Segmentation-Prior Guided Multimodal Learning for Breast Ultrasound Diagnosis

Yuhan Shang*

Southeast University, Nanjing, China

Abstract. Deep learning shows promise for breast ultrasound segmentation and classification, yet modality heterogeneity and scarce annotations often limit performance. We propose Segmentation-Prior Guided Multimodal Learning for Breast Ultrasound Diagnosis, a unified framework for joint lesion segmentation and malignancy classification. An optimized SAM-based SAMUS generates multimodal lesion masks as spatial priors. Guided by these masks, GAB extracts lesion-specific features, while a hybrid fusion module integrates cross-modal multi-head attention for semantic alignment with a dynamic weight generator for sample-adaptive modality weighting. Experiments demonstrate consistent gains, achieving Dice scores of 82.13%/80.98%/73.84% (B-mode/SE/CEUS) and 93.62% classification accuracy.

1 Introduction

Ultrasound (US) is widely adopted for breast screening because it is non-ionizing, portable, and cost-effective[1]. In clinical practice, B-mode images support benign–malignant assessment via morphology, echogenicity, and margin characteristics, yet these cues can be subtle or contradictory in difficult cases, increasing the risk of misdiagnosis. Therefore, B-mode is frequently complemented with strain elastography (SE) and contrast-enhanced ultrasound (CEUS) to provide additional functional evidence[2].

SE offers semi-quantitative characterization of tissue stiffness: benign lesions are typically only slightly stiffer than surrounding tissue, whereas malignant lesions often exhibit higher stiffness and complex boundary mechanics[3]. CEUS visualizes microvascular perfusion: benign lesions usually enhance within the B-mode boundary with relatively clear margins, while malignant lesions often show irregular, heterogeneous enhancement extending beyond the B-mode contour with blurred borders[4]. These modalities are intrinsically complementary—B-mode depicts structure, SE reflects stiffness, and CEUS captures perfusion—making automated multimodal US analysis clinically valuable.

Recent advances in deep learning (DL) have substantially improved breast US segmentation and classification, often surpassing traditional CAD in sensitivity, specificity, and accuracy while reducing radiologists' workload[5]. Early studies predominantly focused on single-modality learning, such as B-mode–based texture/morphology classification[6] or SWE-based stiffness quantification[7]. However, single-modality models cannot fully exploit cross-modal complementarity and frequently overlook practical issues in multimodal settings, including spatial misalignment, redundant information, and unbalanced modality contributions.

To address these limitations, multimodal ultrasound learning has rapidly progressed. Combining B-mode with functional modalities (e.g., CEUS, SWE/SE, Doppler) has been shown to improve benign–malignant discrimination[8, 9]. More advanced fusion designs attempt to model modality importance and alignment: AW3M learns modality weights and compensates for missing modalities across multiple US sources[10], while FAMF-Net introduces region-aware alignment and mutual-attention fusion to better integrate B-mode and SWE features with improved interpretability[11]. Despite these advances, key challenges remain: (i) robust intra-modality multi-scale aggregation under speckle noise and low contrast; (ii) insufficient use of ROI mask priors and region-level constraints, limiting interpretability; (iii) reliance on naive concatenation or static weighting without explicitly modeling modality reliability.

To tackle these challenges, we propose an integrated tri-modal framework for breast ultrasound (B-mode/SE/CEUS), unifying segmentation guidance, global–local feature modeling, cross-modality fusion. Our main contributions are:

Segmentation-prior tri-modal pipeline: We employ SAMUS[12] in three cooperative branches to generate reliable ROI masks as spatial priors for lesion-focused feature learning.

Mask-guided multi-scale aggregation and B-mode–driven adaptive fusion: We redesign GAB[13] as a modality-specific, mask-guided multi-scale aggregator to enhance lesion representations and suppress background. Based on aligned modality tokens, we introduce a B-mode-as-Query attention module for cross-modal fusion, together with a dynamic weight generator that produces sample-specific modality weights for interpretable integration.

*Corresponding author: 1418815675@qq.com

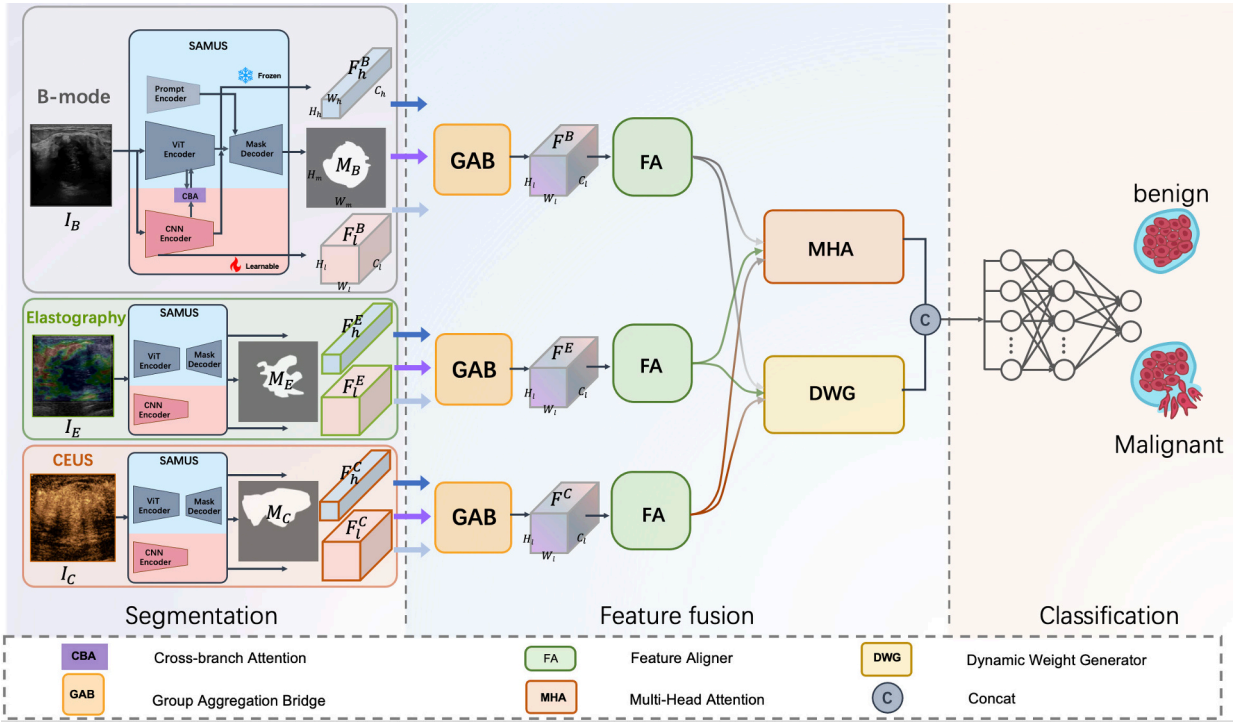

Figure 1. Overview of the proposed tri-modal breast ultrasound segmentation-guided classification framework. It consists of three parallel segmentation branches for B-mode, SE, and CEUS; intra-modal feature aggregation and cross-modal fusion; hybrid fusion for malignancy prediction.

2 Related work

2.1. Baselines and handcrafted features

Early breast ultrasound CAD systems typically follow BI-RADS–driven handcrafted pipelines. Although interpretable, these approaches are highly operator- and device-dependent. Prior work enhanced this paradigm by combining quantitative BI-RADS/morphological descriptors with SVM/RF or nomogram models, and by extending handcrafted features to multimodal settings to improve AUC[14-16]. Nonetheless, predefined feature spaces and mostly linear fusion limit their ability to capture high-order cross-modal interactions.

2.2 Deep learning for single-modal ultrasound

With increased computing power, breast ultrasound research has shifted to end-to-end deep networks. For segmentation, U-Net variants remain a strong baseline; representative designs incorporate residual blocks, depthwise separable convolutions, and attention to better delineate fuzzy boundaries and diverse morphologies[17, 18]. For classification, transfer learning with modern backbones and attention/feature-selection strategies has improved performance and interpretability[6, 19]. Despite notable progress, single-modality models still suffer from limited data, modality bias, weaker cross-device generalization.

2.3 Adapting general pretrained models to ultrasound

The Segment Anything Model (SAM)[20] has accelerated research on promptable, transferable segmentation and has been adapted to medical imaging. For breast ultrasound, BUSSAM[21] modifies SAM with an auxiliary CNN branch and cross-branch attention. MedSAM[22] further promotes SAM-style foundations for broad medical modalities. SAMUS introduces a parallel CNN pathway and cross-branch attention for stronger local-detail modeling[12]. However, most SAM-based adaptations still focus on single-modality scenarios and rarely tackle clinically critical multimodal fusion.

2.4. Multimodal ultrasound fusion learning

As multifunctional ultrasound systems become routine, multimodal fusion has gained increasing attention. Prior studies have shown that combining B-mode with functional modalities can improve segmentation and diagnosis. Misra et al. integrated B-mode and SE in a deep fusion framework for joint segmentation and classification[23]. TMAN employs hierarchical morphological attention across B-mode, elastography, and color Doppler to enhance robustness[24]. AW3M introduces auto-weighting and missing-modality recovery across B-mode, Doppler, SWE, and SE[10], while FAMF-Net combines region-aware alignment with mutual-attention fusion to better integrate B-mode and elastography features[11]. Despite these advances, practical deployment remains challenged by spatial/temporal misalignment.

3 Method

The pipeline comprises three stages: (i) modality-specific encoding and segmentation-prior generation, (ii) intra-modal aggregation and cross-modal fusion, and (iii) feature fusion for malignancy prediction, as shown in Figure 1. All images are resampled to 256×256. To match SAMUS normalization, grayscale B-mode and CEUS are replicated to three channels, while SE is kept as three-channel pseudo-color.

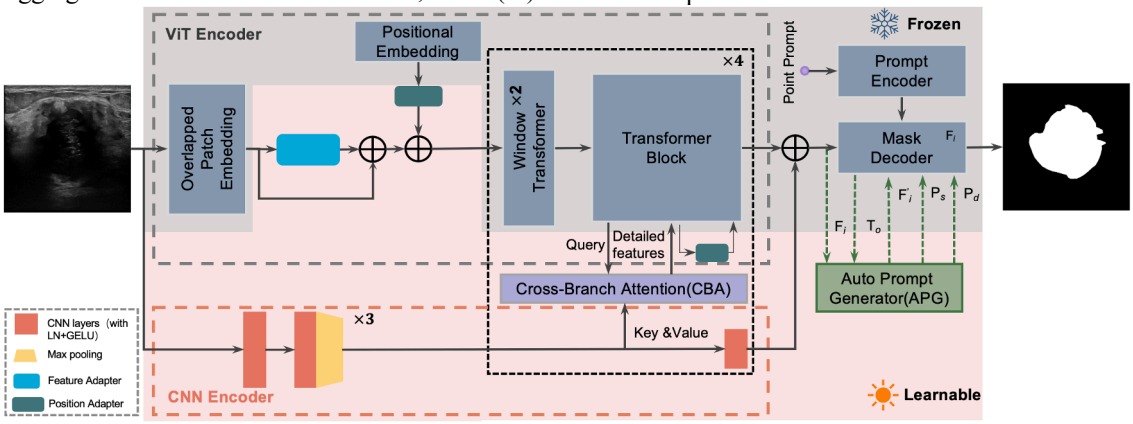

Figure 2. The architecture is adapted from SAMUS, consisting of a ViT encoder and a lightweight CNN encoder connected via a Cross-Branch Attention (CBA) mechanism. Each modality (B-mode, SE, CEUS) employs an independent copy of this module to generate its corresponding lesion ROI mask.

Three independent modality branches preserve modality-specific representations while maintaining sample-level correspondence for downstream fusion.

3.1 Segmentation prior with multi-branch SAMUS

To provide stable and interpretable ROI priors across B-mode, SE, and CEUS, we extend SAMUS into a tri-branch segmentation backbone (Figure. 2). Each modality employs an independent copy of the SAMUS-based module (no parameter sharing), ensuring modality-aware feature learning without mutual interference. The original SAM is prompt-driven and trained on natural images, and its performance can degrade on ultrasound due to speckle noise, low contrast, and ambiguous boundaries. SAMUS is specifically tailored to ultrasound by introducing a parallel CNN stream to capture local texture cues and injecting them into the ViT stream via Cross-Branch Attention, together with lightweight adapter designs that reduce adaptation complexity; AutoSAMUS further incorporates an Automatic Prompt Generator to enable end-to-end automatic segmentation without manual interaction.

This dual-stream design explicitly exploits semantic–texture complementarity: the ViT path captures global, noise-robust semantics, while the CNN path preserves local structures and speckle textures, improving robustness to low contrast and blurred boundaries. For each modality $m \in \{B, E, C\}$, the branch outputs a lesion mask $\boldsymbol{M_m} \in [0,1]^{1\times H_m \times W_m}$ as a spatial prior, together with intermediate features $\boldsymbol{F_h^m} \in R^{C_h \times H_h \times W_h}$ (ViT semantics) and $\boldsymbol{F_l^m} \in R^{C_l \times H_l \times W_l}$ (CNN textures). We retain the original SAMUS architecture and loss design to ensure stable convergence, while extending it into a modality-separated multi-branch prior generator that better supports subsequent cross-modal alignment and fusion.

3.2 Mask-guided intra-modal fusion

After obtaining modality-wise ROI priors in Sec. 3.1, we perform intra-modal semantic–texture aggregation to derive ROI-enhanced representations for each modality. Specifically, given the mask $\boldsymbol{M_m}$, we downsample it to the feature resolution and use a mask-guided Group Aggregation Bridge (GAB) [13] to aggregate complementary semantics and textures (Figure 3).

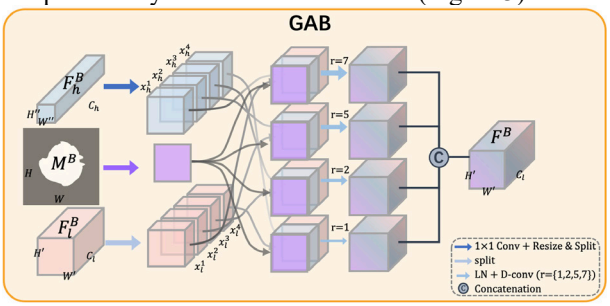

Figure 3. Structure of the Group Aggregation Bridge (GAB) module. Low-level CNN features, high-level ViT features, and the segmentation mask are aligned and split into four groups processed by dilated convolutions with rates {1, 2, 5, 7}.

Let $\boldsymbol{F_h} \in R^{C_h \times H_h \times W_h}$ be the ViT semantic features and $\boldsymbol{F_l} \in R^{C_l \times H_l \times W_l}$ be the CNN shallow features. We first align $\boldsymbol{F_h}$ to $\boldsymbol{F_l}$ in channel and spatial resolution:

$$\widetilde{\boldsymbol{F}}_{\boldsymbol{h}} = \mathrm{Upsample}(\mathrm{Conv}_{1\times1}(\boldsymbol{F_h})), \widetilde{\boldsymbol{F}}_{\boldsymbol{l}} = \boldsymbol{F_l}. \quad (1)$$

To efficiently capture multi-scale context, $\widetilde{\boldsymbol{F}}_{\boldsymbol{h}}$ and $\widetilde{\boldsymbol{F}}_{\boldsymbol{l}}$ are split into N=4 channel groups. For each group i, we concatenate the corresponding features with the mask and apply a dilation-specific branch:

$$Y_i = \mathcal{G}_i\left([\widetilde{\boldsymbol{F}}_{\boldsymbol{h},i}; \widetilde{\boldsymbol{F}}_{\boldsymbol{l},i}; M]\right), \quad (2)$$

where $\mathcal{G}_i(Z) = \mathrm{Conv}_{3\times3}^{d_i}(LN(Z))$ and $d_i = [1,2,5,7]$ in our implementation. The outputs are concatenated and fused:

$$Y_{out} = Conv_{1\times1}(LN[Y_0; Y_1; Y_2; Y_3]), \quad (3)$$

yielding an ROI-enhanced feature map. Compared with the original pixel-level usage of GAB in segmentation networks, here it serves as a modality-specific feature aggregator guided by the lesion mask, producing an ROI-enhanced feature map for subsequent cross-modal fusion.

Table 1. Quantitative Comparison of Tumor Segmentation Performance on the Dataset.

Network	B-mode		SE		CEUS	
	Dice-B-mode (%)↑	IoU-B-mode (%)↑	Dice- SE (%)↑	IoU-SE (%)↑	Dice- CEUS (%)↑	IoU- CEUS (%)↑
UNet	68.53	56.88	61.82	49.85	42.06	31.88
TransUNet	82.00	71.43	78.32	68.65	73.18	**62.26**
BUSSAM	71.94	61.16	57.48	46.11	56.97	47.73
MedSAM	80.54	68.97	75.32	63.37	65.63	51.93
SAMUS	80.79	70.18	77.58	66.16	65.04	60.97
Ours	**82.13**	**71.98**	**80.98**	**70.20**	**73.84**	61.72

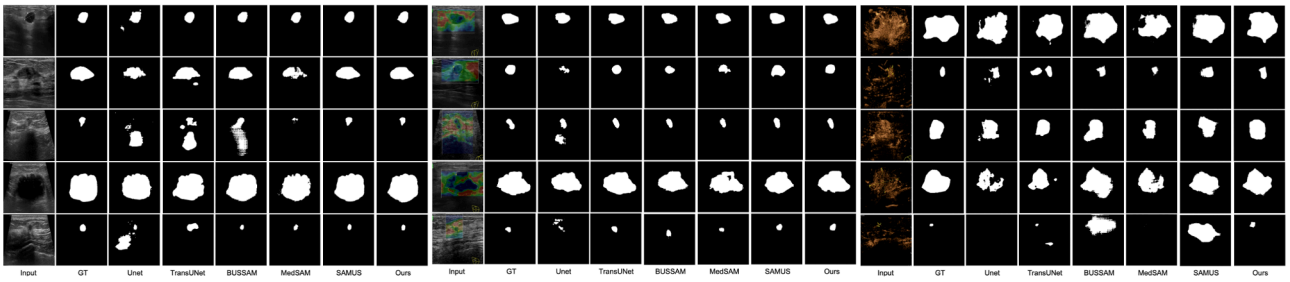

Figure 4. Structure of the Group Aggregation Bridge (GAB) module. Low-level CNN features, high-level ViT features, and the segmentation mask are aligned and split into four groups processed by dilated convolutions with rates {1, 2, 5, 7}

3.3 Cross-modal alignment and fusion

To mitigate cross-modality misalignment and case-dependent modality reliability, we design a Cross-Modal Alignment and Fusion (CMAF) module with two parallel streams: B-mode–guided Multi-Head Cross-Attention(MHA) for semantic alignment and a Dynamic Weight Generator (DWG) for sample-adaptive modality selection. The CMAF module is applied to each branch (Global/ROI) independently (with separate parameters).

B-mode provides the anatomical anchor, while SE/CEUS contribute functional cues. For a given branch, let z_b denote the B-mode embedding and $Z \in R^{3\times D}$ be the stacked modality embeddings. We define:

$$\mathrm{Q}= \mathrm{z}_b, K = V = Stack(\mathrm{z}_b, \mathrm{z}_e, \mathrm{z}_c). \quad (4)$$

Using z_b as the query encourages the model to retrieve functional evidence consistent with anatomical structure. The aligned representation is:

$$\mathrm{z}_{attn} = \mathrm{Softmax}(\frac{QK^T}{\sqrt{D}})V + \mathrm{z}_b, \quad (5)$$

where the residual term preserves anatomical information and stabilizes optimization.

Since modality quality varies across cases, we further introduce a Dynamic Weight Generator (DWG) to assign sample-adaptive modality importance. A lightweight MLP predicts modality logits:

$$I = MLP_{dwg}(Concat[z_b, z_e, z_c]) \in R^3, \quad (6)$$

The final normalized weights w are obtained via Softmax:

$$w = Softmax(I) = [w_b, w_e, w_c]. \quad (7)$$

The weighted fusion embedding is then:

$$z_{fuse} = w_b \cdot z_b + w_e \cdot z_e + w_c \cdot z_c. \quad (8)$$

The final representation is a balanced combination of z_{attn} and z_{fuse}. This yields a robust fused representation that is both anatomically anchored and adaptively weighted for downstream classification.

4 Result

4.1 Comparison with segmentation networks

To evaluate tumor segmentation performance, we compared our method with representative baselines, including UNet, TransUNet, BUSSAM, MedSAM, and SAMUS, under identical training/testing settings on the B-mode, SE, and CEUS datasets. Dice, IoU, and Hausdorff Distance (HD) were used for evaluation, and the results are reported in Table 1.

Overall, our method achieves the best overlap-based performance across modalities. On B-mode, it obtains 82.13% Dice and 71.98% IoU, slightly higher than TransUNet (82.00%/71.43%) and better than SAMUS (80.79%/70.18%). On SE, it delivers a clear gain with 80.98% Dice and 70.20% IoU, outperforming TransUNet and SAMUS. On the most challenging CEUS modality, our model reaches 73.84% Dice, surpassing TransUNet (73.18%) and markedly improving over SAMUS

(65.04%). The CEUS IoU (61.72%) is comparable to TransUNet (62.26%) while substantially exceeding MedSAM (51.93%) and SAMUS (60.97%). Averaged across modalities, our approach achieves 78.98% mean Dice and 67.97% mean IoU, consistently outperforming TransUNet (77.83%/67.45%) and SAMUS (74.47%/65.77%).

Table 2. Quantitative Comparison of Tumor Classification Performance on the Dataset.

Model	Accuracy	Precision	Recall	F1-Score	AUC	Specificity
ResNet-50	0.7079	0.7517	0.6258	0.6718	0.7740	0.7875
VGG-16	0.6794	0.7970	0.4903	0.5974	0.7590	0.8625
DenseNet-121	0.7016	0.8348	0.5161	0.6260	0.8272	0.8812
EfficientNet	0.70797	0.8097	0.5419	0.6420	0.7940	0.8688
ViT	0.7556	0.8373	0.6387	0.7128	0.8454	0.8687
Long et al.2024	0.8677	0.8000	**0.9231**	0.8571	0.8281	0.8235
Ours	**0.9362**	**0.9444**	*0.8947*	**0.9189**	**0.9605**	**0.9643**

Table 3. Ablation Study. GAB: Group Aggregation Bridge; MHA: Multi-Head Attention; DWG: Dynamic Weight Generator

Network Module			Evaluation Index					
GAB	MHA	DWG	Accuracy	Precision	Recall	F1-Score	AUC	Specificity
√			0.8511	0.8333	0.7895	0.8108	0.8835	0.8929
	√		0.8298	0.7895	0.7895	0.7895	0.8346	0.8571
		√	0.8298	0.8235	0.7368	0.7778	0.9135	0.8929
√	√		0.9060	0.8947	0.8947	0.8947	0.9060	0.9286
√		√	0.8723	0.7600	1.0000	0.8636	0.9380	0.7857
	√	√	0.8723	0.8421	0.8421	0.8421	0.9192	0.8929
√	√	√	**0.9362**	**0.9444**	**0.8947**	**0.9189**	**0.9605**	**0.9643**

Figure 4 presents qualitative comparisons on representative cases. On B-mode, U-Net/TransUNet often yield blurred or incomplete contours, while SAM-based methods improve localization but may suffer from boundary breaks; our method produces smoother and more continuous outlines. On SE, where heterogeneous stiffness and weak margins hinder delineation, CNN/Transformer baselines frequently generate fragmented masks or leakage, whereas our results are more complete and anatomically plausible. On CEUS, existing approaches commonly fail under heterogeneous enhancement and speckle noise, producing irregular or discontinuous masks; in contrast, our predictions better preserve lesion integrity with fewer broken components and improved extent alignment. Overall, the segmentation branch provides robust ROI masks across modalities, supporting downstream fusion and classification.

4.2 Comparison with classification networks

To assess the malignancy discrimination ability of the model, we compared it with representative single-modality classifiers (ResNet-50, VGG-16, DenseNet-121, EfficientNet, and ViT) and the two-stage framework of Long et al., under identical training and evaluation protocols. The results are summarized in Table 2.

Overall, our model achieves the best and most balanced performance. It attains an accuracy of 0.9362 and an F1-score of 0.9189, outperforming the strongest competing method (Long et al. [25]). Notably, it achieves the highest AUC of 0.9605, substantially exceeding the best generic baseline (ViT, 0.8454), indicating markedly improved benign–malignant separability. In addition, our model yields the best precision (0.9444) and specificity (0.9643), demonstrating strong false-positive suppression, which is particularly important for clinical triage.

4.3 Ablation study

To quantify the contribution of each core component in our model, we performed stepwise ablations on GAB, MHA, and DWG (Table 3). With only one module enabled, performance is limited: GAB-only reaches 0.8511 accuracy and 0.8835 AUC, showing that intra-modal aggregation alone cannot resolve cross-modal inconsistency; MHA-only performs worst (0.8298/0.8346), indicating attention fusion is noise-sensitive without strong feature extraction; DWG-only improves separability (AUC 0.9135) but yields low accuracy/F1, suggesting weighting requires discriminative representations.

Combining modules consistently improves results. GAB+MHA provides a strong and stable baseline (0.9060 accuracy, 0.8947 F1). GAB+DWG achieves higher recall but at the cost of precision and specificity, while MHA+DWG improves over either alone yet remains inferior to GAB-based variants, highlighting the foundational role of intra-modal aggregation.

Overall, the full model GAB+MHA+DWG performs best, achieving 0.9362 accuracy, 0.9189 F1, and 0.9605 AUC, with the highest precision (0.9444) and specificity (0.9643).

5 Conclusion

This study presents A Multi-Modal Segmentation–Classification Network for breast ultrasound. The model integrates B-mode, SE, and CEUS to enable mask-guided

fusion and multi-task learning, consistently outperforming competing methods in both segmentation and diagnosis.

Several limitations remain. The framework depends on predicted ROI masks, and CEUS is simplified by using a single peak-TIC frame to fit a 2D setting, which may discard temporal perfusion dynamics. In addition, experiments are conducted on a single-center dataset with limited size and relatively homogeneous devices and acquisition protocols, which may restrict generalizability.

Future work will extend evaluation to larger multi-center cohorts and investigate tighter end-to-end coupling so that segmentation and classification can be jointly optimized within a shared representation space.

References

1. Dan, Q., et al., *Ultrasound for Breast Cancer Screening in Resource-Limited Settings: Current Practice and Future Directions.* Cancers (Basel), 2023. **15**(7).
2. Guo, W., et al., *Non-mass Breast Lesions: Could Multimodal Ultrasound Imaging Be Helpful for Their Diagnosis?* Diagnostics (Basel), 2022. **12**(12).
3. Shahzad, R., et al., *Diagnostic value of strain elastography and shear wave elastography in differentiating benign and malignant breast lesions.* Ann Saudi Med, 2022. **42**(5): p. 319-326.
4. Boca Bene, I., S.M. Dudea, and A.I. Ciurea, *Contrast-Enhanced Ultrasonography in the Diagnosis and Treatment Modulation of Breast Cancer.* J Pers Med, 2021. **11**(2).
5. Dan, Q., et al., *Diagnostic performance of deep learning in ultrasound diagnosis of breast cancer: a systematic review.* NPJ Precis Oncol, 2024. **8**(1): p. 21.
6. Jabeen, K., et al., *Breast Cancer Classification from Ultrasound Images Using Probability-Based Optimal Deep Learning Feature Fusion.* Sensors (Basel), 2022. **22**(3).
7. Tang, Y., et al., *Machine learning-based diagnostic evaluation of shear-wave elastography in BI-RADS category 4 breast cancer screening: a multicenter, retrospective study.* Quant Imaging Med Surg, 2022. **12**(2): p. 1223-1234.
8. Li, S.Y., et al., *Determining whether the diagnostic value of B-ultrasound combined with contrast-enhanced ultrasound and shear wave elastography in breast mass-like and non-mass-like lesions differs: a diagnostic test.* Gland Surg, 2023. **12**(2): p. 282-296.
9. Qian, X., et al., *Prospective assessment of breast cancer risk from multimodal multiview ultrasound images via clinically applicable deep learning.* Nat Biomed Eng, 2021. **5**(6): p. 522-532.
10. Huang, R., et al., *AW3M: An auto-weighting and recovery framework for breast cancer diagnosis using multi-modal ultrasound.* Med Image Anal, 2021. **72**: p. 102137.
11. Liu, Y., et al., *FAMF-Net: Feature Alignment Mutual Attention Fusion With Region Awareness for Breast Cancer Diagnosis via Imbalanced Data.* IEEE Trans Med Imaging, 2025. **44**(3): p. 1153-1167.
12. Lin, X., et al. *Beyond adapting SAM: Towards end-to-end ultrasound image segmentation via auto prompting*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2024. Springer.
13. Ruan, J., et al. *EGE-UNet: An Efficient Group Enhanced UNet for Skin Lesion Segmentation*. 2023. Cham: Springer Nature Switzerland.
14. Shan, J., et al., *Computer-Aided Diagnosis for Breast Ultrasound Using Computerized BI-RADS Features and Machine Learning Methods.* Ultrasound Med Biol, 2016. **42**(4): p. 980-8.
15. Yang, K., et al., *Development and validation of a nomogram for discriminating between benign and malignant breast masses by conventional ultrasound and dual-mode elastography: a multicenter study.* Quant Imaging Med Surg, 2023. **13**(2): p. 865-877.
16. Moustafa, A.F., et al., *Color Doppler Ultrasound Improves Machine Learning Diagnosis of Breast Cancer.* Diagnostics (Basel), 2020. **10**(9).
17. Yuan, S., et al., *Rmau-net: Breast tumor segmentation network based on residual depthwise separable convolution and multiscale channel attention gates.* Applied Sciences, 2023. **13**(20): p. 11362.
18. Pramanik, P., et al., *DAU-Net: Dual attention-aided U-Net for segmenting tumor in breast ultrasound images.* PLoS One, 2024. **19**(5): p. e0303670.
19. Latha, M., et al., *Revolutionizing breast ultrasound diagnostics with EfficientNet-B7 and Explainable AI.* BMC Med Imaging, 2024. **24**(1): p. 230.
20. Kirillov, A., et al. *Segment anything*. in *Proceedings of the IEEE/CVF international conference on computer vision*. 2023.
21. Tu, Z., et al., *Ultrasound sam adapter: Adapting sam for breast lesion segmentation in ultrasound images.* arXiv preprint arXiv:2404.14837, 2024.
22. Ma, J., et al., *Segment anything in medical images.* Nat Commun, 2024. **15**(1): p. 654.
23. Misra, S., et al., *Ensemble transfer learning of elastography and B-mode breast ultrasound images.* arXiv preprint arXiv:2102.08567, 2021.
24. Wang, D., M. Xue, and H. Wang, *TMAN: A Triple Morphological Feature Attention Network for Fine-Grained Classification of Breast Ultrasound Images.* Journal of Imaging Informatics in Medicine, 2025.
25. Long, B., Y. Guan, and M. Holden. *A Two-Stage Neural Network Model for Breast Ultrasound Image Classification*. in *2023 IEEE 23rd International Conference on Bioinformatics and Bioengineering (BIBE)*. 2023.