MUX-USCT: A Noise-Robust Neural Network for Ultrasound Computed Tomography

Yuchen Yuan[1], Hanhan Wu[1], Jinyang Li[1], Hanchen Wang[2], Yixuan Wu[3], Youzuo Lin[2], and Lei Yang[1]*

[1] George Mason University, Fairfax, VA, USA
{yyuan21, hwu28, jli56, lyang29}@gmu.edu
[2] The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
{hcwang, yzlin}@unc.edu
[3] Johns Hopkins University, Baltimore, MD, USA
yixuan_wu@jhu.edu

Abstract. Deep neural networks (DNNs) have shown strong potential for ultrasound computed tomography (USCT) reconstruction in ideal noise-free environments, yet existing DNNs are vulnerable to the noisy conditions in clinical practice, as they equally treat inputs that suffer mild, moderate, or severe noise. More challenging, the distributions of noise shift along with the environment, indicating the limited effectiveness of noise-aware training, which injects a specific noise distribution into the training data. We rethink these challenges and observe that the DNN models can become more robust to noise if we know the noise sources and filter them out. This filtering operation is akin to the Multiplexers (or MUX), a fundamental combinational circuit in digital logic design. However, the challenge here is that noise can happen randomly during inference; as a result, the manually predefined MUX cannot work. To address these challenges, we propose MUX-USCT, a novel encoder-decoder DNN architecture that encodes the known acoustic acquisition geometry with an "adaptive MUX" that can automatically identify and filter noise, where the attention mechanism is applied in reconstructing the speed-of-sound map. On the OpenPros benchmark, MUX-USCT reaches 6.88 m/s MAE with 17% fewer parameters than the leading baseline with 7.65 m/s of MAE. Under simulated clinical noise, it remains stable across diverse degradation types that cause geometry-agnostic baselines to fail. Results show that the attention distributions in MUX-USCT reveal patterns that shift in response to transducer signal degradation.

Keywords: USCT · Robustness · DNN · MUX · Attention

1 Introduction

Ultrasound computed tomography (USCT) reconstructs quantitative tissue maps from acoustic waveform data, offering a radiation-free modality for breast screening [12–14, 21], prostate imaging [9, 16] and many others [4–8, 10, 18]. Unlike

* Corresponding author.

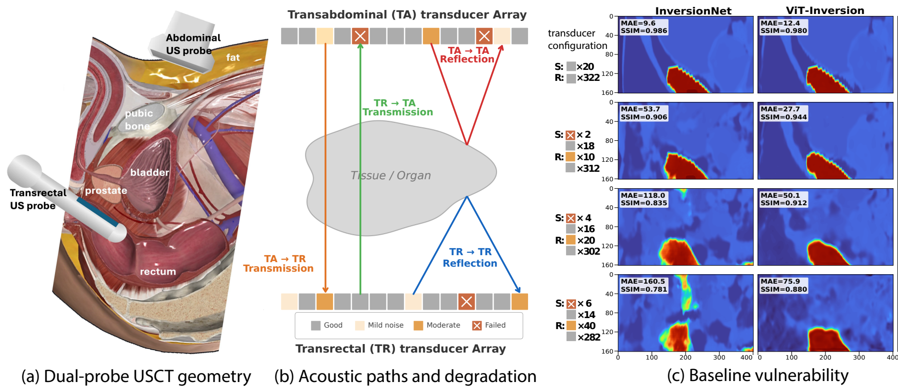

Fig. 1. (a) Dual-probe prostate USCT geometry [16]. (b) Four acoustic paths defined by source and receiver arrays, with transducer degradation types. (c) Reconstruction quality under progressively severe degradation with models trained on clean data.

conventional B-mode ultrasound, USCT exploits the full acoustic wavefield to achieve quantitative accuracy [11], but the underlying inverse problem, full waveform inversion (FWI) [15], is computationally demanding. Deep learning, in particular DNN-based approaches, has shown strong potential for data-driven USCT reconstruction: InversionNet [17], a fully convolutional encoder-decoder, and ViT-Inversion [2, 16], a Vision Transformer variant, both achieve strong accuracy on clean simulated data. Yet these DNNs are vulnerable to the noisy conditions in clinical practice, as they treat the input waveform tensor as a generic feature map, which cannot distinguish noisy data [7, 20].

On the other hand, the hardware degrades and environmental noise commonly exists in practice: Surveys [3, 8] report that 27–40% of probes in routine hospital use carry at least one fault. In addition, clinical environments introduce transducer noise; for example, poor probe-tissue contact attenuates signal amplitude (a.k.a., coupling loss), which can lead to different degradation levels (mild, moderate, and failed in Fig. 1b). Moreover, clinical noise distributions shift across sites and equipment, so noise-aware training that injects a fixed degradation profile offers limited generalization. These practical insights reveal the need for a DNN model that can be trained on clean data while adapting to noise during inference. Ideally, beyond reconstruction, the data-driven approach could further indicate which transducers are degraded, offering diagnostic information that may benefit future clinical workflows.

Unfortunately, the existing DNNs lack these capabilities: First, as shown in Fig. 1c, along with the increase in noise (from 1 failed source and 5 moderately noisy receivers to 3 and 20 of each), reconstruction results using the SOTA DNNs (InversionNet [17] and ViT [2]) trained on clean data will collapse. Second, these DNN models equally treat all input channels regardless of quality; degradation in any single transducer propagates unchecked through the DNN, and the resulting reconstruction offers no clue as to *which* transducer is at fault.

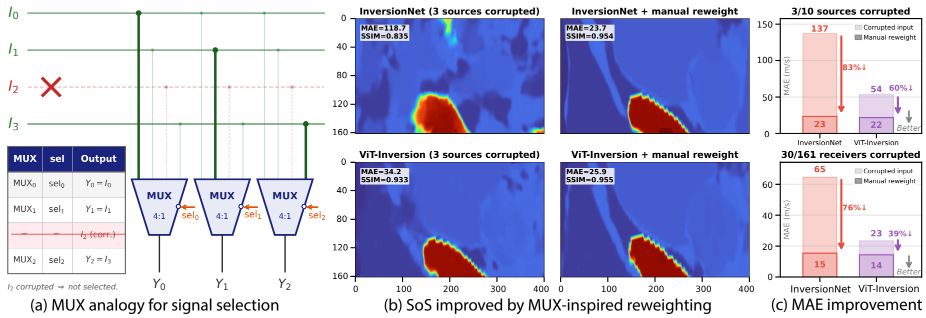

Fig. 2. Motivation of MUX-inspired design. (a) Illustration of an MUX in selecting clean and corrupted inputs. (b) under 3 out of 10 source corruption, SoS improved by MUX-inspired reweighting. (c) MAE improvements by MUX-inspired manual reweight.

To achieve our goal, we first investigate the data composition and observe four different data paths in USCT with different propagation characteristics, as shown in Fig. 1b. By decoupling the data into paths, different paths can be weighted to reconstruct organs (e.g., Bone and Prostate). In addition, inspired by the MUX, a fundamental combinational circuit in digital logic, given the waveform with known noise on sources, we manually filter out the noisy paths as the MUX in Fig. 2a. We observe a significant performance boost on both visualization and quantitative results in Fig. 2b-c. However, in practice, it remains challenging because noise paths are generated at runtime and cannot be identified manually using a preset configuration.

In addressing this practical challenge, we propose MUX-USCT, a novel DNN architecture. In MUX-USCT, an encoder acts as an *adaptive learned MUX* that uses an attention mechanism to assess each transducer by its signal content, which can automatically identify and filter out the corrupted transducers and compress multi-path waveforms into a compact representation. Then, a decoder reverses this process, selectively routing the most relevant path information to each output pixel. The entire architecture is parametric in the number of paths and transducers, which can be set according to the acquisition geometry.

Our contributions include (1) MUX-USCT, a novel DNN architecture with an encoder to map physical acquisition geometry into USCT reconstruction and a position-aware decoder to enable spatially selective path routing; (2) stable robustness under diverse clinically motivated degradations with no noise-augmented training, while matching or improving clean-data accuracy at 17% fewer parameters; (3) attention distributions that shift measurably in response to transducer degradation and vary with tissue type, with potential for automated anomaly detection in future clinical deployment.

2 Method

Problem Formulation. Given a sensing system with a known acquisition geometry, the input can be described as a waveform tensor $\mathbf{X} \in \mathbb{R}^{P\times S\times T\times R}$, where P acoustic paths are defined by the source–receiver probe pairing (e.g., same-probe reflection vs. cross-probe transmission), each path contains S transducer elements that fire sequentially, and each firing is recorded over T time steps at R receivers. Every measurement, therefore, carries two structural tags: *which path* it belongs to (encoding propagation geometry) and *which element* produced it (encoding source identity).

Design Philosophy. Unlike conventional approaches that ignore the structural information by flattening all channels into a single feature map, MUX-USCT retains and exploits it. In an acoustic system, different elements and paths carry unequal information about different spatial regions; a principled reconstruction model should reflect this structure rather than treating all channels identically. MUX-USCT therefore preserves path and element identity throughout the network, allowing attention mechanisms to learn signal relevance in a data-driven manner. Because the model captures the physics of the acquisition, robustness to hardware degradation emerges as a natural byproduct: attention trained on clean data learns to weight channels by their informativeness, thereby tending to downweight corrupted signals. Thus, even if the entire model is trained on *clean data only*, requiring no noise-specific augmentation or retraining, MUX-USCT can filter out or downweight corrupted signals at inference.

Architecture Overview. Fig. 3b shows the MUX-USCT overall network architecture. It contains two key components: (i) a *path-aware encoder* (see Fig. 3a) compresses each path into M regional tokens (i.e., $M\times P$ tokens in total); and (ii) a *position-aware decoder* (see Fig. 3c) takes the output tokens of the encoder and reconstructs a 2-D tissue-property SoS map via per-pixel cross-attention.

Path-Aware Encoder. Each path $\mathbf{X}_p \in \mathbb{R}^{S\times T\times R}$ is independently normalized and processed by a shared convolutional backbone (CNN Stem in Fig. 3a), producing features $\mathbf{h}_{p,s}\in\mathbb{R}^d$. A *transducer attention* mechanism aggregates across the S elements via a learned global query $\mathbf{q}\in\mathbb{R}^d$:

$$\alpha_{p,s} = \frac{\exp\big(\mathbf{q}^\top \mathbf{k}_{p,s}/\sqrt{d}\big)}{\sum_{s'=1}^{S}\exp\big(\mathbf{q}^\top \mathbf{k}_{p,s'}/\sqrt{d}\big)}, \qquad \bar{\mathbf{h}}_p = \sum_{s=1}^{S}\alpha_{p,s}\,\mathbf{v}_{p,s}, \tag{1}$$

where $\mathbf{k}_{p,s}=W_k\mathbf{h}_{p,s}$ and $\mathbf{v}_{p,s}=W_v\mathbf{h}_{p,s}$. The weights $\alpha_{p,s}$ act as *learned select signals* analogous to MUX control lines. Since the model is trained only on clean data, the query–key inner product space reflects the statistics of intact signals; at inference, a corrupted element s^* produces a key $\mathbf{k}_{p,s^*}$ that falls outside this learned subspace, receiving a low score. Softmax then concentrates mass on clean elements, filtering the corrupted channel without an explicit noise detector. Temporal attention pooling along the time axis further compresses each path into M regional tokens, each summarizing a spatially distinct sub-region of that path's acoustic coverage, yielding $\mathbf{Z}\in\mathbb{R}^{M\times P\times d}$. A multi-head self-attention layer over all $M \times P$ tokens then enables information exchange across paths.

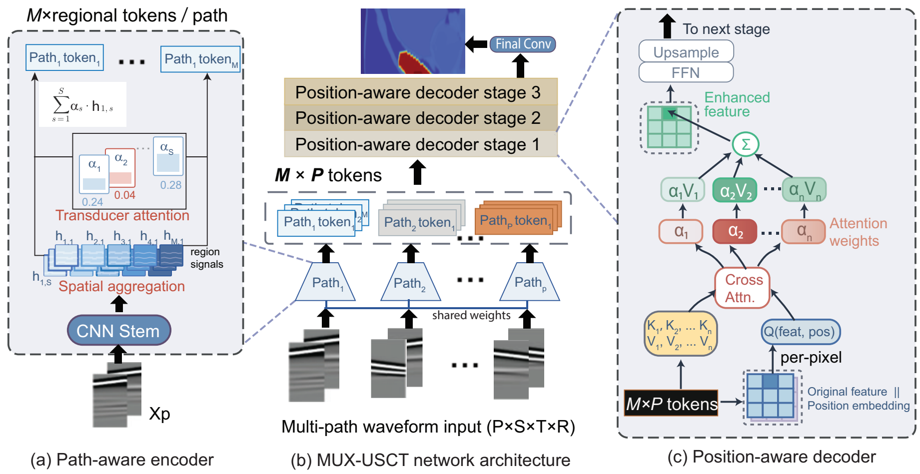

Fig. 3. MUX-USCT architecture: (a) details of MUX-USCT encoder; (b) MUX-USCT overall network architecture; (c) details of MUX-USCT decoder.

Position-Aware Decoder. The decoder reverses the MUX operation: it routes the compressed $M \times P$ tokens back to spatial locations via per-pixel cross-attention [1]. At position (i, j), a learned query $\mathbf{q}_{ij}$ attends to the full token set:

$$\mathbf{o}_{ij} = \sum_{n=1}^{M\times P} \beta_{ij,n}\, W_v' \mathbf{z}_n, \qquad \beta_{ij,n} = \mathrm{softmax}_n\Big(\mathbf{q}_{ij}^{\top} W_k' \mathbf{z}_n / \sqrt{d}\Big). \tag{2}$$

Since acoustic coverage is spatially non-uniform, each location learns to weight different path-region tokens according to its geometry (Fig. 6c). The resulting maps $\beta \in \mathbb{R}^{H\times W\times M\times P}$, reshapable to $[H, W, P, M]$, provide per-path contribution maps at every pixel. A TokenToGrid module [1] first materializes a coarse spatial feature map from the tokens via this cross-attention; subsequent upsampling stages then refine it into the final tissue-property map.

3 Experiments

3.1 Settings. To evaluate the proposed MUX-USCT, we use the OpenPros prostate USCT benchmark [16], which includes 1000 test samples based on Finite Difference Time Domain (FDTD) simulated waveforms. The acquisition yields P=4 acoustic paths, S=10 sources per path, T=1000 time steps, and R=161 receivers (Fig. 1a–b). We set M=5 regional tokens per path ($P \cdot M$=20 total) and embedding dimension d=384; the output is a 401×161 speed-of-sound (SoS) map. According to different clinical failure modes [8, 3], we created 4 simulated degradation scenarios, denoted by S1 to S4, summarized in Tab. 1.

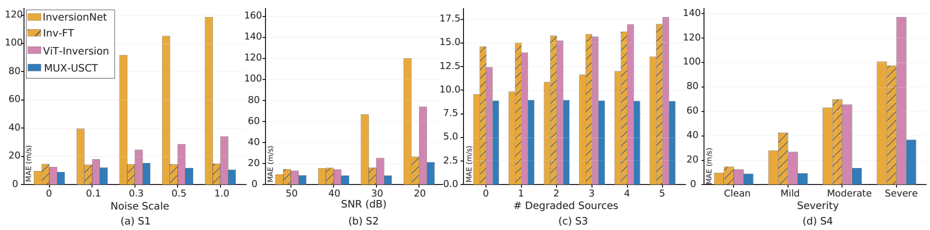

Fig. 4. Robustness across four clinical noise scenarios (S1–S4) in Tab. 1.

For comparison, MUX-USCT has three competitors, including InversionNet, ViT-Inversion, and Inv-FT. InversionNet and ViT-Inversion are obtained from official pretrained checkpoints in [16], which are trained on clean data. Inv-FT applies noise-aware training [19], using InversionNet with additional fine-tuning for 40 epochs on data with noise scenario S1, where 1–5 randomly selected sources per sample are corrupted with additive Gaussian noise ($\sim$0 dB SNR per channel).

3.2 Robustness Evaluation. Table 2 shows quantitative results comparison among MUX-USCT and competitors on clean data and noisy scenario S4. For noise-agnostic DNNs, InversionNet and ViT-Inversion, performance degrades significantly under noise S4, compared with clean data, with SSIM decreasing from 0.9893 to 0.8908. For Inv-FT, as it is finetuned using S1, it also has the same performance degradation on S4. Although MUX-USCT also has performance degradation on S4, its performance on S4 is close to the performance of ViT-Inversion on clean data. Moreover, MUX-USCT has 17% fewer parameters than InversionNet, and it even achieves performance improvement on clean data. These results validate the robustness of MUX-USCT in noisy environments and reveal that exploiting the acquisition structure can improve reconstruction.

The robustness of MUX-USCT can be further verified on varied degradation scenarios. Fig. 4 reports the results across all four scenarios (S1–S4) with different noise levels. Results again show that the noise-agnostic models (InversionNet and ViT-Inversion) suffer significant performance degradation under noise during inference. Such degradation also appears on Inv-FT for S2–S4; however, Inv-FT shows strong robustness in S1, as expected since its fine-tuning data follow the distribution of S1. These results suggest that noise-aware training is effective within the fine-tuned distribution but does not readily generalize to

Table 1. Simulated degradation scenarios.

ID	Degradation	Simulation	Clinical Origin
S1	Source noise	-10 dB Gaussian per element	Faulty transducer element
S2	Receiver noise	Global AWGN, 5–50 dB SNR	Electronic/thermal interference
S3	Coupling loss	10–50% amplitude attenuation	Poor probe-tissue contact
S4	Combined	S1-S3 at mild/moderate/severe levels	Multi-fault co-occurrence

Table 2. Reconstruction quality comparison (best in **bold**). Moderate level is for S4.

Model	Params	Clean			S4: Combined		
		MAE↓	SSIM↑	PCC↑	MAE↓	SSIM↑	PCC↑
InversionNet	20.45M	7.65	0.9893	0.9908	62.94	0.8908	0.9208
Inv-FT	20.45M	12.18	0.9806	0.9766	69.78	0.9115	0.9353
ViT-Inversion	28.33M	14.27	0.9792	0.9760	65.57	0.8988	0.8995
MUX-USCT	**16.92M**	**6.88**	**0.9912**	**0.9923**	**13.49**	**0.9761**	**0.9635**

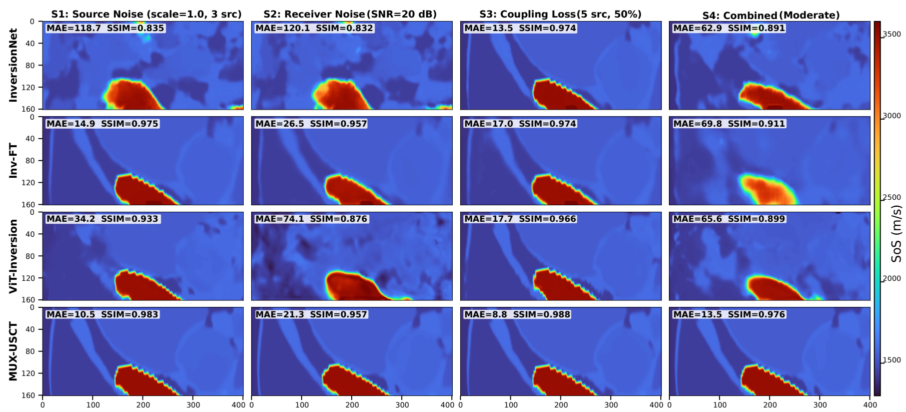

Fig. 5. SoS map visualization results of four models on noise scenarios (S1–S4).

unseen degradation types, and exhaustively covering all clinical noise profiles is impractical.

On the other hand, results in Fig. 4 clearly show that MUX-USCT is robust to different scenarios. We have an interesting observation for "S3: Coupling loss" in Fig. 4c, where MUX-USCT's MAE is nearly unchanged while all baselines rise along with the increase of degraded sources. Such robustness can also be reflected from the visualization results in Fig. 5, where the last row is obtained by MUX-USCT for all scenarios, which preserves tissue boundaries. For comparison, InversionNet and ViT-Inversion produce severe artifacts under source and receiver noise (S1, S2). Inv-FT recovers well under source noise (S1) but degrades sharply under combined degradation (S4). These results show that MUX-USCT enables training on clean data while achieving high performance across varying levels of noise during inference.

3.3 Physics Accordance and Attention Analysis. To analyze why MUX-USCT works, we record the encoder's transducer attention. Fig. 6a compares attention weights on clean data versus data with 5 corrupted sources. The orange boxes highlight weights that drop markedly relative to the clean baseline, consistently corresponding to the corrupted sources, indicating that the encoder effectively downweights these channels. This confirms that transducer attention

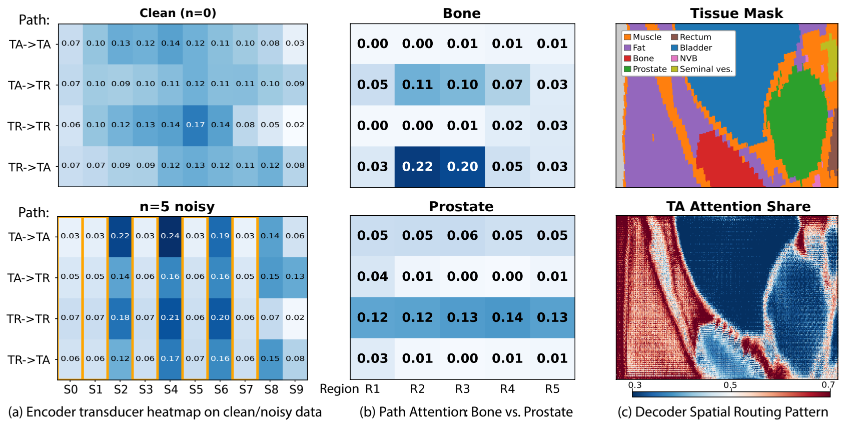

Fig. 6. Attention analysis: (a) Encoder transducer attention on clean vs. corrupted data; (b) Per-tissue path attention for bone vs. prostate; (c) Decoder spatial routing: TA attention share varies smoothly across the image.

Table 3. Ablation study. Noise columns: $n/10$ sources corrupted ($-10\,\text{dB}$).

Configuration	Clean MAE	n=3	n=5	n=7
MUX-USCT (full)	**6.88**	**10.51**	**18.18**	**42.53**
w/o transducer attention	7.66	21.05	45.81	82.93
w/o decoder cross-attn	30.93	26.93	33.21	55.07

functions as the learned selective filter described in Section 2, consistently suppressing corrupted channels without explicit noise labels.

Fig. 6b reveals that bone and prostate tissues rely on distinct path-region tokens, indicating that the decoder's routing is tissue-dependent.

Fig. 6c shows the decoder's transabdominal (TA) attention share, which varies smoothly across the image: different spatial locations draw on different paths, reflecting position-dependent routing learned without explicit supervision.

3.4 Ablation Study. We conduct an ablation study to demonstrate the importance of attention stages in MUX-USCT, as shown in Tab. 3. From the results, removing transducer attention in encoder barely affects clean MAE ($+0.78\,\text{m/s}$) but sharply increases noise sensitivity ($2.5\times$ at n=5): the channel-weighting it provides is modest for clean signals yet critical when some channels degrade. Disabling decoder cross-attention raises clean MAE over fourfold to $30.93\,\text{m/s}$, confirming per-pixel token routing as the core reconstruction mechanism. Together, the two attention stages are complementary: the encoder selects *what* to listen to, and the decoder decides *where* each signal contributes.

4 Conclusion

MUX-USCT demonstrates that encoding the physical acquisition geometry into the network architecture improves both clean-data accuracy and robustness under diverse clinical degradations. Although this paper uses a dual-probe prostate configuration, the design requires only that the probe layout and resulting path structure are known, a condition met by all synthetic aperture USCT systems. The principle of learned selective routing of structured sensor channels can extend to other multi-channel sensing modalities (e.g., multi-coil MRI). Our code is available at `https://github.com/TheYuchen/mux-usct-miccai2026`.

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

References

1. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers. In: European Conference on Computer Vision (ECCV). LNCS, vol. 12346, pp. 213–229. Springer (2020)
2. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: transformers for image recognition at scale. In: International Conference on Learning Representations (ICLR) (2021)
3. Dudley, N.J., Woolley, D.J.: A multicentre survey of the condition of ultrasound probes. Ultrasound **24**(4), 190–197 (2016)
4. Fradi, M., Zahzah, E.h., Machhout, M.: Real-time application based cnn architecture for automatic usct bone image segmentation. Biomedical Signal Processing and Control **71**, 103123 (2022)
5. Lin, Y., Feng, S., Theiler, J., Chen, Y., Villa, U., Rao, J., Greenhall, J., Pantea, C., Anastasio, M., Wohlberg, B.: Survey of deep learning and physics-based approaches in computational wave imaging (2025), `https://arxiv.org/abs/2410.08329`
6. Long, X., Chen, J., Liu, W., Tian, C.: Deep learning ultrasound computed tomography under sparse sampling. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control **70**(9), 1084–1100 (2023)
7. Lozenski, L., Wang, H., Li, F., Anastasio, M., Wohlberg, B., Lin, Y., Villa, U.: Learned full waveform inversion incorporating task information for ultrasound computed tomography. IEEE transactions on computational imaging **10**, 69–82 (2024)
8. Mårtensson, M., Olsson, M., Segall, B., Fraser, A.G., Winter, R., Brodin, L.Å.: High incidence of defective ultrasound transducers in use in routine clinical practice. European Journal of Echocardiography **10**(3), 389–394 (2009)
9. Mohsen, N.: Role of mri, ultrasound, and computed tomography in the management of prostate cancer. Pet Clinics **17**(4), 565–583 (2022)
10. Pfistner, P., Zapf, M., Ruiter, N.V.: Improved transducer performance for 3d usct. In: 2024 IEEE Ultrasonics, Ferroelectrics, and Frequency Control Joint Symposium (UFFC-JS). pp. 1–3. IEEE (2024)
11. Ruiter, N.V., Zapf, M., Hopp, T., Gemmeke, H.: Ultrasound tomography. In: Quantitative Ultrasound in Soft Tissues, pp. 171–200. Springer (2023)

12. Sandhu, G.Y., Li, C., Roy, O., Schmidt, S., Duric, N.: Frequency domain ultrasound waveform tomography: breast imaging using a ring transducer. Physics in Medicine & Biology **60**(14), 5381–5398 (2015)
13. Sheng, Y., Wang, H., Liu, Y., Yang, J., Jiang, W., Lin, Y., Yang, L.: Aps-usct: Ultrasound computed tomography on sparse data via ai-physic synergy. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 99–108. Springer (2024)
14. Vargas, A., Hernandez, N., Ramirez, A.B., Pertuz, S.: Breast cancer detection from ultrasound computed tomography imaging using radiomic analysis: in silico trial. Medical Physics **52**(4), 2465–2474 (2025)
15. Virieux, J., Operto, S.: An overview of full-waveform inversion in exploration geophysics. Geophysics **74**(6), WCC1–WCC26 (2009)
16. Wang, H., Wu, Y., Feng, Y., Jin, P., Zhang, L., Feng, S., Wiskin, J., Turkbey, B., Pinto, P.A., Wood, B.J., Luo, S., Chen, Y., Boctor, E., Lin, Y.: OpenPros: A large-scale dataset for limited view prostate ultrasound computed tomography. arXiv preprint arXiv:2505.12261 (2025)
17. Wu, Y., Lin, Y.: InversionNet: An efficient and accurate data-driven full waveform inversion. IEEE Transactions on Computational Imaging **6**, 419–433 (2020)
18. Xie, X., Yang, M.: Usct-unet: Rethinking the semantic gap in u-net network from u-shaped skip connections with multichannel fusion transformer. IEEE Transactions on Neural Systems and Rehabilitation Engineering **32**, 3782–3793 (2024)
19. Zhang, Y., Borisov, D., Abbasi, S., Hoch, A., Miller, R.D., Sloan, S.D.: Enhancing machine learning based seismic inversion with noise-augmented training datasets. Geophysical Journal International **244**(1), ggaf429 (2026). https://doi.org/10.1093/gji/ggaf429
20. Zhang, Z., Lin, Y.: Data-driven seismic waveform inversion: A study on the robustness and generalization. IEEE Transactions on Geoscience and Remote sensing **58**(10), 6900–6913 (2020)
21. Zhou, L., Zhang, Q., Wu, Y., Liu, Z., Wu, Y., Li, X., Qiu, W., Lou, C., Ding, M., Yuchi, M.: A coupling, stabilizing, and shaping strategy for breast ultrasound computed tomography (usct) with a ring array transducer. Ultrasonics **138**, 107212 (2024)