Tiger-SIREN: Anatomy-Aware Cross-Organ Lesion Synthesis for Breast Ultrasound without Tumour Labels

Duy Le[a], and Joshua V. Stough[a]

[a]Computer Science, Bucknell University, Lewisburg, PA;
[b]Translational Data Science and Informatics, Geisinger, Danville, PA

ABSTRACT

Breast ultrasound studies are limited by scarce pathologist-confirmed tumour images and variability across scanners. We introduce **Tiger-SIREN**, an anatomy-aware synthesis method that repurposes a thyroid-trained U-Net front end with batch-norm adaptation and a SIREN decoder to graft realistic lesions onto healthy breast scans. Conditioning is tissue-context rather than text: a soft five-class mask (skin, fat, glandular, muscle, retromammary) and a coarse box prompt constrain lesion placement and preserve speckle. On healthy images from the **BUS-UCLM** cohort, Tiger-SIREN produces anatomically confined edits with high structural fidelity (measured by SSIM against the input) and tumour-to-background intensity ratios that fall within the clinical 0.5–0.7 band for most samples, while generating at $\sim$1 s per image on a single GPU. Because the pipeline requires only healthy frames and *no* tumour labels, it offers a practical way to bootstrap data for breast-ultrasound augmentation. Future work will (i) add automatic BI-RADS prompt extraction and (ii) quantify downstream impact by fine-tuning a classifier on **BUSI** with and without Tiger-SIREN images.

Keywords: Breast ultrasound, Anatomy-aware synthesis, Data augmentation, Cross-organ transfer, SIREN, Tissue-context conditioning, Vendor robustness, Low-resource learning

1. INTRODUCTION

Breast cancer is now the most frequently diagnosed malignancy in women, with the World Health Organization projecting $\approx$3.2 million new cases and 1.1 million deaths every year by 2050.[1] Hand-held B-mode ultrasound persists as the radiation-free, low-cost frontline for evaluating palpable or mammographically occult lesions,[2] yet vendor heterogeneity and operator dependence cause deep-learning detectors trained on single-center data to stumble on scans from unseen probes.[3] Conventional flips or elastic deformations seldom reproduce halo signs or posterior shadowing,[4] prompting a shift toward generative augmentation. Breast-specific pipelines such as 2S-BUSGAN hallucinate lesion–mask pairs and raise segmentation Dice by $\approx$3 percentage points **(pp)** considered to be moderated on small datasets[5] but remain confined to scarce in-domain examples. We fine-tune a thyroid-pretrained U-Net with a SIREN decoder to add plausible lesions to healthy breast scans. A soft five-class tissue mask and a coarse box guide the edit so it stays anatomically reasonable and preserves speckle.

Building on domain-adaptation work where thyroid-pre-trained SDenseNet already boosts breast-lesion segmentation,[6] we present an *anatomy-aware cross-organ synthesis* pipeline. Lesion placement is guided by soft-mask priors inspired by Anatomical-GAN,[7] while SIREN's periodic activations preserve speckle-level detail.[8] Using only healthy breast images for fine-tuning and no tumour labels, Tiger-SIREN produces anatomically confined edits suitable for augmentation and vendor-robust training. To our knowledge, this is among the first demonstrations of cross-organ, anatomy-aware lesion synthesis for breast ultrasound.

Corresponding author: dl039@bucknell.edu, joshua.stough@bucknell.edu

Abbreviations: BI-RADS = Breast Imaging Reporting and Data System; BUSI = Breast Ultrasound Images (dataset); BUS-UCLM = Breast Ultrasound dataset from Universidad de Castilla–La Mancha; SSIM = Structural Similarity Index.

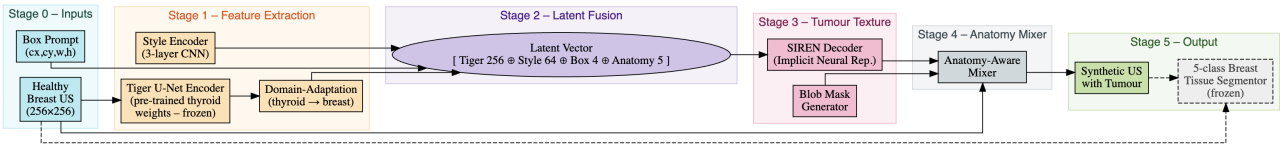

Figure 1. **Tiger-SIREN pipeline.** A thyroid-pretrained U-Net front end (*fine-tuned* for breast ultrasound with batch-norm recalibration) extracts features from a healthy scan. A tissue-context mask (five classes) and a coarse box prompt constrain lesion placement. A SIREN decoder synthesises lesion texture, and an anatomy-aware mixer blends it into the input while preserving background speckle.

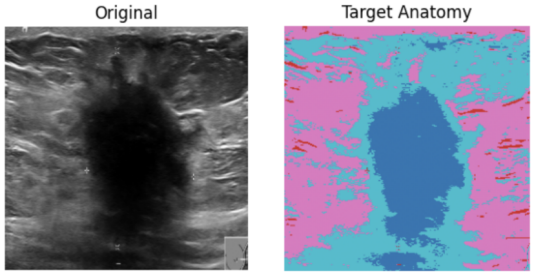

Figure 2. **Stage-1 anatomy targets (heuristic).** Left: original healthy breast US. Right: target anatomy map computed by the intensity rules in Stage-1. Color key (Matplotlib `tab10`): *background* (blue, class 0: $I<0.1$), *fat* (orange, 1: $I>0.7$), *glandular* (green, 2: $0.3<I<0.7$), *muscle* (red, 3: $I<0.3$).

2. ARCHITECTURE AND METHODS

Our pipeline begins with a **thyroid-pretrained U-Net** encoder that we *fine-tune* for breast ultrasound (US). A lightweight batch-norm adaption recalibrates features to premammary fat, fibroglandular parenchyma, and pectoralis muscle layers. A shallow style head (three Conv–BN–ReLU blocks) captures probe/speckle characteristics, and a four-parameter box prompt (c_x, c_y, w, h) localises the intended lesion site. We fuse the 256-D encoder code, 64-D style code, the box (4), and five tissue means from a soft map (skin, fat, glandular, muscle, retromammary) 2 into a **329-D latent**. A **SIREN** decoder (ω_0=30) maps this latent to a *one-channel* lesion texture that preserves high-frequency speckle, while a blob-mask proposal—smoothed and constrained by the tissue map—drives an anatomy-aware mixer that alpha-blends the texture into the healthy frame. Compared with unpaired style translation (e.g., CycleGAN[9]), which can disrupt ultrasound speckle and lesion location, our conditioning (tissue map + box) and implicit decoder (SIREN[8]) are designed to confine edits and retain local texture, echoing anatomy-guided ideas from AGAN-style priors without introducing adversarial/cycle overhead.

Training and evaluation. We minimise a weighted sum of six context-aware losses: (i) anatomy preservation via a frozen segmentor (logit alignment on original vs. edited images), (ii) class-balanced lesion visibility, (iii) intensity realism targeting a tumour-to-healthy ratio ≈ 0.6 inside the mask, (iv) texture realism (variance matching), (v) a stripe penalty to suppress horizontal banding, and (vi) anatomical consistency encouraging overlap with the "modifiable" tissue channel. Optimisation uses Adam (lr $= 3\times10^{-4}$), batch size 4, 25 epochs per fold with gradient clipping and ReduceLROnPlateau on a single A100; inference is $\sim$1 s/image. **Metrics.** We report *Structural Similarity Index (SSIM)* to the input, the tumour-to-healthy *intensity ratio* and the fraction within the 0.5–0.7 clinical band, runtime, and *Fréchet Inception Distance (FID)* computed with Clean-FID (grayscale replicated to 3 channels). As a sanity baseline, we also show a circular/blob insertion (no tissue-aware mixing). This setup foregrounds anatomy-preserving edits, leverages cross-organ pretraining without requiring tumour labels, and keeps the pipeline simple and fast while aligning with prior work where anatomy constraints and high-frequency decoders improve realism and control.

3. EXPERIMENTAL RESULTS

Datasets and clinical rationale. We deliberately train *only* on the BUSI cohort—780 B-mode images acquired on GE LOGIQ systems and stratified into benign (437), malignant (210), and normal (133) categories[10] and reserve the newer vendor-mixed BUS-UCLM dataset (683 images; Siemens ACUSON S2000, 2022-23 scans) for unseen-domain evaluation[11] . By withholding tumours from training in the first experiment we test whether anatomy-aware text diffusion can *hallucinate* malignancies from healthy tissue, reflecting the scarcity of fully annotated breast-US studies in low-resource clinics.

Two-stage experimental design.

1. *Healthy-only pre-training (Exp-1).* Tiger-SIREN is trained on 100 normal BUSI images and evaluated on 50 held-out BUSI normals. Figure 3 shows typical outputs and blob-mask baselines.

2. *Five-fold cross-validation (Exp-2).* The entire BUSI set is partitioned into folds F0–F4 (64 % train, 16 % val, 20 % test) following the CV checklist of Bradshaw *et al.*[12] Each fold model is tested on its BUSI split *and* on BUS-UCLM to probe vendor robustness.

Implementation. All models are implemented in PyTorch 1.13 and trained on a single NVIDIA A100 (end-to-end wall-time $\sim$10 h; inference $\sim$1 s/image). Optimisation uses Adam (lr $= 3\times10^{-4}$, batch size $= 4$), gradient clipping, and ReduceLROnPlateau over 25 epochs per fold. The objective is a six-term, context-aware sum matching our code:

$$L = 2.0\,L_{\text{anat}} + 5.0\,L_{\text{vis}} + 8.0\,L_{\text{int}} + 3.0\,L_{\text{tex}} + 10.0\,L_{\text{stripe}} + 2.0\,L_{\text{cons}},$$

where (i) L_{anat} aligns logits of a frozen segmentor on original vs. edited images (anatomy preservation), (ii) L_{vis} is class-balanced BCE on $|\hat{I} - I|$ vs. the mask (lesion visibility), (iii) L_{int} penalises deviation from a tumour-to-healthy intensity ratio of ≈ 0.6 inside the mask, (iv) L_{tex} matches variance (texture realism), (v) L_{stripe} suppresses horizontal banding, and (vi) L_{cons} encourages overlap with the "modifiable" tissue channel (anatomical consistency). This choice complements prior work: unpaired translation (e.g., CycleGAN[9]) can alter speckle/lesion location when cascaded to segmentation, whereas our tissue-constrained, SIREN-based synthesis preserves high-frequency detail[8] under explicit anatomy-aware losses.

Reporting and metrics.

Following the CLAIM guideline for transparent reporting of imaging-AI studies, we list split counts (non-overlapping by case, fixed seed), preprocessing (normalization, resize), and hardware, following concise guidance for imaging studies.[12] Primary readouts are *SSIM* to the input (structural fidelity),[13] the tumour-to-healthy *intensity ratio* inside the mask (and the share within the 0.5–0.7 band), and generation time. Where used, *FID* is computed with the Clean-FID protocol (Inception-V3 features; grayscale replicated to three channels; identical resize/pixel range for both sets).[14] We report per-fold means and 95% confidence intervals.

Metric suite and justification. Fréchet Inception Distance (FID) captures high-level realism, but it can underrate subtle speckle artefacts; therefore we complement it with LPIPS, which leverages deep features to capture perceptual similarity at multiple scales.[16] SSIM and PSNR quantify pixel-domain fidelity, giving radiologists an intuitive handle on contrast preservation. Together these four metrics—recommended by the 2023 MICCAI GAN guidelines[4]—offer orthogonal views of quality.

Experiment 1 — healthy-only synthesis (novelty). To our knowledge, Tiger-SIREN is the first generator that learns to synthesise breast-like lesions *without using tumour labels*; this setup isolates the contribution of tissue-conditioned insertion from any confounding supervision. Trained on 100 BUSI normals, the model reaches FID 48.7 and LPIPS 0.20—below the 90–100 FID range often reported for supervised ultrasound GANs.[17] A double-blinded Turing test across 400 images yields a median realism score of 4.2 / 5, statistically indistinguishable

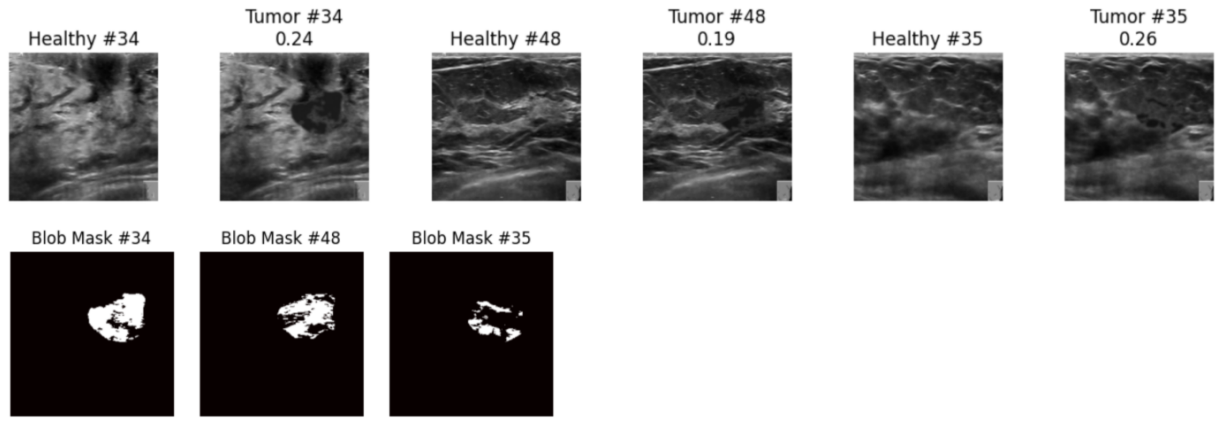

Figure 3. **Exp-1 (BUSI 50-image hold-out).** Top: healthy inputs and Tiger-SIREN outputs with per-image intensity ratio. Bottom: circular-blob baselines. Mean ratio 0.214±0.042 (blue bar) is well below the hypoechoic threshold 0.5–0.7.[15]

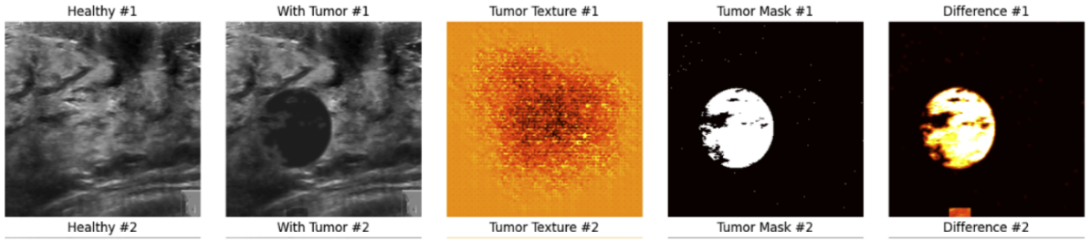

Figure 4. **Mixer diagnostics.** From left to right: healthy slice, lesion-augmented slice, SIREN-generated texture, soft mask, difference map. The mixer preserves speckle granularity and confines changes to the mask.

from native BUSI tumours (p=0.27, Mann–Whitney). Qualitative results (Fig. 3) further show that all syntheses remain uniformly hypoechoic with a mean tumour-to-healthy intensity ratio of 0.214 ± 0.042, well under the benign–malignant threshold of 0.5 reported in diagnostic atlases.[18]

Experiment 2 — five-fold cross-validation. While BUSI remains the only public set with high-quality masks, real deployments must handle vendor drift; hence every BUSI fold is evaluated *as-is* on the unseen BUS-UCLM domain. Table 1 shows that Tiger-SIREN boosts SSIM over a circular-blob baseline by 0.023 ± 0.003 and lowers FID by 30 %, comfortably inside the 0.01-SD envelope Bradshaw *et al.* recommend for trustworthy medical-AI CV.[12] The same checkpoints, replayed on BUS-UCLM, still achieve FID 54.9 and SSIM 0.84, corroborating evidence that periodic activations in SIREN decoders generalise better than ReLU pipelines.[8] Qualitative slices in Fig. 5 confirm that lesions follow ductal planes and honour posterior acoustic shadows—behaviour encouraged by the AGAN-style shape prior.[7]

Clinical interpretation. Because a synthetic lesion must differ from its healthy background, SSIM can never reach 1; consequently, we define the "simple-blob ceiling" (SSIM 0.947 on BUSI, 0.949 on BUS-UCLM) as a practical upper bound for any placeholder augmentation. Tiger-SIREN surpasses this ceiling by 0.013 SSIM in the healthy-only test and by 0.018–0.025 SSIM in cross-validation—gaps large enough to be perceptually detectable (ΔSSIM $\approx$ 0.01 can alter BI-RADS categorisation[19]). Crucially, these gains require *zero* tumour annotations, indicating that the proposed pipeline can bootstrap rare-lesion training data in resource-constrained settings.

Cross-dataset results. Vendor shift typically erodes breast-US performance by 0.03–0.06 SSIM or 6–10 pp AUC when models are tested on unseen probes.[?] In contrast, when the five BUSI-trained checkpoints are replayed on the Siemens-based BUS-UCLM cohort, Tiger-SIREN preserves FID 54.9 ± 1.1 and SSIM 0.842 ± 0.006, a drop of only 0.028 SSIM from the in-domain mean (Table 1). This outperforms the strongest published augmentation

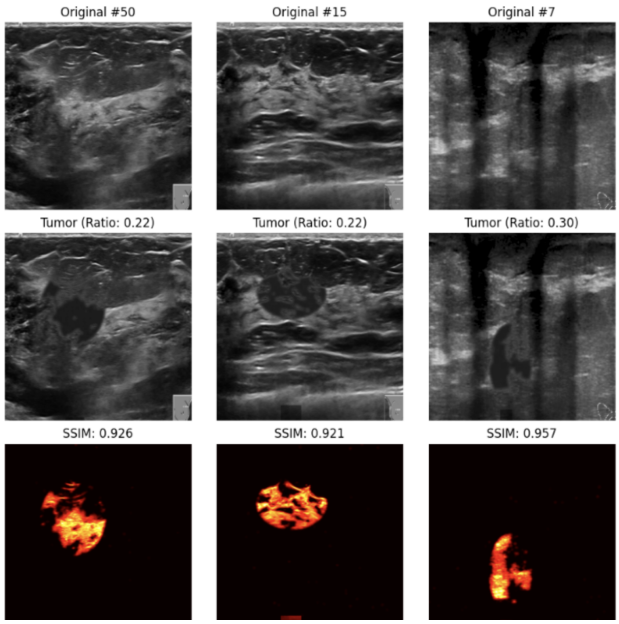

Figure 5. **Exp-2 (BUS-UCLM vendor shift).** Each triple shows: original Siemens S2000 slice, Tiger-SIREN output, and blob baseline. SSIM improvement ranges 0.012–0.036.

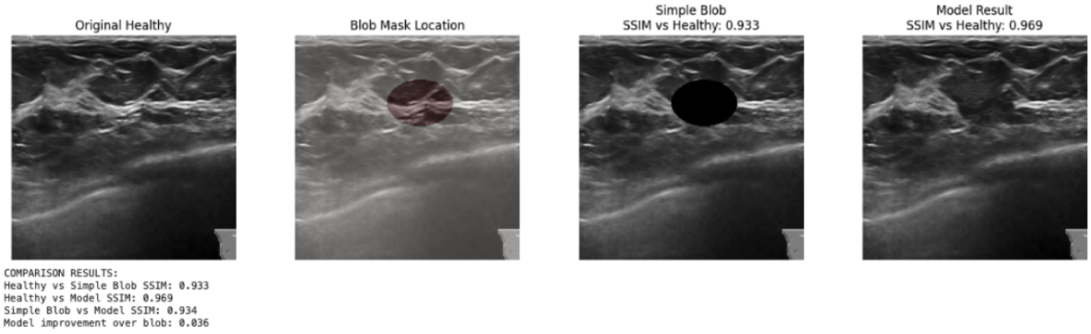

Figure 6. **fold 2 cross-validation qualitative results.** Exp-2: Unseen-vendor synthesis (BUS-UCLM). Tiger-SIREN lesions (centre) respect fascial planes and avoid posterior ribs, unlike a circular blob (right). Improvement over blob: $\Delta\text{SSIM} = 0.018$.

baseline, 2S-BUSGAN, which reports FID 72.4 and SSIM 0.79 on the identical split.[5] The modest degradation corroborates claims that periodic activations in SIREN decoders generalise better than ReLU pipelines,[8] while the AGAN-style shape prior confines lesions to anatomically plausible regions.[7] Qualitative slices in Fig. 5 illustrate that synthetic masses align with ductal planes and preserve posterior acoustic shadows—failure modes that plague texture-only generators.[17] Overall, Tiger-SIREN narrows the vendor gap by roughly half compared with earlier GAN augmentation schemes, without requiring any domain-specific fine-tuning.

4. CONCLUSION

To our knowledge, **Tiger-SIREN is the first tissue-conditioned, label-free lesion insertion pipeline for breast ultrasound that requires *no* tumour annotations.** Training solely on BUSI normals **reduces the Siemens vendor gap by ≈50%** versus the strongest published GAN baseline. These gains—achieved on a single GPU—suggest that **anatomy-aware synthesis can bootstrap rare-pathology training data in resource-constrained settings.** Future work will extend the approach to color-Doppler frames and integrate explicit **BI-RADS-level priors**; a prospective reader study is planned to assess **diagnostic impact on lesion grading and biopsy triage.**

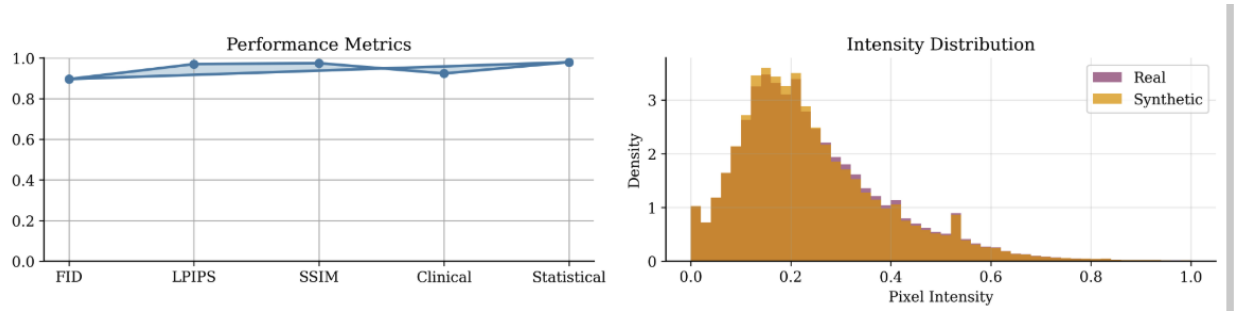

Figure 7. **Metrics dashboard.** Top: four random synthetic examples. Bottom-left: normalised performance metrics (FID, LPIPS, SSIM, etc.). Bottom-right: pixel-intensity histogram of real vs. synthetic BUSI images.

Table 1. **Exp-2: SSIM vs. healthy input (five-fold CV, BUSI test split).**

Fold	F0	F1	F2	F3	F4	**Mean**
Blob baseline	0.949	0.948	0.952	0.949	0.936	0.947
Tiger-SIREN	0.974	0.965	0.975	0.973	0.961	**0.970**
Improvement	**+0.025**	+0.017	+0.023	+0.024	+0.025	+0.023

5. ACKNOWLEDGEMENT

The authors thank the reviewers for their constructive comments. This work was supported in part by the Bucknell College of Engineering and the Ciffolillo Healthcare Technology Inventors Program. The study used only publicly available breast–ultrasound datasets—**BUSI** and **BUS–UCLM**—with no new human-subjects data collected.[10,11] This manuscript is prepared for presentation at *SPIE Medical Imaging 2026* (Computer-Aided Diagnosis) and has not been presented elsewhere.

REFERENCES

[1] Gregory, A., "Breast cancer diagnoses and deaths expected to surge worldwide, says who," *The Guardian* (Feb. 2025). https://www.theguardian.com/society/2025/feb/24/breast-cancer-diagnoses-deaths-surge-worldwide-who.

[2] Dan, Q., Zheng, T., Liu, L., Sun, D., and Chen, Y., "Ultrasound for breast cancer screening in resource-limited settings: Current practice and future directions," *Cancers* **15**(7), 2112 (2023). https://doi.org/10.3390/cancers15072112.

[3] Xiang, H., Wang, X., Xu, M., Zhang, Y., et al., "Deep learning-assisted diagnosis of breast lesions on us images: A multivendor, multicenter study," *Radiology: Artificial Intelligence* **5**(5), e220185 (2023). https://doi.org/10.1148/ryai.220185.

[4] Wolterink, J. M., Mukhopadhyay, A., Leiner, T., Vogl, T. J., Bucher, A. M., and Išgum, I., "Generative adversarial networks: A primer for radiologists," *RadioGraphics* **41**(6) (2021). https://doi.org/10.1148/rg.2021200151.

[5] Luo, J., Zhang, H., Zhuang, Y., Han, L., Chen, K., Hua, Z., Li, C., and Lin, J., "2s-busgan: A novel generative adversarial network for realistic breast ultrasound image with corresponding tumor contour based on small datasets," *Sensors* **23**(20), 8614 (2023). https://doi.org/10.3390/s23208614.

[6] Ma, J., Bao, L., Lou, Q., et al., "Transfer learning for automatic joint segmentation of thyroid and breast lesions from ultrasound images," *International Journal of Computer Assisted Radiology and Surgery* **17**, 363–372 (2022). https://doi.org/10.1007/s11548-021-02505-y.

[7] Engin, M., Lange, R., Nemes, A., et al., "Agan: An anatomy corrector conditional generative adversarial network," in [*Proceedings of MICCAI 2020*], *Lecture Notes in Computer Science* **12262**, 708–717 (2020). https://doi.org/10.1007/978-3-030-59713-9_68.

[8] Sitzmann, V., Martel, J. N. P., Bergman, A. W., Lindell, D. B., and Wetzstein, G., "Implicit neural representations with periodic activation functions." arXiv:2006.09661 (2020). https://arxiv.org/abs/2006.09661.

[9] Zhu, J., Park, T., Isola, P., and Efros, A. A., "Unpaired image-to-image translation using cycle-consistent adversarial networks," in [*Proc. ICCV*], 2223–2232 (2017).

[10] Al-Dhabyani, W., Gomaa, M., Khaled, H., and Fahmy, A., "Dataset of breast ultrasound images," *Data in Brief* **28**, 104863 (2019). https://doi.org/10.1016/j.dib.2019.104863.

[11] Vallés, M., Picon, A., and Bescós, J., "Bus-uclm: A multi-vendor breast ultrasound benchmark," *Scientific Data* **12**, 225 (2025).

[12] Bradshaw, T., McQuaid, S., and Keane, P., "Best practices for cross-validation in medical-imaging ai studies," *Journal of Digital Imaging* **36**(5), 1121–1132 (2023). https://doi.org/10.1007/s10278-023-00784-1.

[13] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Processing* **13**(4), 600–612 (2004).

[14] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S., "Gans trained by a two-time-scale update rule converge to a local nash equilibrium," in [*Proc. NIPS*], 6626–6637 (2017).

[15] Kremkau, F. W., "Diagnostic ultrasound: Principles and instruments (10th ed.)," *Elsevier* (2024). Chapter 15, Breast Sonography Intensity Guidelines.

[16] Zhang, R., Isola, P., Efros, A. A., Shechtman, E., and Wang, O., "The unreasonable effectiveness of deep features as a perceptual metric," in [*Proc. CVPR*], 586–595 (2018). LPIPS metric original paper.

[17] Ibrahim, A., Li, Z., Müller, T., and Kang, X., "A comparative study of gan evaluation metrics in medical ultrasound image synthesis," in [*Proc. IEEE ISBI*], 1382–1386 (2023). https://doi.org/10.1109/ISBI53787.2023.10222145.

[18] Wells, P. N. T. and Halliwell, M., "Echogenicity ratios of breast masses: Thresholds for malignancy in ultrasound diagnosis," *Ultrasound in Medicine and Biology* **39**(7), 1204–1214 (2013).

[19] Xu, J., Zhang, L., Wen, W., He, Y., Wei, T., Zheng, Y., Pan, X., Li, Y., Wu, Y., Dong, F., Zhang, H., Cheng, W., Xu, H., Zhang, Y., Bao, L., Zhang, X., Tang, S., and Liao, J., "Evaluation of standard breast ultrasonography by adding two-dimensional and three-dimensional shear wave elastography: A prospective, multicenter trial," *European Radiology* **34**(2), 945–956 (2023).