

Article

Not peer-reviewed version

LBA-Net: Lightweight Boundary-Aware Network for Efficient Breast Ultrasound Image Segmentation

[Ye Deng](#), Meng Chen, Jieguang Liu, [Qi Cheng](#), [Xiaopeng Xu](#)^{*}, [Yali Qu](#)^{*}

Posted Date: 1 December 2025

doi: 10.20944/preprints202509.2349.v3

Keywords: breast ultrasound; lightweight network; boundary-aware



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

LBA-Net: Lightweight Boundary-Aware Network for Efficient Breast Ultrasound Image Segmentation

Ye Deng ¹, Meng Chen ², Jieguang Liu ³, Qi Cheng ¹, Xiaopeng Xu ^{4,*} and Yali Qu ^{5,*}

¹ Science and Education Section, Wuhan Jinyintan Hospital, Tongji Medical College of Huazhong University of Science and Technology, Wuhan, Hubei 430023, China

² Institute of Aging, Wenzhou Medical University, Wenzhou 325027, China

³ Clinical Biological Biospecimen Resource Center, Wuhan Jinyintan Hospital, Tongji Medical College of Huazhong University of Science and Technology, Wuhan, Hubei 430023, China

⁴ Key Laboratory of Intelligent Control and Optimization for Industrial Equipment of Ministry of Education, School of Control Science and Engineering, Dalian University of Technology, Dalian, 116024, China

⁵ Ultrasound Department, Wuhan Jinyintan Hospital, Tongji Medical College of Huazhong University of Science and Technology, Wuhan, Hubei 430023, China

* Correspondence: xuxp@mail.dlut.edu.cn (X.X.); 49714972@qq.com (Y.Q.)

Abstract

Breast ultrasound segmentation is challenged by strong noise, low contrast, and ambiguous lesion boundaries. Although deep models achieve high accuracy, their heavy computational cost limits deployment on portable ultrasound devices. In contrast, lightweight networks often struggle to preserve fine boundary details. To address this gap, we propose the lightweight boundary-aware network. A MobileNetV3-based encoder with the atrous spatial pyramid pooling is integrated for efficient multi-scale representation learning. The applied the lightweight boundary-aware block uses an adaptive fusion to combine efficient channel attention and depthwise spatial attention to enhance discriminative capability with minimal computational overhead. A boundary-guided dual-head decoding scheme injects explicit boundary priors and enforces boundary consistency to sharpen and stabilize margin delineation. Experiments on curated BUSI* and BUET* datasets demonstrate that the proposed network achieves 82.8% Dice, 38 px HD95, and real-time inference speeds (123 FPS GPU / 19 FPS CPU) using only 1.76M parameters. They show that this proposed network can offer a highly favorable balance between accuracy and efficiency.

Keywords: breast ultrasound; lightweight network; boundary-aware

I. Introduction

Globally, breast cancer continues to be a major cause of cancer-related mortality among women. Early detection and precise diagnosis are essential for enhancing survival rates [1]. Among various imaging modalities, Breast ultrasound segmentation (BUS) has emerged as a vital tool in clinical practice, due to its non-invasive nature, lack of radiation, and relatively low cost [2,3]. However, BUS images are often characterized by strong speckle noise, low contrast, uneven lesion appearance, and blurred boundaries, which pose significant challenges for automatic tumor segmentation in computer-aided diagnosis (CAD) systems.

Deep learning-based segmentation methods have completely transformed BUS image analysis. Architectures such as U-Net [4] and its variants—including U-Net++ [5] and Attention U-Net [6] have achieved remarkable success by leveraging encoder-decoder structures with skip connections. Recently, transformer-based methods and hybrid CNN-state space models have demonstrated outstanding global context modeling capabilities, further enhancing segmentation accuracy. Although they are very effective, these models usually involve a large number of parameters and

high computational costs, which limits their practicality in resource-constrained environments, such as portable ultrasound devices used in point-of-care Settings.

To improve efficiency in breast ultrasound segmentation, several studies have adopted lightweight networks that replace heavy backbones with efficient architectures such as MobileNet [7] and EfficientNet [8]. While these approaches substantially decrease computational costs, they often suffer from reduced robustness to noise and inadequate delineation of lesion boundaries—a critical aspect in BUS segmentation. Moreover, although attention mechanisms have demonstrated effectiveness in emphasizing salient features and suppressing background interference, prevalent modules tend to introduce considerable computational overhead, hindering their applicability in real-time clinical settings.

Existing BUS segmentation methods therefore face three key limitations. First, heavy models, although achieving high accuracy, are unsuitable for real-time deployment on portable scanners due to their large computational and memory demands. Second, lightweight models often lack sufficient boundary sensitivity and tend to degrade severely on noisy or low-contrast images, which are common in BUS. Third, most existing approaches do not explicitly incorporate boundary priors, even though accurate margin delineation is clinically essential for tumor assessment and treatment planning.

To address these challenges, we propose Lightweight Boundary-Aware Network (LBA-Net) based on the U-Net architecture. It uses a MobileNetV3-Small encoder and an Atrous Spatial Pyramid Pooling (ASPP) bottleneck to capture multi-scale features. The applied Lightweight Boundary-Aware Block (LBA-Block) combines channel and spatial attention to enhance feature discrimination. A boundary-guided decoding scheme injects boundary priors into the decoder to refine skip connections. A dual-head supervision strategy optimizes both segmentation and boundary predictions by a hybrid loss.

The key contributions of this work are as follows:

1. We propose LBA-Net, a lightweight boundary-aware network for efficiency breast ultrasound image segmentation. By integrating a MobileNetV3 encoder with LBA-Blocks, it achieves high accuracy with minimal computational cost, ideal for resource-constrained environments.
2. We introduce the Boundary Guidance Module (BGM), which enhances boundary delineation by generating attention maps that guide the decoder, improving lesion boundary accuracy, especially in noisy and low-contrast images.
3. The Lightweight Boundary-Aware (LBA) Block combines channel and spatial attention mechanisms to refine feature representations, ensuring better sensitivity to key features in ultrasound images.
4. A dual-head supervision mechanism is proposed, where one head focuses on segmentation and the other on boundary prediction, leading to more precise and stable segmentation results.

II. Related Work

Automatic segmentation of breast lesions in ultrasound images is an essential step in CAD systems [9]. It assists early tumor detection and supports quantitative analysis. Deep learning has greatly advanced this task. Convolutional neural networks (CNNs) have played a central role [10]. U-Net is one of the most influential architectures. Ronneberger et al. [4] introduced an encoder-decoder structure with skip connections. It preserves spatial details and captures high-level semantics. Several important variants followed. SegNet uses pooling indices for efficient upsampling and reduces computational cost [11]. ResUNet adds residual connections and improves information flow in deeper networks [12]. DeepLabv3+ expands the receptive field using atrous convolutions and ASPP for multi-scale context extraction [13]. Transformer-based models such as TransUNet [14] and Swin-UNet [15] also enhance long-range dependency modeling and achieve strong segmentation accuracy.

Although these models perform well, they often contain many parameters. Their computational cost is high. This makes them unsuitable for real-time clinical deployment. Portable ultrasound

devices usually operate under strict hardware limitations. This situation has increased interest in lightweight segmentation networks for BUS images.

Many recent BUS studies replace heavy encoders with compact networks. SqueezeNet [16] and its variants are common choices. These backbones greatly reduce parameters and speed up inference. However, two problems remain. First, depth-wise separable convolutions reduce the receptive field. This weakens fine-grained boundary features and causes performance drops on low-contrast or noisy images. Second, most lightweight encoders do not include explicit boundary enhancement or noise-suppression modules. As a result, decoders must recover sharp boundaries from degraded feature maps. Existing lightweight models often lose boundary accuracy or require extra post-processing, which reduces their real-time advantage.

Attention mechanisms have become widely used in medical image segmentation. Channel attention modules such as Squeeze-and-Excitation (SE) [17] and Efficient Channel Attention (ECA) [18] adjust channel responses through global statistics. Spatial attention highlights important regions in the image. However, many attention designs rely on large kernels or dense pooling. These operations increase computational cost and memory usage, which is problematic for edge devices. Lightweight designs such as LAED-Net [19] use depth-wise convolutions to reduce overhead. Even so, they do not specifically address speckle noise or unclear boundaries in BUS images.

Accurate boundary extraction is important in breast ultrasound. Tumor shape and margin characteristics often guide malignancy assessment and surgical planning. Several networks introduce auxiliary boundary prediction. Examples include Boundary-Aware Networks [20,21] and BES-Net [22]. They produce sharper boundaries using an additional edge branch. However, these models still depend on heavy encoders or lack built-in noise suppression mechanisms.

Some methods focus on robustness to ultrasound artifacts. HCMNet [23,24] combines CNNs, Mamba structures, and wavelet transforms to handle noise and domain shifts. These designs improve stability, but they add multiple processing branches. As a result, inference time becomes much longer, which conflicts with real-time requirements.

Overall, current BUS segmentation methods face a clear gap. Many accurate models are too heavy, while lightweight models often struggle with unclear boundaries and noise. In summary, existing BUS segmentation methods fall into two extremes:

1. heavy models with good accuracy but poor deployability, and lightweight models with fast inference but weak boundary awareness and robustness to speckle noise.
2. To our knowledge, there is still a lack of a segmentation framework that is simultaneously lightweight, boundary-aware, and efficient to ultrasound-specific artifacts.

Our work aims to fill this gap by introducing a lightweight network with explicit boundary guidance and lightweight attention mechanisms.

III. METHODOLOGY

A. Overall Architecture

To achieve a favorable trade-off between segmentation accuracy and computational efficiency, LBA-Net integrates several lightweight strategies into its architecture. First, a MobileNetV3-Small [25] backbone is adopted as the encoder to extract multi-scale representations. By leveraging depth-wise separable convolutions, squeeze-and-excitation modules, and efficient activation functions, MobileNetV3-Small substantially reduces parameter count and floating-point operations (FLOPs), leading to more than 90% reduction in complexity compared with conventional U-Net encoders.

Second, we introduce the LBA-Block, which enhances feature discriminability with negligible computational overhead. Unlike conventional attention modules that rely on heavy convolutions and fully connected layers, the LBA-Block fuses two lightweight components: (i) ECA, which models cross-channel interactions through a 1D convolution without dimensionality reduction, thereby avoiding parameter inflation; and (ii) spatial attention based on a 3×3 depth-wise separable convolution, which highlights salient spatial regions. Moreover, we introduce adaptive fusion

weights (α , β) to dynamically balance the contributions of channel and spatial attention, allowing the model to emphasize either modality depending on the input context.

Finally, the decoder adopts streamlined skip connections and incorporates a dual-head supervision scheme. This design not only facilitates precise boundary localization but also maintains a compact model size by avoiding redundant layers. LBA-Net adopts an encoder–decoder architecture with skip connections, following the U-Net paradigm. The overall pipeline consists of four main components (Figure 1):

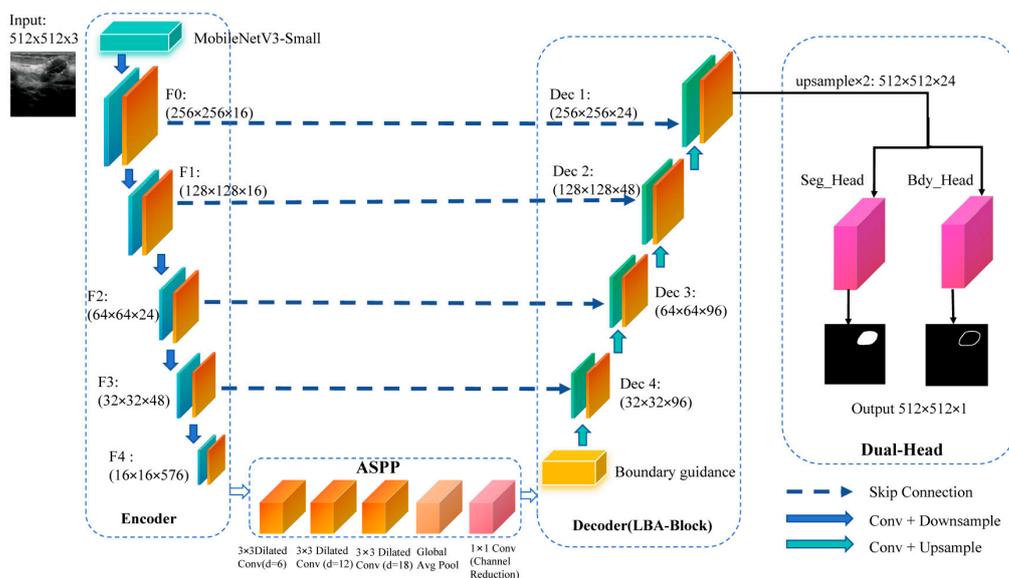


Figure 1. The structure diagram of LBA-Net.

The proposed network is built upon a lightweight encoder-decoder framework. A pre-trained MobileNetV3-Small model is adopted as the encoder to extract hierarchical multi-scale features $\{F_0, F_1, F_2, F_3, F_4\}$ at progressively reduced spatial resolutions. At the bottleneck, the ASPP module [13] is incorporated to capture multi-scale contextual information over diverse receptive fields. In the decoder path, feature maps are gradually upsampled and fused with corresponding encoder features via skip connections. At each stage of the decoder, a Lightweight Boundary-aware block is applied to enhance discriminative feature learning while suppressing irrelevant background responses. Furthermore, the network employs a dual-head output structure: one head produces the segmentation mask, while the other generates a boundary map. In the decoder path, feature maps are progressively upsampled and fused with encoder features via skip connections. At the bottleneck, an ASPP module first produces multi-scale features, which are further transformed by a boundary guidance module into a boundary attention map. This attention is injected into the first decoder block, where it modulates the corresponding skip feature through a learnable guidance branch, enhancing responses along potential lesion boundaries while suppressing background noise.

At each decoding stage, an LBA-Block refines the fused features via lightweight channel–spatial attention. After the final decoder layer, a dual-head output structure is employed: one head predicts the segmentation mask and the other predicts a boundary probability map, both from the same decoded feature representation. The boundary head is supervised by edge labels obtained from morphological gradients of the ground-truth masks, and a consistency loss additionally aligns the internal boundary attention map with these edge labels. Notation: Given an input ultrasound image $I \in \mathbb{R}^{3 \times H \times W}$, the backbone produces a feature map $F_{enc} \in \mathbb{R}^{3 \times H \times W}$. All feature maps in this paper follow the convention where C , H , and W denote the channel, height and width, respectively.

Algorithm 1: Main structure of LBA-Net**Input:** $I \in \mathbb{R}^{(H \times W \times C)}$ **Output:** Segmentation mask S , boundary map B , total loss L_{total} 1: $\{F_0, F_1, F_2, F_3\} \leftarrow \text{ENCODER}(I)$ 2: $F_{aspp} \leftarrow \text{ASPP_MODULE}(F_4)$ 3: $A_b \leftarrow \text{BOUNDARY_GUIDANCE}(F_{aspp})$ 4: $X_{dec} \leftarrow \text{DECODER}(F_{aspp}, \{F_3, F_2, F_1, F_0\}, A_b)$ 5: $S, B, S_{logits}, B_{logits} \leftarrow \text{HEADS}(X_{dec})$ 6: $L_{total} \leftarrow \text{LOSS}(S_{logits}, B_{logits}, A_b, G, Bgt)$ 7: return S, B, L_{total} **B. MobileNetV3-based Encoder with Atrous Spatial Pyramid Pooling**

The overall encoder configuration is illustrated in Figure 2. To efficiently extract discriminative representations from ultrasound images, the encoder of LBA-Net adopts MobileNetV3 as the backbone due to its lightweight and high-performance design. The backbone generates a hierarchical feature maps $\{F_0, F_1, F_2, F_3, F_4\}$ through multiple inverted residual bottlenecks with stride-2 downsampling. Compared to conventional convolutions, the inverted residual structure with depth-wise separable convolution significantly reduces computation overhead while preserving local semantic features. Additionally, MobileNetV3 incorporates squeeze-and-excitation (SE) and hard-swish (HS) activation, which enhance channel dynamics and nonlinear capacity, respectively, thus improving robustness against ultrasound noise and artifacts.

However, breast tumors often exhibit large scale variations and blurred boundaries, which require sufficient receptive field. To address this issue, the highest-level feature F_4 is fed into an ASPP module for enlarged multi-scale context aggregation. ASPP consists of four parallel branches: a 1×1 convolution, three 3×3 dilated convolutions with dilation rates $\{6, 12, 18\}$, and a global average pooling branch. The outputs are concatenated and projected to form an enhanced representation with rich local-global context:

$$F_{aspp} = \phi(\{Conv_{1 \times 1}(F_4), Conv_{3 \times 3}^{r=6}(F_4), Conv_{3 \times 3}^{r=12}(F_4), Conv_{3 \times 3}^{r=18}(F_4), GAP(F_4)\}) \quad (1)$$

where $\phi(\cdot)$ denotes concatenation followed by a 1×1 projection.

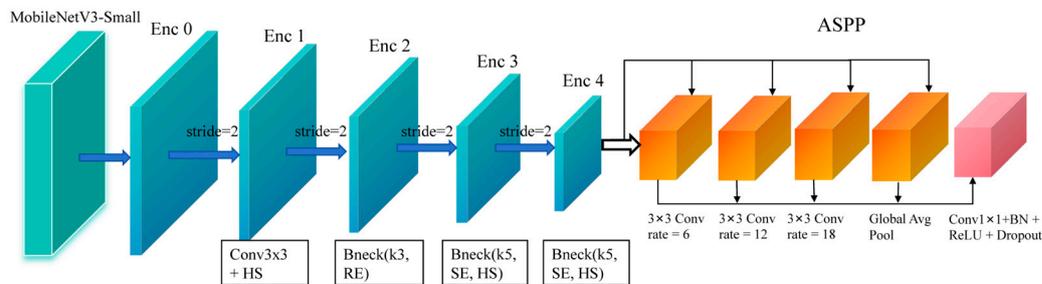


Figure 2. Encoder structure of LBA-Net using MobileNetV3 backbone and ASPP module.

To extract multi-scale features with high discriminative power while minimizing computational overhead, we adopt MobileNetV3-Small.

(a) Channel Attention

Given feature map $F \in \mathbb{R}^{H \times W \times C}$, we apply ECA or a squeeze-excitation (SE) mechanism:

$$\omega_c = \sigma(\text{Conv1D}(\text{GAP}(F))), \quad (2)$$

where GAP denotes global average pooling, Conv1D is a 1D convolution with kernel size k , and σ is the sigmoid function.

The channel-refined feature is:

$$F_c = F \otimes \omega_c, \quad (3)$$

where \otimes denotes channel-wise multiplication.

(b) Spatial Attention

Spatial attention focuses on salient regions using depth-wise separable convolution:

$$\omega_s = \sigma(\text{DWConv}(F)), \quad (4)$$

where depth-wise separable convolution is with kernel size 3×3 .

The spatially refined feature is:

$$F_s = F \otimes \omega_s, \quad (5)$$

(c) Fusion

The final output is obtained by combining both attentions:

$$Y = \alpha F_c + \beta F_s, \quad (6)$$

where α and β are learnable scalar weights initialized as 0.5.

C. Boundary Guidance

Ultrasound images often present weak and ambiguous lesion boundaries. To explicitly inject boundary priors into the decoder for sharper lesion delineation. Figure 3 illustrates the Boundary Guidance Module (BGM). The ASPP feature is first transformed by a lightweight convolutional branch to generate an attention map highlighting potential tumor boundaries. This attention is then injected into the skip connection of the first decoder stage using a modulation network, allowing boundary cues to influence early decoding where spatial details are most critical.

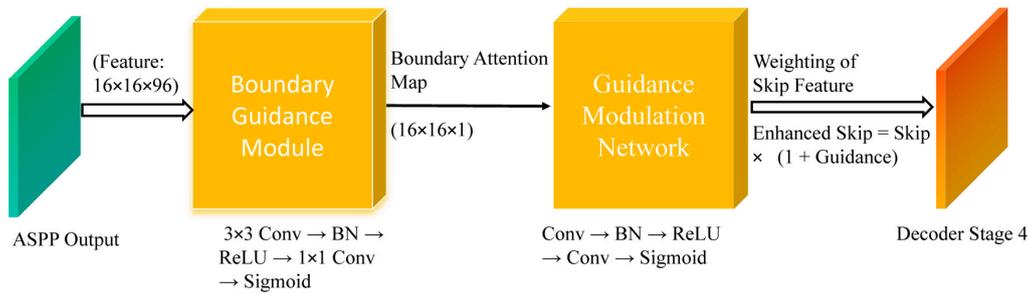


Figure 3. Boundary Guidance structure.

(a) Boundary Guidance Modulation

Given the ASPP feature $F_{aspp} \in \mathbb{R}^{C \times H \times W}$, BGM predicts a boundary-attention map through a compact convolutional subnetwork:

$$F_b = \sigma(\text{Conv}_{1 \times 1}(\text{ReLU}(\text{BN}(\text{Conv}_{3 \times 3}(F_{aspp}))))), \quad (7)$$

Where $\sigma(\cdot)$ denotes the sigmoid function

(b) Boundary Modulation Network

Let $G(F_b)$ represent the modulation weight obtained from the boundary attention map F_b :

$$G(F_b) = \sigma(\text{Conv}_{1 \times 1}(\text{ReLU}(\text{BN}(\text{Conv}_{3 \times 3}(F_b))))), \quad (8)$$

let $F_{skip} \in \mathbb{R}^{B \times C_{skip} \times H \times W}$ represent the skip connection feature map from the decoder, where C_{skip} is the number of channels in the skip connection. For each input, the boundary-guided feature modulation is applied as follows:

$$F_{skip}^{mod} = F_{skip}(1 + G(F_b)), \quad (9)$$

where F_{skip} is the skip connection feature map.

(c) Boundary Consistency Loss:

$$L_b = \| \text{Up}(F_b) - B_{gt} \|_2^2, \quad (10)$$

where B_{gt} is the ground-truth boundary map derived from the segmentation mask, and $\text{Up}(\cdot)$ indicates bilinear upsampling to the original image size.

D. Boundary-Aware LBA-Block

To enhance feature discriminability while maintaining computational efficiency, we propose the Lightweight Boundary-aware Attention Block, an attention module that adaptively refines feature representations by fusing channel-wise and spatial-wise attention mechanisms with minimal parameter overhead. Unlike conventional attention modules that rely on dense convolutions and global pooling operations, the LBA-Block is specifically designed for edge deployment in resource-constrained clinical environments.

As illustrated in Figure 4, LBA-Block first reweights channels via ECA [26], then generates a 3×3 depth-wise spatial attention map, and finally fuses both branches through learnable α , β to yield the refined feature.

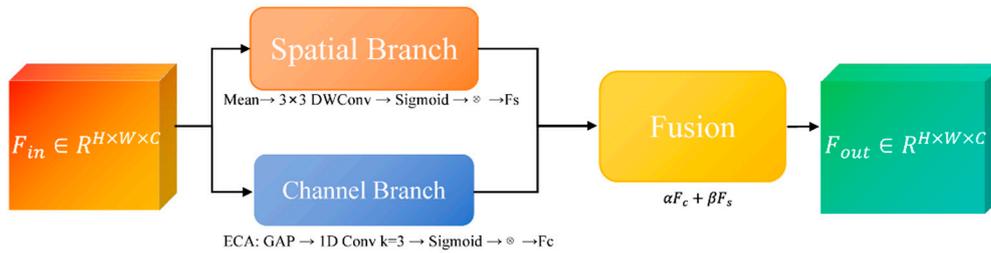


Figure 4. LBA-Block structure. Structure of the proposed Lightweight Boundary-aware Attention Block, consisting of channel attention, spatial attention, and adaptive fusion modules.

(a) Channel Branch

Channel attention aims to recalibrate feature channels by emphasizing informative responses and suppressing less useful ones. We adopt the ECA mechanism due to its parameter-free design and strong performance in low-resource settings. Given input $F_{in} \in R^{C \times H \times W}$, global average pooling (GAP) is first applied to produce channel-wise descriptors:

$$Z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W F_{in}^c(i, j), \quad (11)$$

where Z_c represents the mean activation of channel c .

A 1D convolution with kernel size k (set to 3 in our implementation) is then applied to capture cross-channel interactions without dimensionality reduction:

$$\omega = \sigma(\text{Conv1D}_k(z)), \quad (12)$$

where σ denotes the sigmoid activation function. The channel-refined feature F_c is obtained via channel-wise multiplication:

$$F_c = \omega \otimes F_{in}, \quad (13)$$

This design eliminates channel dimensionality reduction and significantly reduces parameters while effectively modeling inter-channel dependencies.

(b) Spatial Branch

Spatial attention focuses on identifying salient regions within each feature map. To minimize computational cost, we employ depth-wise separable convolution with a 3×3 kernel instead of standard convolutions or large-kernel spatial attention. The spatial attention map A_s is computed as:

$$A_s = \sigma(\text{DWConv}_{3 \times 3}(F_{in})) \in R^{1 \times H \times W}, \quad (14)$$

where DWConv denotes depth-wise convolution followed by point-wise projection to a single channel. The spatially refined feature F_s is then:

$$F_s = A_s \otimes F_{in}, \quad (15)$$

This lightweight design substantially reduces the spatial attention computation while effectively preserving fine-grained lesion boundaries and suppressing background clutter in ultrasound images.

(c) Adaptive Fusion

To effectively integrate semantic enhancement from the channel branch and boundary refinement from the spatial branch, we introduce learnable fusion weights α and β , initialized to 0.5 and jointly optimized in an end-to-end manner:

$$F_{out} = \alpha \cdot F_c + \beta \cdot F_s, \quad (16)$$

This adaptive strategy enables the network to dynamically emphasize channel semantics in noisy regions, or spatial boundary cues in low-contrast tumor edges, while keeping computational overhead negligible.

As illustrated in Figure 5, the proposed boundary-aware attention mechanism yields accurate and visually coherent breast lesion segmentation results in ultrasound images. The original images in Column (A) contain lesions with complex textures and heterogeneous backgrounds, while the corresponding ground truth masks in Column (B) provide precise lesion annotations. The boundary-

attention heatmaps shown in Column (C), together with the predicted masks in Column (D), indicate that the model effectively concentrates on lesion boundaries and produces accurate delineations, achieving Dice scores of 0.976 and 0.953. Moreover, the overlay images in Column (E), obtained by superimposing the predicted masks onto the original images, intuitively demonstrate the spatial consistency between the predictions and the underlying anatomy, thereby further confirming the effectiveness of the proposed model in delineating breast lesions in ultrasound images.

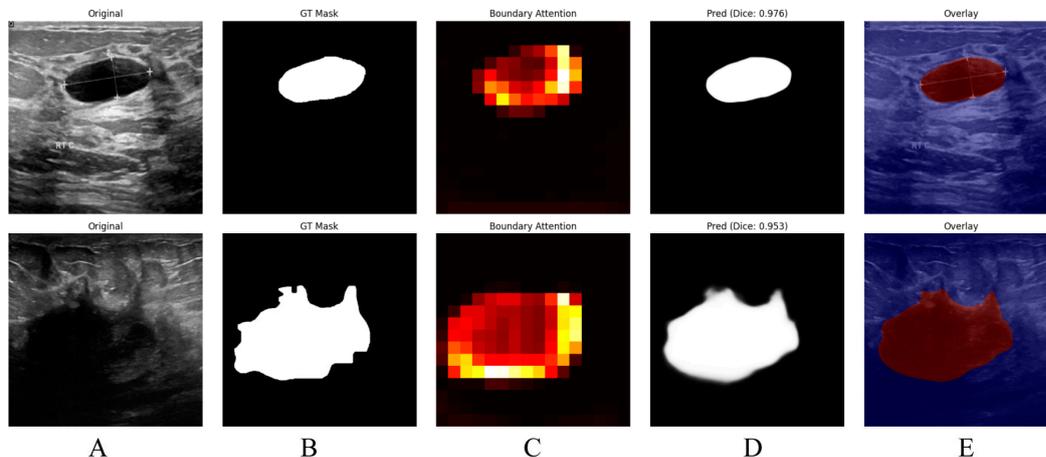


Figure 5. Visualization of Breast Lesion Segmentation Results in Ultrasound Images with Boundary-Aware Attention by using the BUSI. This Figure shows breast lesion segmentation results in ultrasound images using a boundary-aware attention model. A (Original): Original ultrasound images. B (GT Mask): Ground truth lesion masks. C (Boundary Attention): Heatmaps of boundary attention (warmer colors = high boundary focus). D (Pred): Predicted segmentation masks with Dice similarity coefficient. F (Overlay): Predicted masks overlaid on original images. Two cases are presented, with Dice scores of 0.976 and 0.953, illustrating accurate lesion delineation.

E. Dual-Task Joint Supervision Head

To enhance boundary delineation in ultrasound segmentation, LBA-Net adopts a dual-head architecture in Figure 6:

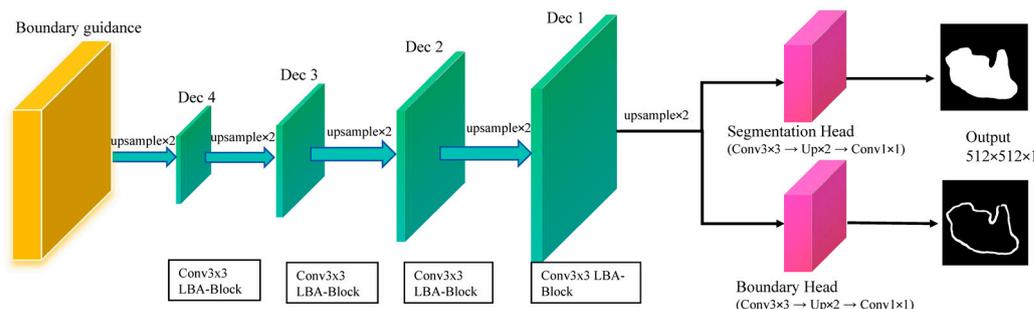


Figure 6. LBA-Block powered decoder with dual-head boundary-aware prediction.

LBA-decoder: four-step upsampling of multi-scale encodings (Dec1-Dec4) to dual-head outputs; segmentation logits S refined by 3×3 weight W_b inferred from boundary logits B ; losses: Dice+BCE on S , Tversky on B .

(a) Segmentation Head:

Outputs a binary segmentation mask $S \in [0,1]^{H \times W}$ where $H \times W$ denotes the spatial resolution of the input image which consistent with the 512×512 input size. A pixel value of 1 represents the tumor region, while 0 represents the background which are normal breast tissue or artifacts.

(b) Boundary Head:

Outputs a probability map $B \in [0,1]^{H \times W}$ highlighting object boundaries, with ground-truth labels derived via morphological gradient operations on S_{gt} . The boundary labels are computed as:

$$B_{gt} = Dilate(S_{gt}) - Erode(S_{gt}), \quad (17)$$

where $Dilate(\cdot)$ and $Erode(\cdot)$ denote morphological dilation and erosion, respectively. The subtraction result is then binarized to ensure a consistent supervision target:

$$B_{gt}(x,y) = \begin{cases} 1, & \text{if } B_{gt}(x,y) > 0 \\ 0, & \text{otherwise} \end{cases} \quad (18)$$

F. Loss Function

The overall training objective combines segmentation loss, boundary loss, and a boundary-consistency term:

$$L_{total} = L_{seg} + \lambda_{bdy} L_{bdy} + \lambda_{cons} L_{cons}, \quad (19)$$

where L_{seg} supervises the mask predictions, L_{bdy} supervises the boundary head, and L_{cons} enforces consistency between the internal boundary attention and the morphology-derived boundary labels. In our implementation, we set $\lambda_{bdy}=0.3$ and $\lambda_{cons}=0.1$.

(a) Segmentation Loss:

$$L_{seg} = L_{BCE}(S_{logits}, G) + L_{Dice}(S_{logits}, G), \quad (20)$$

Where L_{BCE} (Binary Cross-Entropy Loss) measures pixel-wise differences between the predicted mask S_{logits} and the ground-truth mask G , and L_{Dice} (Dice Loss) evaluates the overlap between the predicted and ground-truth masks, which is especially useful for handling class imbalance between foreground and background pixels.

(b) Boundary Loss:

We adopt Tversky loss to address class imbalance in boundaries:

$$L_{bdy} = 1 - \frac{TP}{TP + \alpha FP + \beta FN}, \quad (21)$$

where TP, FP, FN are the numbers of true positives, false positives, and false negatives along the boundary. We set $\alpha=0.3$ and $\beta=0.7$ to emphasize recall, penalizing false negatives more strongly.

(c) Boundary-consistency loss

Finally, we encourage the boundary attention map A_b predicted by the boundary guidance module to be consistent with B_{gt} via a mean squared error (MSE):

$$L_{cons} = MSE(Resize(A_b, size(B_{gt})), B_{gt}),$$

where the attention map is interpolated to the resolution of the boundary labels. This term links the implicit attention in the encoder-decoder with the explicit boundary supervision, leading to sharper and more stable boundary predictions.

G. Evaluation Metrics

To comprehensively assess segmentation performance, four commonly used metrics are employed:

$$Dice = \frac{2|P \cap G|}{|P| + |G|}, \quad (22)$$

$$IoU = \frac{|P \cap G|}{|P \cup G|}, \quad (23)$$

$$Recall = \frac{|P \cap G|}{|G|}, \quad (24)$$

$$HD95(P, G) = \max \left\{ \text{perc}_{95} \left(\min_{g \in G} \|p - g\|_2 \mid p \in P \right), \text{perc}_{95} \left(\min_{p \in P} \|g - p\|_2 \mid g \in G \right) \right\}, \quad (25)$$

Where P and G denote the predicted and ground-truth tumor regions, $\|\cdot\|_2$ denotes the Euclidean distance, perc_{95} represents the 95th percentile. Lower HD95 indicates better boundary accuracy. A smaller HD95 indicates closer spatial agreement between the predicted and ground-truth boundaries, corresponding to higher boundary accuracy. HD95 is reported in pixel units since physical pixel spacing was unavailable; thus the value should be interpreted as a relative metric. Furthermore, we chose the model's parameter count and computational cost (GFLOPs) as critical metrics to evaluate the balance between its lightweight design and performance.

H. Training Strategy

All images are resized to 512×512, normalized using ImageNet statistics, and augmented with random horizontal flips (0.5), vertical flips (0.3), $\pm 30^\circ$ rotations, brightness–contrast adjustment, and 3×3 Gaussian blur. Training is performed using mixed-precision (FP16) with a batch size of 24 and gradient clipping for optimization stability. We employ the AdamW [27] optimizer with module-specific learning rates: 1×10^{-4} for the MobileNetV3 backbone, 2×10^{-3} for the decoder, and 3×10^{-3} for the boundary-guidance and head modules. A One-Cycle learning rate schedule is used over 300 epochs, consisting of a 10% warm-up phase followed by cosine decay. The MobileNetV3-Small encoder is initialized with ImageNet pre-trained weights.

IV. Experiments

A. Datasets

We evaluated LBA-Net on the publicly available breast ultrasound dataset. The first dataset was the BUSI dataset [28] contains 780 images, including 437 benign cases, 210 malignant cases and 133 normal cases, and the second dataset utilized in this study was established at the medical center of the Bangladesh University of Engineering and Technology (BUET) [29], Dhaka, Bangladesh, between 2012 and 2013. Breast ultrasound data were acquired using a Sonix-Touch Research system equipped with a linear L14-5/38 transducer, operating at a center frequency of 10 MHz and a sampling frequency of 40 MHz. It comprises 260 images from 223 female patients (aged 13–75 years), including 190 benign and 70 malignant cases. The image resolutions vary from 298×159 to 800×600 pixels. Each image is accompanied by pixel-level truth segmentation masks from multiple patients.

A chief physician with 15 years of experience in radiology and an attending physician with 3 years of experience in radiology independently read the films using a double-blind method. According to the ACR TI-RADS quality clause, the original ultrasound sequences were scored at five levels. Only cases with a score ≥ 4 and the two agreed ($k = 0.87$) were retained. The differences were arbitrated by chief physicians with 15 years of experience. Eventually, as shown in Figure 7, their assessment identified a subset of low-quality breast ultrasound images, which exhibited issues including misclassification, text overlay obscuring tissue, the presence of axillary tissue, indistinct lesion boundaries, and substantial acoustic shadowing. These examples demonstrate typical failure cases where reliable annotation is impossible, justifying their exclusion during dataset curation.

We selected 570 cases (380 benign, 190 malignant) from BUSI and 230 cases (174 benign, 56 malignant) from BUET for our analysis. To ensure clarity and comparability, these quality-screened subsets are denoted as BUSI* and BUET* throughout this paper, with the asterisk symbol (*) in tables and figures indicating the curated datasets used in this study.

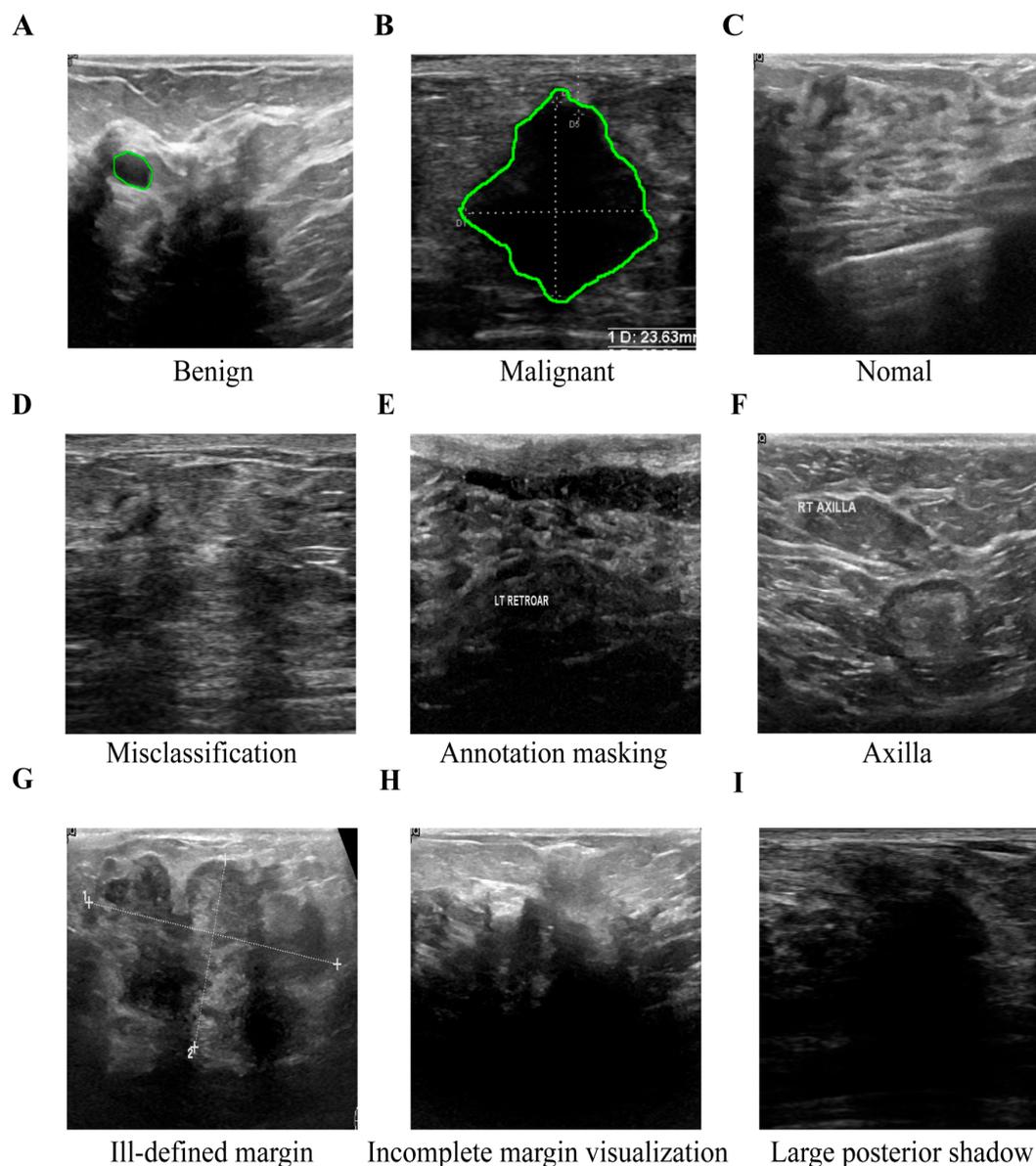


Figure 7. Example of some sub-optimal images from the BUSI and BUET datasets. To ensure data quality, images exhibiting critical flaws were excluded from this study and the green circle indicates GT mask. The retained training images (A-C) show clinically validated, high-resolution examples of benign, malignant, and normal breast tissues. The excluded problematic images (D-I) illustrate specific exclusion criteria: (D) misclassification, (E) text occlusion, (F) inclusion of irrelevant axillary tissue, (G) indistinct boundaries, (H) missing boundaries, and (I) extensive acoustic shadowing.

B. Implementation Details

All experiments are implemented in PyTorch 1.12 and conducted on an NVIDIA GeForce RTX 4090 D GPU (25 GB) with TensorRT and Automatic Mixed Precision (AMP) enabled, as well as a Core i7-14700KF CPU (3.40 GHz). Notably, all comparative models are also subject to the same TensorRT acceleration and AMP settings to ensure fair performance comparison. Model efficiency is evaluated in terms of parameters, FLOPs, and inference speed on both the aforementioned GPU and CPU. Segmentation performance is assessed using Dice, IoU, precision, recall, and HD95. For both BUSI* and BUET* datasets, we adopt 5-fold stratified cross-validation with identical preprocessing and evaluation protocols across all folds.

As summarized in Table 1, the proposed LBA-Net achieves strong segmentation performance on BUSI, BUSI*, BUET, and BUET*, with the refined subsets (BUSI* and BUET*) yielding notably

higher Dice, IoU, and Recall scores as well as substantially lower HD95 values. These improvements highlight the advantages of higher-quality annotations and demonstrate the model's ability to produce more accurate and reliable lesion boundary localization across datasets of varying complexity.

Table 1. Quantitative comparison of segmentation performance on the BUSI, BUSI*, BUET, and BUET* datasets.

Dataset	Dice (%)	IoU (%)	Recall(%)	HD95(px)
BUSI	78.39 ± 2.43	69.52 ± 2.57	79.35 ± 2.51	67.81 ± 6.94
BUSI*	82.79 ± 2.48	74.96 ± 2.43	85.02 ± 2.09	45.96 ± 9.82
BUET	83.93 ± 1.59	74.86 ± 1.88	84.82 ± 1.79	63.27 ± 8.54
BUET*	86.59 ± 1.87	77.92 ± 1.52	87.54 ± 1.38	52.60 ± 7.35

C. Comparison with State-of-the-art

To comprehensively evaluate the effectiveness of the proposed LBA-Net, we compare it with several state-of-the-art CNN-based segmentation models, including UNet, UNet++, FPN, DeepLabv3+, AttentionUNet, SegNet, and ResUNet. Tables 2 and 3 summarize the quantitative results on the BUSI* and BUET* datasets at an input resolution of 512×512.

(a) Performance on the BUSI*

On the BUSI* dataset, LBA-Net achieves a Dice score of 83.35% and IoU of 76.49%, which are competitive with or better than most CNN-based baselines. While DeepLabv3+ achieves a slightly higher Dice score (84.96%), LBA-Net surpasses all models in terms of efficiency, requiring only 3.34 GFLOPs and 1.76M parameters, which are dramatically lower than all competing architectures. The HD95 of LBA-Net (38.02 px) is also significantly better than most classical CNN models and substantially better than heavy architectures such as AttentionUNet, SegNet, and ResUNet.

In addition, LBA-Net achieves 123.43 FPS on GPU and 19.42 FPS on CPU, far exceeding all CNN-based baselines. This indicates that LBA-Net maintains strong segmentation accuracy while being exceptionally lightweight and fast, making it suitable for real-time or resource-constrained clinical applications.

Table 2. Performance comparison of each model on the BUSI* dataset @ 512×512 Input.

Model Category	Model Name	Dice (%)	IoU (%)	Recall(%)	HD95(px)	FLOPs (G)	Params (M)	GPU FPS	CPU FPS
CNN-based	UNet	80.44 ± 1.65	73.96 ± 1.58	88.53 ± 1.77	23.86 ± 2.15	54.61	31.03	34.98	0.99
		82.32 ± 1.68	76.10 ± 1.62	87.26 ± 1.74	19.69 ± 1.97	39.63	29.16	32.60	0.17
CNN-based	FPN	83.29 ± 2.70	75.66 ± 2.60	84.84 ± 1.70	50.99 ± 3.06	36.50	26.85	136.77	3.97
CNN-based	DeepLabv3+	84.96 ± 1.61	76.18 ± 3.24	86.54 ± 1.61	50.54 ± 6.29	43.57	27.17	108.51	3.15
CNN-based	AttentionUNet	76.12 ± 2.58	64.84 ± 2.42	78.66 ± 1.57	82.13 ± 4.93	63.38	39.51	28.65	0.16
		73.82 ± 1.54	58.57 ± 1.35	75.90 ± 1.52	97.61 ± 5.86	36.01	15.27	46.24	0.46

Model Category	Model Name	Dice (%)	IoU (%)	Recall(%)	HD95(px)	FLOPs (G)	Params (M)	GPU FPS	CPU FPS
CNN-based	ResUnet	81.16 ± 1.66	72.47 ± 1.55	82.78 ± 1.66	64.61 ± 3.88	41.17	25.01	69.15	0.26
		83.35 ± 2.42	76.49 ± 1.55	86.41 ± 1.73	38.02 ± 2.66				
Proposed	LBA-Net					3.74	1.76	123.43	19.42

(b) Performance on the BUET*

The performance trend on BUET* is consistent with that on BUSI*. LBA-Net again delivers a strong balance between segmentation accuracy and computational efficiency. Although DeepLabv3+ records the highest Dice score (87.42%), LBA-Net maintains competitive performance with a Dice score of 86.59%, while requiring up to 10^{x-70} fewer FLOPs and 10^{-20} fewer parameters than traditional CNN architectures.

Furthermore, LBA-Net preserves its extremely high inference speed across hardware platforms, obtaining 123.43 GPU FPS and 19.42 CPU FPS, which are significantly higher than all state-of-the-art competitors. This demonstrates that LBA-Net not only performs efficiency on different ultrasound datasets but also generalizes well with consistently low latency and minimal computational overhead.

Table 3. Performance comparison of each model on the BUET* dataset @ 512×512 Input.

Model Category	Model Name	Dice (%)	IoU (%)	Recall (%)	HD95(px)	FLOPs (G)	Params (M)	GPU FPS	CPU FPS
CNN-based	UNet	81.15 ± 3.46	71.77 ± 4.26	82.66 ± 3.23	86.57 ± 18.52	54.61	31.03	34.98	0.99
		81.38 ± 3.59	71.87 ± 4.42	83.18 ± 2.35	91.77 ± 18.18				
CNN-based	UNet++	86.73 ± 2.16	78.69 ± 2.55	87.14 ± 2.76	43.97 ± 7.85	36.50	26.85	136.77	3.97
		87.42 ± 2.78	79.84 ± 3.21	87.92 ± 3.71	38.12 ± 7.21				
CNN-based	DeepLabv3+	80.49 ± 3.40	70.91 ± 4.02	81.49 ± 3.73	89.86 ± 14.93	63.38	39.51	28.65	0.16
		81.64 ± 3.58	72.58 ± 4.13	81.52 ± 3.46	76.66 ± 14.20				
CNN-based	SegNet	80.28 ± 3.18	70.67 ± 3.78	81.89 ± 2.21	85.63 ± 19.82	41.17	25.01	69.15	0.26
		86.59 ± 1.87	77.92 ± 1.52	87.54 ± 1.38	52.60 ± 7.35				
Proposed	LBA-Net					3.74	1.76	123.43	19.42

Figures 8 and 9 compare segmentation outputs across different CNN architectures. LBA-Net consistently produces masks that adhere more closely to ground-truth boundaries, particularly in low-contrast or irregular-shaped tumors. Competing models such as SegNet and ResUNet tend to produce over-segmented or under-segmented results, as highlighted by the false-positive (red) and false-negative (green) regions. As shown in Figure 10, LBA-Net attains competitive Dice and IoU

scores while exhibiting the lowest computational cost (FLOPs) and the highest inference speed (FPS) among all compared methods. Such a favorable accuracy–efficiency trade-off underscores its suitability for real-time clinical deployment, particularly on resource-constrained ultrasound devices.

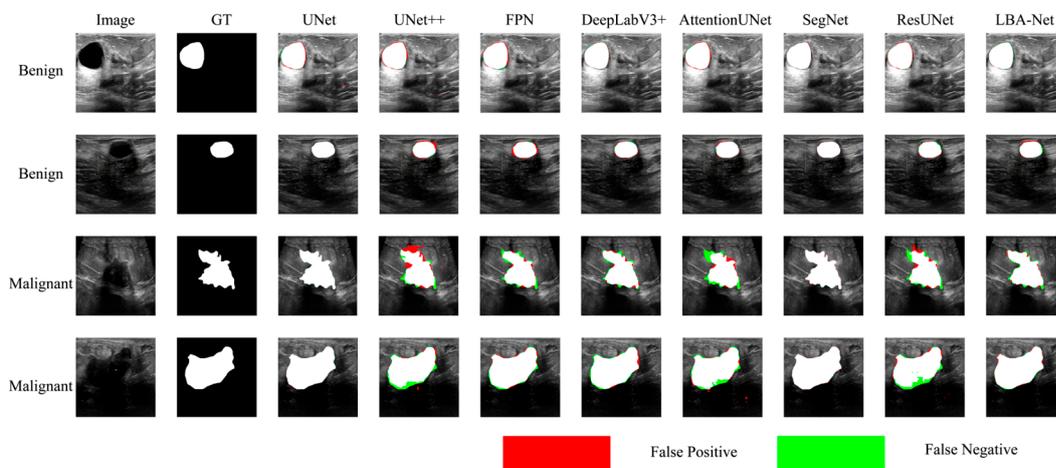


Figure 8. Comparison of Segmentation Results on BUSI* datasets.

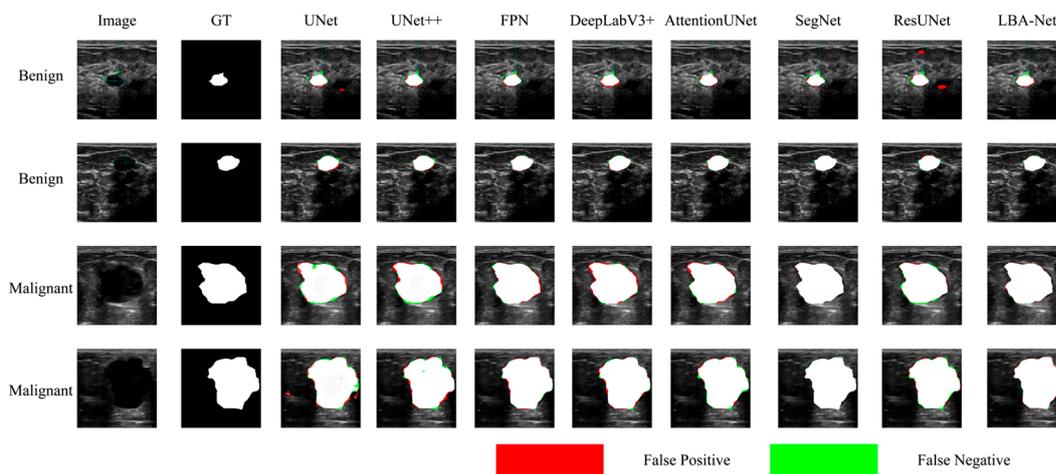


Figure 9. Comparison of Segmentation Results on BUET* datasets.

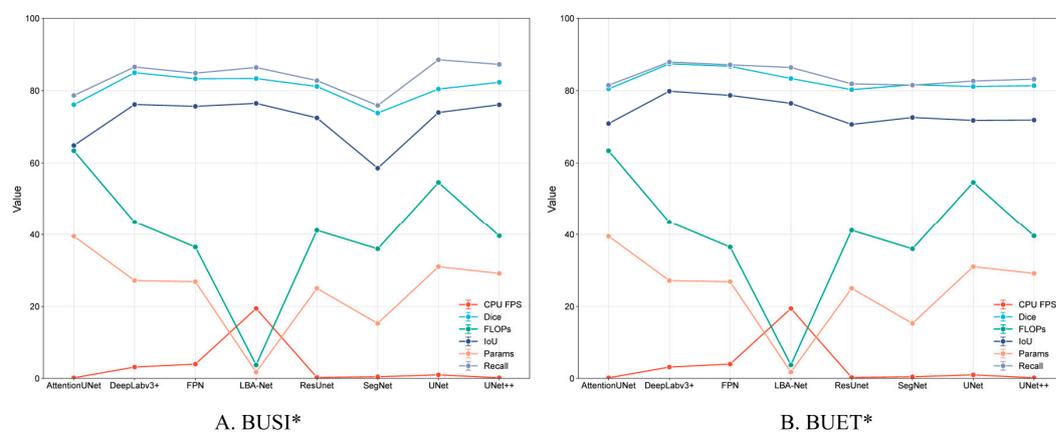


Figure 10. Performance Comparison of Segmentation Models on BUSI* and BUET* Datasets.

D. Ablation Studies

To quantitatively validate the contribution of each design in LBA-Net, we performed a lightweight yet complete ablation study on the BUSI* dataset. Starting from the full model, we progressively ablated or substituted the key components, resulting in five configurations:

- (1) **Full (LBA-Net)**: all proposed modules are active.
- (2) **w/o Boundary Guidance**: The boundary guidance module is removed, but the LBA-Blocks and boundary head (with boundary loss and consistency loss) remain.
- (3) **w/o Boundary Head (Single-head)**: The boundary head is removed, leaving only the segmentation head, without boundary loss or consistency loss.
- (4) **w/o Boundary-Consistency Loss**: The boundary-consistency loss is removed, but both segmentation and boundary heads are trained with their respective losses.
- (5) **w/o LBA-Block**: LBA-Blocks are replaced with standard SE-enhanced convolutions, with ASPP, boundary guidance, and dual-head supervision retained
- (6) **w/o ASPP**: The ASPP module is removed, and the deepest encoder feature is directly passed to the decoder without multi-scale context aggregation.

All models were trained under identical hyper-parameters (512×512 input, 300 epochs, AdamW, cosine LR), without test-time augmentation.

To quantify the contribution of each component in LBA-Net, we performed a series of ablation experiments by removing one module at a time, as reported in Table 4 and visualized in Figure 11. The results show that each component contributes to both accuracy and boundary stability.

Table 4. Ablation results of LBA-Net on BUSI* dataset at 512×512 Input.

Model Variant	Dice (%)	IoU (%)	Recall (%)	HD95 (px)	FLOPs(G)	Params (M)	GPU FPS	CPU FPS
Full (LBA-Net)	82.79 ± 2.48	74.96 ± 2.43	85.02 ± 2.09	45.96 ± 9.82	3.74	1.76	123.43	19.42
w/o Boundary Guidance	82.02 ± 2.87	72.90 ± 3.15	82.89 ± 4.42	57.70 ± 8.95	3.73	1.71	131.62	22.07
w/o Boundary Head (single-head)	82.12 ± 2.93	73.25 ± 3.20	84.07 ± 3.30	58.34 ± 9.01	3.51	1.74	123.86	21.51
w/o Boundary-Consistency Loss	81.76 ± 3.05	73.01 ± 3.28	82.39 ± 3.51	55.97 ± 9.11	3.74	1.75	122.94	23.86
w/o LBA-Block	82.03 ± 2.58	73.17 ± 2.94	84.01 ± 1.80	58.20 ± 8.89	3.73	1.75	138.78	27.05
w/o ASPP	81.81 ± 2.99	73.07 ± 3.24	83.53 ± 3.39	56.65 ± 9.24	4.11	1.95	132.31	24.64

Removing the Boundary Guidance Module leads to a noticeable degradation in segmentation quality, with Dice dropping from 82.79% to 82.02% and HD95 increasing by +11.7 px. This trend, clearly reflected in the radar chart in Figure 11A, demonstrates that injecting boundary priors at an early decoding stage is crucial for restoring sharper lesion margins under ultrasound noise.

The single-head variant (without the Boundary Head) shows a similar deterioration pattern (Dice 82.12%, HD95 +12.4 px), indicating that explicit boundary supervision provides stable structural cues that complement mask prediction. Meanwhile, removing the Boundary-Consistency Loss produces the largest Dice reduction among boundary-related components (-1.03%), as seen in

both Table 4 and Figure 11B, confirming that enforcing consistency between implicit attention and morphology-derived boundaries is essential for precise edge recovery.

Eliminating the LBA-Block also decreases boundary accuracy substantially (HD95 +12.2 px), highlighting the importance of lightweight channel–spatial attention for refining local features, particularly in low-contrast or fuzzy regions typical of breast ultrasound.

Finally, removing the ASPP module yields the lowest Dice score (81.81%) and increases the model size, as the multi-branch dilated convolutions are replaced by standard convolutions. This demonstrates that multi-scale context aggregation is necessary for capturing lesions of varying sizes and shapes.

Overall, both Table 4 and Figure 11 consistently show that the full LBA-Net achieves the most balanced performance across Dice, IoU, HD95, FLOPs, and inference speed, validating the complementary roles of all proposed components.

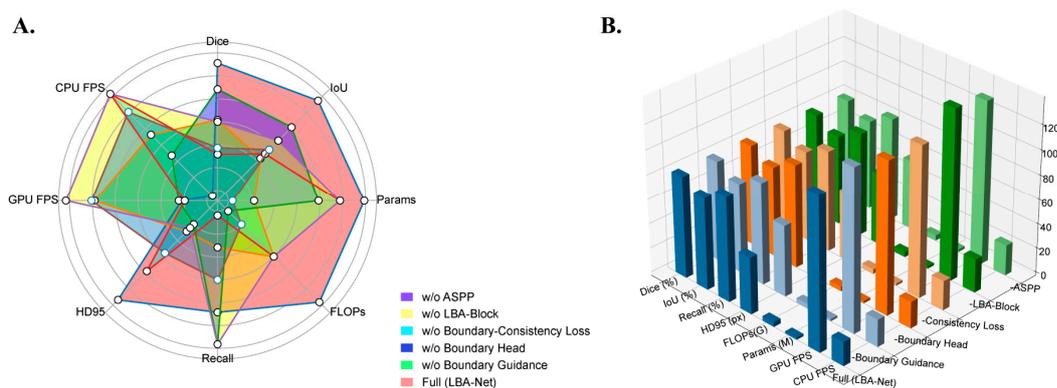


Figure 11. Visualization of LBA-Net ablation study results: A. Radar comparison across Dice, IoU, Recall, HD95, FLOPs, Params, GPU FPS, and CPU FPS; B. 3D bar chart illustrating metric differences among ablated variants.

V. Discussion

LBA-Net offers an efficient solution for breast ultrasound segmentation by balancing high boundary accuracy (HD95) with low computational cost. Leveraging a MobileNetV3 encoder, it significantly reduces parameters and FLOPs compared to traditional models, while maintaining strong segmentation performance through the integration of LBA-Blocks for refined feature representation. The dual-head supervision, with both segmentation and boundary heads, enhances lesion boundary localization, crucial for dealing with the noisy and low-contrast nature of ultrasound images. Its lightweight design ensures real-time performance on both GPU and CPU, making it ideal for deployment in portable ultrasound devices, especially in resource-constrained environments. The model's efficiency and accuracy position it as a valuable tool for real-time clinical use, particularly in point-of-care settings, where efficient, reliable lesion delineation is essential for improving diagnostic workflows and patient outcomes. Future work could explore multi-center validation, hybrid context modeling, and semi-supervised learning to further enhance its applicability.

VI. Conclusions

This work presents LBA-Net, a lightweight boundary-aware framework for breast ultrasound segmentation. By combining a MobileNetV3 encoder, LBA-Blocks, and dual-head boundary supervision, the proposed model achieves competitive accuracy with only 1.76M parameters and real-time inference. Experiments on BUSI* and BUET* validate its robustness to noise and fuzzy edges.

Funding: Not applicable.

Institutional Review Board Statement: This study did not use animal experiments.

Informed Consent Statement: Experiments were conducted on publicly available dataset, and it is cited properly.

Data Availability Statement: The BUSI* and BUET* dataset used in this study is publicly available at https://drive.google.com/drive/folders/1JtyfHpYnyqVuishj1sHxVE_RMWXc5Q9e?usp=sharing, https://drive.google.com/drive/folders/1SAZn1bRqpNbpANhk6P7kJL_IT27dEPHn?usp=sharing. The code of LBA-Net will be made available publicly at <https://github.com/DY221/LBA-Net>.

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

1. H. M. Bizuayehu et al., "Global disparities of cancer and its projected burden in 2050," *JAMA Netw. Open*, vol. 7, no. 11, p. e2443198, 2024. <https://doi.org/10.1001/jamanetworkopen.2024.43198>.
2. F. Chen, Y. Chen, and Y. Hu, "Multimodal ultrasound imaging in the diagnosis of primary giant cell tumor of the breast: A case report and literature review," *J. Clin. Ultrasound*, vol. 53, no. 4, pp. 885–892, 2025. <https://doi.org/10.1002/jcu.23902>.
3. B. Liu, S. Liu, Z. Cao, J. Zhang, X. Pu, and J. Yu, "Accurate classification of benign and malignant breast tumors in ultrasound imaging with an enhanced deep learning model," *Front. Bioeng. Biotechnol.*, vol. 13, p. 1526260, 2025. <https://doi.org/10.3389/fbioe.2025.1526260>.
4. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput.-Assist. Interv. (MICCAI)*, 2015, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.
5. R. A. Praba and L. Suganthi, "Human activity recognition utilizing optimized attention induced multihead convolutional neural network with MobileNet V1 from mobile health data," *Network*, vol. 36, no. 2, pp. 294–321, 2025. <https://doi.org/10.1080/0954898X.2024.2438967>.
6. O. Oktay et al., "Attention U-Net: Learning where to look for the pancreas," *arXiv*, arXiv:1804.03999, 2018.
7. A. G. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," *arXiv*, arXiv:1704.04861, 2017.
8. M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn. (ICML)*, 2019, pp. 6105–6114.
9. M. H. Yap et al., "Breast ultrasound lesions recognition: end-to-end deep learning approaches," *J. Med. Imag.*, vol. 6, no. 1, p. 011007, 2019. <https://doi.org/10.1117/1.JMI.6.1.011007>.
10. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 1097–1105.
11. V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017. <https://doi.org/10.1109/TPAMI.2016.2644615>.
12. Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-Net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, May 2018. <https://doi.org/10.1109/LGRS.2018.2802944>.
13. L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 833–851. https://doi.org/10.1007/978-3-030-01234-2_49.
14. J. Chen et al., "TransUNet: Transformers make strong encoders for medical image segmentation," *arXiv*, arXiv:2102.04306, 2021. <https://doi.org/10.48550/arXiv.2102.04306>.
15. H. Cao et al., "Swin-Unet: Unet-like pure transformer for medical image segmentation," in *Lect. Notes Comput. Sci.*, vol. 13803, 2022, pp. 205–218. https://doi.org/10.1007/978-3-031-25066-8_9.
16. F. N. Iandola et al., "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size," *arXiv*, arXiv:1602.07360, 2016.
17. J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 7132–7141.

18. Q. Lou, Y. Li, Y. Qian, F. Lu, and J. Ma, "Mammogram classification based on a novel convolutional neural network with efficient channel attention," *Comput. Biol. Med.*, vol. 150, p. 106082, 2022. <https://doi.org/10.1016/j.combiomed.2022.106082>.
19. Q. Zhou, Q. Wang, Y. Bao, L. Kong, X. Jin, and W. Ou, "LAEDNet: A lightweight attention encoder–decoder network for ultrasound medical image segmentation," *Comput. Electr. Eng.*, vol. 99, p. 107777, 2022. <https://doi.org/10.1016/j.compeleceng.2022.107777>.
20. A. Hatamizadeh, D. Terzopoulos, and A. Myronenko, "Boundary aware networks for medical image segmentation," in *Proc. Mach. Learn. Med. Imag. (MLMI)*, 2019, pp. 187–194. https://doi.org/10.1007/978-3-030-32692-0_22.
21. X. Qin et al., "Boundary-aware segmentation network for mobile and web applications," arXiv, arXiv:2101.04704, 2021.
22. F. Chen, H. Liu, Z. Zeng, X. Zhou, and X. Tan, "BES-Net: Boundary enhancing semantic context network for high-resolution image semantic segmentation," *Remote Sens.*, vol. 14, no. 7, p. 1638, Mar. 2022. <https://doi.org/10.3390/rs14071638>.
23. Y. Xiong, X. Shu, Q. Liu, and D. Yuan, "HCMNet: A hybrid CNN-Mamba network for breast ultrasound segmentation for consumer assisted diagnosis," *IEEE Trans. Consum. Electron.*, vol. 71, no. 3, pp. 8045–8054, 2025. <https://doi.org/10.1109/TCE.2025.3593784>.
24. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," arXiv, arXiv:1512.03385, 2015. <https://doi.org/10.48550/arXiv.1512.03385>.
25. A. Howard et al., "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2019, pp. 1314–1324. <https://doi.org/10.1109/ICCV.2019.00140>.
26. Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2020.11534–11542.
27. I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," arXiv, arXiv:1711.05101, 2019. <https://doi.org/10.48550/arXiv.1711.05101>.
28. W. Al-Dhabyani, M. Goma, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data Brief*, vol. 28, p. 104863, 2020. <https://doi.org/10.1016/j.dib.2019.104863>.
29. J. Tasnim and M. K. Hasan, "CAM-QUS guided self-tuning modular CNNs with multi-loss functions for fully automated breast lesion classification in ultrasound images," *Phys. Med. Biol.*, vol. 69, no. 1, p. 015018, 2024. <https://doi.org/10.1088/1361-6560/ad1319>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.