

Ultrasound Image Enhancement with the Variance of Diffusion Models

Yuxin Zhang*, Clément Huneau*, Jérôme Idier* and Diana Mateus*

*Nantes Université, École Centrale Nantes, LS2N, CNRS, UMR 6004, F-44000 Nantes, France

Email: yuxin.zhang@ls2n.fr

Abstract—Ultrasound imaging, despite its widespread use in medicine, often suffers from various sources of noise and artifacts that impact the signal-to-noise ratio and overall image quality. Enhancing ultrasound images requires a delicate balance between contrast, resolution, and speckle preservation. This paper introduces a novel approach that integrates adaptive beamforming with denoising diffusion-based variance imaging to address this challenge. By applying Eigenspace-Based Minimum Variance (EBMV) beamforming and employing a denoising diffusion model fine-tuned on ultrasound data, our method computes the variance across multiple diffusion-denoised samples to produce high-quality despeckled images. This approach leverages both the inherent multiplicative noise of ultrasound and the stochastic nature of diffusion models. Experimental results on a publicly available dataset demonstrate the effectiveness of our method in achieving superior image reconstructions from single plane-wave acquisitions. The code is available at: https://github.com/Yuxin-Zhang-Jasmine/IUS2024_Diffusion.

Index Terms—Diffusion models, denoising, despeckling, ultrasound imaging

I. INTRODUCTION

Ultrasound (US) imaging is a widely used diagnostic tool due to its real-time capabilities, affordability, portability, and non-ionizing nature. These advantages make it a preferred option over Magnetic Resonance (MR) and Computed Tomography (CT) in fields such as cardiology and obstetrics. Despite these benefits, US imaging is often challenged by electronic noise, acoustic attenuation, and artifacts like reverberation, shadowing, and speckle.

The standard Delay-and-Sum (DAS) [1] beamforming algorithm converts channel signals into B-mode images, prioritizing speed over image quality. To enhance US image quality, various advanced techniques have been developed, including adaptive beamforming methods [2]–[4], model-based approaches [5]–[7], and (physics-informed) deep learning (DL) techniques [8]–[12]. Although these methods can produce high-quality reconstructions, they often regard US speckle as useful information since speckle can assist in motion tracking by indicating tissue movement. However, in applications such as organ and tumor segmentation or classification, speckle can be problematic [13].

Speckle noise in US imaging is caused by the interference of coherent US waves scattered by small tissue structures, creating a granular pattern. Various despeckling techniques

have been developed, such as Anisotropic Diffusion (AD)-based methods [14]–[16], wavelet-based techniques [17]–[19], the Bilateral filter [20], Nonlocal Means [21], and more recently, DL-based approaches [13], [22]. However, these methods often overlook the presence of electronic noise, which can be significant, especially for ultrafast unfocused emissions such as Plane Waves (PWs). Additionally, they typically operate on processed US images, such as enveloped absolute or log-compressed data, rather than the original signed signals, limiting the preservation of signal characteristics.

Stochastic diffusion models, renowned for image synthesis [23], have recently demonstrated their strong capability in US imaging [12], [24]–[27]. In our previous work [28], we have demonstrated that computing the variance across multiple diffusion-reconstructed samples effectively achieves despeckling without over-smoothing. Specifically, US image reconstruction was framed as an inverse problem with the DAS-beamformed Radio-Frequency (RF) image serving as the measurement. Using the Denoising Diffusion Restoration Models (DDRM) [29] framework, each run estimated the tissue reflectivity, and the variance of multiple independent estimations was used to produce the final despeckled image. However, the high complexity of the model matrix slowed diffusion sampling and made the Singular Value Decomposition (SVD) required by DDRM both memory- and time-intensive.

In the present work, we use a denoising model rather than an inverse problem model and replace the DAS-beamformed image with an adaptively beamformed image. By applying diffusion denoising multiple times and calculating the variance across these denoised samples, we can produce a high-resolution despeckled image more efficiently, addressing the complexity and computational challenges observed in our previous work. Our contributions are: 1) Introducing an adaptive beamforming-based diffusion variance imaging, which achieves faster sampling and competitive image quality. 2) Showing the complementary effects of combining pixel-wise beamforming with denoising diffusion variance imaging, particularly for resolution improvement and background recovery.

II. METHOD

The proposed method aims to remove both electronic and speckle noise from the received signals after a single PW transmission, see Fig. 1 for an overview. Initially, an adaptive pixel-wise beamformer is applied to convert the acquired time-domain RF channel data into the spatial domain. Subsequently,

This work has been supported by the European Regional Development Fund (FEDER), the Pays de la Loire Region on the Connect Talent scheme (MILCOM Project) and Nantes Métropole (Convention 2017-10470).

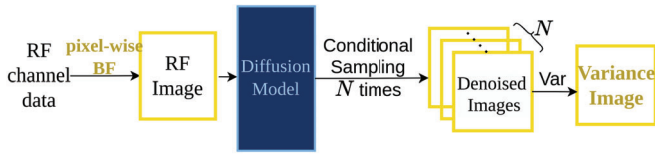


Fig. 1. Method overview

by running the conditional diffusion generative process multiple times and calculating the variance of the generated samples, an enhanced image is produced.

A. Adaptive Beamforming

Convolutions between the excitation pulse, transducer impulse responses (both transmit and receive), and the spatial impulse responses result in a spatially variant point-spread effect in the DAS-beamformed image. This convolutional effect is carefully modeled in [28] to achieve highly accurate reconstruction, albeit at the cost of increased computational complexity. In contrast, the present work mitigates this degradation by using the adaptively beamformed RF image as the measurement, simplifying the diffusion sampling process from solving an inverse problem to straightforward denoising. This approach enables faster diffusion sampling while maintaining image quality. Among the various pixel-wise adaptive beamforming approaches [2]–[4], this work employs the Eigenspace-Based Minimum Variance (EBMV) [3] technique.

The EBMV beamformer builds upon the Minimum Variance (MV) beamformer [2]. Let the delayed RF channel data for one pixel be denoted as $\mathbf{y} \in \mathbb{R}^{Ne \times Np}$, where Ne and Np represent the number of channels and the number of time samples, respectively. The channel weights for each pixel in MV beamforming are calculated as follows:

$$\mathbf{w}_{MV} = \frac{\mathbf{R}^{-1}\mathbf{1}}{\mathbf{1}^H\mathbf{R}^{-1}\mathbf{1}}, \quad (1)$$

where $\mathbf{w}_{MV} \in \mathbb{R}^{Ne \times 1}$ and \mathbf{R} denotes the expected covariance matrix of \mathbf{y} . To ensure the robustness of the inverse of \mathbf{R} , subarray averaging and diagonal loading techniques are employed.

The EBMV beamformer decomposes \mathbf{R} through eigenvalue decomposition:

$$\mathbf{R} = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^H, \quad (2)$$

where $\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{Ne}]$ comprises the eigenvectors, and $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues. The beamforming weights in EBMV are computed by projecting \mathbf{w}_{MV} onto the signal subspace, which consists of the eigenvectors associated with the largest eigenvalues. These weights are given by:

$$\mathbf{w}_{EBMV} = \mathbf{E}_s \mathbf{E}_s^H \mathbf{w}_{MV}, \quad (3)$$

where $\mathbf{E}_s = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{Num}]$ with $Num < Ne$. The EBMV beamformed pixel value is then formulated as:

$$\mathbf{x}_{EBMV} = \mathbf{w}_{EBMV}^H \mathbf{y}. \quad (4)$$

Although speckle noise persists and image background completeness is often disrupted, EBMV effectively enhances spatial resolution. For simplicity, the notation \mathbf{x} will be used throughout the remainder of this paper to represent the entire RF image reconstructed using EBMV.

B. Denoising Diffusion Variance Imaging

Diffusion models estimate clean images from noisy ones using both the measurement and the prior knowledge learned from training images [29]. The measurement guides the sampling path, while the diffusion prior infers the information affected by the additive noise. Just as the diffusion sampling process can be modeled by a Stochastic Differential Equation (SDE) [30], it involves numerous drift steps, so each sampling produces a different result. Although the generated estimations may appear quite similar under the same conditions, there will always be a certain degree of variation.

Typically, the mean or the median of multiple samples, rather than a single sample, is used as the final result [31] to enhance reliability, with the variance calculated to indicate the uncertainty of the sampling procedure. Interestingly, the variance map consistently shows a non-uniform pattern. In tasks where the ground truth lacks multiplicative noise, such as natural image restoration, the variance map tends to highlight the image boundaries [32]. In contrast, in the presence of multiplicative noise, the variance map closely resembles an image with reduced noise [28].

In US imaging, speckle, which is intrinsic to US images, can be modeled as multiplicative noise [10], [33]:

$$\mathbf{o} = \mathbf{m} \odot \mathbf{p}, \quad (5)$$

where \mathbf{o} denotes the tissue reflectivity, \mathbf{m} represents the standard Gaussian multiplicative noise, and \mathbf{p} refers to the echogenicity map. Assuming that the EBMV-beamformed RF image is the sum of tissue reflectivity and additive noise, the denoising model for diffusion sampling can be formulated as:

$$\mathbf{x} = \mathbf{m} \odot \mathbf{p} + \mathbf{n}, \quad (6)$$

where \mathbf{x} represents the adaptively beamformed measurement, and $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \gamma^2 \mathbf{I})$ denotes additive noise with a standard deviation of γ .

Given the measurement \mathbf{x} and the noise level γ , denoising diffusion sampling can generate multiple estimations of \mathbf{o} (with the c -th sample denoted as $\hat{\mathbf{o}}_c$). We proposed an empirical model in [28] that characterizes the output of a diffusion-based reconstruction as a function of the echogenicity \mathbf{p} :

$$\hat{\mathbf{o}}_c = \mathbf{m} \odot \mathbf{p} + \mathbf{p}^{0.5} \odot \mathbf{G}_c. \quad (7)$$

In this model, \mathbf{G}_c follows a standard normal distribution to account for stochasticity in the generative process. Relying on such a model, it can be easily verified that the variance of several diffusion-based US reconstructions provides information on the echogenicity.

III. EXPERIMENTAL VALIDATION ON REAL DATA

We focus on reconstructing US images from a single PW acquisition, using the results of coherent compounding from 75 DAS-beamformed PW images as the reference. The proposed method was evaluated on the Plane Wave Imaging Challenge in Medical Ultrasound (PICMUS) [34] dataset. To demonstrate its performance, both qualitative and quantitative analyses were conducted using the *in-vitro* Experimental Contrast (EC) dataset. Additionally, we present qualitative results on the *in-vivo* Carotid Cross-sectional (CC) dataset.

EBMV beamforming was performed using the Ultra-Sound ToolBox (USTB) [35], with the following settings: *active_element_criterium* (determining whether a transducer element is used) = 0.16, *L_element* (subarray size) = 80, *K_in_lambda* (temporal averaging factor) = 0, *refCoef* (diagonal loading constant) = 0.01, and *gamma* (signal space criterion) = 0.05. All other parameters were kept at their default values. The EBMV beamformed RF signal was normalized to $[-1, 1]$ to match the value range of the training set.

The employed diffusion model, initially pre-trained on ImageNet [36], was fine-tuned on an in-house US dataset. The fine-tuning dataset consists of 3551 (2,339 *in vitro* + 1212 *in vivo*) high-quality RF images with each reconstructed from 101 PWs using DAS. These images were collected using a TPAC Pioneer machine from the CIRS 040GSE phantom and a volunteer. Note that different from conventional autoencoder-based models for denoising [37], diffusion-based models do not need paired data, only high-quality images are expected.

Each variance image was produced from 10 samples, and each sampling process (DUS) involved 50 iterations. The qualitative evaluation on the *in-vitro* PICMUS-EC dataset is shown in Fig. 2. It consists of seven sets of images, each with a full view and a zoomed-in view to detail the scatterer region. The top-left is the reference reconstructed from 75 PWs using DAS, while the other images were reconstructed from a single PW. In the top row, the second image used DAS beamforming, and the following two images applied denoising diffusion to this DAS result: one shows the median of multiple samples, and the other the variance. In the bottom row, the first image was reconstructed with EBMV, and it served as the measurement for denoising diffusion. The second image displays the median result, whereas the fourth image, which utilizes the proposed EBMV+DUSvar method, illustrates the variance.

A comparison of the anechoic regions and the zoomed-in scatterer views reveals that EBMV+DUSvar significantly outperforms DAS1+DUSvar in contrast and spatial resolution. Additionally, while the completeness of the background is compromised in images reconstructed with EBMV or EBMV+DUSmedian (as indicated by the red arrows in Fig. 2), the background is effectively restored in the image of EBMV+DUSvar.

The improvement in contrast and spatial resolution entirely relies on EBMV. The background recovery phenomenon is attributed to the generative stochasticity resulting from data

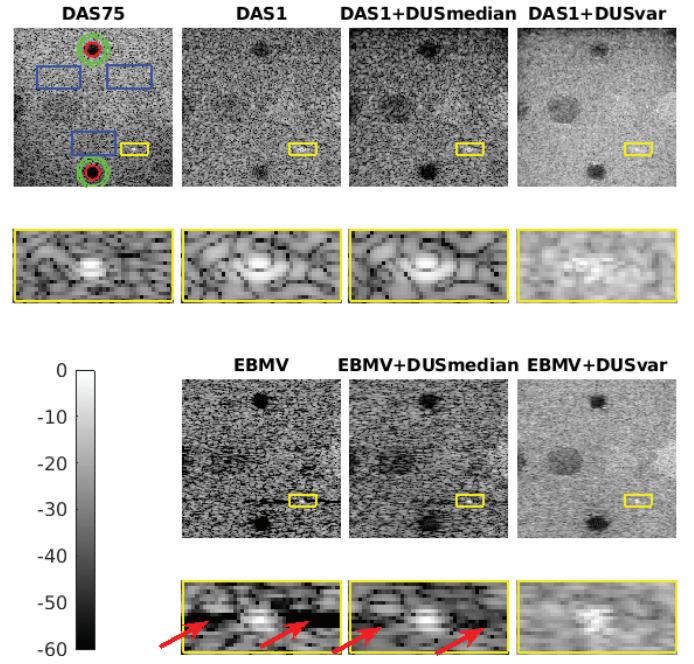


Fig. 2. Comparison of reconstructed images on the PICMUS-EC dataset. The colored boundaries outline the regions where the evaluation metrics are calculated.

mismatch. Specifically, the stochasticity in regions rendered anechoic by EBMV arises not from multiplicative noise but from discrepancies between the EBMV image and the training database. In the EBMV-beamformed image, the values in regions on either side of the scatterer are nearly zero, so denoising diffusion reconstruction for these regions relies entirely on the learned prior knowledge, as there are no measurements to guide the sampling process. As illustrated in Fig. 3, since the diffusion model was fine-tuned on high-quality US images with complete backgrounds, it tends to generate images with similarly complete backgrounds.

Quantitative results evaluated in the colored regions on the DAS75 image are shown in Fig. 4. Spatial resolution for the bright scatterer was assessed using the Full Width at Half Maximum (FWHM) in both axial and lateral directions. Contrast for the anechoic regions was evaluated using the generalized Contrast-to-Noise Ratio (gCNR) [38]. Background cleanliness was evaluated using the Signal-to-Noise Ratio (SNR).

As illustrated in Fig. 4, the quantitative scores are consistent with the qualitative observations in Fig. 2. This alignment confirms the significant improvement in spatial resolution and contrast of using EBMV compared to DAS1 for denoising diffusion variance imaging.

The qualitative evaluation on the *in-vivo* PICMUS-CC dataset, shown in Fig. 5, further confirms the feasibility of the proposed approach. Specifically, EBMV+DUSvar effectively smooths the image background compared to EBMV, while maintaining good spatial resolution and avoiding detail loss relative to DAS1+DUSvar.

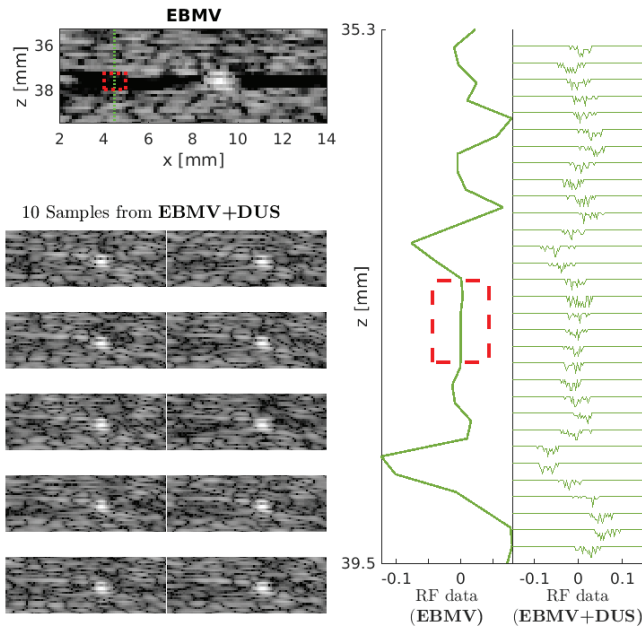


Fig. 3. Statistical behavior of EBMV+DUS compared to EBMV on the PICMUS-EC dataset. The values at the position indicated by the dotted green line in the top-left image are compared in the right green plots. It shows that the variance of EBMV+DUS in the zero region enclosed by the dashed red box is non-zero. All grayscale images are in decibels with a dynamic range [-60,0].

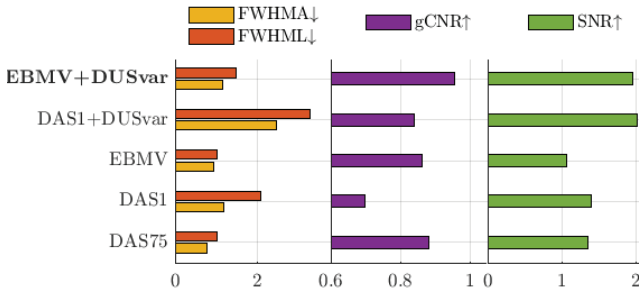


Fig. 4. Quantitative comparison of the PICMUS-EC dataset. A and L denote axial and lateral directions, respectively.

IV. CONCLUSION

This work proposes a model-based DL method combining EBMV beamforming with denoising diffusion variance imaging to address both electronic and speckle noise, enhancing the quality of single-PW US images. Due to the high computational burden of solving inverse problems [28], this work adopts a simpler denoising model to achieve faster sampling. To overcome the limited spatial resolution of DAS, EBMV beamforming is used instead. Note that while EBMV is utilized here, other adaptive beamformers could also be considered. Experimental evaluations on the PICMUS-EC and CC datasets highlight the capacity of our method to balance spatial resolution, contrast, and background SNR. Furthermore, our model effectively recovers regions that are incorrectly rendered anechoic by EBMV, confirming its effectiveness.

REFERENCES

- [1] V. Perrot, M. Polichetti, F. Varray, and D. Garcia, "So you think you can DAS?" *Ultrasonics*, vol. 111, p. 106309, 2021.
- [2] J. F. Synnevag, A. Austeng, and S. Holm, "Adaptive beamforming applied to medical ultrasound imaging," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 54, no. 8, pp. 1606–1613, 2007.
- [3] B. M. Asl and A. Mahloojifar, "Eigenspace-based minimum variance beamforming applied to medical ultrasound imaging," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 57, no. 11, pp. 2381–2390, 2010.
- [4] T. Chernyakova, D. Cohen, M. Shoham, and Y. C. Eldar, "iMAP beamforming for high-quality high frame rate imaging," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 66, 2019.
- [5] E. Ozkan, V. Vishnevsky, and O. Goksel, "Inverse problem of ultrasound beamforming with sparsity constraints and regularization," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 65, pp. 356–365, 2018.
- [6] S. Goudarzi, A. Basarab, and H. Rivaz, "Inverse problem of ultrasound beamforming with denoising-based regularized solutions," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 69, 2022.
- [7] N. Laroche, S. Bourguignon, J. Idier, E. Carcreff, and A. Duclos, "Fast non-stationary deconvolution of ultrasonic beamformed images for nondestructive testing," *IEEE Trans. Comput. Imaging*, vol. 7, 2021.
- [8] B. Luijten *et al.*, "Adaptive ultrasound beamforming using deep learning," *IEEE Trans. Med. Imaging*, vol. 39, 2020.
- [9] J. Zhang, Q. He, Y. Xiao, H. Zheng, C. Wang, and J. Luo, "Ultrasound image reconstruction from plane wave radio-frequency data by self-supervised deep neural network," *Med. Image Anal.*, vol. 70, 2021.
- [10] D. Perdios, M. Vonlanthen, F. Martinez, M. Arditi, and J.-P. Thiran, "NN-based image reconstruction method for ultrafast ultrasound imaging," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 69, pp. 1154–1168, 2022.
- [11] R. J. G. van Sloun, R. Cohen, and Y. C. Eldar, "Deep learning in ultrasound imaging," *Proc. IEEE*, vol. 108, pp. 11–29, 2020.
- [12] Y. Zhang, C. Huneau, J. Idier, and D. Mateus, "Ultrasound image reconstruction with denoising diffusion restoration models," in *DGM4MICCAI*, 2023, pp. 193–203.
- [13] D. Mishra, S. Tyagi, S. Chaudhury, M. Sarkar, and A. S. Soin, "Despeckling CNN with ensembles of classical outputs," in *IEEE ICPR*, 2018, pp. 3802–3807.
- [14] S. Aja-Fernandez and C. Alberola-Lopez, "On the estimation of the coefficient of variation for anisotropic diffusion speckle filtering," *IEEE Trans. Image Process.*, vol. 15, no. 9, pp. 2694–2701, 2006.
- [15] K. Krissian, C.-F. Westin, R. Kikinis, and K. G. Vosburgh, "Oriented speckle reducing anisotropic diffusion," *IEEE Trans. Image Process.*, vol. 16, no. 5, pp. 1412–1424, 2007.
- [16] G. Ramos-Llordén, G. Vegas-Sánchez-Ferrero, M. Martín-Fernández, C. Alberola-López, and S. Aja-Fernández, "Anisotropic diffusion filter with memory based on speckle statistics for ultrasound images," *IEEE Trans. Image Process.*, vol. 24, pp. 345–358, 2014.
- [17] Y. Yue, M. Croitoru, A. Bidani, J. Zwischenberger, and J. Clark, "Nonlinear multiscale wavelet diffusion for speckle suppression and edge enhancement in ultrasound images," *IEEE Trans. Med. Imaging*, vol. 25, no. 3, pp. 297–311, 2006.
- [18] M. I. H. Bhuiyan, M. O. Ahmad, and M. N. S. Swamy, "New spatially adaptive wavelet-based method for the despeckling of medical ultrasound images," in *2007 IEEE ISCAS*, 2007, pp. 2347–2350.
- [19] S. Esakkirajan, C. T. Vimalraj, R. Muhammed, and G. Subramanian, "Adaptive wavelet packet-based de-speckling of ultrasound images with bilateral filter," *Ultrasound Med. Biol.*, vol. 39, no. 12, pp. 2463–2476, 2013.
- [20] S. Balocco, C. Gatta, O. Pujol, J. Mauri, and P. Radeva, "SRBF: Speckle reducing bilateral filtering," *Ultrasound Med. Biol.*, vol. 36, no. 8, pp. 1353–1363, 2010.
- [21] P. Coupe, P. Hellier, C. Kervrann, and C. Barillot, "Nonlocal means-based speckle filtering for ultrasound images," *IEEE Trans. Image Process.*, vol. 18, no. 10, pp. 2221–2229, 2009.
- [22] C.-C. Shen and J.-E. Yang, "Estimation of ultrasound echogenicity map from B-mode images using convolutional neural network," *Sensors*, vol. 20, no. 17, p. 4931, 2020.
- [23] P. Dhariwal and A. Nichol, "Diffusion models beat GANs on image synthesis," *NeurIPS*, vol. 34, pp. 8780–8794, 2021.

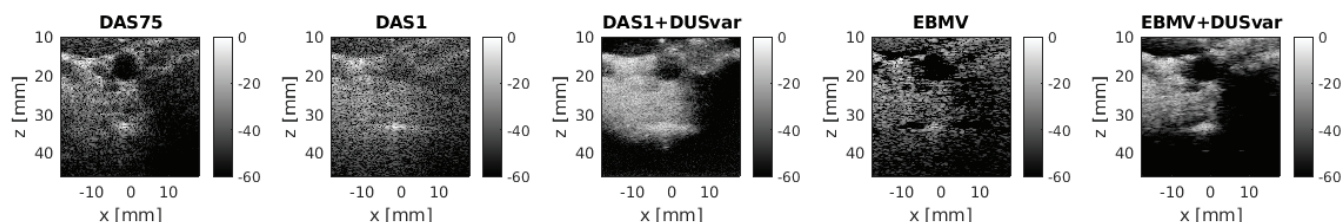


Fig. 5. Comparison of reconstructed images on the PICMUS-CC dataset.

- [24] T. S. Stevens, F. C. Meral, J. Yu, I. Z. Apostolakis, J.-L. Robert, and R. J. Van Sloun, "Dehazing ultrasound using diffusion models," *IEEE Trans. Med. Imaging*, 2024.
- [25] H. Asgariandehkordi, S. Goudarzi, A. Basarab, and H. Rivaz, "Deep ultrasound denoising using diffusion probabilistic models," in *IEEE IUS*, 2023.
- [26] H. Asgariandehkordi, S. Goudarzi, M. Sharifzadeh, A. Basarab, and H. Rivaz, "Denoising plane wave ultrasound images using diffusion probabilistic models," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, 2024.
- [27] H. Lan, Z. Li, Q. He, and J. Luo, "Fast sampling generative model for ultrasound image reconstruction," *arXiv preprint*, 2023.
- [28] Y. Zhang, C. Huneau, J. Idier, and D. Mateus, "Ultrasound imaging based on the variance of a diffusion restoration model," *arXiv preprint*, 2024.
- [29] B. Kawar, M. Elad, S. Ermon, and J. Song, "Denoising diffusion restoration models," *NeurIPS*, vol. 35, pp. 23 593–23 606, 2022.
- [30] Y. Song, L. Shen, L. Xing, and S. Ermon, "Solving inverse problems in medical imaging with score-based generative models," in *ICLR*, 2022.
- [31] H. Chung and J. C. Ye, "Score-based diffusion models for accelerated MRI," *Med. Image Anal.*, vol. 80, p. 102479, 2022.
- [32] B. Kawar, G. Vaksman, and M. Elad, "SNIPS: Solving noisy inverse problems stochastically," *NeurIPS*, vol. 34, pp. 21 757–21 769, 2021.
- [33] J. Ng, R. Prager, N. Kingsbury, G. Treece, and A. Gee, "Wavelet restoration of medical pulse-echo ultrasound images in an EM framework," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 54, 2007.
- [34] H. Liebgott, A. Rodriguez-Molares, F. Cervenansky, J. Jensen, and O. Bernard, "Plane-wave imaging challenge in medical ultrasound," in *IEEE IUS*, 2016.
- [35] A. Rodriguez-Molares *et al.*, "The ultrasound toolbox," in *IEEE IUS*, 2017.
- [36] O. Russakovsky *et al.*, "Imagenet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, pp. 211–252, 2015.
- [37] L. Gondara, "Medical image denoising using convolutional denoising autoencoders," in *IEEE ICDMW*, 2016, pp. 241–246.
- [38] A. Rodriguez-Molares *et al.*, "The generalized contrast-to-noise ratio: A formal definition for lesion detectability," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 67, pp. 745–759, 2020.