FREQTRANS: ADAPTIVE FREQUENCY FEATURE FUSION FOR ROBUST BREAST ULTRASOUND IMAGE ANALYSIS

Hongyang Zhao[1] *Dan Lu*[1,3] *Yingnan Zhao*[1,3,*] *Yanchen Xu*[1] *Yi Lin*[2,*]

[1] College of Computer Science and Technology, Harbin Engineering University, China
[2] School of Interdisciplinary Medicine and Engineering, Harbin Medical University, China
[3] National Engineering Laboratory for Modeling and Emulation in E-Government
Harbin Engineering University, Harbin, 150001, China

ABSTRACT

Breast ultrasound imaging is widely used for early detection of breast cancer but is challenged by low contrast, blurred lesion boundaries, and speckle noise, which complicate the identification of subtle tumors. Here, we present FreqTrans, a frequency-domain model that decomposes images into low- and high-frequency components and applies attention mechanisms to enhance diagnostically relevant features. By integrating local and global cues and fusing multi-level representations, FreqTrans improves edge delineation and suppresses noise, enabling robust lesion detection. On public breast ultrasound datasets, FreqTrans outperforms standard convolutional and vision Transformer models, achieving higher sensitivity and overall detection accuracy, highlighting its potential for clinical application.

***Index Terms*—** Breast ultrasound, Tumor detection, Frequency-domain analysis, Adaptive fusion, Deep learning.

1. INTRODUCTION

Accurate breast cancer detection from medical images is crucial for early diagnosis and effective treatment [1]. However, breast ultrasound images pose significant challenges for automated analysis due to low contrast, blurred lesion boundaries, and severe speckle noise, particularly in small or early-stage lesions [2, 3]. Moreover, variations in imaging conditions, acquisition protocols, and patient anatomy, together with the limited availability of well-annotated datasets, further hinder robust deep learning performance, even for recent vision transformer (ViT) based models that rely heavily on large-scale data and global self-attention [4–6].

Existing deep learning models have achieved promising results but still exhibit notable limitations in breast ultrasound analysis. Convolutional neural networks (CNNs), such as VGG16 [7], primarily operate in the spatial domain and are effective at capturing local patterns, yet they struggle with long-range dependencies and generalization under noisy ultrasound conditions [8]. Transformer-based models, including HoVerTrans [9], improve global context modeling via self-attention but remain limited in representing fine-grained lesion boundaries and disentangling frequency-specific noise in heterogeneous ultrasound images [10–12].

Breast ultrasound images exhibit complementary frequency characteristics: low-frequency components encode global tissue structure and lesion distribution, while high-frequency components emphasize lesion boundaries and fine-grained texture details. Effectively leveraging such information requires explicit frequency modeling and adaptive fusion. To this end, we propose **FreqTrans**, a hybrid framework that integrates spatial and frequency-domain representations in an end-to-end manner [11]. Low-frequency features are used to capture global tissue architecture [12], while high-frequency features highlight lesion margins and subtle structural details [15]. An Adaptive Frequency Fusion (AFF) module dynamically weights low- and high-frequency information to integrate complementary features [16]. A multi-scale feature extraction module further refines frequency-augmented embeddings, followed by a merge module for final classification [17, 18]. Unlike graph-based spectral methods that implicitly learn frequency responses through graph filters, FreqTrans performs explicit frequency-band decomposition in the signal domain, enabling more interpretable and controllable fusion.

By balancing global context and fine-grained details, FreqTrans improves sensitivity to small or low-contrast lesions. Experimental results demonstrate consistent improvements in recall and F1-score compared with representative CNN and transformer baselines on breast ultrasound datasets. To summarize, the main contributions of this work are as follows:

- Enhancing low-contrast lesion visibility: By decomposing images into low- and high-frequency components, FreqTrans improves the representation of blurred tumor boundaries and subtle lesion regions.
- Suppressing noise and reducing frequency redundancy: The proposed frequency-domain attention selectively emphasizes informative channels while suppressing speckle noise and irrelevant spectral components.
- Balancing local and global feature integration: Through adaptive frequency fusion and a multi-scale feature extraction method, FreqTrans effectively combines fine edge details with global tissue context for more robust classification.

2. METHODS

2.1. Overview

The proposed FreqTrans framework is illustrated in Fig. 1. The pipeline consists of four major components: (1) Fourier Preprocessing, which decomposes the input into low- and high-frequency signals; (2) Frequency-Domain Attention, which adaptively enhances spectral channels most relevant for tumor detection; (3) Adaptive Frequency Fusion, which dynamically balances complementary frequency cues; (4) Multi-Scale Feature Extraction with Merge Module, which integrates local lesion edges and global tissue structures for robust classification.

Unlike conventional CNNs or Vision Transformers that process

*Corresponding authors

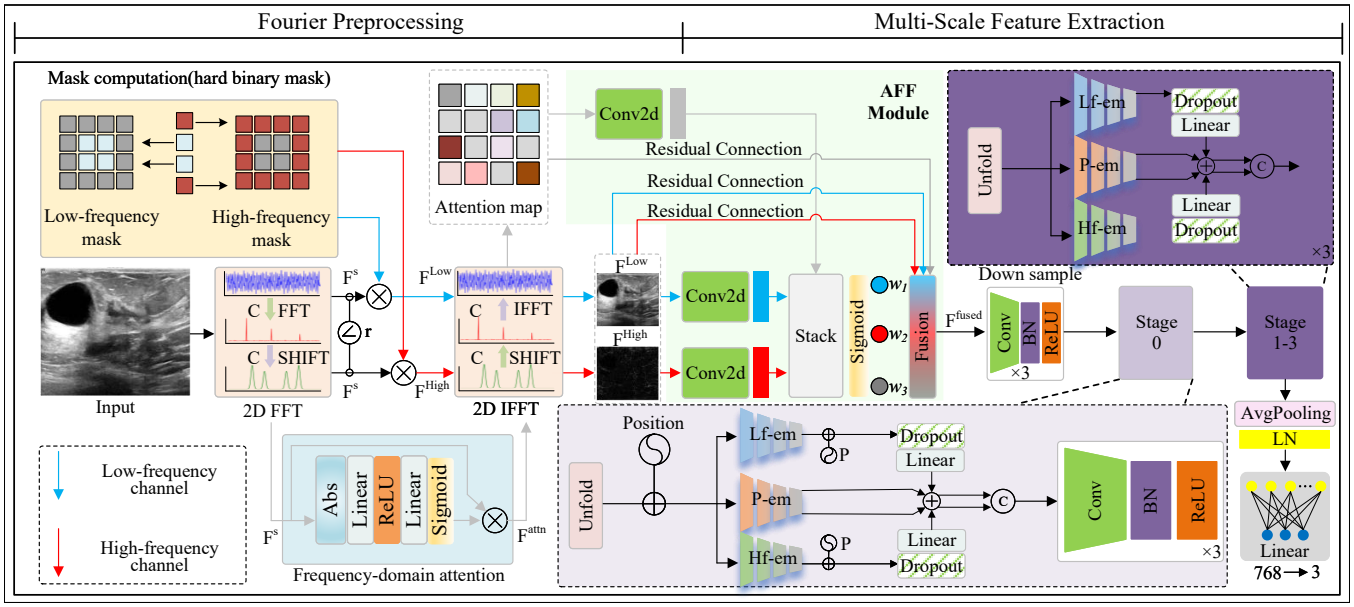

Fig. 1. Overall architecture of the proposed FreqTrans framework. It consists of four main components: (1) Fourier Preprocessing, which decomposes the input image into low- and high-frequency representations; (2) Frequency-Domain Attention, which selectively enhances informative spectral channels; (3) Adaptive Frequency Fusion (AFF), which combines low-frequency, high-frequency, and attention-enhanced features to produce F^{fused}; (4) Multi-Scale Feature Extraction, which constructs low-frequency, high-frequency, and position-aware embeddings and processes them through multi-scale transformer streams to obtain F_{ms}. Finally, F_{fused}, F_{ms} are merged via a convolutional fusion block for classification.

only spatial features, FreqTrans explicitly models frequency-domain information, making it particularly effective for capturing both fine-grained boundaries (from high-frequency channels) and global context (from low-frequency channels).

2.2. Fourier Preprocessing

Breast ultrasound images often contain subtle lesions that are difficult to detect in the raw spatial domain. To alleviate this, we employ a Fourier transform to convert the image into the frequency domain:

$$F^s = \text{fftshift}(\mathscr{F}(x)), \tag{1}$$

where the shift operation centers the low-frequency spectrum.

After frequency shifting, we define the normalized frequency radius $r(u,v) \in [0,1]$ as the Euclidean distance to the spectrum center, where 0 corresponds to the DC component and 1 corresponds to the Nyquist frequency at the spectrum corners. Based on this definition, we construct two hard binary circular masks:

$$M_{low}(u,v) = \mathbb{I}\big(r(u,v) \le r_{low}\big), \quad M_{high}(u,v) = \mathbb{I}\big(r(u,v) \ge r_{high}\big), \tag{2}$$

where $\mathbb{I}(\cdot)$ denotes the indicator function. In our implementation, we set $r_{low} = 0.7$ and $r_{high} = 0.3$, which creates an overlap band $r \in [r_{high}, r_{low}]$ to preserve transitional spectral information between low- and high-frequency components. Both masks are implemented as hard binary masks in the frequency domain.

Applying the masks yields:

$$F_{low} = F^s \odot M_{low}, \quad F_{high} = F^s \odot M_{high}. \tag{3}$$

Inverse FFT recovers spatial maps x_{low} and x_{high}, which explicitly highlight global tissue distribution and fine lesion boundaries, respectively, as shown in Fig. 2. Note that the overlapping frequency components are not manually discarded but are adaptively re-weighted in subsequent fusion modules, allowing the network to learn an appropriate balance between global context and local structural details.

2.3. Frequency-Domain Attention

Although frequency decomposition provides rich information, not all channels contribute equally to lesion classification. Some frequency bands may contain irrelevant noise, while others emphasize diagnostically important structures. Inspired by prior studies on channel attention, we extend this idea to the frequency domain [19]. To address this, we compute per-channel frequency statistics:

$$s_c = \frac{1}{HW} \sum_{i,j} |F^s(c,i,j)|, \tag{4}$$

which measure the overall energy of each channel. These statistics are passed through a small MLP followed by a sigmoid activation to produce channel-wise attention weights $\alpha_c \in [0,1]$. Let $\alpha^{attn} = [\alpha_c]_c$ denote the resulting attention vector.

The frequency attention mechanism enhances discriminative channels (e.g., strong edge responses around tumor boundaries) while suppressing uninformative ones. The enhanced spectrum is:

$$F^{attn}(c,i,j) = \alpha_c F^s(c,i,j), \tag{5}$$

where α_c is broadcast along the spatial dimensions. This module tackles noise interference and frequency redundancy by selectively amplifying informative channels (e.g., edges or calcification signals) while suppressing channels dominated by noise, thus reducing false activations.

2.4. Adaptive Frequency Fusion (AFF)

Low-, high-, and attention-enhanced frequency features provide complementary information for breast ultrasound analysis. Their relative importance varies across lesion types: for example, microcalcifications require stronger high-frequency emphasis, whereas diffuse low-contrast lesions benefit more from low-frequency context [20, 21].

To adaptively integrate these cues, we fuse the three feature streams using learnable weights:

$$F^{fused} = \alpha^{low} \cdot F_{low} + \beta^{high} \cdot F_{high} + F^{attn}. \quad (6)$$

Here, α^{low} and β^{high} are scalar fusion weights learned for each image, which control the relative importance of low- and high-frequency features.

2.5. Multi-Scale Feature Extraction

After adaptive frequency fusion, the network constructs multiple complementary embeddings rather than relying on a single fused representation. Specifically, three types of embeddings are formed: low-frequency embeddings (X^{lf}), high-frequency embeddings (X^{hf}), and position-aware embeddings (X^{p}), which respectively encode global tissue structure, fine lesion boundaries, and spatial context.

All embeddings are projected into patch tokens and processed by S parallel Transformer streams with different receptive fields. For the s-th scale, feature refinement is performed as:

$$X'_s = X_s + \mathrm{MSA}(\mathrm{LN}(X_s)) + \mathrm{FFN}(\mathrm{LN}(X_s)), \quad s = 1, 2, \ldots, S. \quad (7)$$

To integrate information across scales, the outputs of all streams are aggregated via learnable weighted fusion:

$$Z = \sum_{s=1}^{S} \gamma_s \cdot X'_s, \quad \sum_{s=1}^{S} \gamma_s = 1, \quad (8)$$

where γ_s controls the relative contribution of each scale. The aggregated representation Z is reshaped into spatial maps and refined using convolution:

$$F_{ms} = \mathrm{ConvBlock}(Z). \quad (9)$$

This multi-scale design enables the model to jointly preserve global structural context and fine-grained boundary details, producing a unified spatial representation F_{ms} that complements frequency-domain features for robust lesion classification.

2.6. Merge Module

Finally, the outputs from the two main branches are merged to generate the final prediction. The frequency branch provides $Y_1 \equiv F^{fused}$ (adaptive frequency-enhanced representation), while the multi-scale spatial branch provides $Y_2 \equiv F_{ms}$ (multi-scale abstraction from low-, high-frequency, and position-aware embeddings). Both Y_1 and Y_2 are reshaped into spatial feature maps, concatenated, and passed through a convolutional fusion block:

$$F_{out} = \mathrm{ConvBlock}([Y_1, Y_2]). \quad (10)$$

Global average pooling then aggregates the fused representation, and a fully connected layer outputs the classification logits:

$$\hat{y} = \mathrm{FC}(\mathrm{GAP}(F_{out})). \quad (11)$$

This hierarchical merging ensures that frequency-based cues (captured in F^{fused}) and spatial multi-scale abstractions (captured in F_{ms}) complement each other. The final representation thus leverages both enhanced spectral information and refined spatial context, leading to robust lesion classification even under low-contrast or noisy imaging conditions.

2.7. Summary

In summary, FreqTrans combines Fourier decomposition, adaptive spectral enhancement, and multi-scale feature extraction into a unified framework. Each module is tightly connected: Fourier preprocessing disentangles signals, frequency attention enhances them, AFF balances them, and the transformer learns complex interactions. This synergy enables the model to detect subtle and low-contrast lesions more effectively than conventional CNN or ViT baselines.

3. EXPERIMENTS AND ANALYSIS

3.1. Dataset

We evaluate our method on two widely used breast ultrasound datasets: UDIAT [22] and BUSI [23]. UDIAT contains 163 images, and BUSI consists of 647 images, including benign, malignant, and normal cases. For fair evaluation, we adopt a 4-fold cross-validation strategy on both datasets, with all images resized to 256×256. Data augmentation, such as horizontal flipping and random rotation, is applied to improve generalization.

3.2. Implementation Details

All experiments are conducted on an NVIDIA RTX 4070. The proposed FreqTrans framework is trained from scratch without pretraining. We use the Adam optimizer with $\beta = (0.9, 0.999)$. The batch size is set to 32 and the learning rate is 1×10^{-6}. Training proceeds for 250 epochs, and the best model is selected based on validation F1-score. For fair comparison, all baseline CNN and Transformer models are trained under the same settings. The best results are highlighted in bold in the tables. All experiments are repeated five times with different random seeds.

3.3. Overall Comparison

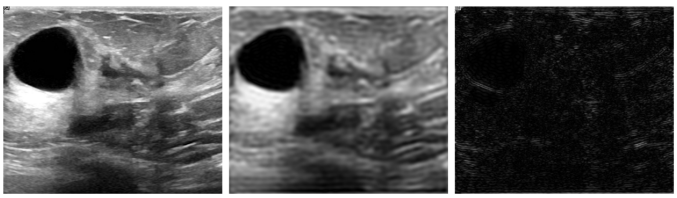

Fig. 2. From left to right: original image, low-frequency representation, and high-frequency representation.

We first compare our proposed FreqTrans with representative CNN- and Transformer-based baselines on two public breast ultrasound datasets: UDIAT and BUSI. As illustrated in Fig. 2, the original image, its low-frequency representation, and its high-frequency representation are presented from left to right. This visualization highlights how frequency decomposition separates global tissue structures from fine lesion boundaries, providing intuitive evidence for the effectiveness of our frequency-domain design and demonstrating why both components are essential for accurate lesion detection.

Tables 1 and 2 present the performance metrics of FreqTrans and baseline methods. Our model consistently achieves higher recall and F1-score, indicating its strength in detecting subtle lesions and maintaining balanced sensitivity and precision.

Table 1. Comparison of different methods on UDIAT. VGG16, ResNet50 [24], EfficientNet-B3 [25], Vision Transformer (ViT), GSM [26], HoVer-Trans, and FABRF-Net [27] are used as baseline methods.

Method	Acc	Auc	Rec	F1-score
VGG16	0.7632	0.8710	0.7400	0.7614
ResNet50	0.7885	0.9067	0.7251	0.7513
ViT	0.7838	0.8730	0.7460	0.7625
EfficientNet-B3	0.7568	0.8946	0.6333	0.7227
GSM	0.8108	0.8612	0.7667	0.7729
HoVer-Trans	0.8000	0.8640	0.7534	0.7682
FABRF-Net	0.8141	**0.9207**	0.7399	0.7921
FreqTrans (Ours)	**0.8158**	0.8783	**0.8274**	**0.8183**

Table 2. Comparison of different methods on BUSI. VGG16, ResNet50, EfficientNet-B3, ViT, GSM, HoVer-Trans, and FABRF-Net are used as baseline methods.

Method	Acc	Auc	Rec	F1-score
VGG16	0.7944	0.8774	0.7677	0.7775
ResNet50	0.7895	0.8952	0.7575	0.7920
ViT	0.8000	0.9306	0.8181	0.7944
EfficientNet-B3	0.8129	**0.9690**	0.6901	0.7681
GSM	0.8258	0.9335	0.7435	0.7845
HoVer-Trans	0.8158	0.8615	0.8261	0.8297
FABRF-Net	0.8158	0.9400	0.7788	0.8179
FreqTrans (Ours)	**0.8462**	0.9156	**0.8274**	**0.8372**

In both datasets, FreqTrans achieves the best recall, which is particularly important for clinical applications, as missing malignant lesions can lead to severe consequences. Compared with recent medical Transformer models (e.g., HoVer-Trans) and frequency-aware architectures (e.g., FABRF-Net), FreqTrans consistently attains higher sensitivity, indicating its effectiveness in capturing subtle and low-contrast lesions. While the AUC does not always increase proportionally, this outcome results from the design of our model. Specifically, FreqTrans emphasizes enhancing high-frequency components that capture subtle lesion boundaries and low-frequency components that highlight global tissue context. This design inherently prioritizes sensitivity to small or low-contrast lesions, leading the model to predict more positive cases. Although this improves recall, it may also introduce additional false positives, which in turn reduces specificity and slightly lowers the overall AUC.

Such a trade-off between sensitivity and AUC has been widely discussed in prior medical imaging studies. In highly imbalanced datasets, or when the clinical priority is to avoid missing any positive cases, models often deliberately bias toward high recall at the expense of precision [28]. As several works have noted, ROC-based AUC can remain relatively stable even when false positives increase, whereas precision–recall metrics provide a more sensitive reflection of performance under class imbalance [28, 29]. Consequently, our results are consistent with this phenomenon: FreqTrans achieves higher sensitivity by capturing subtle lesion cues, while a moderate decrease in specificity slightly lowers AUC but aligns with the overarching clinical objective of minimizing false negatives.

3.4. Ablation Study

To evaluate the contribution of each module in the proposed framework, we conduct ablation experiments on the BUSI dataset. We compare the following variants: (1) backbone only; (2) backbone with Fourier Preprocessing (FP) [30, 31]; (3) backbone with FP and Adaptive Frequency Fusion (AFF); (4) backbone with FP, AFF, and Frequency Attention (FA) [32]; and (5) the complete FreqTrans model. The results indicate that each component contributes positively: FP improves representation by introducing explicit frequency decomposition, AFF further enhances recall by balancing low- and high-frequency cues, and FA emphasizes discriminative spectral regions. The full model achieves the best overall performance, demonstrating the complementary effects of all modules.

Table 3. Ablation study on Dataset BUSI.

Method	Acc	Rec	F1-score
Baseline (Backbone only)	0.7483	0.6944	0.6944
+ FP	0.7631	0.7231	0.7588
+ FP + AFF	0.8000	0.7534	0.7682
+ FP+ AFF +FA	0.8102	0.7765	0.7873
+ Full Model (Ours)	**0.8205**	**0.8274**	**0.8183**

The results in Table 3 show a clear and consistent performance improvement as each module is progressively added. Fourier Preprocessing (FP) enhances feature representation by introducing explicit frequency decomposition. Adaptive Frequency Fusion (AFF) further improves recall by balancing global low-frequency context and fine-grained high-frequency details. Frequency Attention (FA) provides additional gains by emphasizing discriminative spectral components and suppressing noise-dominated channels. Overall, the full FreqTrans model achieves the best performance, indicating that FP, AFF, and FA contribute complementary benefits and collectively validate the design rationale of the proposed framework.

4. CONCLUSION

In this paper, we proposed FreqTrans, a frequency-aware transformer framework that explicitly integrates multi-scale spatial features with frequency-domain representations for breast ultrasound image classification. By decomposing images into low- and high-frequency components, the proposed framework jointly captures global tissue context and fine-grained lesion details. The Adaptive Frequency Fusion (AFF) and Frequency Attention (FA) modules further facilitate effective integration of complementary spectral information.

Experiments on two public datasets (UDIAT and BUSI) demonstrate that FreqTrans achieves consistent improvements in recall and F1-score compared with representative CNN and vision transformer baselines, which is particularly important for detecting subtle and low-contrast lesions in clinical screening scenarios. Although some baselines obtain higher AUC in specific cases, FreqTrans provides a favorable trade-off between sensitivity and precision, indicating robust performance under diverse imaging conditions.

In summary, the proposed frequency-domain design offers an effective strategy for integrating global spectral context with local spatial cues in medical image analysis. Future work will explore extending FreqTrans to larger-scale datasets and other medical imaging modalities, as well as investigating its integration with hierarchical medical transformers (e.g., Swin-UNETR) to combine frequency-domain modeling with strong spatial representations.

5. REFERENCES

[1] J. G. Elmore, S. M. Reisch, L. Carney, K. Abraham, A. B. Fosse, B. M. Longton, and C. D. Nelson, "Diagnostic Concordance Among Pathologists Interpreting Breast Biopsy Specimens," *JAMA*, vol. 313, no. 11, pp. 1122–1132, 2015.

[2] Q. Huang, Y. Luo, and C. Zhang, "Breast Ultrasound Image Segmentation: A Survey," *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, pp. 197–208, 2020.

[3] H. R. Koo, Y. Cho, J. Song, and H. Park, "Noise Reduction in Breast Ultrasound Imaging: A Review," *Ultrasonography*, vol. 38, no. 3, pp. 244–252, 2019.

[4] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in *International Conference on Learning Representations (ICLR)*, 2021.

[5] X. Chen, S. Xie, and K. He, "An Empirical Study of Training Self-Supervised Vision Transformers," in *International Conference on Computer Vision (ICCV)*, pp. 9640–9649, 2021.

[6] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, "A Survey on Deep Learning in Medical Image Analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[7] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *International Conference on Learning Representations (ICLR)*, 2015.

[8] S. Wang, Y. Zhang, and X. Li, "Noise-Robustness Test for Ultrasound Breast Nodule Neural Network," *IEEE Transactions on Medical Imaging*, vol. 42, no. 5, pp. 1301–1312, 2023.

[9] Yuhao Mo, Chu Han, Yu Liu, Min Liu, Zhenwei Shi, Jiatai Lin, Bingchao Zhao, Chunwang Huang, Bingjiang Qiu, Yanfen Cui, Lei Wu, Xipeng Pan, Zeyan Xu, Xiaomei Huang, Zaiyi Liu, Ying Wang, Changhong Liang, "HoVer-Trans: Anatomy-aware HoVer-Transformer for ROI-free Breast Cancer Diagnosis in Ultrasound Images," *IEEE Transactions on Medical Imaging*, vol. 42, no. 5, pp. 1696–1706, 2023.

[10] Y. Feng, H. Zhang, X. Li, and J. Wang, "EH-former: Regional easy-hard-aware transformer for breast lesion segmentation," *Expert Systems with Applications*, vol. 202, p. 117254, 2024.

[11] F. Liu, M. Tan, and J. Zhou, "Ultrasound Signal Processing: From Models to Deep Learning," *Ultrasonics*, vol. 122, p. 106768, 2022.

[12] Y. Li, X. Zhang, and J. Wang, "A survey on deep learning in medical ultrasound imaging," *Frontiers in Physics*, vol. 12, p. 1398393, 2024.

[13] R. Acharya, O. Faust, S. Molinari, L. S. Sree, and J. S. Suri, "Nonlinear Analysis of Ultrasound Images for Breast Cancer Diagnosis," *Medical and Biological Engineering and Computing*, vol. 46, no. 12, pp. 999–1009, 2008.

[14] Y. Xu, T. Mo, Q. Feng, P. Zhong, M. Lai, and E. I.-C. Chang, "Deep Learning of Feature Representation with Multiple Instance Learning for Medical Image Analysis," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 12, pp. 2794–2805, 2017.

[15] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual Blur and Ringing Metrics: Application to JPEG2000," *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 163–172, 2004.

[16] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 3–19, 2018.

[17] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 10012–10022, 2021.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[19] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7132–7141, 2018.

[20] Y. Ouyang, Z. Zhou, W. Wu, J. Tian, F. Xu, S. Wu, and P.-H. Tsui, "A review of ultrasound detection methods for breast microcalcification," *Mathematical Biosciences and Engineering*, vol. 16, no. 4, pp. 1761–1785, 2019.

[21] K. Frick, T. M. L. van der, A. J. M. Nooijen-Parker, et al., "Preoperative localization of breast microcalcification using high-frequency ultrasound," *Clinical Radiology*, vol. 52, no. 12, pp. 924-926, 1997.

[22] A. Oliver, J. Freixenet, J. Martí, E. Pérez, J. Pont, E. R. Denton, and R. Zwiggelaar, "A Novel Breast Tissue Density Classification Methodology," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 1, pp. 55–65, 2008.

[23] M. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of Breast Ultrasound Images," *Data in Brief*, vol. 28, p. 104863, 2020.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[25] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 6105–6114, 2019.

[26] Y. Zhang, X. Li, and S. Wang, "GSM: Global Similarity Modeling for Medical Image Classification," *IEEE Transactions on Medical Imaging*, vol. 41, no. 6, pp. 1452–1463, 2022.

[27] Y. Liu, H. Zhang, X. Li, and J. Wang, "FABRF-Net: A Frequency-Aware Boundary and Region Fusion Network for Breast Ultrasound Image Segmentation," *Expert Systems with Applications*, vol. 236, p. 121142, 2024.

[28] Saito, T. and Rehmsmeier, M., "The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets," *PLOS ONE*, vol. 10, no. 3, e0118432, 2015.

[29] Lever, J., Krzywinski, M., and Altman, N., "Classification Evaluation: Part 2 — ROC, Precision-Recall and Imbalanced Data," *Nature Methods*, vol. 13, pp. 132–133, 2016.

[30] Y. Chen, H. Li, Z. Wang, and X. Liu, "Syn-Net: A Synchronous Frequency-Perception Fusion Network for Breast Tumor Segmentation in Ultrasound Images," *Computers in Biology and Medicine*, vol. 165, p. 107393, 2023.

[31] S. Ying, X. Song, and H. Wang, "High-frequency-based Multispectral Attention for Domain Generalization," *Artificial Intelligence Review*, pp. 1–23, 2025.

[32] M. Zhang, L. Zhao, and Q. Huang, "FA-Unet: A Deep Learning Method with Fusion of Frequency Domain Features for Fruit Leaf Disease Identification," *Horticulturae*, vol. 11, no. 7, p. 783, 2023.