Semantic Diffusion Posterior Sampling for Cardiac Ultrasound Dehazing

Tristan S.W. Stevens, Oisín Nolan, and Ruud J.G. van Sloun

Eindhoven University of Technology, the Netherlands
{t.s.w.stevens,o.i.nolan,r.j.g.v.sloun}@tue.nl

Abstract. Echocardiography plays a central role in cardiac imaging, offering dynamic views of the heart that are essential for diagnosis and monitoring. However, image quality can be significantly degraded by haze arising from multipath reverberations, particularly in difficult-to-image patients. In this work, we propose a semantic-guided, diffusion-based dehazing algorithm developed for the MICCAI Dehazing Echocardiography Challenge (DehazingEcho2025). Our method integrates a pixel-wise noise model, derived from semantic segmentation of hazy inputs into a diffusion posterior sampling framework guided by a generative prior trained on clean ultrasound data. Quantitative evaluation on the challenge dataset demonstrates strong performance across contrast and fidelity metrics. Code for the submitted algorithm is available on GitHub.[1].

Keywords: Cardiac Ultrasound Imaging · Dehazing · Diffusion Models.

1 Introduction

Ultrasound is a popular modality for cardiac imaging due to its high temporal resolution, cost effectiveness, and real-time imaging capabilities, enabling the detection of a variety of cardiac abnormalities [2]. An ongoing challenge in echocardiography is that of clutter or haze resulting from multipath reverberations [12,14], which can prevent accurate measurement from B-Mode images. This has motivated the development *dehazing algorithms*, which aim to recover clean images $\mathbf{x}$ from hazy input images $\mathbf{y}$.

Recently, a number of dehazing algorithms leveraging deep generative models (DGMs) have been proposed, using prior knowledge of the clean image distribution to infer sets of clean images corresponding to observed hazy images. One such approach involves using Generative Adversarial Networks (GANs) to perform domain adaptation, wherein the style of one dataset is transferred to samples from another [17,9] while retaining structural contents. Other approaches opt for diffusion models (DMs) [8,13], which are known to represent the state-of-the-art in image synthesis [6]. One such method, introduced by Stevens *et al.*, involves using DMs to learn models of the distributions of both clean images and haze, which are then used to separate the clean tissue and haze components of

[1] https://github.com/tristan-deep/semantic-diffusion-echo-dehazing

the input hazy image [14]. However, in this work, the signal model employed is defined on pre-envelope-detected signals, which are in some cases not available. This motivates the development of diffusion-based dehazing algorithms that operate in the image domain. In this paper, we propose such an algorithm, *Semantic Diffusion Posterior Sampling*, which first computes a semantic segmentation map estimating the haze content of each pixel, and then uses the Diffusion Posterior Sampling (DPS) algorithm to generate posterior samples from the clean image distribution given the hazy measurement. The estimated haze map serves to control the strength of the conditional guidance during the image generation process, according more with the measurements in clean regions, and less in hazy regions.

2 Challenge

This work was developed in the context of the MICCAI Dehazing Echocardiography Challenge (DehazingEcho2025)[2], which aims to enhance the quality of transthoracic echocardiographic images acquired from difficult-to-image patients. The dataset provided for this challenge comprises two subsets: (1) a clean set of 4,376 frames obtained from 75 easy-to-image subjects, and (2) a noisy set of 2,324 frames acquired from 40 difficult-to-image subjects. Each image in the dataset is part of a 60-frame four-chamber view cine-loop.

For quantitative benchmarking, a hidden test set of 536 noisy frames is used for online evaluation on the Grand Challenge platform [4]. The evaluation protocol incorporates multiple complementary metrics designed to capture different aspects of image quality and utility. Specifically:

- **Fréchet Inception Distance (FID)** assesses perceptual similarity between denoised and clean images.
- **CNR** and **gCNR** quantify contrast between myocardial tissue (septum) and noise-prone regions (left ventricular).
- **Kolmogorov–Smirnov (KS)** test measures distributional similarity to assess structure preservation of the septum and noise removal in the ventrical.
- **Dice coefficient** and **Average Surface Distance (ASD)** evaluate the compatibility of denoised images with downstream segmentation tasks, using a pre-trained universal ultrasound foundation model (USFM) [10] targeting the left ventricle.

The *final score* is derived from a weighted aggregation of the above metrics, balancing denoising performance, structural preservation, and downstream task impact with respective weights of 5:3:2.

3 Method

The algorithm consists of two primary steps: first, generating a semantic segmentation mask based on the input hazy image, and second, using that mask to

[2] https://dehazingecho2025.grand-challenge.org/

guide a diffusion model towards generating a dehazed image. We frame the task of dehazing as a Bayesian inverse problem with the following forward model:

$$\mathbf{y} = \mathbf{x} + \mathbf{h}, \qquad \mathbf{h} \sim \mathcal{N}(0, \boldsymbol{\Sigma}), \quad \mathbf{x} \sim p_{\text{generative model}}(\mathbf{x}) \tag{1}$$

where $\mathbf{y}, \mathbf{x}, \mathbf{h} \in \mathbb{R}^n$ denote the hazy measurement, clean image, and residual haze respectively. The haze is modeled as additive, zero-mean Gaussian noise with a spatially varying diagonal covariance matrix:

$$\boldsymbol{\Sigma}^{-1} = \text{diag}(\sigma_1^{-2}, \sigma_2^{-2}, \ldots, \sigma_n^{-2}) = \text{diag}(\mathbf{p}), \tag{2}$$

where each variance σ_i^2 is determined by the pixel-wise semantic segmentation map $\mathbf{p}$ as outlined in Section 3.2. To solve (1), we employ a diffusion prior, which we detail in the following section.

3.1 Deep generative prior for clean cardiac ultrasound

We start with leveraging a deep generative model (DGM) to model the distribution $p(\mathbf{x})$ of clean echocardiography images $\mathbf{x}$. Specifically, we train a diffusion model (DM), and subsequently perform posterior sampling $p(\mathbf{x}|\mathbf{y})$ conditioned on the hazy inputs $\mathbf{y}$.

Training: Diffusion models learn to approximate the gradient of the log-density (i.e., the score function) of the data distribution by denoising progressively noised inputs. The training objective is based on denoising score matching (DSM), which seeks to minimize the discrepancy between predicted and true noise across different noise levels:

$$\mathcal{L}_{\text{DSM}}(\theta) = \mathbb{E}_{\mathbf{x}_0 \sim p(\mathbf{x}_0), \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \tau \sim \mathcal{U}(0, \mathcal{T})} \left[\left\| \epsilon_\theta(\mathbf{x}_\tau, \tau) - \epsilon \right\|^2 \right], \tag{3}$$

where $\mathbf{x}_\tau = \alpha_\tau \mathbf{x}_0 + \sigma_\tau \epsilon$ denotes a corrupted version of a sample from our clean dataset $\mathbf{x}_0$ at a continuous noise level τ, with a pre-defined noise schedule, parameterized by α_τ and σ_τ. The model $\epsilon_\theta(\mathbf{x}_\tau, \tau)$ is trained to predict the noise component $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ from the corrupted input $\mathbf{x}_\tau$, effectively learning the score $\epsilon_\theta(\mathbf{x}_\tau, \tau) \approx -\sigma_\tau \nabla_{\mathbf{x}_\tau} \log p(\mathbf{x}_\tau)$. To further improve the perceptual quality of generated samples (note the FID objective of the challenge in Section 2), we additionally incorporate a Kernel Inception Distance (KID) loss [3], computed between generated and real clean images. Unlike FID, KID is unbiased and better suited for comparing small datasets. Formally, the KID loss is defined as the squared maximum mean discrepancy between the feature embeddings of a pretrained InceptionV3 network of generated and real samples. This encourages the model to produce samples that align more closely with the distribution of clean dataset. The total training loss is then given by:

$$\mathcal{L}(\theta) = \mathcal{L}_{\text{DSM}}(\theta) + \lambda_{\text{KID}} \cdot \mathcal{L}_{\text{KID}}(\theta), \tag{4}$$

where λ_{KID} is a weighting factor controlling the influence of the perceptual loss. The diffusion model is pre-trained on the publicly available EchoNet-LVH

dataset [7], and finetuned on the clean partition of the DehazingEcho2025 challenge dataset. For more details see Table 1.

3.2 Semantic Segmentation

The first step of the algorithm involves generating a segmentation mask from the hazy image which estimates the haze content of each pixel. This segmentation mask provides a *noise level* for each pixel, defining a forward model for Diffusion Posterior Sampling (DPS) [5] wherein high-signal pixels provide strong guidance, closely matching the measured pixels, and high-noise pixels provide weak guidance, falling back towards the prior distribution on dehazed images.

In order to construct these semantic segmentation masks, a combination of learned and classical segmentation methods was used.

- **Ventricle and Septum Segmentation**: In order to identify the ventricles and septum, a DeepLabV3+ [11] model was trained on manually annotated regions of interest corresponding to the septum and ventricle, provided in the challenge dataset. We denote the resulting masks as $v(\mathbf{y})$ and $s(\mathbf{y})$, identifying the ventricle and septum from the hazy image $\mathbf{y}$. The DeepLabV3+ model outputs a map of logits, which are thresholded by a parameter θ to create a binary mask, which is then blurred using a Gaussian kernel with standard deviation σ_{blur}.
- **Tissue Segmentation**: In order segment regions of tissue, the hazy image was skeletonized using `skimage.morphology.skeletonize`[3] [16], a function which implements the thinning algorithm proposed by Zhang *et al.* [18], mapping the hazy input image to thin skeleton tracing the tissue. The skeleton mask is denoted $t(\mathbf{y})$.
- **Fixed pixels**: Two bands of pixels, at the top and bottom of the image, are kept fixed to ensure the preservation of details from the DICOM overlay.
- **Background pixels**: Any pixel not present in $v(\mathbf{y})$, $s(\mathbf{y})$, or $t(\mathbf{y})$ is considered a *background pixel*, $b(\mathbf{y})$.
- **Dark pixels**: Any pixel for which $\mathbf{y} < 1e^{-6}$ is segmented as a *dark pixel*, $d(\mathbf{y})$.

The final mask is then created by taking a weighted sum of these individual masks, producing the precision vector $\mathbf{p}$ populating the diagonal of the precision matrix $\boldsymbol{\Sigma}^{-1}$ used in the DPS forward model:

$$\mathbf{p} = \omega b(\mathbf{y}) + \omega_v v(\mathbf{y}) + \omega_s (s(\mathbf{y}) + t(\mathbf{y}) + d(\mathbf{y})) \tag{5}$$

Finally, we handle the edge case where the skeleton $t(\mathbf{y})$ crosses the ventricle $v(\mathbf{y})$, splitting it in two, by setting $\omega_v = 0$. An overview of the individual segmentation steps, and the resulting guidance weighting map is shown in Fig. 1.

[3] https://scikit-image.org/docs/0.25.x/api/skimage.morphology.html#skimage.morphology.skeletonize

Notably, imperfections introduced at one stage, such as over-segmentation or missed structures, can often be mitigated by complementary information from other segmentation steps, resulting in a more robust pixel-wise noise estimation.

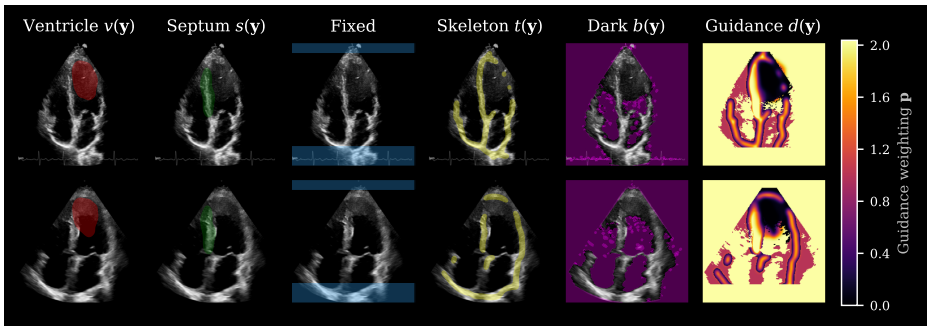

Fig. 1. Visualization of the individual components for the semantic segmentation and the resulting constructed guidance weighting map $\mathbf{p}$ for two different patients.

3.3 Semantic Diffusion Posterior Sampling

Sampling from the posterior distribution $\mathbf{x}_0 \sim p(\mathbf{x}_0|\mathbf{y})$ is achieved by initiating $\mathbf{x}_\mathcal{T}$ as a Gaussian random vector, and then applying an iterative denoising process that progressively refines the sample across decreasing noise levels as follows:

1. Estimate the clean image from the current corrupted input:

$$\hat{\mathbf{x}}_0 \leftarrow \frac{1}{\alpha_\tau}\left(\mathbf{x}_\tau - \sigma_\tau \epsilon_\theta(\mathbf{x}_\tau, \tau)\right). \tag{6}$$

2. Guide the sample towards our hazy measurement with the forward model defined in (1), along with a penalty on a smoothed L1 norm[4], parameterized by β, on the pixels in the ventricle.

$$\hat{\mathbf{x}}_0 \leftarrow \hat{\mathbf{x}}_0 - \frac{1}{2}\nabla_{\mathbf{x}_\tau}(\mathbf{y} - \hat{\mathbf{x}}_0)^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \hat{\mathbf{x}}_0) - \eta\nabla_{\mathbf{x}_\tau}|v(\mathbf{y}) \odot \hat{\mathbf{x}}_0|_\beta. \tag{7}$$

3. Predict the next corrupted sample at a lower noise level $\tau' < \tau$:

$$\mathbf{x}_{\tau'} \leftarrow \alpha_{\tau'}\hat{\mathbf{x}}_0 + \sigma_{\tau'}\epsilon_\theta(\mathbf{x}_\tau, \tau). \tag{8}$$

This deterministic denoising process is repeated until $\tau' = 0$, yielding a sample from the clean data distribution, conditioned on hazy measurement $\mathbf{y}$.

[4] https://docs.pytorch.org/docs/stable/generated/torch.nn.SmoothL1Loss.html

Table 1. Overview of model components, architectures, training/inference settings, and datasets used.

Model	Architecture	Inference	Training	Dataset
Diffusion	UNet	$N = 480$ $\eta = 0.007$ $\omega = 1$, $\omega_s = 2$ $\omega_v = 0.3$ $\beta = 1.6$	$\lambda_{\text{KID}} = 0.05$ $N_{\text{KID}} = 15$ ema $= 0.999$ $\text{lr}_{\text{pre}} = 10^{-4}$ $\text{lr}_{\text{fine}} = 10^{-5}$	EchoNet-LVH (pretrain) DehazingEcho2025 (clean subset, finetune)
Segmentation	DeepLabV3+	$\theta = 0.176$ $\sigma_{\text{blur}} = 4.2$	$\text{lr} = 5e-4$	DehazingEcho2025 (noisy subset w/ ROIs)

3.4 Algorithm details

A summary of the most important parameters used in each component of the dehazing algorithm is listed in Table 1. Hyperparameters were optimized with `Optuna` [1] using a subset of 237 images from the available noisy set that have corresponding masks. The optimization objective was set to the final score as specified in Section 2. A plot of the hyperparameter sweep is shown in Fig. 2. The final scores are slightly underestimated due to the FID's sensitivity to the smaller sample size, approximately 2× lower than the online challenge test set.

The algorithm is implemented using `Keras 3` with the `JAX` backend for accelerated inference through JIT compilation. Furthermore, our implementation heavily relies on `zea`, a toolbox for cognitive ultrasound imaging [15][5]. Code with the complete algorithm implementation is available on GitHub [1].

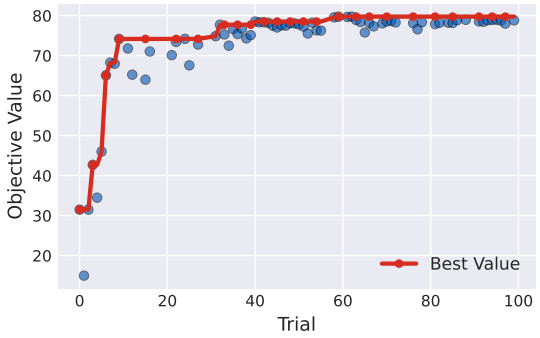

Fig. 2. Hyperparameter optimization of 100 trials for the inference parameters listed in Table 1, with the challenge's final score as objective.

[5] https://zea.readthedocs.io/

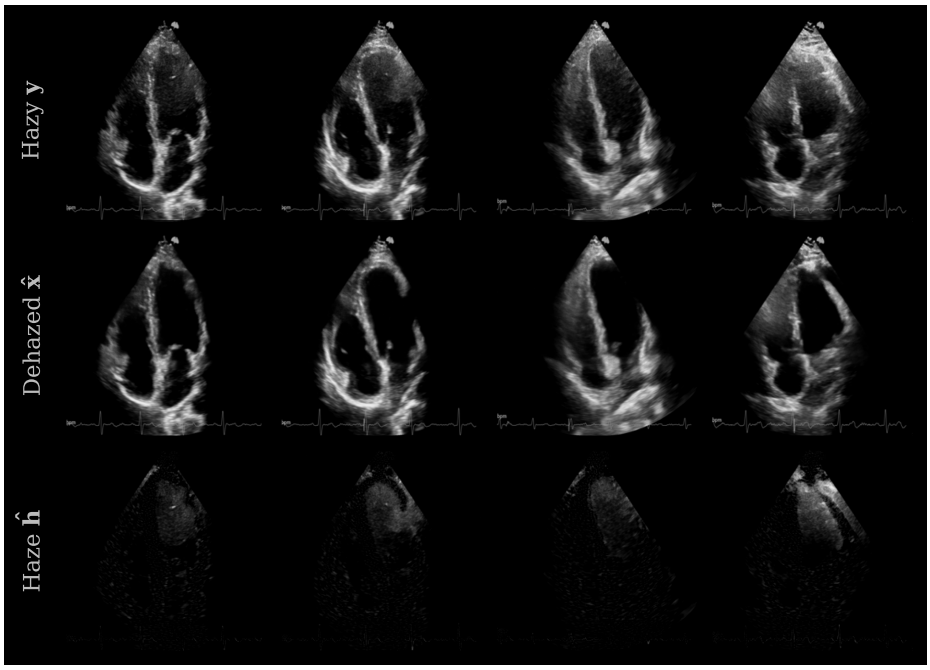

Fig. 3. Hazy echocardiographic images $\mathbf{y}$ and their decomposition into dehaze doutputs $\hat{\mathbf{x}}$ and haze estimates $\hat{\mathbf{h}}$ as a result of the submitted algorithm.

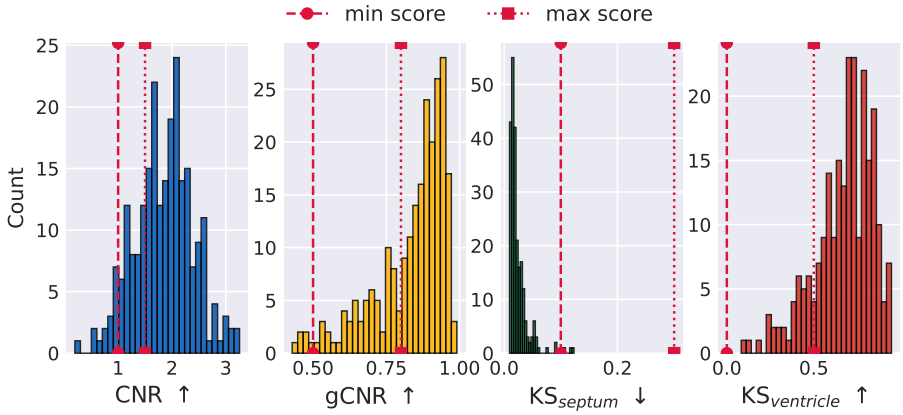

Fig. 4. Contrast metrics and KS statistics for the dehazed results from the submitted algorithm. Minimum and maximum obtainable scores as set by the challenge organizers are marked in red.

4 Results and discussion

A subset of samples from the DehazingEcho2025 noisy set, and corresponding dehazed outputs and haze estimates is shown in Fig. 3. A qualitative analysis of the results is shown in Fig. 4. Both contrast metrics and KS statistics values are plotted for the 237 images in the noisy set that have corresponding ROI masks. The reported FID obtained under these settings is 61.3.

One interesting observation is that the hyperparameters yielding the highest challenge score did not necessarily produce the best visual quality, suggesting a misalignment between the evaluation metrics and perceptual fidelity of the dehazed images. For instance, the metrics appear to incentivize an almost binary contrast between the ventricle and septum, which, while improving numerical scores, leads to a loss of subtle structural nuances. This overly sharp separation diminishes the natural appearance of the tissue, which clinicians may find misleading or diagnostically unhelpful. Rather than merely removing haze, the desired outcome is to reveal underlying tissue structures that were previously occluded or obscured. While FID partially captures this goal, future work should explore more perceptually and clinically aligned evaluation metrics that better reflect meaningful dehazing of echocardiographic images. In the current approach, the ventricle guidance parameter ω_v and haze prior weighting parameter η can be increased and decreased, respectively, to reduce the amount of dehazing in the left ventricle area.

A final observation is that haze reduction is primarily focused on the left ventricle, driven by both the challenge metrics, centered on septum–ventricle contrast, and the available labels, which include ROIs only for these two regions. As a result, areas like the right ventricle receive less attention. Extending the segmentation model to include the right ventricle would allow for more comprehensive dehazing.

5 Conclusion

We present a diffusion-based dehazing algorithm guided by semantic segmentation maps that adaptively modulate the influence of hazy measurements during posterior sampling. Developed in the context of the MICCAI DehazingEcho2025 challenge, our method achieves strong performance across the contrast and fidelity metrics. Preliminary results suggest a potential gap between the existing challenge metrics and perceptual quality, particularly in preserving subtle anatomical details. Nonetheless, the challenge has provided a valuable platform to advance and benchmark dehazing approaches. In particular, our generative modeling approach is able to effectively reduce haze by leveraging semantic guidance to preserve anatomical details. Future work can focus on improved semantic segmentation and development of evaluation metrics that better capture perceptual quality and clinical relevance.

References

1. Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M.: Optuna: A next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 2623–2631 (2019)
2. Barry, T., Farina, J.M., Chao, C.J., Ayoub, C., Jeong, J., Patel, B.N., Banerjee, I., Arsanjani, R.: The role of artificial intelligence in echocardiography. Journal of imaging **9**(2), 50 (2023)
3. Bińkowski, M., Sutherland, D.J., Arbel, M., Gretton, A.: Demystifying mmd gans. In: International Conference on Learning Representations (2018)
4. Challenge, G.: Grand challenge—a platform for end-to-end development of machine learning solutions in biomedical imaging (2021)
5. Chung, H., Kim, J., McCann, M.T., Klasky, M.L., Ye, J.C.: Diffusion Posterior Sampling for General Noisy Inverse Problems. In: The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023. OpenReview.net (2023), https://openreview.net/forum?id=OnD9zGAGT0k
6. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. Advances in neural information processing systems **34**, 8780–8794 (2021)
7. Duffy, G., Cheng, P.P., Yuan, N., He, B., Kwan, A.C., Shun-Shin, M.J., Alexander, K.M., Ebinger, J., Lungren, M.P., Rader, F., et al.: High-throughput precision phenotyping of left ventricular hypertrophy with cardiovascular deep learning. JAMA cardiology **7**(4), 386–395 (2022)
8. Ho, J., Jain, A., Abbeel, P.: Denoising Diffusion Probabilistic Models. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H. (eds.) Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual (2020), https://proceedings.neurips.cc/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html
9. Huang, L., Zhou, Z., Guo, Y., Wang, Y.: A stability-enhanced cyclegan for effective domain transformation of unpaired ultrasound images. Biomedical Signal Processing and Control **77**, 103831 (2022)
10. Jiao, J., Zhou, J., Li, X., Xia, M., Huang, Y., Huang, L., Wang, N., Zhang, X., Zhou, S., Wang, Y., et al.: Usfm: A universal ultrasound foundation model generalized to tasks and organs towards label efficient image analysis. Medical image analysis **96**, 103202 (2024)
11. Peng, H., Xue, C., Shao, Y., Chen, K., Xiong, J., Xie, Z., Zhang, L.: Semantic segmentation of litchi branches using deeplabv3+ model. Ieee Access **8**, 164546–164555 (2020)
12. Sjoerdsma, M., Bouwmeester, S., Houthuizen, P., van de Vosse, F.N., Lopata, R.G.: A spatial near-field clutter reduction filter preserving tissue speckle in echocardiography. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control **68**(4), 979–992 (2020)
13. Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based Generative Modeling through Stochastic Differential Equations. In: 9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021. OpenReview.net (2021), https://openreview.net/forum?id=PxTIG12RRHS
14. Stevens, T.S.W., Meral, F.C., Yu, J., Apostolakis, I.Z., Robert, J., van Sloun, R.J.G.: Dehazing Ultrasound Using Diffusion Models. IEEE Trans. Medical Imaging **43**(10), 3546–3558 (2024). https://doi.org/10.1109/TMI.2024.3363460

15. Stevens, T.S.W., van Nierop, W.L., Luijten, B., van de Schaft, V., Nolan, O.I., Federici, B., van Harten, L.D., Penninga, S.W., Schueler, N.I., van Sloun, R.J.: zea: A Toolbox for Cognitive Ultrasound Imaging (Jul 2025), https://github.com/tue-bmd/zea
16. Van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N., Gouillart, E., Yu, T.: scikit-image: image processing in python. PeerJ **2**, e453 (2014)
17. Xia, M., Yang, H., Qu, Y., Guo, Y., Zhou, G., Zhang, F., Wang, Y.: Multilevel structure-preserved gan for domain adaptation in intravascular ultrasound analysis. Medical Image Analysis **82**, 102614 (2022)
18. Zhang, T.Y., Suen, C.Y.: A fast parallel algorithm for thinning digital patterns. Communications of the ACM **27**(3), 236–239 (1984)