Diffusion as Sound Propagation: Physics-inspired Model for Ultrasound Image Generation

Marina Domínguez[1,2*], Yordanka Velikova[1,2*], Nassir Navab[1,2], and Mohammad Farid Azampour[1,2,3]

[1] Computer Aided Medical Procedures, Technical University of Munich, Germany
[2] Munich Center for Machine Learning, Munich, Germany
[3] Department of Electrical Engineering, Sharif University of Technology, Tehran, Iran

Abstract. Deep learning (DL) methods typically require large datasets to effectively learn data distributions. However, in the medical field, data is often limited in quantity, and acquiring labeled data can be costly. To mitigate this data scarcity, data augmentation techniques are commonly employed. Among these techniques, generative models play a pivotal role in expanding datasets. However, when it comes to ultrasound (US) imaging, the authenticity of generated data often diminishes due to the oversight of ultrasound physics.
We propose a novel approach to improve the quality of generated US images by introducing a physics-based diffusion model that is specifically designed for this image modality. The proposed model incorporates an US-specific scheduler scheme that mimics the natural behavior of sound wave propagation in ultrasound imaging. Our analysis demonstrates how the proposed method aids in modeling the attenuation dynamics in US imaging. We present both qualitative and quantitative results based on standard generative model metrics, showing that our proposed method results in overall more plausible images. Our code is available at github.com/marinadominguez/diffusion-for-us-images.

Keywords: Ultrasound · Synthetic Image Generation · Diffusion Models

1 Introduction

The scarcity of labeled medical data poses a significant challenge for training deep learning models, thereby encouraging the exploration of alternative solutions. Generative models have emerged as a popular approach to address this issue, allowing the generation of synthetic data that complements the limited available labeled examples [10]. By producing synthetic samples, generative models, such as diffusion models, can significantly reduce the problem of data scarcity, protect patient privacy, and address class imbalance [9].

Contrasting to natural images, where diffusion models have shown great success, ultrasound images are formed through the interpretation of echo patterns.

*Shared first authorship.

Therefore, generating realistic synthetic US images requires accounting for their underlying physics. B-mode ultrasound formation process involves emitting ultrasound pulses and capturing returning echoes generated by the interaction of sound waves with tissues [19]. This interaction leads to reflection, refraction, and attenuation, posing challenges in capturing internal structures [6]. Understanding these challenges, particularly attenuation, is crucial, as the upper regions appear more defined and brighter due to the stronger signal, while the lower regions become darker as the signal diminishes with depth. Addressing this gradual attenuation is key to enhancing the realism of synthetic US images.

Diffusion models in US imaging were initially used for tasks such as denoising and image generation. Initial studies have focused on reducing speckle noise and improving image clarity [5,25,16]. More recent research propose methods that reduce noise but preserve the speckle texture, enhancing image quality [1]. Experimental results from these studies show that such methods outperform traditional denoising techniques in both Peak Signal to Noise Ratio (PSNR) and Generalized Contrast to Noise Ratio (GCNR) [15]. Currently, the use of these models extends beyond denoising tasks. This includes semi-supervised learning for US segmentation [20] and image generation from semantic maps [18], both works showcasing significant improvements in segmentation accuracy. These studies demonstrate the power of diffusion models in enhancing US imaging and the potential for boosting DL models in tasks like image segmentation [2].

While diffusion models have shown remarkable success in generating high-quality images across various domains, their direct application to US imaging overlooks its physical properties critical to this modality [21,23,22]. Considering the significant differences between B-Mode ultrasound and natural images, applying the same synthetic image generation methods to both is impractical. Consequently, we have developed an approach that adapts standard diffusion models to better align with the actual process of US image generation.

Contributions This paper presents a novel approach to diffusion models designed specifically for US image generation. We propose a new noise scheduler inspired by the natural behavior of sound wave propagation. This scheduler simulates the attenuation of echoes returning to a US receiver. We consider the changes in depth-dependent US resolution and put more emphasis on regions closer to the probe, where images inherently show greater clarity, detail, and reliability of internal structures. We evaluate and compare the generated images qualitatively and quantitatively against a baseline with a conventional noise scheduler, both with and without semantic labels, on publicly available datasets.

2 Methodology

This section details the adaption of diffusion models for US synthesis by integrating a novel noise scheduler: the B-maps. We show that by introducing this scheduler, which is designed to mimic the natural attenuation of sound waves interacting with tissues, we are able to generate more plausible B-mode images.

2.1 Background

Forward process. DDPM [8] defines the forward diffusion process as a Markov chain where Gaussian noise is added in successive steps to obtain a set of noisy samples. Consider $q(x_0)$ as the uncorrupted (original) data distribution. Given a data sample $x_0 \sim q(x_0)$, a forward noising process p which produces latent x_1 through x_T by adding Gaussian noise at time t is defined as follows [1]:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t} \cdot x_{t-1}, \beta_t \cdot I), \quad \forall t \in \{1, \ldots, T\}, \tag{1}$$

where T represent the number of diffusion steps and $\beta_1, \ldots, \beta_T \in [0, 1]$ the noise scheduler across diffusion steps [9]. Considering $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{t=1}^{T} \alpha_t$, by applying the parametrization trick: $x_t = \sqrt{\alpha_t} x_0 + \sqrt{1-\alpha_t}\varepsilon$, t times, one can directly sample a step of the noised latent conditioned on the input x_0 [13]:

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t} x_0, (1-\bar{\alpha}_t) I) \tag{2}$$

Reverse process. The reverse process seeks to approximate a sample from the original data distribution $q(x_0)$ by starting from a standard Gaussian distribution $p(x_T) = \mathcal{N}(x_T; 0, I)$ and iteratively denoising towards x_0. To this end, we can parameterize this reverse process as follows:

$$p_\theta(x_{0:T}) = p(x_T) \prod_{t=1}^{T} p_\theta(x_{t-1}|x_t) \tag{3}$$

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(t)), \tag{4}$$

employing learned parameters to guide the reverse diffusion towards accurate reconstruction of the original data. This dual-phase approach can create high-quality images from noise, laying a solid foundation for our method.

2.2 B-Maps Definition

Central to our approach is the introduction of B-Maps; these matrices, have the same dimensions as the US images, and allow for precise control of the noise level at each pixel. In this way, we change how the standard DDPMs introduce noise in the image. While DDPM, originally designed for natural images, applies noise uniformly in the image, the B-Maps scale the noise across the vertical axis of the image, simulating the top-to-bottom image construction of US imaging. Our novel adaptation introduces more noise—and thus, faster convergence towards a standard Gaussian distribution—at the bottom of the image than at the top. As a result, we define a diffusion model that focuses on learning the distribution of the upper region of the image first before addressing the inherently noisier lower region, as the sound wave loses strength the deeper it propagates into the body. This prioritized learning process ensures that the most reliable details—those closer to the probe—are captured with higher fidelity and we also prevent coming up with artifacts and non-anatomically plausible features. An illustration of the definition of B-Maps is shown in Figure 1.

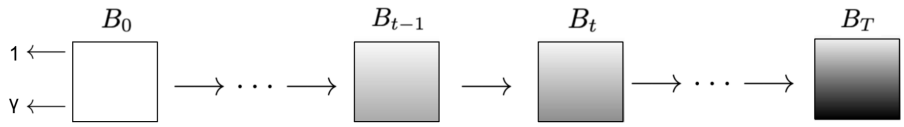

Fig. 1. Evolution of B-maps across time-steps. In every timestep, the values in the B-Maps decrease top-to-bottom from 1 to a number, γ. As the timestep increases, γ goes from 1 to $1-\epsilon$, with ϵ being a small fixed value in the interval $(0, 1)$.

2.3 Denoising Diffusion Probabilistic Models with B-Maps

Forward Process. We modify the standard forward pass (Eq. 2) that introduces the noise in the image centered in $\bar{\alpha_t} x_0$ by integrating noise in a way that better reflects US image formation. By incorporating our B-maps noise scheduler, we specifically adjust the distribution's mean and variance towards the image's lower part. This adjustment is done by point-wise multiplying the existing noise schedule α_t and the preceding image x_{t-1} with B-map scheduler B_t at time t. This leads to our proposed new forward distribution:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t \cdot \bar{B}_t} x_{t-1}, (1 - \bar{\alpha}_t \cdot \bar{B}_t)\mathbf{I}) \tag{5}$$

with: $\bar{\alpha}_t = \prod_{t=1}^{T} \alpha_t$ and $\bar{B}_t = \prod_{t=1}^{T} B_t$ and $\cdot$ denoting the point-wise multiplication. This method guides the diffusion model's behavior, introducing noise in the forward process by simulating the progressive attenuation of echo intensity. The visual representation of this modified process is shown in Fig. 2.

Reverse Process. Incorporating B-maps into our model alters traditional distributions (Eq. 4), leading to new derivations for the reverse process, detailed in the Supplementary Material. Thus, the newly derived posterior distribution for our diffusion model becomes:

$$q(x_{t-1}|x_t, x_0) \propto \mathcal{N}\left(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(t)\right),$$

where

$$\mu_\theta(x_t, t) = \frac{\sqrt{\alpha_t \cdot B_t(\mathbf{1} - \bar{\alpha}_{t-1} \cdot \bar{B}_{t-1})} x_t + \sqrt{\bar{\alpha}_{t-1} \cdot \bar{B}_{t-1}(\mathbf{1} - \alpha_t \cdot B_t)} x_0}{\mathbf{1} - \bar{\alpha}_t \cdot \bar{B}_t}$$

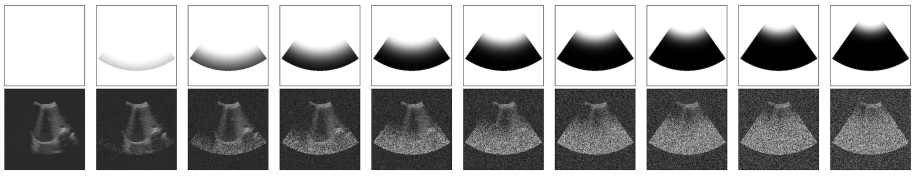

Fig. 2. Forward pass: Noise addition from bottom to top. Linearly-scheduled cone-shaped B-Maps on the top row and the visualization of the noising process of the US image in the bottom row. B-Maps are applied at each step, making the gaussian distribution converge earlier on the bottom than on the top.

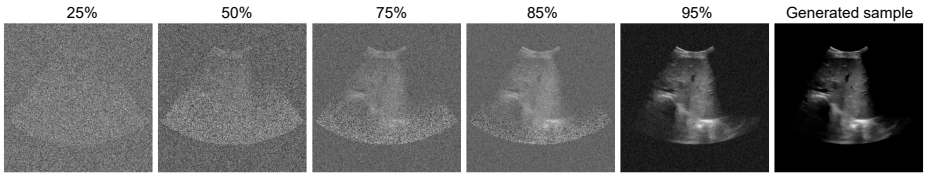

Fig. 3. Reverse Process: denoising the image. Initially focusing on the area near the probe, the model progresses to denoise the image toward the bottom, mimicking the way US images are generated.

and

$$\Sigma_\theta(t) = \frac{(\mathbf{1} - \alpha_t \cdot B_t)(\mathbf{1} - (\bar{\alpha}_{t-1} \cdot \bar{B}_{t-1}))}{\mathbf{1} - (\bar{\alpha}_t \cdot \bar{B}_t)}\mathbf{I}$$

This reverse distribution considers the varying noise levels influenced by B-maps, creating a more natural top-to-bottom reconstruction of US images, as we can see in Figure 3.

In our work, we introduced this novel noise scheduler into Guided-Diffusion (GD) [4] and Semantic Diffusion Models (SDM) [24]. Guided-Diffusion is based on the improved DDPMs [14] and enhances the generation process by utilizing guidance mechanisms. SDMs, on the other hand, extend the diffusion framework by incorporating semantic information. SDMs include the integration and utilization of semantic labels during the generation process and have greater control over the output characteristics by utilizing semantic information.

We modified these models by replacing the original equations that define the diffusion process with those incorporating B-Maps. Specifically, the equation that introduces noise in the forward pass of these baselines was replaced with our derived Eq. (5). Additionally, we defined our noise scheduler, the B-Maps, and adjusted the reverse process to incorporate them. Consequently, for the original variance in the models, we now use our newly derived variance, $\Sigma_\theta(t)$. These adaptations allow us to generate US images from datasets with and without semantic labels, using SDM and Guided-Diff correspondingly. By incorporating our B-Maps into these pre-existing frameworks, we can compare the effectiveness of our approach against the original models without B-Maps.

3 Experimental Setup

3.1 Datasets

SegThy dataset contains annotated 3D US images of the thyroid [11] from 28 healthy volunteers, acquired with Siemens Acuson NX-3 US machine with a 12MHz VF12-4 probe. We extracted 2D slices and labels from the 3D US scans and removed the images without a thyroid label or with empty labels, totalling 2,250 images, where 512 were used for validation. The dataset is available here.

CAMUS dataset includes 400 patient images for training and 50 for validation [12]. Each patient contributes four images at both end-diastole (ED) and end-systole (ES) across two- and four-chamber views. In total, 1600 training and 200 validation images. Following Stojanovski et al. [18], we applied five random affine and elastic deformations. This augmented the dataset to 8000 training and 1000 validation images. The dataset can be found here.

Liver images were acquired in-house using a tracked probe ACUSON Juniper (Siemens Healthineers, Erlangen, Germany) with a 5C1 convex probe. We scanned 14 volunteers aged between 22 and 34. After excluding images with more than 50% shadow, we ended up with 6,900 2D slices, where 1000 were used for validation. The images were horizontally padded to become square and resized to 256 for our experiment.

3.2 Experiments

We train our model separately on each dataset and analyse the image outputs both qualitatively and quantitatively. The qualitative evaluation involves a visual comparison of synthetic images produced by our approach against those generated by well-established baseline models, GD and SDM, to visually highlight the advancements our model offers in terms of image realism and fidelity.

Metrics Quantitatively, we calculate standard metrics in image generation model evaluation: Fréchet Inception Distance (FID), Learned Perceptual Image Patch Similarity (LPIPS), Structural Similarity Index Measure (SSIM), and Peak Signal-to-Noise Ratio (PSNR). In measuring FID [7], we choose to use a different feature layer of the Inception Network instead of the default *pool3* layer. We use the *first-max-pooling* and *second-max-pooling* layers that capture fundamental image features without relying on the ImageNet-specific learned parameters, offering a more relevant evaluation for US images. The metrics of LPIPS, SSIM, and PSNR are computed for the CAMUS and Thyroid datasets, given their availability of semantic labels. These metrics are calculated for all corresponding pairs of images of the synthetical data and the original data. To compare the results and draw conclusions, we calculate the average, standard deviation, and range of these metrics.

Training and Hyper-parameters We performed the training with a batch size of 4, learning rate of 0.0001 and 2000 diffusion steps. The image resolutions varied: 128 for CAMUS and SegThy and 256 for Liver. The number of training iterations also differed, with 28,000 for CAMUS, 36,000 for Liver, and 50,000 for SegThy, stopping earlier for some datasets as they achieved convergence. Following previous works [4], we selected a cosine scheduler for the α values and a square root scheduler for the B-maps. Through our experiments, we determined the optimal minimum values for the B-maps to be 0.96 and 0.97, which yielded the best results.

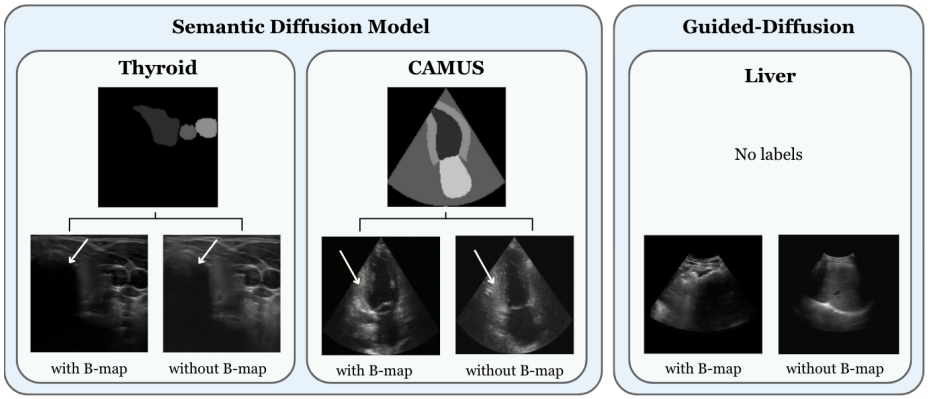

Fig. 4. Qualitative comparison: The top row displays label maps of SegThy and CAMUS datasets. For the liver dataset, no labels were available, and the Guided-Diffusion model was used. The bottom row shows the US images generated with B-Maps (left) versus without B-Maps (right) for each dataset. The overlaying arrows highlight key differences: in the thyroid dataset, the tracheal area appears darker in the image with the B-Map, consistent with the presence of air; in the CAMUS dataset, the borders of the heart chambers are better visualized in the image with the B-Map rather than without. No arrows were added to the liver images as they are not paired due to the absence of labels.

4 Results and Discussion

4.1 Qualitative Analysis

Visually, the integration of B-Maps has demonstrated a notable improvement in the synthetic images generated using both SDM and GD. Figure 4 showcases examples from each dataset, illustrating the improvement in image quality and realism, especially in terms of contrast in the upper regions of the images, aligning with the known characteristics of US scans. Additional visual comparisons are available in Supplementary Material.

4.2 Quantitative Evaluation

The improvement achieved by B-Maps is also evident in the quantitative metrics used to evaluate image quality. FID scores, as shown in Table 1, calculated with pytorch-fid [17], underscore the superiority of our proposed method over the baseline across all datasets. The scores from the 1^{st} and 2^{nd} max-pooling layers of the Inception Network for all synthetically generated images from the validation set demonstrate our method's ability to produce images that are closer to real US images, with a notable reduction in the FID scores. This improvement indicates that our generated images have higher fidelity and are statistically closer to the distribution of real US images.

In our evaluation, the LPIPS metric, in Table 2, computed using Perceptual-Similarity [26], provides insight into the perceptual quality of generated images,

Table 1. FID Scores across Datasets. Features for FID calculations are extracted from the *first-max-pooling* (1st MP) and *second-max-pooling* (2nd MP) layers of Inception. Results indicate that our method surpasses the baseline in generating datasets with more realistic images, as evidenced by the significantly lower FID scores.

	Baseline		**Proposed method**	
Dataset	1st MP	2nd MP	1st MP	2nd MP
Thyroid	4.259	14.315	**0.619**	**2.769**
CAMUS	3.581	12.769	**0.204**	**0.959**
Liver	20.746	73.447	**0.192**	**0.867**

reflecting how closely the synthetic data resembles real US scans. The results indicate that our method consistently yields lower LPIPS scores across Thyroid and CAMUS datasets, implying that images generated by our method are more realistic, aligning closely with the perceptual properties of real US scans.

Our SSIM and PSNR analyses, also detailed in Table 2 and computed using torchmetrics [3], offer additional insight into the image quality improvements achieved through our approach. While SSIM values show minimal differences from baseline methods—indicating comparable structural integrity—the PSNR values are significantly higher. This suggests that our method enhances image precision by improving the signal-to-noise ratio, thus generating clearer and sharper US images.

Table 2. LPIPS, SSIM, and PSNR Metrics Comparison: This table evaluates the quality of generated US images for Thyroid and CAMUS datasets using LPIPS, SSIM, and PSNR metrics. Lower LPIPS scores indicate closer resemblance to real images. Additionally, our approach achieves higher SSIM and PSNR values compared to baselines for both datasets, indicating better preservation of structural details and improved clarity, reflecting improved image quality.

		Baseline		**Our Method**	
	Metric	**Thyroid**	**CAMUS**	**Thyroid**	**CAMUS**
LPIPS	Mean	0.362	0.234	**0.316**	**0.161**
	Std. Dev.	0.129	0.127	**0.066**	**0.047**
	Range	0.628	0.534	**0.321**	**0.266**
SSIM	Mean	0.279	0.265	**0.292**	**0.297**
	Std. Dev.	0.137	0.131	**0.094**	**0.092**
	Range	0.523	0.587	**0.451**	**0.523**
PSNR	Mean	14.302	13.118	**16.798**	**15.871**
	Std. Dev.	5.526	4.511	**2.468**	**2.545**
	Range	19.707	18.899	**11.549**	**10.420**

5 Conclusion

This study presents a novel approach to US image synthesis by adapting diffusion models with B-Maps. Our method introduces a customized noise schedule that reflects the natural attenuation of US waves. This innovation significantly enhances the realism of synthetic US images, as supported by our comprehensive evaluation across several datasets.

While our model showcases promising advancements in synthetic US generation, it also opens the door to exploring more sophisticated models that further incorporate ultrasound's physical properties. Future directions could involve developing models that estimate attenuation maps at each diffusion step, offering even more precise control over the synthetic image generation process.

Acknowledgments. We thank the team at ImFusion (ImFusion GmbH, Munich, Germany) for providing their software to us.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Asgariandehkordi, H., Goudarzi, S., Basarab, A., Rivaz, H.: Deep ultrasound denoising using diffusion probabilistic models (2023)
2. Bi, Y., Jiang, Z., Clarenbach, R., Ghotbi, R., Karlas, A., Navab, N.: Mi-segnet: Mutual information-based us segmentation for unseen domain generalization (03 2023)
3. Detlefsen, N.S., Borovec, J., Schock, J., Harsh, A., Koker, T., Di Liello, L., Stancl, D., Quan, C., Grechkin, M., Falcon, W.: TorchMetrics - Measuring Reproducibility in PyTorch (Feb 2022). https://doi.org/10.21105/joss.04101, https://www.pytorchlightning.ai, repository code available at https://github.com/Lightning-AI/torchmetrics
4. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis (2021)
5. Goudarzi, S., Rivaz, H.: Deep ultrasound denoising without clean data. In: Medical Imaging 2023: Ultrasonic Imaging and Tomography. vol. 12470, pp. 131–136. SPIE (2023)
6. Grogan, S., Mount, C.: Ultrasound physics and instrumentation (Mar 2023), updated 2023 Mar 27. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK570593/
7. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. arXiv preprint arXiv:1706.08500 (Dec 2017)
8. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models (2020)
9. Kazerouni, A., Aghdam, E.K., Heidari, M., Azad, R., Fayyaz, M., Hacihaliloglu, I., Merhof, D.: Diffusion models for medical image analysis: A comprehensive survey (2022)
10. Kebaili, A., Lapuyade-Lahorgue, J., Ruan, S.: Deep learning approaches for data augmentation in medical imaging: A review. Journal of Imaging (2023), correspondence: su.ruan@univ-rouen.fr

11. Kroenke, M., Eilers, C., Dimova, D., Köhler, M., Buschner, G., Schweiger, L., Konstantinidou, L., Makowski, M., Nagarajah, J., Navab, N., Weber, W., Wendler, T.: Tracked 3d ultrasound and deep neural network-based thyroid segmentation reduce interobserver variability in thyroid volumetry. PLOS ONE **17**, e0268550 (07 2022). https://doi.org/10.1371/journal.pone.0268550
12. Leclerc, S., Smistad, E., Pedrosa, J., Østvik, A., Cervenansky, F., Espinosa, F., Espeland, T., Berg, E.A.R., Jodoin, P.M., Grenier, T., Lartizien, C., D'hooge, J., Løvstakken, L., Bernard, O.: Deep learning for segmentation using an open large-scale dataset in 2d echocardiography. IEEE Transactions on Medical Imaging **38**, 2198–2210 (2019), https://api.semanticscholar.org/CorpusID:73510235
13. Luo, C.: Understanding diffusion models: A unified perspective (2022)
14. Nichol, A., Dhariwal, P.: Improved denoising diffusion probabilistic models (2021)
15. Rodriguez-Molares, A., Rindal, O., D'hooge, J., Masoy, S., Austeng, A., Lediju Bell, M., Torp, H.: The generalized contrast-to-noise ratio: A formal definition for lesion detectability. IEEE Trans Ultrason Ferroelectr Freq Control **67**(4), 745–759 (Apr 2020). https://doi.org/10.1109/TUFFC.2019.2956855, epub 2019 Nov 29. PMID: 31796398; PMCID: PMC8354776
16. van de Schaft, V., van Sloun, R.J.: Ultrasound speckle suppression and denoising using mri-derived normalizing flow priors. arXiv preprint arXiv:2112.13110 (2021)
17. Seitzer, M.: pytorch-fid: FID Score for PyTorch. https://github.com/mseitzer/pytorch-fid (August 2020), version 0.3.0
18. Stojanovski, D., Hermida, U., Lamata, P., Beqiri, A., Gomez, A.: Echo from noise: Synthetic ultrasound image generation using diffusion models for real image segmentation (2022)
19. Szabo, T.L.: Essentials of Ultrasound Imaging. Elsevier (2021)
20. Tang, F., Ding, J., Wang, L., Xian, M., Ning, C.: Multi-level global context cross consistency model for semi-supervised ultrasound image segmentation with diffusion model (2023)
21. Tirindelli, M., Eilers, C., Simson, W., Paschali, M., Azampour, M.F., Navab, N.: Rethinking ultrasound augmentation: A physics-inspired approach. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24. pp. 690–700. Springer (2021)
22. Velikova, Y., Azampour, M.F., Simson, W., Gonzalez Duque, V., Navab, N.: Lotus: learning to optimize task-based us representations. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 435–445. Springer Nature Switzerland Cham (2023)
23. Velikova, Y., Simson, W., Azampour, M.F., Paprottka, P., Navab, N.: Cactuss: Common anatomical ct-us space for us examinations. International Journal of Computer Assisted Radiology and Surgery pp. 1–9 (2024)
24. Wang, W., Bao, J., Zhou, W., Chen, D., Chen, D., Yuan, L., Li, H.: Semantic image synthesis via diffusion models (2022)
25. Zhang, L., Zhang, J.: Ultrasound image denoising using generative adversarial networks with residual dense connectivity and weighted joint loss. PeerJ Computer Science **8**, e873 (2022)
26. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: CVPR (2018)