Ultrasound Bone Image Generation via Monte Carlo Simulation and VR-CycleGAN

Ruize Li*

Southeast University, Nanjing, China

Abstract. To address the performance bottleneck in deep learning caused by the scarcity of paired orthopedic ultrasound data, this paper proposes a hybrid data generation framework fusing physical acoustic simulation with Generative Adversarial Networks (GANs). We first utilize Monte Carlo simulation with generalized parameter sampling to construct structured synthetic data covering diverse anatomical variations. This is followed by a proposed VR-CycleGAN for Sim-to-Real domain adaptation. Unlike standard CycleGAN variants that are prone to geometric distortions, our model integrates a VAE module to explicitly decouple stochastic texture synthesis from anatomical structure. This design imparts realistic clinical textures while rigorously preserving anatomical integrity using unpaired data. Experimental results demonstrate that the generated images not only achieve superior perceptual quality but also exhibit Nakagami distribution parameters ($m = 0.5856$, $\Omega = 0.0537$) that closely align with real ultrasound backscattering statistics ($m = 0.5991$, $\Omega = 0.0608$). This quantitative evidence confirms the framework's capability to reproduce complex physical speckle features and validates the potential of integrating physical priors with data-driven approaches for high-fidelity medical image synthesis.

1 Introduction

With the rapid advancement of deep learning technologies, Computer-Aided Diagnosis (CAD) has achieved significant breakthroughs in medical ultrasound imaging. Particularly in orthopedic applications, intelligent segmentation and detection algorithms significantly assist clinicians in fracture identification and anatomical measurement. However, the success of deep learning models relies heavily on the availability of massive, diverse, and precisely annotated training data.

In practical clinical scenarios, acquiring large-scale paired ultrasound data faces substantial challenges. On one hand, due to intrinsic speckle noise and artifacts, manual annotation is labor-intensive and highly susceptible to subjective variability. On the other hand, clinical data often exhibit a long-tail distribution, clustering around common pathologies while failing to cover rare lesions or extreme imaging angles. Consequently, data scarcity has become a primary bottleneck constraining the generalization capability of medical AI algorithms.

Generating synthetic data via computer simulation offers an effective solution to these issues. Physical models can produce virtual images with perfect Ground Truth at a low cost. However, existing physical simulation methods struggle to perfectly reproduce the complex tissue textures and nonlinear acoustic effects of real ultrasound, leading to a significant distribution discrepancy between the simulation domain (Source Domain) and the real clinical domain (Target Domain), known as the "Sim-to-Real Gap." Directly training models on coarse simulation data often fails to yield satisfactory performance on real-world tasks. Therefore, bridging this gap to leverage both the structural diversity of simulation and the realistic texture of clinical scans remains a critical research focus.

To address these challenges, this paper proposes a Sim-to-Real data generation framework designed to combine the strengths of physical priors and data-driven methods. Specifically, we first construct a Monte Carlo-based stochastic simulation environment, ensuring data distribution diversity through generalized sampling of physical parameters. Subsequently, an improved VR-CycleGAN is introduced to perform domain adaptation, mapping simulation images into the real style space. Unlike traditional methods, we emphasize not only visual realism but also the consistency of physical statistical properties in the generated data.

The main contributions of this paper are summarized as follows:

1.Proposed a hybrid simulation pipeline: We effectively integrate the geometric precision of Monte Carlo simulation with the texture generation capability of VR-CycleGAN, addressing the challenge of high-quality data acquisition under small-sample conditions.

2.Validated physical statistical consistency: Experiments demonstrate that the generated images not only excel in perceptual metrics (e.g., FID, LPIPS) but also exhibit grayscale distributions and Nakagami speckle statistics that highly approximate real ultrasound physics, proving the physical fidelity of our method.

*Corresponding author: 2506979476@qq.com

3.Constructed a high-quality data augmentation framework: This work establishes a low-cost and scalable paradigm for generating synthetic ultrasound data, which can potentially support downstream segmentation and detection tasks.

2 Related work

Existing research on ultrasound simulation can generally be categorized into three classes: physics-based numerical simulation, ray-tracing-based geometric simulation, and deep learning-based data-driven methods.

2.1 Physics-based numerical simulation

Early ultrasound simulations focused on high-fidelity modeling of acoustic propagation processes. The Field II framework proposed by Jensen et al., based on linear acoustics and pulse-echo models, is widely regarded as seminal work in this domain [1, 2]. To simulate nonlinear propagation and wave behavior in complex heterogeneous media, Treeby et al. introduced the k-Wave toolbox based on the finite-difference time-domain (FDTD) method, which significantly improved acoustic accuracy [3]. Although the SIMUS platform proposed by Garcia et al. further optimized numerical efficiency [4–6], such methods generally suffer from high computational costs. Furthermore, the generated images are often based on idealized assumptions, making it difficult to reproduce the complex speckle textures and artifacts found in real clinical ultrasound, resulting in insufficient "visual realism."

2.2 Ray-tracing and CT-driven simulation

To balance simulation speed and realism, researchers have introduced ray-acoustic approximations. The Monte-Carlo Ray-tracing (MCRT) method proposed by Mattausch et al. achieved interactive simulation by modeling the refraction and attenuation of sound beams within tissues [7]. Building on this, Salehi et al. combined convolutional ray-tracing techniques to realize CT-driven patient-specific ultrasound generation[8,9]. Such methods have demonstrated potential in surgical navigation scenarios; for instance, Li et al. utilized diagnostic CT to generate realistic ultrasound images to assist robotic spinal surgery[10]. While these methods excel in geometric structural accuracy and patient specificity, they remain limited by physical simplifications when handling complex acoustic field scattering effects (e.g., full-wave characteristics).

2.3 Deep learning-based simulation

With the rapid development of deep learning, directly utilizing neural networks for ultrasound image generation has emerged as a new trend. Hu et al. achieved B-mode image synthesis correlated with probe pose using conditional Generative Adversarial Networks (cGAN) [11], while Magnetti et al. further explored the direct mapping from patient anatomy to ultrasound appearance[12]. Recently, Stojanovski et al. introduced diffusion models into this domain, enhancing the diversity of generated samples [13]. Furthermore, hybrid strategies have gained attention, where GANs are employed to perform style transfer and texture enhancement on results from traditional ray-based or acoustic simulations[14,15]. Although purely data-driven methods possess significant advantages in texture generation, they often lack strict physical constraints, making them prone to anatomical deformations or hallucinations.

Summary In conclusion, existing methods struggle to simultaneously balance the diversity of physical distributions, the accuracy of anatomical structures, and the realism of textural appearance. To address this limitation, our proposed method aiming to achieve a unification of high physical fidelity and visual realism.

3 Method

3.1 Improved monte Carlo ray tracing simulation

To address the limitations of traditional optical-based MCRT methods in modeling acoustic interfaces (e.g., soft tissue-bone boundaries), we propose a physics-enhanced simulation pipeline. We introduce an acoustic impedance-based reflection model and a linear Time Gain Compensation (TGC) mechanism into the standard Mattausch & Goksel framework.

3.1.1 Acoustic impedance reflection model

Unlike optical models that rely on surface roughness, ultrasound reflection is governed by the acoustic impedance mismatch between tissues. When a ray strikes an interface between medium 1 and medium 2 with incident angle θ_i and transmission angle θ_t, the reflection coefficient R is formulated as:

$$R = \frac{Z_2 \cos\theta_i - Z_1 \cos\theta_t}{Z_2 \cos\theta_i + Z_1 \cos\theta_t} \quad (1)$$

where Z_1 and Z_2 represent the acoustic impedance of the respective tissues. This modification ensures accurately high reflectivity at bone surfaces and suppresses unrealistic energy penetration into the acoustic shadow regions. The schematic illustration of ultrasound reflection behavior at the acoustic interface is presented in Figure 1.

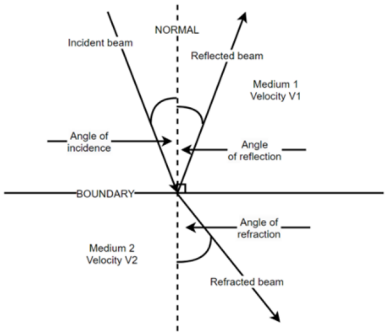

Figure 1. Schematic illustration of ultrasound reflection and propagation at the acoustic impedance interface.

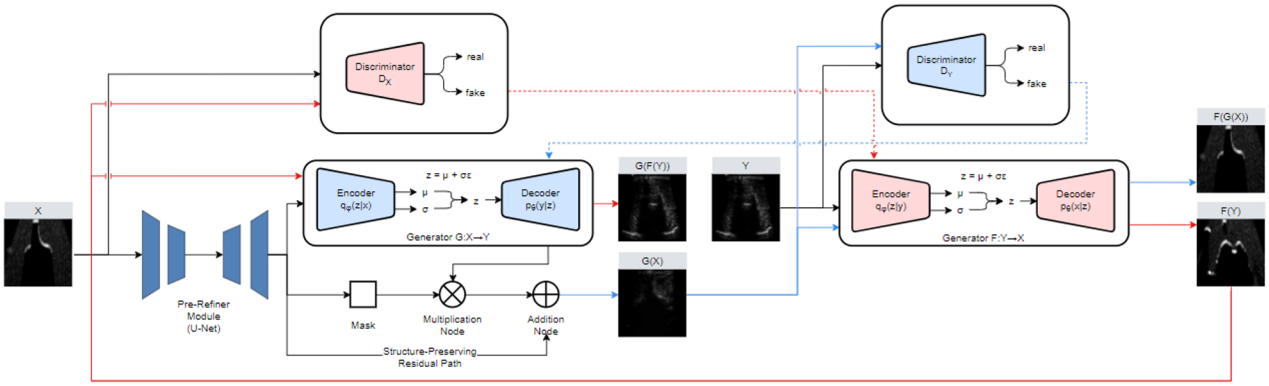

Figure 2. The framework consists of two mapping functions: generators G and F that translate between domain X and domain Y ($X \rightarrow Y$ and $Y \rightarrow X$, respectively). Adversarial discriminators D_X and D_Y are employed to enforce domain alignment, while the cycle-consistency losses ($F(G(X)) \approx X$ and $G(F(Y)) \approx Y$) ensure the preservation of anatomical structures during translation.

3.1.2 Linear time gain compensation (TGC)

To simulate the frequency-dependent attenuation of ultrasound energy over depth, we apply a linear gain compensation to the received energy signal. The gain function $G(x)$ with respect to propagation depth x is defined as:

$$G(x) = kx + b \tag{2}$$

where k is the gain slope and b is the initial bias. This aligns the simulated brightness distribution with clinical B-mode images.

3.1.3 Simulation setup

The overall pipeline consists of: (i) mapping CT volumes to acoustic grids (density/speed of sound); (ii) stochastic ray emission and tracing; and (iii) TGC correction and B-mode reconstruction. The key simulation parameters derived from the setup are detailed in Table 1.

Table 1. Simulation Parameters Configuration

Parameter	Value	Parameter	Value
Sound Speed	1500 m/s	Transducer Radius	3 cm
Center Freq.	4.5 MHz	Opening Angle	60°
Element Count	512	Axial Resolution	0.32 mm
Grid Resolution	145 μm	Imaging Depth	15 cm

3.2 VR-CycleGAN network architecture

To bridge the domain gap between the physics-based simulation and real clinical ultrasound, we propose the VR-CycleGAN. As illustrated in Figure 2 , this generator cascades a deterministic structural refinement module with a stochastic textural synthesis module, fused via an anatomy-guided gating mechanism.

3.2.1 Pre-Refinement module

The raw simulated image (x_{sim}) typically exhibits binary-like sharp edges, which differs significantly from the soft tissue transitions in clinical scans. The Pre-Refiner serves as a structural aligner to transform x_{sim} into a natural biological structural anchor, denoted as $x_{refined}$.

We employ a U-Net architecture with long-range skip connections. Let f_l be the feature map at layer l. The basic convolutional block is defined as:

$$f_l = \sigma\big(\mathrm{IN}(W_l * f_{l-1} + b_l)\big) \tag{3}$$

where IN(·) denotes Instance Normalization and σ(·) is the LeakyReLU activation. To preserve high-frequency geometric information (e.g., bone boundaries) during downsampling, we concatenate the encoder features E_i with the upsampled decoder features D_i:

$$D_{out}^{(i)} = \mathcal{F}_{dec}\big([E_i \oplus \mathrm{Upsample}(D_{in}^{(i)})]\big) \tag{4}$$

The output $x_{refined}$ retains the precise anatomical topology defined by the Monte Carlo simulation while adapting the feature distribution to the soft tissue domain.

3.2.2 Global-VAE generator with FC-bottleneck

To address the limitations of deterministic texture synthesis in standard CNNs, we model the generation process as a conditional probabilistic distribution $P(y_{tex}|z, x_{refined})$, as depicted in Figure 3. We introduce a Fully-Connected (FC) Bottleneck to strictly decouple texture from spatial structure. Specifically, the encoder features are flattened to extract global statistics, which are then projected to latent parameters:

$$(\mu, log\,\sigma^2) = \mathrm{FC}\left(\mathrm{Flatten}\left(\mathcal{E}_{conv}(x_{refined})\right)\right) \tag{5}$$

The latent variable z is subsequently sampled via the reparameterization trick $z = \mu + \sigma \odot \epsilon$, with $\epsilon \sim \mathcal{N}(0, I)$. This mechanism forces the network to ignore local spatial coordinates and focus solely on global stylistic distributions (e.g., speckle density) to reconstruct the texture residual y_{tex}.

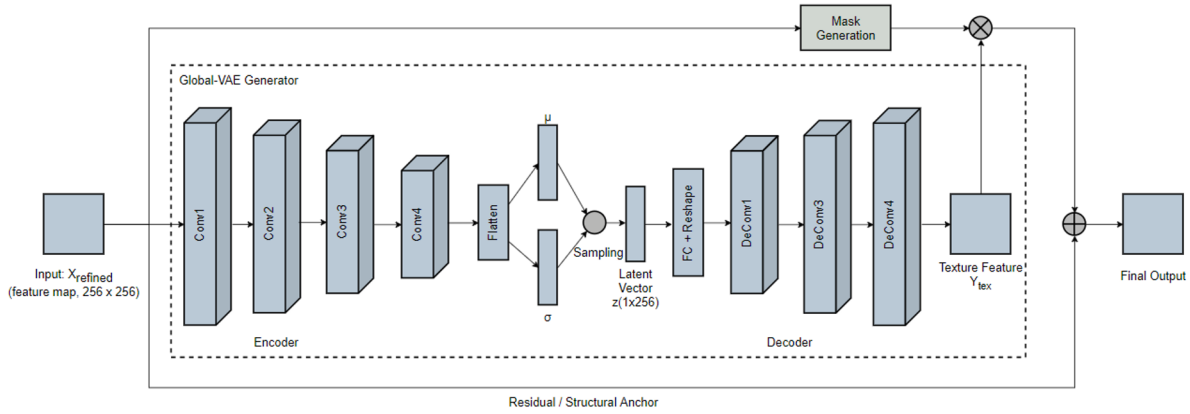

Figure 3. This module employs a Fully-Connected (FC) bottleneck to extract global statistical representations from input features. By sampling the latent variable z via the reparameterization trick, the decoder synthesizes stochastic speckle textures (y_{tex}) that are decoupled from spatial anatomical structures.

4 Result

4.1 Validation of MCRT simulation

Based on the proposed acoustic parameter modeling and the improved MCRT pipeline, we generated synthetic ultrasound data of the spine. Figure 4 illustrates a representative comparison between the baseline and improved simulation.

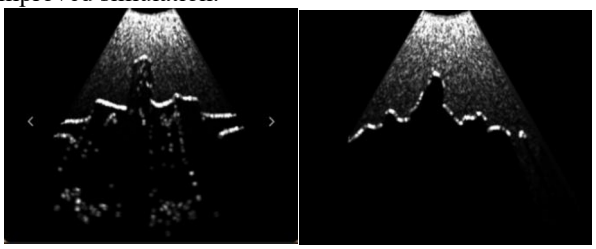

Figure 4. Comparison of Monte Carlo simulation results before (Left) and after (Right) improvement. The images correspond to different cross-sections of the spine.

The original MCRT simulation exhibits weak cortical reflections and insufficient attenuation at depth, failing to reproduce realistic acoustic shadowing. In contrast, the improved model, incorporating the acoustic impedance reflection model and TGC, exhibits prominent hyperechoic signals at bone boundaries and distinct acoustic shadows. The energy distribution follows a physically realistic decay trend, and the speckle texture statistically aligns with clinical acquisitions. These improvements ensure that the source domain data possesses both precise anatomical control and high physical fidelity for downstream tasks.

4.2 Comparative analysis

To validate the effectiveness of VR-CycleGAN, we compared it against state-of-the-art unsupervised image translation models: CycleGAN[16], deep-attention cycleGAN [17], and CUT[18]. Visual comparisons are shown in Figure 5, and quantitative metrics are detailed in Table 2.

As observed in Figure 4, standard CycleGAN and CUT both suffer from severe geometric distortions. Specifically, in low-SNR regions, anatomical structures are often blurred or hallucinated with unrealistic artifacts. Although deep-attention cycleGAN improves local texture via the attention mechanism, it leads to speckle over-enhancement, resulting in unnatural bright spots. In contrast, our method demonstrates superior structural preservation. It accurately maintains the morphology of the lamina and spinous processes while rendering clear, continuous echoes at tissue-bone interfaces.

As reported in Table 2, our method outperforms baselines across all metrics.

- Perceptual Quality: We achieved the best FID (135.81) and KID (0.1295), indicating that our generated distribution is statistically closest to the real clinical domain.
- Physical Consistency: The Nakagami parameters of our images ($m = 0.5856, \Omega = 0.0537$) show the lowest deviation from real ultrasound statistics, validating the efficacy of the physics-constrained generation strategy.

4.3 Ablation study

To verify the contribution of internal components, we conducted an ablation study with two variants: (1) removing the Pre-Refinement module (**w/o Refiner**), and (2) replacing the Global-VAE with a standard ResNet generator (**w/o VAE**).

The quantitative results in Table 3 demonstrate the impact of each module:

- w/o Refiner: Without structural refinement, the model struggles to adapt to the sharp edges of the simulation. This is evidenced by a significant degradation in FID (rising to 212.51), indicating a loss of realistic textural details.
- w/o VAE: Replacing the variational module leads to mode collapse. The generated speckle patterns lack diversity, resulting in a substantial deviation from the true Nakagami distribution parameters compared to the full model.

Table 2. Quantitative comparison with state-of-the-art methods

Metric	Sim-US	CycleGAN	Proposed	deep-attention cycleGAN	CUT	Real US
Global Distribution (↓ lower is better)						
FID	353.51	137.42	**135.81**	204.81	297.48	–
KID ($\times 10^{-1}$)	4.32	1.44	**1.29**	2.05	3.74	–
Brightness & Uniformity (closer to Real US is better)						
SNU (Std)	214.52	144.24	**140.13**	144.69	204.66	133.19
SNU (Range)	409.31	375.31	**374.65**	321.64	428.40	363.14
Speckle Consistency (closer to Real US is better)						
Nakagami-m	0.5653	0.4880	**0.5856**	0.7277	0.4313	0.5991
Nakagami-Ω	0.0846	0.0487	**0.0537**	0.0649	0.0708	0.0608

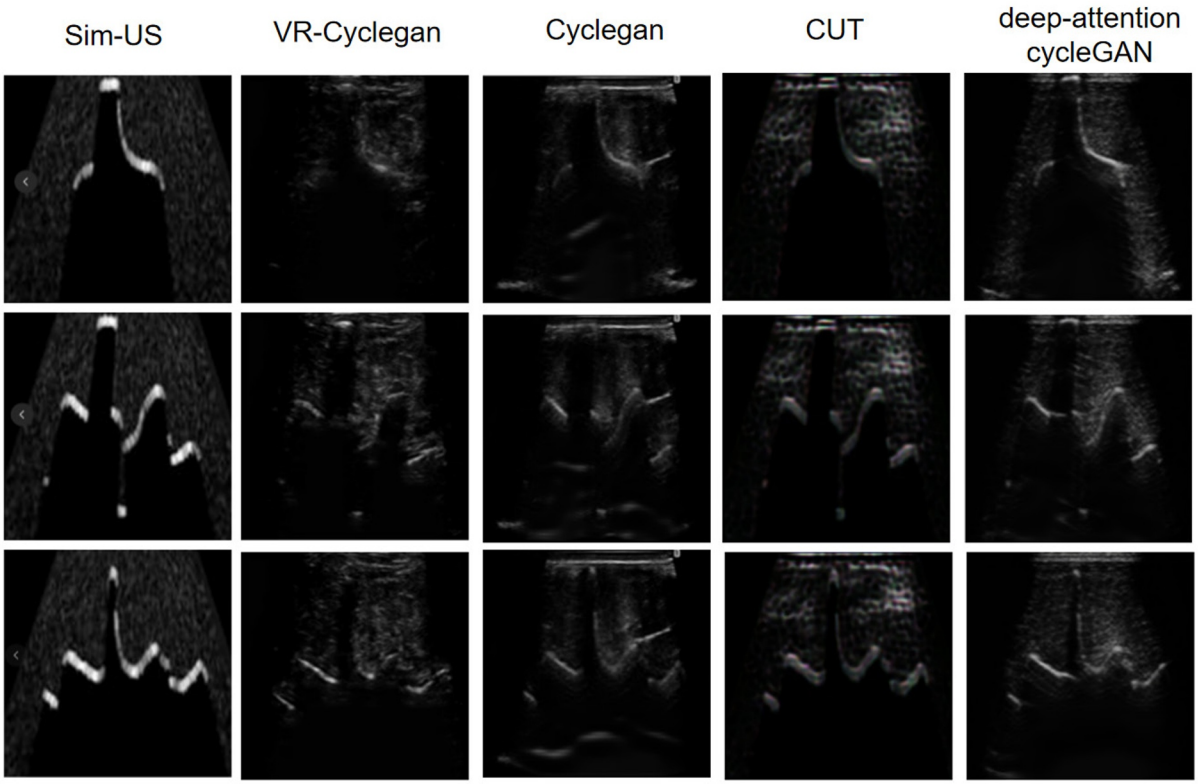

Figure 5. Comparison of ultrasound style translation results obtained by different models. From left to right: simulated ultrasound (Simulated US), VR-CycleGAN, CycleGAN, CUT, and CycleGAN with AG.

Table 3. Ablation study of key components

Metric	Proposed	w/o Refiner	w/o VAE
Perceptual Quality (↓)			
FID	**135.81**	212.51	180.47
KID	**0.1295**	0.2157	0.1757
Brightness (Ref: Std=133.19, Range=363.14)			
SNU (Std)	**140.13**	117.21	114.04
SNU (Range)	**374.65**	327.23	304.61
Speckle (Ref: m=0.5991, Ω=0.0608)			
Nakagami-m	**0.5856**	0.6971	0.7963
Nakagami-Ω	**0.0537**	0.0686	0.0698

In conclusion, the Pre-Refiner is essential for anatomical stability, while the Global-VAE is crucial for synthesizing physically realistic stochastic textures.

5 Conclusion

In this paper, we proposed a hybrid Sim-to-Real ultrasound generation framework, VR-CycleGAN, bridging the gap between physical simulation and clinical reality. Addressing the data scarcity and the "synthetic artifact" issues, we first established a robust source domain using an improved Monte Carlo ray-tracing pipeline incorporating acoustic impedance modeling and TGC. Subsequently, we introduced a cascaded generative network featuring a Pre-Refiner and a Global-VAE,

integrated via an anatomy-guided gated residual mechanism. Experimental results demonstrate that our approach not only outperforms state-of-the-art unsupervised translation methods in perceptual metrics (FID/KID) but also ensures strict physical consistency, as evidenced by the alignment of Nakagami speckle statistics and the preservation of osseous geometries. Future work will focus on two directions: extending the framework to 3D volumetric data generation for spatial reconstruction tasks, and deploying the synthesized datasets to train downstream segmentation networks, thereby validating the clinical utility of our method in computer-assisted orthopedic diagnosis.

References

1. J.A. Jensen, FIELD: A program for simulating ultrasound systems, *Proceedings (Suppl.) of Medical & Biological Engineering & Computing* **34**, 351–353 (1996)
2. J.A. Jensen, Simulation of advanced ultrasound systems using Field II, *IEEE Int. Symp. Biomed. Imaging (ISBI)* **2004**, 636–639 (2004)
3. B.E. Treeby, B.T. Cox, k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields, *J. Biomed. Opt.* **15**, 021314 (2010)
4. D. Garcia, SIMUS: An open-source simulator for medical ultrasound imaging. Part I: Theory & examples, *Comput. Methods Programs Biomed.* **218**, 106726 (2022)
5. A. Cigier, F. Varray, D. Garcia, SIMUS: An open-source simulator for medical ultrasound imaging. Part II: Comparison with four simulators, *Comput. Methods Programs Biomed.* **220**, 106774 (2022)
6. D. Garcia, F. Varray, SIMUS3: An open-source simulator for 3-D ultrasound imaging, *Comput. Methods Programs Biomed.* **250**, 108169 (2024)
7. O. Mattausch, O. Goksel, Monte-Carlo ray-tracing for realistic interactive ultrasound simulation, *Eurographics Workshop on Visual Computing for Biology and Medicine (VCBM)* **2016**, 173–181 (2016)
8. M. Salehi, S.-A. Ahmadi, R. Prevost, N. Navab, W. Wein, Patient-specific 3D ultrasound simulation based on convolutional ray-tracing and appearance optimization, *Lecture Notes in Computer Science* **9350**, 510–518 (2015)
9. K. Szostek, A. Piórkowski, Real-time simulation of ultrasound refraction phenomena using ray-trace based wavefront construction method, *Comput. Methods Programs Biomed.* **135**, 187–197 (2016)
10. A. Li, J. Han, Y. Zhao, K. Li, L. Liu, Realistic ultrasound synthesis based on diagnostic CT to facilitate ultrasound-guided robotic spine surgery, *IEEE Trans. Med. Robot. Bionics* **5**, 879–889 (2023)
11. Y. Hu, E. Gibson, L.-L. Lee, T. Vercauteren, J.A. Noble, Freehand ultrasound image simulation with spatially-conditioned generative adversarial networks, *arXiv preprint arXiv:1707.05392* (2017)
12. C. Magnetti, V. Zimmer, N. Ghavami, et al., Deep generative models to simulate 2D patient-specific ultrasound images in real time, *arXiv preprint arXiv:2005.04931* (2020)
13. D. Stojanovski, et al., Echo from noise: Synthetic ultrasound image generation via denoising diffusion probabilistic models, *Simplifying Medical Ultrasound (ASMUS 2023)* **2023**, 34–43 (2023)
14. S. Vitale, et al., Improving realism in abdominal ultrasound simulation using generative adversarial networks, *Med. Phys.* **52**, 4540–4556 (2025)
15. S. Shi, H. Li, Y. Zhang, X. Wang, Semantic information-guided attentional GAN-based ultrasound image synthesis method, *Biomed. Signal Process. Control* **102**, 107273 (2025)
16. J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)* **2017**, 2242–2251 (2017)
17. Y. Liu, Y. Lei, T. Wang, Y. Fu, X. Tang, W.J. Curran, T. Liu, P. Patel, X. Yang, CBCT-based synthetic CT generation using deep-attention cycleGAN for pancreatic adaptive radiotherapy, *Med. Phys.* **47**, 2472–2483 (2020)
18. T. Park, A.A. Efros, R. Zhang, J.-Y. Zhu, Contrastive learning for unpaired image-to-image translation, *Lecture Notes in Computer Science* **12354**, 319–345 (2020)