advancement in ultrasound imaging by leveraging a CycleGAN model enhanced with perceptual loss to bridge the quality gap between images from different devices. By processing nonregistered image pairs,

the model not only enhances visual quality but also ensures the preservation of essential anatomical structures, crucial for accurate diagnosis. This approach holds the potential to democratize high-quality ultrasound imaging, making it accessible through low-cost portable devices, thereby improving health care outcomes, particularly in resource-limited settings. Future research will focus on further validation and optimization for clinical use.

(*JMIR Biomed Eng* 2024;9:e58911) doi: [10.2196/58911](https://doi.org/10.2196/58911)

## KEYWORDS

generative networks; cycle generative adversarial network; image enhancement; perceptual loss; ultrasound scans; ultrasound images; imaging; machine learning; portable handheld devices

## Introduction

Ultrasound imaging is crucial in medical diagnostics due to its noninvasive nature and high accuracy. It provides point-of-care assessments that have been increasingly adopted by health care professionals [1,2]. Historically, technology has been limited to large, expensive devices typically found in specialized medical settings. However, there has been a transformative shift toward the development and adoption of compact, handheld ultrasound devices [3,4]. These smaller devices promise to democratize access to medical imaging by making it more affordable and widely available. Yet, the miniaturization and cost-effectiveness often come at the expense of image quality, a trade-off primarily attributable to hardware constraints [5-7].

Machine learning algorithms have been explored to enhance low-quality images without the need for hardware improvements [8]. For instance, generative adversarial networks (GANs) [9] have been used to create high-quality reconstructions of ultrasound images and videos, providing a cost-efficient avenue for the enhancement of portable ultrasound devices [10-12]. The cycle-consistent generative adversarial network (CycleGAN) framework, which is particularly useful for image-to-image translations without requiring paired data, has become increasingly popular [13]. The technology has been applied across a spectrum of tasks including, style transfer [14], where the appearance of one image is transformed to match another style, and object transfiguration [13,15,16], which involves changing 1 object in an image into another while retaining the overall structure. In medical imaging, CycleGANs have been used in tasks such as pixel-wise translation in echocardiography [17]. CycleGANs have also been applied in cross-modality medical image translation such as converting computed tomography to magnetic resonance imaging [18]. The architecture has even found use in histopathology to standardize microscopy staining for more accurate diagnoses [19].

We hypothesize that the integration of computational algorithms, particularly CycleGAN, can mitigate the disparities in images acquired from different medical imaging devices. Traditional training approaches for these models artificially introduce corruption into medical images to create pixel-wise pairs [20-22]. However, these methods typically fail to encapsulate the different characteristics of images acquired using different devices. Acquiring paired images using different devices leads to technical issues as images are captured at different time instances with varying orientations, leading to structural changes that cannot be completely resolved using image registration.

In this work, we benchmark several key models that are highly relevant to our task of ultrasound image enhancement. Pix2Pix [4] uses conditional adversarial networks for paired image-to-image translation, making it effective for directly comparing low- and high-quality images. CycleGAN [5] enables unpaired image-to-image translation, which is crucial when paired datasets are not available. Registration GAN (RegGAN) [6] focuses on medical image translation by aligning structural content using a registration network, and multilevel structure-preserved GAN (MSPGAN) [7] introduces a multilevel structure-preserved GAN for domain adaptation in intravascular ultrasound analysis. However, the current state-of-the-art is the stability-enhanced CycleGAN [1], which specifically addresses domain transformation challenges in unpaired ultrasound images, making it particularly relevant and effective for our specific application.

Evaluation metrics play a critical role in assessing the effectiveness of image enhancement models. Commonly used metrics include structural similarity index (SSI), peak signal-to-noise ratio (PSNR), and locally normalized cross-correlation (LNCC) [5,10-12,23]. While these metrics are widely accepted, they primarily capture low-frequency information and may not adequately reflect true image quality, particularly in preserving high-frequency details, which are crucial for medical diagnostics. Models that perform well on these traditional metrics may produce visually appealing images but fail to retain essential high-frequency content, leading to a loss of critical diagnostic information and perceptual quality [24]. To address this limitation, we incorporate the learned perceptual image patch similarity (LPIPS) [24] metric in our evaluations. LPIPS is designed to capture perceptual differences that align more closely with human visual perception, ensuring that our method not only performs well quantitatively but also produces qualitatively superior images, preserving both low- and high-frequency details essential for accurate medical analysis.

To overcome these challenges, our approach leverages perceptual loss, which can eliminate the need for registration and more accurately relate images from disparate domains. Traditional loss functions used in CycleGAN can result in hallucinated features in the enhanced images [25]. By incorporating perceptual loss [24], more interpretable images are generated that are more robust to registration artifacts [26]. This method can enhance the reliability and consistency of images from handheld ultrasound devices to bridge the gap with expensive high-end systems for greater equity in access to health care.

## Methods

### Study Design

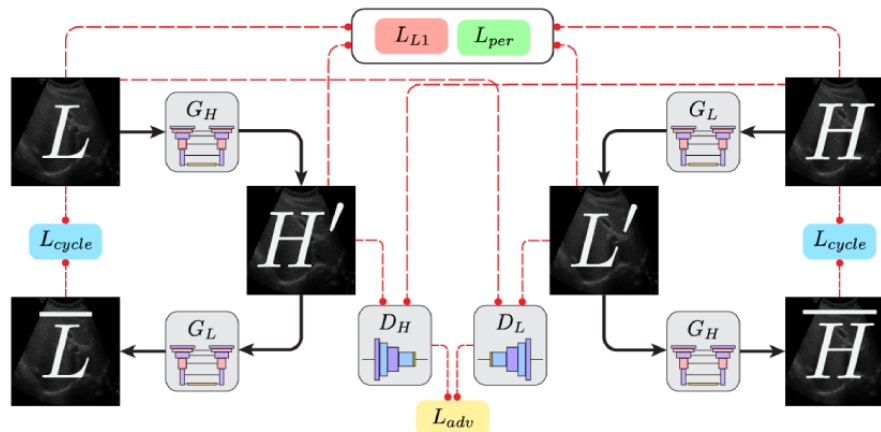
In this study, we aim to address the challenge of enhancing ultrasound image quality, particularly for images captured by compact, portable devices that often suffer from lower quality due to hardware limitations. To achieve this, we used a CycleGAN framework enhanced with perceptual loss. This approach focuses on processing nonregistered image pairs from various organ systems, ensuring that the enhanced images retain anatomical integrity and closely mimic high-resolution outputs. Our method is designed to be robust, versatile, and applicable across diverse clinical settings.

### Model Overview

Our framework for generating high-quality images is a modification of the CycleGAN architecture, designed to map between 2 distinct imaging domains. In ultrasound image enhancement, these domains correspond to low-quality (domain  $L$ ) and high-quality (domain  $H$ ) images. The model uses 2 generators,  $G_L$  and  $G_H$ , and 2 discriminators,  $D_L$  and  $D_H$  (Figure 1). Note that the generators  $G_L$  and  $G_H$  share the same model architecture. Similarly, the discriminators  $D_L$  and  $D_H$  share the

same model architecture. The generator  $G_L$  is responsible for converting an image from domain  $H$ , which represents high-quality images, to domain  $L$ , characterized by low-quality images. Conversely, the generator  $G_H$  performs the opposite transformation, taking an image from domain  $L$  and converting it to align with domain  $H$ . This bidirectional transformation process is essential for the task of image enhancement, as it allows for the improvement of low-quality images by translating them into their high-quality counterparts. The discriminators aim to distinguish real images in their respective domains from those transformed by the generators. A unique feature of this approach is the cycle consistency loss [13], which plays a crucial role in image quality enhancement. This loss ensures that when an image is translated to the other domain and then reverted to its original domain, it closely resembles the original image. Specifically, for enhancing low-quality images to high-quality images, the cycle consistency loss maintains the integrity of the image content throughout the transformation process. This prevents the introduction of artifacts and ensures that the enhanced image retains the essential features of the original low-quality image, resulting in a high-quality output that remains true to the source. After training, the  $G_H$  generator is used to enhance images, maintaining essential structural attributes while improving clarity and resolution.

**Figure 1.** An overview of the cycle generative adversarial network model training and loss computation framework. The solid black arrows indicate the flow of data. The dashed red arrows indicate the flow of information for loss computation.



### Model Description

GANs have seen transformative advancements, with CycleGAN [13] representing a significant milestone in facilitating unsupervised image-to-image translations. The GAN architecture comprises 2 primary modules: the generator and the discriminator.

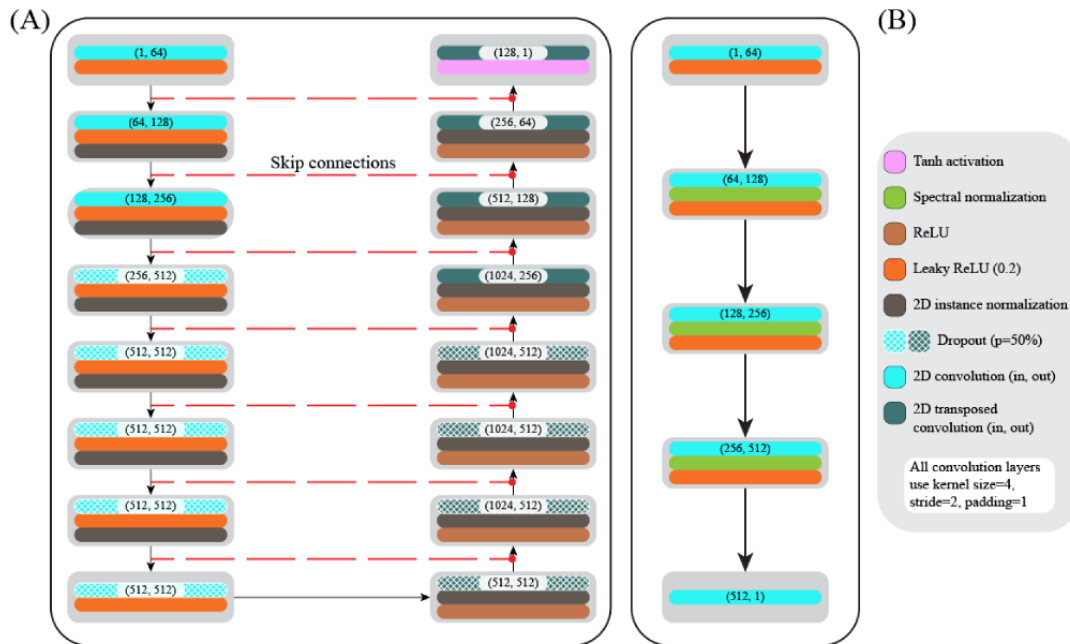
The generator (Figure 2A) architecture is inspired by the generator used by Isola et al [27]. The generator network is structured as a UNet [28], divided into encoding and decoding phases, incorporating detailed mechanisms for both down-sampling and up-sampling the input data. The encoder initiates with a 64-channel 2D convolutional layer designed to capture broad contextual details. This phase uses multiple down-sampling layers, each comprising a convolutional layer with instance normalization and leaky rectified linear unit (ReLU) activation functions. The instance normalization layers

stabilize the training process by normalizing the feature maps, while leaky ReLU activations introduce nonlinearity and mitigate the vanishing gradient problem. To enhance model robustness, dropout layers are included in deeper layers of the encoder. The down-sampling process reduces the spatial dimensions while increasing the depth, thereby emphasizing hierarchical feature extraction. In the decoding phase, the model uses transposed convolutional layers for up-sampling, which restores the spatial dimensions. Each up-sampling step involves skipping connections from the corresponding down-sampling layers, preserving detailed features from earlier stages. These layers also incorporate instance normalization and ReLU activations, where ReLU functions introduce nonlinearity, promoting sparse activations and efficient learning. The final layer uses a tanh activation function, scaling the output values to  $[-1, 1]$ , suitable for image generation tasks. This design ensures effective image enhancement by maintaining high-quality feature extraction and reconstruction.

The discriminator (Figure 2B) distinguishes between real and generated images. We use spectral normalization [29] to ensure stability during training. Its architecture begins with a convolutional layer that compresses spatial information and expands depth followed by a leaky ReLU activation layer. Subsequent layers maintain the use of spectral normalization to ensure 1 – Lipschitz continuity. This constraint on the spectral norm of each layer’s weights helps to balance the generator and

discriminator during training. The PatchGAN [27,30] style discriminator output is a 30×30 grid with a depth of 1, which provides a spatial map indicating the likelihood of each region in the input image being real or generated. This final classification output allows for a more detailed evaluation of the image, helping to distinguish between authentic and synthetic content across different spatial locations.

**Figure 2.** The model architectures. The (A) generator and (B) discriminator model architectures. The figure legend lists the different layers in the models. ReLU: rectified linear unit.



**Loss Function**

**Perceptual Loss**

Conventional methodologies like mean squared error (MSE) and SSI rely on pixel-wise alignment which makes them unsuitable for nonregistered image pairs acquired using different devices. The LPIPS metric addresses these constraints by evaluating the perceptual similarity between images [24]. LPIPS leverages deep features extracted from pretrained convolutional networks, such as the visual geometry group network [31]. The LPIPS metric comparing images  $X$  and  $Y$  is given by the following equation.

$$\Phi(X, Y) = \sum_i w_i \cdot ||F_i(X) - F_i(Y)||_2$$

where  $F_i$  and  $w_i$  denotes the feature maps and optimized weights from the  $i^{th}$  layer of the pretrained network. Deep feature maps are systematically extracted from every layer within the network, ensuring a comprehensive reflection of the multi-scale characteristics of human perceptual judgment. These features are then unified through linear combination, optimizing the weights to align with perceptual judgments assessed by human evaluators. The LPIPS metric consistently outranks traditional metrics, showcasing superior performance across an array of perceptual judgment tasks [26]. This loss is calculated between

real images  $L$  and  $H$ , and those generated through the CycleGAN framework’s generators as follows.

$$L_{per}(G_H, G_L) = \Phi(H, H') + \Phi(L, L')$$

where  $H'$  represents  $G_H(L)$  and  $L'$  represents  $G_L(H)$ .

**Generator Loss**

The generator’s loss function is a linear combination of several distinct loss terms, each playing a pivotal role in optimizing image translation between the 2 domains. First, adversarial loss  $L_{adv}(D_H, D_L)$  induces the discriminators to perceive generated images as genuine, whether they are translated from low to high quality or vice versa.

$$L_{adv}(D_H, D_L) = MSE(I, D_H) + MSE(I, D_L)$$

Specifically, MSE calculates the discrepancy between the discriminator’s predictions and an array of ones. The array of ones represents the target output for real images, indicating that the discriminator should classify these images as genuine. By comparing the discriminator’s predictions to this ideal output, the MSE helps measure how far the generated images are from being perceived as real. These terms push the generator to produce images that can convince the discriminator they belong to the high-quality domain, thereby improving the realism and quality of the generated images. Using an array of ones ensures that the generator is continuously driven to reduce the difference

between its output and real high-quality images, enhancing its performance over time. The cycle loss  $L_{cycle}(G_H, G_L)$  prevents the loss of critical image features by ensuring that an image translated to the other domain and back yields the original image.

$$L_{cycle}(G_H, G_L) = \|H - \bar{H}\|_1 + \|L - \bar{L}\|_1$$

where  $\bar{H} = G_H(G_L(H))$ , and  $\bar{L} = G_L(G_H(L))$ . The  $L1$  loss  $L_{L1}(G_H, G_L)$  ensures that the generated images are closer to true high-quality images in an  $L1$  distance sense.

$$L_{L1}(G_H, G_L) = \|H - G_H(L)\|_1 + \|L - G_L(H)\|_1$$

Finally, the aggregate generator loss,  $L(G_H, G_L, D_H, D_L)$ , is computed by combining all individual loss terms weighed by their respective lambda constants.

$$\begin{aligned} L(G_H, G_L, D_H, D_L) &= \lambda_{adv} L_{adv}(D_H, D_L) + \lambda_{cycle} L_{cycle}(G_H, G_L) + \lambda_{L1} L_{L1}(G_H, G_L) \\ &+ \lambda_{per} L_{per}(G_H, G_L) \end{aligned}$$

By using this multifaceted loss function, the model ensures that the generators achieve high-quality image translations while preserving the intrinsic characteristics of the source domain.

### Discriminator Loss

The discriminator loss function is designed to evaluate the authenticity of images, incorporating the principle of label smoothing to further enhance the model's generalizability. The discriminator is tasked with distinguishing between real images from the dataset and synthetic images generated by the corresponding generator.

$$L(D_H, D_L) = \begin{cases} MSE(1, D_H) + MSE(1, D_L), & \text{if } \text{mean}(D_H) \text{ and } \text{mean}(D_L) < 0.9 \\ MSE(0.9, D_H) + MSE(0.9, D_L), & \text{otherwise} \end{cases}$$

For each domain, the discriminator computes scores for both real and synthetic images. Conventionally, discriminators are trained using hard labels, where real images are labeled as "1" and synthetic images as "0." However, hard labels can cause vulnerability to adversarial perturbations—small, intentional changes to the input that can deceive the model into making incorrect predictions—and lead to overconfidence, where the discriminator becomes excessively certain in its predictions. Label smoothing improves the generalization and robustness of neural networks by preventing overconfidence in predictions. Szegedy et al [32] demonstrated its effectiveness in reducing overfitting and enhancing performance in image classification. Similarly, Salimans et al [33] applied 1-sided label smoothing in the training of GANs, which helped stabilize training and improve the quality of generated images. These studies support the use of label smoothing as a strategy to mitigate the negative effects of hard labels. In our framework, if the mean scores of both real and synthetic images for the high-quality domain are less than 0.9, the label 1.0 is used. Otherwise, a smoothing factor of 0.9 is applied, meaning the real images are given a target value slightly less than 1, to prevent overconfidence and promote model robustness. The total discriminator loss,  $L(D_H, D_L)$ , is

then computed by aggregating the individual MSE losses  $L_{D_H}$  and  $L_{D_L}$  for high and low-quality domains, respectively.

### Implementation Details

All models were trained for 300 epochs with a batch size of 4 images. We used the Adam optimizer for model optimization, with a learning rate ( $3 \times 10^{-4}$ ) set for both the generators and the discriminators. A beta value of 0.9 for the first and 0.999 for the second moments were used in each optimizer. A learning rate scheduler reduced learning rates by half ( $\gamma=0.5$ ) every 100 epochs, to allow adaptability during training. Weights were assigned to each loss term:  $\lambda_{adv}=1$  for adversarial loss,  $\lambda_{cycle}=10$  for cycle-consistency loss,  $\lambda_{L1}=2$  for  $L1$  loss, and  $\lambda_{per}=10$  for perceptual loss. The overall dataset was split with 70% for model training, 10% for model validation, and 20% for the hold-out test set. Gradient scaling was used to optimize the model's precision and speed. Code implementation will be made publicly available.

### Evaluation

The synthetic high-quality images generated by the model are evaluated using 4 key metrics: SSI, LNCC, PSNR, and LPIPS. Each of these metrics provides a unique perspective on the fidelity and quality of the generated images compared to the ground truth.

- SSI evaluates the structural fidelity between the generated image  $H'$  and the actual high-quality image  $H$ . It considers 3 aspects: luminance, contrast, and structure. The SSI is computed as follows.

$$SSI(H, H') = \frac{(2\mu_H\mu_{H'} + C_1)(2\sigma_{HH'} + C_2)}{(\mu_H^2 + \mu_{H'}^2 + C_1)(\sigma_H^2 + \sigma_{H'}^2 + C_2)}$$

where  $\mu_H$  and  $\mu_{H'}$  are the mean intensities,  $\sigma_H^2$  and  $\sigma_{H'}^2$  are the variances, and  $\sigma_{HH'}$  is the covariance between  $H'$  and  $H$ . The constants  $C_1$  and  $C_2$  are used to stabilize the division.

- LNCC measures the local similarity in intensity patterns between the generated and actual high-quality images. This metric is particularly sensitive to local differences in intensity, making it useful for detecting fine-grained discrepancies. LNCC is calculated by dividing the cross-correlation of local image patches by the product of their local SDs.

$$LNCC(H, H') = \frac{\sum_{i,j} (H_{i,j} - \bar{H}_{i,j})(H'_{i,j} - \bar{H}'_{i,j})}{\sqrt{\sum_{i,j} (H_{i,j} - \bar{H}_{i,j})^2 \sum_{i,j} (H'_{i,j} - \bar{H}'_{i,j})^2}}$$

where  $\bar{H}_{i,j}$  and  $\bar{H}'_{i,j}$  are the local means of the patches.

- PSNR quantifies the error signal strength between the generated and actual images, derived from the MSE. It is defined as follows.

$$PSNR(H, H') = 10 \cdot \log_{10} \left( \frac{MAX_I^2}{MSE(H, H')} \right)$$

where  $MAX_I$  is the maximum possible pixel value of the image, and  $MSE(H, H')$  is the mean squared error between  $H'$  and  $H$ . A higher PSNR indicates that the generated image is closer to the high-quality reference.

- LPIPS [24] is a perceptual metric that compares deep features extracted from a neural network rather than directly comparing pixel values. This approach better aligns with human perception of image similarity. LPIPS is computed by passing both the generated and actual images through a pretrained deep network and measuring the distance between their respective feature representations. Lower LPIPS values indicate higher perceptual similarity.

Together, these metrics provide a comprehensive evaluation of the generated images, which is particularly relevant for our image enhancement task, where unregistered low-high quality image pairs are compared. SSI and LNCC assess how well the structural and intensity patterns are preserved during enhancement, even when images are not perfectly aligned. PSNR quantifies the reduction in error relative to the original image, indicating overall fidelity. LPIPS, on the other hand, evaluates perceptual quality, ensuring that the enhancement appears natural and realistic to human observers, even in challenging scenarios with misaligned inputs.

### Ethical Considerations

This study did not involve the collection of new data from human participants. The dataset used is publicly available and provided as part of the 26th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2023) USEnhance Challenge [34,35]. Therefore, no ethics review or approval was required for this study. As the study used publicly available data provided by the organizers of the challenge, informed consent specific to this research was

not required. It is assumed that the original informed consent for data collection includes provisions for secondary analysis without requiring additional consent. All images used in this study are fully deidentified with no personal health information included. The dataset provided by the challenge organizers ensured anonymity, thus protecting the privacy and confidentiality of any potential human participants. There was no direct interaction with human participants in this study; hence, no compensation was provided. No images in this study or its supplementary materials allow for the identification of individual participants. All data are deidentified and anonymous, ensuring that no individual can be recognized from the images used.

## Results

### Overview

This study uses a dataset consisting of 2100 ultrasound images, including 1050 pairs of low- and high-quality images (Table 1 [34]). These images were collected from 131 patients with suspected thyroid tumors, carotid plaque, or breast cancer, along with healthy participants. During scans, volunteers were instructed to hold their breath for approximately 10 seconds to minimize deformation, and landmark points were noted for nonrigid registration to ensure the creation of accurate data pairs. This well-curated dataset provides a robust foundation for this study. This dataset was provided by the organizers of the MICCAI 2023 USEnhance Challenge [34]. Our baseline compares low-quality images directly to high-quality images without any enhancement or learning-based processing, serving as the starting point for evaluating the effectiveness of various models, including our approach. The models benchmarked in this study include Pix2Pix [27], MSPGAN [11], CycleGAN [13], RegGAN [23], and stable CycleGAN [12]. Among these, MSPGAN, RegGAN, and stable CycleGAN are the most recent advancements and are considered state-of-the-art for this task. To rigorously assess the improvements offered by our method, we computed the statistical significance of our results using the 1-sided Wilcoxon signed rank test.

**Table 1.** Dataset summary across different ultrasound devices and organs.

Organ	Low-end device	High-end device	Patients (n=131), n (%)	Ultrasound image pairs, n (%)		
				Training (n=840)	Testing (n=210)	Total (n=1050)
Thyroid	mSonics MU1	Toshiba Aplio 500	33 (25.2)	233 (27.7)	58 (27.6)	291 (27.7)
Carotid	SSUN	Toshiba Aplio 500	54 (41.2)	229 (27.3)	57 (27.1)	286 (27.2)
Abdomen	SSUN	General Electric LOGIQ E9	21 (16)	217 (25.8)	54 (25.7)	271 (25.8)
Breast	mSonics MU1	Aixplorer ultrasound system (SuperSonic Imaging SA)	23 (17.6)	161 (19.2)	41 (19.5)	202 (19.2)

### Quantitative Results

In the evaluation of the SSI, our proposed method achieved a score of 0.2889 (Table 2), surpassing the reference low baseline (0.2363;  $P < .001$ ), as well as CycleGAN (0.2622;  $P < .001$ ) and stable CycleGAN (0.2502;  $P < .001$ ). This places our method on par with the top-performing models like Pix2Pix (0.2862;

$PP > .99$ ), MSPGAN (0.2796;  $P < .001$ ), and RegGAN (0.2809;  $P < .001$ ). Among the methods evaluated, stable CycleGAN exhibited the lowest SSI score, indicating the least effective structural preservation. Pix2Pix, on the other hand, performed slightly better than MSPGAN and RegGAN, highlighting its strength in maintaining structural details. For LNCC, our method recorded a score of 0.8454, which is significantly higher than

the reference low baseline (0.7836;  $P < .001$ ) and comparable to the scores achieved by MSPGAN (0.8535;  $P > .99$ ) and Pix2Pix (0.8491;  $P > .99$ ). While MSPGAN led in LNCC, the differences between the top performing methods are minimal, underscoring the similar performance levels across these models. Notably, CycleGAN and stable CycleGAN scored 0.8271 ( $P < .001$ ) and 0.8145 ( $P < .001$ ), respectively, showing lower but still competitive performance.

In terms of PSNR, the proposed method achieved a score of 15.8935, which is a marked improvement over the reference low baseline (14.2978;  $P < .001$ ). Although Pix2Pix (16.3914;  $P > .99$ ) and MSPGAN (16.2602;  $P > .99$ ) reported higher PSNR values, indicating lower overall error between the generated and high-quality images, the differences between these models and our approach are modest. RegGAN also performed well

with a score of (16.2721;  $P > .99$ ), while CycleGAN (14.9126;  $P < .001$ ) and stable CycleGAN (14.9430;  $P < .001$ ) had lower PSNR values, indicating higher error rates. Finally, for the LPIPS metric, our method demonstrated the best performance with a score of 0.4490, significantly lower than Pix2Pix (0.4664;  $P < .001$ ), MSPGAN (0.4709;  $P < .001$ ), and RegGAN (0.4855;  $P < .001$ ). This indicates that our method produced images that were perceptually closer to high-quality outputs. CycleGAN and stable CycleGAN reported LPIPS scores of 0.4828 ( $P < .001$ ) and 0.5005 ( $P < .001$ ), respectively, with stable CycleGAN showing the least favorable performance among all models in terms of perceptual quality. Across these metrics, while certain models like Pix2Pix and MSPGAN excel in specific metrics such as LNCC and PSNR, our approach consistently delivers competitive performance, particularly in SSI and LPIPS, making it a robust framework for ultrasound image enhancement.

**Table 2.** Performance evaluation of models on the test set.

Model configurations	LNCC <sup>a</sup> ↑ <sup>b</sup>	SSI <sup>c</sup> ↑	PSNR <sup>d</sup> ↑	LPIPS <sup>e</sup> ↓ <sup>f</sup>
Reference low	0.7836 <sup>g</sup>	0.2363 <sup>g</sup>	14.2978 <sup>g</sup>	0.5080 <sup>g</sup>
Pix2Pix [27]	0.8491	0.2862	16.3914	0.4664 <sup>g</sup>
MSPGAN <sup>h</sup> [11]	0.8535	0.2796 <sup>g</sup>	16.2602	0.4709 <sup>g</sup>
CycleGAN <sup>i</sup> [13]	0.8271 <sup>g</sup>	0.2622 <sup>g</sup>	14.9126 <sup>g</sup>	0.4828 <sup>g</sup>
RegGAN <sup>j</sup> [23]	0.8475	0.2809 <sup>g</sup>	16.2721	0.4855 <sup>g</sup>
Stable CycleGAN [12]	0.8145 <sup>g</sup>	0.2502 <sup>g</sup>	14.9430 <sup>g</sup>	0.5005 <sup>g</sup>
Proposed method	0.8454	0.2889	15.8935	0.4490

<sup>a</sup>LNCC: locally normalized cross-correlation.

<sup>b</sup>↑: higher scores are better.

<sup>c</sup>SSI: structural similarity index.

<sup>d</sup>PSNR: peak signal-to-noise ratio.

<sup>e</sup>LPIPS: learned perceptual image patch similarity

<sup>f</sup>↓: lower scores are better.

<sup>g</sup>Statistically significant improvement.

<sup>h</sup>MSPGAN: multilevel structure-preserved generative adversarial network.

<sup>i</sup>CycleGAN: cycle-consistent generative adversarial network.

<sup>j</sup>RegGAN: registration generative adversarial network.

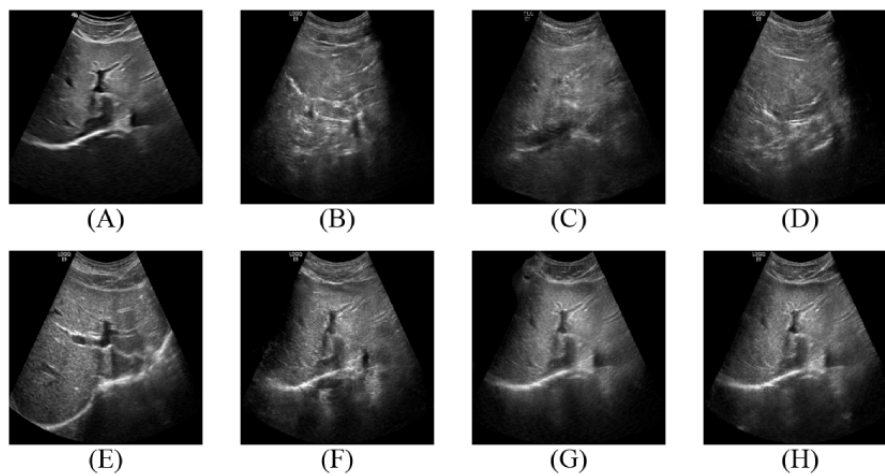
## Qualitative Results

Qualitative analysis further illustrates the differences in the generated images across the models. As shown in [Figures 3 and 4](#), methods such as Pix2Pix, MSPGAN, and RegGAN, despite their higher scores in SSI, LNCC, and PSNR, often produce images that lack anatomical detail and introduce distortions that may affect clinical interpretation. In contrast, methods like CycleGAN, Stable CycleGAN, and our proposed approach

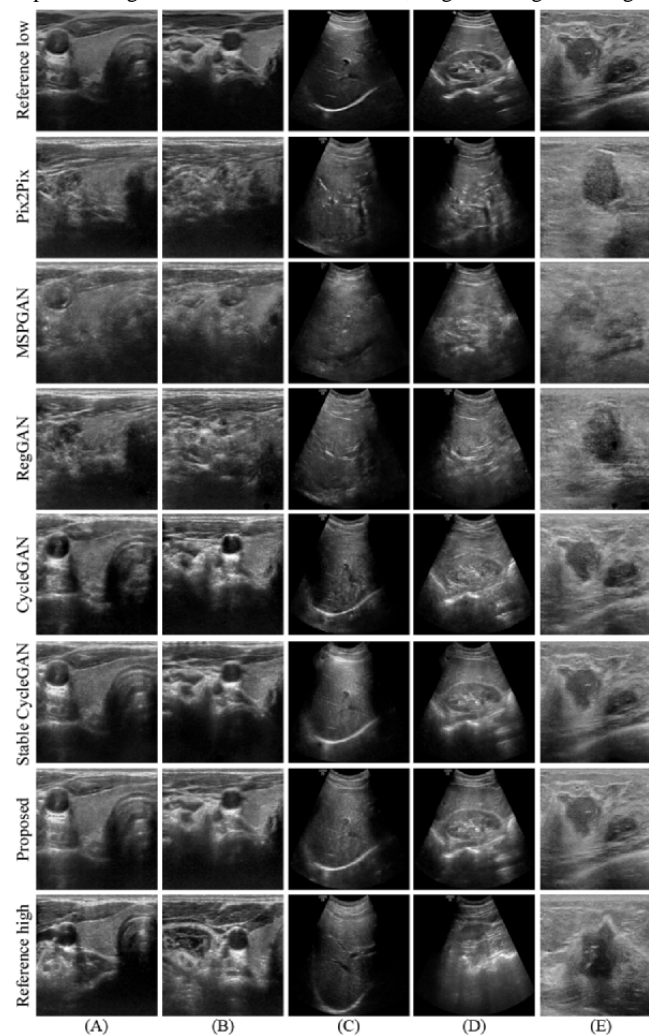
maintain the integrity of anatomical structures, ensuring that the generated images closely resemble the original high-quality images. Our approach is particularly effective in preventing the loss of critical diagnostic information, which is essential for accurate medical assessments. While quantitative metrics provide a useful evaluation framework, the qualitative results underscore the importance of preserving anatomical integrity, an area where our method excels.



**Figure 3.** A comparative visualization of ultrasound scans from the test set, showcasing the performance of different enhancement frameworks on the same high-low quality image pair. Each subfigure corresponds to a different model's output, allowing for a direct comparison of the anatomical preservation and image quality achieved by each approach. (A) reference low, (B) Pix2Pix [27], (C) MSPGAN [11], (D) RegGAN [23], (E) reference high, (F) CycleGAN [13], (G) stable CycleGAN [12], and (H) proposed method. CycleGAN: cycle-consistent generative adversarial network; MSPGAN: multilevel structure-preserved generative adversarial network; RegGAN: registration generative adversarial network.



**Figure 4.** A comparative visualization of ultrasound scans from the test set, showcasing the performance of different enhancement frameworks on the same high-low quality image pairs. (A) Thyroid, (B) carotid, (C) liver, (D) kidney, and (E) breast. CycleGAN: cycle-consistent generative adversarial network; MSPGAN: multilevel structure-preserved generative adversarial network; RegGAN: registration generative adversarial network.



## Discussion

In this study, we developed and evaluated an advanced CycleGAN framework enhanced with perceptual loss to address the challenge of varying image quality in ultrasound imaging across different devices. Our primary objective was to improve the quality of low-resolution ultrasound images captured by portable devices while preserving anatomical integrity, which is critical for accurate clinical diagnostics. The results from our evaluation demonstrated that the integration of perceptual loss enhanced the quality of the generated images, achieving strong performance in key metrics such as SSI and LPIPS, though slightly lower in LNCC and PSNR compared to other models. These outcomes suggest that our approach represents a significant step toward bridging the gap between low- and high-quality ultrasound images, making it particularly beneficial for portable, handheld devices that often struggle with image quality due to hardware limitations.

The use of perceptual loss in our model allowed for a more direct and meaningful comparison between low- and high-quality images, which contrasts with previous studies that treated these domains as independent [12]. By leveraging paired images from different devices, our model was able to learn the nuances of quality differences in a manner that closely mirrors real-world clinical scenarios. This pairing led to significant improvements in metrics such as SSI and LPIPS, indicating that our model preserves structural fidelity and local intensity patterns more effectively than current state-of-the-art approaches. However, it is important to note that while some models, such as Pix2Pix, MSPGAN, and RegGAN, achieve high scores in SSI, LNCC, and PSNR, they often do so at the expense of anatomical integrity. These models tend to remove or alter critical anatomical structures, leading to a loss of valuable diagnostic information. In contrast, our approach retains the anatomical content while producing comparable performance in these metrics and outperforming all other models in LPIPS,

which measures perceptual quality. This balance between maintaining anatomical fidelity and achieving high image quality is a significant strength of our method, making it more suitable for clinical applications where accuracy is paramount.

Despite these promising results, there are some limitations to our approach that need to be addressed in future work. The reliance on perceptual loss, while beneficial for maintaining image fidelity, introduces additional computational complexity, leading to longer training times. This requirement could be a limitation in scenarios where computational resources are limited or rapid model deployment is necessary. Additionally, while our model has demonstrated strong performance across a well-curated dataset, the findings need to be validated through extensive real-world applications across diverse datasets and imaging conditions to ensure robustness and generalizability. Furthermore, the current model is designed to work across various organ systems and diseases, but future research could explore the development of more specialized models tailored to specific clinical contexts, potentially optimizing performance for targeted diagnostic tasks.

To conclude, this work introduced an advanced CycleGAN-based framework that effectively enhances ultrasound image quality across devices by using perceptual loss to train on paired images. Our findings demonstrate the feasibility of bridging the image quality gap between low- and high-quality ultrasound images, thereby improving the accessibility and equity of high-quality diagnostic imaging. As we move forward, it will be crucial to conduct clinical validation of this approach across a wide range of medical scenarios and explore its application to other imaging modalities. This result could pave the way for integrating our model into routine clinical practice, ultimately enhancing diagnostic accuracy and improving patient outcomes. By making high-quality imaging more accessible, particularly through portable ultrasound devices, our approach holds the potential to significantly impact health care delivery and patient care on a global scale.

## Acknowledgments

The authors thank the editor and the reviewers for their valuable input, suggestions, and comments. The dataset used for this study was provided by the organizers of the MICCAI (Medical Image Computing and Computer Assisted Intervention) 2023 USEnhance Challenge [34]. This work was supported by the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under award R21EB030691 and a University of California Los Angeles Radiology Exploratory Research Grant.

## Data Availability

The datasets generated during and/or analyzed during this study are available in the MICCAI (Medical Image Computing and Computer Assisted Intervention) 2023 USEnhance Challenge repository [34].

## Authors' Contributions

SA, AR, VI, VS, CA, and WS contributed to the overall conception, design, oversight, and guidance of the work. SA contributed to the CycleGAN (cycle-consistent generative adversarial network) framework development, training, and further analyses. SA drafted the initial manuscript. All authors contributed to the critical revision of the manuscript, approved the final draft for submission and publication, and agreed to be accountable for the work.

## Conflicts of Interest

None declared.

## References

1. Moore CL, Copel JA. Point-of-care ultrasonography. *N Engl J Med*. 2011;364(8):749-757. [doi: [10.1056/NEJMra0909487](https://doi.org/10.1056/NEJMra0909487)] [Medline: [21345104](https://pubmed.ncbi.nlm.nih.gov/21345104/)]
2. Martocchia A, Bentivegna E, Sergi D, Luciani M, Barlattani M, Notarangelo MF, et al. The point-of-care ultrasound (POCUS) by the handheld ultrasound devices (HUDs) in the COVID-19 scenario: a review of the literature. *SN Compr Clin Med*. 2023;5(1):1. [FREE Full text] [doi: [10.1007/s42399-022-01316-9](https://doi.org/10.1007/s42399-022-01316-9)] [Medline: [36407770](https://pubmed.ncbi.nlm.nih.gov/36407770/)]
3. Le PT, Voigt L, Nathanson R, Maw AM, Johnson G, Dancel R, et al. Comparison of four handheld point-of-care ultrasound devices by expert users. *Ultrasound J*. 2022;14(1):27. [doi: [10.1186/s13089-022-00274-6](https://doi.org/10.1186/s13089-022-00274-6)] [Medline: [35796842](https://pubmed.ncbi.nlm.nih.gov/35796842/)]
4. Laastad Sørensen M, Oterhals K, Pönitz V, Morken IM. Point-of-care examinations using handheld ultrasound devices performed by intensive care nurses in a cardiac intensive care unit. *Eur J Cardiovasc Nurs*. 2023;22(5):482-488. [doi: [10.1093/eurjcn/zvac089](https://doi.org/10.1093/eurjcn/zvac089)] [Medline: [36149811](https://pubmed.ncbi.nlm.nih.gov/36149811/)]
5. Zhou Z, Guo Y, Wang Y. Handheld ultrasound video high-quality reconstruction using a low-rank representation multipathway generative adversarial network. *IEEE Trans Neural Netw Learning Syst*. 2021;32(2):575-588. [doi: [10.1109/tnnls.2020.3025380](https://doi.org/10.1109/tnnls.2020.3025380)]
6. van der Pol HGA, van Karnenbeek LM, Wijkhuizen M, Geldof F, Dashtbozorg B. Deep learning for point-of-care ultrasound image quality enhancement: a review. *Appl Sci*. 2024;14(16):7132. [doi: [10.3390/app14167132](https://doi.org/10.3390/app14167132)]
7. Loizou C, Pattichis SC, Istepanian R, Pantziaris M, Kyriakou E, Tyllis T, et al. Ultrasound image quality evaluation. 2003. Presented at: 4th International IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine, 2003; April 24-26, 2003; Birmingham, United Kingdom. [doi: [10.1109/itab.2003.1222492](https://doi.org/10.1109/itab.2003.1222492)]
8. Chen C, Raymond C, Speier W, Jin X, Cloughesy TF, Enzmann D, et al. Synthesizing MR image contrast enhancement using 3D high-resolution ConvNets. *IEEE Trans Biomed Eng*. 2023;70(2):401-412. [doi: [10.1109/tbme.2022.3192309](https://doi.org/10.1109/tbme.2022.3192309)]
9. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. *Commun ACM*. 2020;63(11):139-144. [FREE Full text] [doi: [10.1145/3422622](https://doi.org/10.1145/3422622)]
10. Zhou Z, Wang Y, Guo Y, Qi Y, Yu J. Image quality improvement of hand-held ultrasound devices with a two-stage generative adversarial network. *IEEE Trans Biomed Eng*. 2020;67(1):298-311. [doi: [10.1109/tbme.2019.2912986](https://doi.org/10.1109/tbme.2019.2912986)]
11. Xia M, Yang H, Qu Y, Guo Y, Zhou G, Zhang F, et al. Multilevel structure-preserved GAN for domain adaptation in intravascular ultrasound analysis. *Med Image Anal*. 2022;82:102614. [doi: [10.1016/j.media.2022.102614](https://doi.org/10.1016/j.media.2022.102614)] [Medline: [36115099](https://pubmed.ncbi.nlm.nih.gov/36115099/)]
12. Huang L, Zhou Z, Guo Y, Wang Y. A stability-enhanced CycleGAN for effective domain transformation of unpaired ultrasound images. *Biomed Signal Process Control*. 2022;77:103831. [doi: [10.1016/j.bspc.2022.103831](https://doi.org/10.1016/j.bspc.2022.103831)]
13. Zhu J, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. 2017. Presented at: IEEE International Conference on Computer Vision (ICCV); October 22-29, 2017; Venice, Italy. [doi: [10.1109/iccv.2017.244](https://doi.org/10.1109/iccv.2017.244)]
14. Chang H, Lu J, Yu F, Finkelstein A. PairedCycleGAN: asymmetric style transfer for applying and removing makeup. 2018. Presented at: IEEE/CVF Conference on Computer Vision and Pattern Recognition; June 18-23, 2018; Salt Lake City, UT. [doi: [10.1109/cvpr.2018.00012](https://doi.org/10.1109/cvpr.2018.00012)]
15. Ye Z, Lyu F, Li L, Sun Y, Fu Q, Hu F. Unsupervised object transfiguration with attention. *Cogn Comput*. 2019;11(6):869-878. [doi: [10.1007/s12559-019-09633-3](https://doi.org/10.1007/s12559-019-09633-3)]
16. Chen X, Xu C, Yang X, Tao D. Attention-Gan for object transfiguration in wild images. 2018. Presented at: European Conference on Computer Vision 2018 (ECCV 2018); September 8-14, 2018:164; Munich, Germany. [doi: [10.1007/978-3-030-01216-8\\_11](https://doi.org/10.1007/978-3-030-01216-8_11)]
17. Teng L, Fu Z, Yao Y. Interactive translation in echocardiography training system with enhanced cycle-GAN. *IEEE Access*. 2020;8:106147-106156. [doi: [10.1109/access.2020.3000666](https://doi.org/10.1109/access.2020.3000666)]
18. Zhang Z, Yang L, Zheng Y. Translating and segmenting multimodal medical volumes with cycle- and shape-consistency generative adversarial network. 2018. Presented at: IEEE/CVF Conference on Computer Vision and Pattern Recognition; June 18-23, 2018; Salt Lake City, UT. [doi: [10.1109/cvpr.2018.00963](https://doi.org/10.1109/cvpr.2018.00963)]
19. Gadermayr M, Gupta L, Appel V, Boor P, Klinkhammer BM, Merhof D. Generative adversarial networks for facilitating stain-independent supervised and unsupervised segmentation: a study on kidney histology. *IEEE Trans Med Imaging*. 2019;38(10):2293-2302. [doi: [10.1109/tmi.2019.2899364](https://doi.org/10.1109/tmi.2019.2899364)]
20. Geng M, Deng Y, Zhao Q, Xie Q, Zeng D, Zeng D, et al. Unsupervised/semi-supervised deep learning for low-dose CT enhancement. *ArXiv*. Preprint posted online on August 8, 2018. [doi: [10.48550/ARXIV.1808.02603](https://doi.org/10.48550/ARXIV.1808.02603)]
21. Chen H, Zhang Y, Zhang W, Liao P, Li K, Zhou J, et al. Low-dose CT via convolutional neural network. *Biomed Opt Express*. 2017;8(2):679. [doi: [10.1364/boe.8.000679](https://doi.org/10.1364/boe.8.000679)]
22. Eun D, Jang R, Ha WS, Lee H, Jung SC, Kim N. Deep-learning-based image quality enhancement of compressed sensing magnetic resonance imaging of vessel wall: comparison of self-supervised and unsupervised approaches. *Sci Rep*. Aug 18, 2020;10(1):13950. [FREE Full text] [doi: [10.1038/s41598-020-69932-w](https://doi.org/10.1038/s41598-020-69932-w)] [Medline: [32811848](https://pubmed.ncbi.nlm.nih.gov/32811848/)]

23. Kong L, Lian C, Huang D, Li Z, Hu Y, Zhou Q. Breaking the dilemma of medical image-to-image translation. 2021. Presented at: NIPS'21: 35th International Conference on Neural Information Processing Systems; December 6-14, 2021; Virtual-only conference. URL: <https://openreview.net/forum?id=COGmZH2RnVR>
24. Zhang R, Isola P, Efros AA, Shechtman E, Wang O. The Unreasonable effectiveness of deep features as a perceptual metric. 2018. Presented at: IEEE/CVF Conference on Computer Vision and Pattern Recognition; June 18-23, 2018; Salt Lake City, UT. [doi: [10.1109/cvpr.2018.00068](https://doi.org/10.1109/cvpr.2018.00068)]
25. Cohen JP, Luck M, Honari S. Distribution matching losses can hallucinate features in medical image translation. 2018. Presented at: 21st International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2018); September 16-20, 2018:529-536; Granada, Spain. [doi: [10.1007/978-3-030-00928-1\\_60](https://doi.org/10.1007/978-3-030-00928-1_60)]
26. Armanious K, Jiang C, Fischer M, Küstner T, Hepp T, Nikolaou K, et al. MedGAN: medical image translation using GANs. *Comput Med Imaging Graph*. 2020;79:101684. [doi: [10.1016/j.compmedimag.2019.101684](https://doi.org/10.1016/j.compmedimag.2019.101684)] [Medline: [31812132](https://pubmed.ncbi.nlm.nih.gov/31812132/)]
27. Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. 2017. Presented at: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); July 21-26, 2017; Honolulu, HI. [doi: [10.1109/cvpr.2017.632](https://doi.org/10.1109/cvpr.2017.632)]
28. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. 2015. Presented at: 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015); October 5-9, 2015:234-241; Munich, Germany. [doi: [10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)]
29. Miyato T, Kataoka T, Koyama M, Yoshida Y. Spectral normalization for generative adversarial networks. 2018. Presented at: 6th International Conference on Learning Representations (ICLR 2018); April 30-May 3, 2018; Vancouver, BC. URL: <https://openreview.net/forum?id=B1QRgziT->
30. Li C, Wand M. Precomputed real-time texture synthesis with Markovian generative adversarial networks. Cham. Springer; 2016. Presented at: European Conference on Computer Vision 2016 (ECCV 2016); October 11-14, 2016:702-716; Amsterdam, The Netherlands. [doi: [10.1007/978-3-319-46487-9\\_43](https://doi.org/10.1007/978-3-319-46487-9_43)]
31. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. ArXiv. Preprint posted online on September 4, 2014. [[FREE Full text](#)]
32. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. 2016. Presented at: IEEE Conference on Computer Vision and Pattern Recognition; June 27-30, 2016; Las Vegas, NV. [doi: [10.1109/cvpr.2016.308](https://doi.org/10.1109/cvpr.2016.308)]
33. Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A, Chen X. Improved techniques for training Gans. Cambridge, MA. *Adv Neural Inf Process Syst*; 2016. Presented at: 30th International Conference on Neural Information Processing Systems (NIPS'16); December 5-10, 2016; Barcelona, Spain.
34. Guo Y, Shichong Z, Shi J, Wang Y. Ultrasound image enhancement challenge 2023. Zenodo. URL: <https://zenodo.org/records/7841250> [accessed 2023-04-08]
35. USenhance 2023: ultrasound image enhancement challenge (multiple organs). Grand Challenge. URL: <https://ultrasound.enhance2023.grand-challenge.org/datasets/> [accessed 2023-10-12]

## Abbreviations

- CycleGAN:** cycle-consistent generative adversarial network
- GAN:** generative adversarial network
- LNCC:** locally normalized cross-correlation
- LPIPS:** learned perceptual image patch similarity
- MICCAI:** Medical Image Computing and Computer Assisted Intervention
- MSE:** mean squared error
- MSPGAN:** multilevel structure-preserved generative adversarial network
- PSNR:** peak signal-to-noise ratio
- RegGAN:** registration generative adversarial network
- ReLU:** rectified linear unit
- SSI:** structural similarity index

*Edited by S Rizvi; submitted 27.03.24; peer-reviewed by N Puzhakkal, J Peng, B-S Shin; comments to author 15.06.24; revised version received 24.08.24; accepted 13.09.24; published 17.12.24*

*Please cite as:*

*Athreya S, Radhachandran A, Ivezić V, Sant VR, Arnold CW, Speier W*

*Enhancing Ultrasound Image Quality Across Disease Domains: Application of Cycle-Consistent Generative Adversarial Network and Perceptual Loss*

*JMIR Biomed Eng 2024;9:e58911*

*URL: <https://biomedeng.jmir.org/2024/1/e58911>*

*doi: [10.2196/58911](https://doi.org/10.2196/58911)*

*PMID:*

©Shreeram Athreya, Ashwath Radhachandran, Vedrana Ivezić, Vivek R Sant, Corey W Arnold, William Speier. Originally published in JMIR Biomedical Engineering (<http://biomsedeng.jmir.org>), 17.12.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Biomedical Engineering, is properly cited. The complete bibliographic information, a link to the original publication on <https://biomedeng.jmir.org/>, as well as this copyright and license information must be included.