

BADF-Net: A Method for Enhancing Abdominal Ultrasound Diagnosis via Boundary-Aware Dynamic Fusion Network-Based Multi-Anatomical Structure Segmentation

Md. Rashed ¹ , Muhammad Jamil ^{2,3} , Md. Imran Hossain ^{1,*} , Adnan Kavak ^{2,3} , Md. Riad Hassan ⁴ , Ohidujjaman ⁵ , Md. Sarwar Hosain ¹ , Özgür Çakır ⁶  and Hossein Fotouhi ^{7,*} 

- ¹ Department of Information and Communication Engineering, Pabna University of Science and Technology, Pabna 6600, Bangladesh; rashedulislam.ice.pust@gmail.com (M.R.); imran05ice@pust.edu.bd (M.I.H.); sarwar.ice@pust.ac.bd (M.S.H.)
- ² Department of Computer Engineering, Kocaeli University, 41001 İzmit, Türkiye; jamil138.amin@gmail.com (M.J.); akavak@kocaeli.edu.tr (A.K.)
- ³ Wireless Information and Intelligent Systems (WINS) Research Center, Kocaeli University, 41001 İzmit, Türkiye
- ⁴ Department of Computer Science and Engineering, Green University of Bangladesh, Dhaka, Bangladesh; riad@cse.green.edu.bd (M.R.H.)
- ⁵ Department of Computer Science and Engineering, United International University, Dhaka 1212, Bangladesh; ohidujjaman@cse.uiu.ac.bd (O.)
- ⁶ Faculty of Medicine, Department of Radiology, Research and Application Hospital, Kocaeli University, 41001 İzmit, Türkiye; ozgur.cakir@kocaeli.edu.tr (Ö.Ç.)
- ⁷ Department of Computer Science and Engineering, Mälardalen University (MDU), 72123 Västerås, Sweden; hossein.fotouhi@mdu.se (H.F.)
- * Correspondence: imran05ice@pust.edu.bd (M.I.H.); akavak@kocaeli.edu.tr (A.K.); hossein.fotouhi@mdu.se (H.F.)

Abstract

Background: Automated segmentation of multiple anatomical structures in abdominal ultrasound could enhance clinical decision-making by providing consistent quantitative measurements for disease monitoring. However, ultrasound-specific artifacts including speckle noise, low contrast and weak boundaries challenge existing deep learning models. **Methods:** We propose BADF-Net, an architecture that integrates a Boundary-Aware Dynamic Attention (BADA) module with Boundary Confidence Estimation (BCE) at every stage of a squeeze-and-excitation (SE) recalibrated ResNet-34 encoder, a Cross-Scale Boundary Fusion (CSBF) module that adaptively combines multi-scale boundary cues into a shared guidance feature and a boundary-guided decoder with attention-gated skip fusion, to improve boundary localization and tissue discrimination. **Results:** On a multi-anatomical kidney and spleen ultrasound dataset, BADF-Net achieved a Dice score of 0.8174, Pixel Accuracy of 0.9703 and Mean Pixel Accuracy of 0.8521, outperforming 11 baselines with the lowest HD95 (11.49), and showing the largest gains in low-contrast structures such as kidney cortex and medulla. Agreement with expert radiologist annotations was strongest for well-defined structures such as the spleen (0.9786). By reducing operator dependency and inter-observer variability, BADF-Net can serve as a reliable second-reader tool in high-volume clinical workflows, processing images in 0.016 seconds for real-time integration. **Conclusions:** Our approach demonstrates how dynamic boundary attention, confidence-weighted feature refinement and channel recalibration jointly address key ultrasound segmentation challenges with direct clinical applicability.

Keywords: Ultrasound segmentation; Multi-anatomical structure segmentation; Boundary-aware attention; Squeeze-and-excitation; ResNet-34; Cross-Scale Boundary Fusion; Kidney and spleen segmentation; Clinical decision support;

Received:

Revised:

Accepted:

Published:

Copyright: © 2026 by the authors.

Submitted to *Journal Not Specified* for possible open access publication under the terms and conditions of the

[Creative Commons Attribution \(CC BY\) license](#).

1. Introduction

Ultrasound imaging is one of the most widely used diagnostic tools worldwide, alongside X-ray, CT, MRI and PET [1,2]. Its advantages include real-time imaging, portability and no radiation exposure, making it ideal for routine abdominal exams and repeated monitoring [3]. However, interpreting ultrasound images remains challenging even for experts [4–6]. The images contain speckle noise, acoustic shadows and low contrast between tissues [7,8]. These artifacts make it difficult to identify organ boundaries consistently.

Accurate multi-anatomical segmentation in abdominal ultrasound enables quantitative assessment of organ morphology, which is crucial for diagnosing and monitoring conditions like chronic kidney disease, spleen enlargement and abdominal pathologies. Automated tools could standardize measurements currently subject to inter-observer variability, potentially improving diagnostic consistency and reducing radiologist workload.

In clinical practice, radiologists often need to measure multiple anatomical structures during abdominal scans. Accurate segmentation of kidney substructures (cortex, medulla) and spleen regions supports quantitative assessment of diseases like chronic kidney disease [10,11]. Manual segmentation is time-consuming and varies between operators [12]. Automated tools could reduce this variability and save valuable clinical time.

Despite recent advances in deep learning, multi-anatomical structure (MAS) segmentation in ultrasound remains difficult. Models trained on cleaner modalities like MRI often fail on ultrasound due to its unique artifacts [9]. Even specialized ultrasound segmentation methods struggle with key problems: missing boundaries in low-contrast regions, leakage into background areas and confusion between visually similar tissues [13–15]. These limitations stem from inadequate handling of both spatial context and channel-wise feature importance [17,18].

Compared with MRI, CT and ultrasound show substantial speckle noise, acoustic shadowing, signal dropouts, intensity changes and weak boundaries between tissues [7,8]. This affects how well models can find organ boundaries. Standard models often produce masks with missing pieces. Performance also varies across patients and imaging devices [9]. The challenge is greater in MAS settings because anatomical regions may look similar, vary in size and appear partially hidden. This requires robust multi-scale representation while keeping fine spatial details.

This study has two main motivations. First, MAS segmentation in abdominal ultrasound can support clinically meaningful measurements like organ localization, thickness estimation and volume approximation. These are relevant for disease screening, treatment planning and monitoring over time [10,11]. Second, reliable MAS segmentation reduces differences between operators in manual drawing. This improves consistency and enables computer-aided diagnosis and decision support without adding much work for radiologists [12]. Specifically, accurate segmentation of kidney substructures (cortex and medulla) and spleen regions can help in assessing chronic kidney disease, inflammation and other conditions where organ shape and echo patterns matter clinically.

From an application view, we aim for radiology workflow automation and point-of-care ultrasound. Here, fast and stable results are needed across different machines and operator skill levels. In such settings, segmentation models must handle low signal-to-noise ratios and keep clinically important boundaries clear. They also need to be efficient enough for hospital systems or edge devices. So, robustness to artifacts and balanced accuracy across classes are key needs, along with good overall performance on easier regions.

Many studies have explored deep learning for ultrasound segmentation, including U-Net variants, multi-scale decoders and attention-based designs [13–16]. Recent methods improve context gathering or efficiency through multi-scale attention, lightweight decoders, or transformer-like parts. They show strong results in specific organ segmentation tasks. However, many current solutions are made for single-organ cases and do not fully address MAS needs in ultrasound. Models that focus on global context may lose boundary details. Models that focus only on local features may still be sensitive to speckle and background noise. Attention designs help with localization but often do not adjust channel importance well enough. This can lead to unstable results when multiple similar regions must be separated in one prediction [17,18]. Also, class imbalance (lots of background versus small tissue regions) can bias training toward easier cases. This hurts segmentation of smaller structures unless the loss function is designed to counter this effect [19,20].

Despite recent progress, existing multi-anatomical-structure (MAS) segmentation methods for ultrasound imaging exhibit four recurring limitations: (i) incomplete boundary recovery under low-contrast, noisy conditions; (ii) background leakage caused by poor filtering of irrelevant skip features; (iii) reduced accuracy across classes when adjacent tissues appear visually similar; and (iv) limited robustness across different imaging conditions and label sets. Addressing these limitations requires an architecture that jointly performs *spatial focus* (where to look, particularly at structure boundaries), *channel adjustment* (what to emphasize) and *multi-scale context aggregation*, within an encoder-decoder framework trained with a loss that explicitly handles class imbalance.

Based on these gaps, our paper answers the following research questions:

- **RQ1:** How can we design a segmentation architecture to improve boundary finding and reduce background errors for MAS in ultrasound?
- **RQ2:** Can a boundary-sensitive dynamic attention mechanism, combined with channel recalibration in the encoder, improve balanced accuracy across different anatomical regions?
- **RQ3:** How should we set the training loss to handle both overlap quality and class imbalance in multi-class ultrasound segmentation?
- **RQ4:** Does the proposed framework show consistent gains over strong baselines under the same experimental conditions?

Our study objectives are: (i) to develop a robust MAS segmentation method for ultrasound that improves feature learning and boundary localization under noise and weak edges; (ii) to create a combined multi-class dataset by merging kidney and spleen ultrasound labels for better MAS evaluation; (iii) to design a training approach and loss function that boost stability and class-balanced accuracy; and (iv) to compare thoroughly against current segmentation networks using the same preprocessing, training and metrics.

Table 1 summarizes how each architectural component maps to the limitations it is designed to address.

This paper contributes the following:

- BADF-Net for multi-class ultrasound segmentation, combining Boundary-Aware Dynamic Attention (BADA) with Boundary Confidence Estimation (BCE) at every stage of an SE-recalibrated ResNet-34 encoder, a Cross-Scale Boundary Fusion (CSBF) module for adaptive multi-scale boundary guidance and an attention-gated, boundary-guided decoder.
- A merged MAS dataset by combining kidney and spleen ultrasound labels into six classes.
- Training with a Dice and focal combined loss to manage overlap quality and class imbalance.
- Comparison with 11 baselines using identical training and test setups.

Table 1. Mapping between BADF-Net components and the limitations they address.

Component	Mechanism	Limitation Addressed	Ad-
SE channel recalibration (encoder)	Rescales channel activations via global context, per residual block	(iii) Tissue class confusion	
BADA module	Fuses standard + dilated conv branches into a sigmoid-gated attention map M	(i) Boundary loss; (ii) Background leakage	
Boundary Confidence Estimation (BCE)	Weights BADA-refined features by a learned per-pixel confidence map before fusion	(i) Boundary loss	
Cross-Scale Boundary Fusion (CSBF)	Adaptively weights and fuses BCE outputs across four encoder scales into a shared decoder-guidance feature	(iv) Robustness across conditions	
Dice-Focal loss	Weighted overlap + hard-example-focused loss (0.5/0.5)	Class imbalance	

The rest of the paper is organized as follows. Section 2 reviews related work. Section 3 presents the BADF-Net design. Section 4 describes experiments and evaluation. Section 5 presents results. Section 6 discusses findings. Section 7 concludes the paper.

2. Related Works

Deep learning has significantly improved ultrasound image segmentation in recent years. Many studies have proposed methods to increase accuracy and robustness under ultrasound-specific artifacts. Chen et al. (2024) [13] proposed MSCA-UNet for medical ultrasound image segmentation. The model uses a multi-scale module to generate feature maps at different scales and applies channel attention to adjust channel weights. A global feature module helps preserve edge details. The network improved Dice and IoU on several datasets. The authors also highlighted common ultrasound challenges, including noise, low contrast and uneven grayscale distribution, which make edge extraction difficult. In addition, the dataset size was limited. The study in [14] proposed a two-stage framework for breast ultrasound image segmentation. The first U-Net detects the lesion region and the second U-Net segments the lesion within the cropped region. This method performed better than a single-stage U-Net in their experiments. However, the final segmentation depends on correct detection in the first stage. If detection fails, the second-stage segmentation also fails. The method is also limited by the small size of the available dataset.

EMCAD [15] introduced an efficient attention decoder that reduces model complexity for multi-organ ultrasound segmentation. It reported a segmentation accuracy of 83.63% and outperformed several baseline methods. However, the study did not include real-world clinical validation. It also provided limited discussion of robustness under common clinical ultrasound artifacts. MISSFormer [16] and ConvTrans-Net [21] combined transformer-based designs with CNN components to capture global and local context. These architectures can model complex anatomical structures in ultrasound imaging. However, discriminating visually similar structures remains challenging in some cases. Another study [17] developed an Attention U-Net model for detecting carotid plaques in ultrasound images of the internal and common carotid arteries. It reported better performance than baseline models such as U-Net, U-Net++ and U-Net3P. However, the authors high-

lighted data selection limitations and suggested using multicenter datasets and stronger augmentation to improve generalization.

U-Net and Attention U-Net were originally designed for MRI and can achieve strong boundary delineation in kidney segmentation [22]. These architectures are also relevant for ultrasound segmentation. However, performance can degrade in ultrasound due to thick edges, low contrast boundaries and variations in organ size. CMUNeXt [18] was designed to capture global image context efficiently. It uses inductive bias, large kernels and inverted bottlenecks to combine distant and local spatial information. CMUNeXt reported strong performance compared with both heavyweight and lightweight segmentation models. Its efficiency can be useful when training data are limited. UNETR++ [19] is a hybrid model for 3D medical image segmentation. It introduced an efficient paired-attention block with a linear spatial attention module and shared parameters to achieve high performance with reduced computational cost. The authors noted limitations when segmenting multiple structures with rare abdominal shapes and suggested future work using improved augmentation strategies.

ResUNet++ [23] was proposed to improve polyp segmentation in colonoscopy data and achieved strong performance on that domain. However, it does not explicitly address the typical artifacts found in ultrasound images. FCNN [24], originally developed for CT, has been adapted to other modalities such as MRI and achieved strong results in multi-organ segmentation. However, it can still struggle on ultrasound images due to speckle noise and low contrast. The deep supervised CNN framework in [25] incorporated multi-scale inputs and pyramid pooling for kidney ultrasound segmentation and achieved high Dice scores. However, it was not designed for multi-organ or multi-anatomical structure segmentation.

BACANet [26] is a liver ultrasound segmentation model that integrates convolutional attention, boundary supervision and large kernels to achieve high accuracy with a lightweight design. However, capturing small anatomical structures remains challenging. Residual U-Net [27] introduced residual blocks and attention fusion to improve breast tumor ultrasound segmentation. It improved segmentation accuracy and gradient stability during training. However, the work does not clearly describe how the method handles difficult cases with strong ultrasound artifacts. The multi-modality framework in [28] distinguished between single-modality and multi-modality information through feature fusion. However, limitations can arise during alignment and high-level semantic fusion may reduce modality-specific details.

Kolmogorov-Arnold Networks [29] addressed limitations of linear modeling in standard transformer and CNN architectures. The method introduced a two-phase encoder-decoder design, with a convolutional phase for feature extraction and a tokenized KAN phase for modeling complex patterns. It achieved an IoU of 78.69% and an F1 score of 87.22%. However, a key gap remains in integrating KANs into general-purpose and real-time systems beyond validation tasks. MCAFT [30] combined transformer-based attention with frequency-domain analysis using a Discrete Wavelet Transform to capture both low-frequency structures and high-frequency edge details. It achieved a Dice score of 83.87 on the Synapse abdominal multi-organ dataset. However, processing high-resolution images can introduce substantial computational overhead.

PMFSNet [31] aimed to balance global and local features while reducing redundancy by simplifying the U-Net design and self-attention mechanisms. It achieved strong performance across several public datasets. However, reduced complexity may limit the ability to capture very fine anatomical details. Karimi et al. [32] designed a two-stage framework for spleen ultrasound segmentation. First, a SegFormerB0 model generates a coarse segmentation. Then, a Pix2Pix model refines boundaries and reduces errors. The

framework achieved a mean Dice of 96.82% and a mean IoU of 94.17%, outperforming traditional U-Net and SS-Net baselines. However, the approach focuses on a limited region of interest and may not generalize well to broader MAS settings.

In a nutshell, these studies show steady progress in ultrasound segmentation, but challenges remain in robust boundary delineation, feature discrimination across similar tissues and generalization under diverse acquisition conditions. Table 2 summarizes representative methods, their main contributions, limitations and reported performance; we address these limitations directly in the architecture proposed below.

Synthesis of Research Gaps

Reviewing these methods reveals a clear progression in ultrasound segmentation. Early CNN-based approaches like U-Net [22] established strong baselines but lacked mechanisms to handle ultrasound-specific artifacts. Attention mechanisms [17] and transformer-based designs [16,21] improved context modeling but often at high computational cost or with limited focus on channel discrimination. Several gaps remain unaddressed for multi-anatomical segmentation in ultrasound. First, most methods treat spatial attention as a single fixed-receptive-field operation, without dynamically fusing local and boundary-sensitive context and rarely combine this with channel recalibration in a unified encoder pathway [17,18]. Second, many approaches are validated on single-organ tasks but struggle with the class imbalance and anatomical diversity of MAS segmentation [19,20]. Third, few studies systematically evaluate both overlap metrics and boundary accuracy under consistent experimental conditions [13,15].

Our work addresses these gaps by proposing *BADF-Net*, which integrates squeeze-and-excitation-based channel recalibration directly into the ResNet-34 encoder with a dedicated Boundary-Aware Dynamic Attention (BADA) module at every encoder stage, a Cross-Scale Boundary Fusion (CSBF) module that adaptively aggregates boundary cues across scales and a boundary-guided decoder that propagates this fused signal to every decoding stage. Rather than applying a single fixed spatial filter, BADA dynamically fuses a standard convolution branch with a dilated convolution branch to jointly capture fine boundary detail and broader anatomical context, producing an attention map that suppresses irrelevant background before decoder fusion. This design allows the model to recalibrate *what* to emphasize at the channel level while dynamically adapting *where* to focus at the boundary level, within a single coherent encoder-decoder pathway. We validate this design on a challenging merged multi-anatomical (kidney and spleen) MAS dataset with rigorous comparisons against multiple baselines.

3. Proposed Methodology

We propose *BADF-Net*, a Boundary-Aware Dynamic Fusion Network for multi-anatomical structure segmentation in abdominal ultrasound images. The network follows an encoder-decoder architecture and integrates boundary-aware feature enhancement, confidence estimation, adaptive cross-scale fusion, attention-controlled skip connections and boundary-guided decoder refinement.

The overall framework consists of five principal components shown in figure 1: (i) an ImageNet-pretrained ResNet-34 encoder, (ii) Boundary-Aware Dynamic Attention (BADA) followed by Boundary Confidence Estimation (BCE) at four encoder levels, (iii) Cross-Scale Boundary Fusion (CSBF), (iv) a four-stage adaptive decoder and (v) a segmentation head followed by full-resolution interpolation.

Table 2. Summary of recent AI-based segmentation models, highlighting their main contributions, limitations, application domains and reported performance/efficiency.

References	Method	Key contributions and limitations	Dataset	Performance / efficiency
[13]	MSCA-UNet	Multi-scale feature extraction with channel attention and depth-separable convolutions to reduce parameters. Limitation: robustness may decrease under severe ultrasound noise and low-contrast boundaries.	Multi-organ ultrasound (4 datasets)	Outperformed prior methods; 29.82% fewer parameters than U-Net
[14]	Two-stage U-Net pipeline	Two U-Nets are used sequentially for lesion detection and fine segmentation in cropped regions. Limitation: segmentation performance depends on successful detection in the first stage; limited by small dataset size.	Breast ultrasound segmentation	Dice improvement: 1.8% overall; 14.5% for hard cases (Dice < 70%)
[15]	EMCAD	Efficient attention decoder with depth-wise convolutions and gated attention for reduced complexity. Limitation: lacks real-world clinical validation and provides limited analysis of robustness under typical clinical artifacts.	Multi-organ ultrasound (12 datasets)	1.91M parameters, 0.381G FLOPs; ~80% reduction in parameters/FLOPs
[16]	MISSFormer	Transformer-based segmentation that captures global and local context. Limitation: feature discrimination between visually similar structures can remain challenging.	Multi-organ and cardiac segmentation	State-of-the-art performance across tasks
[17]	Attention U-Net (cardiovascular)	Attention-augmented U-Net for carotid plaque segmentation in ultrasound. Limitation: generalization may be affected by data selection; multicenter data and stronger augmentation are recommended.	Carotid ultrasound	Correlation: 0.96; AUC: 0.97; surpassed U-Net, U-Net++ and U-Net3P
[18]	CMUNeXt	Efficient global-context modeling using large kernels and inverted bottlenecks. Limitation: may require careful tuning on very small datasets to avoid overfitting.	General medical image segmentation	Effective global context capture with low computational cost
[19]	UNETR++	Hybrid 3D segmentation model with an efficient paired-attention block and linear spatial attention. Limitation: multi-structure segmentation can be difficult for rare abdominal shapes; improved augmentation is suggested.	Synapse, BTCV, ACDC, BraTS, Decathlon Lung	Dice: 87.2%
[21]	ConvTrans-Net	CNN-Transformer hybrid with attention and MLP-based feature preselection. Limitation: separating visually similar tissues can still be challenging in ultrasound.	Breast ultrasound	Jaccard: 85.21%; Precision: 85.17%; Recall: 89.65%
[22]	U-Net and Attention U-Net (MRI)	Benchmark U-Net variants evaluated with standard and focal cross-entropy on MRI kidney segmentation. Limitation: performance can degrade in ultrasound due to thick edges and low-contrast boundaries.	Kidney segmentation in MRI	Dice: 0.966; strong boundary delineation
[23]	ResUNet++	Residual U-Net variant that improves boundary modeling for polyp segmentation. Limitation: not tailored to ultrasound-specific artifacts such as speckle and shadowing.	Kvasir-SEG, CVC-612	Outperformed U-Net and ResUNet; high Dice and mIoU
[24]	FCNN (CT → MRI)	Modality-robust FCNN originally developed for CT and adapted to MRI. Limitation: performance may degrade on ultrasound due to speckle noise and low contrast.	Multi-organ MRI segmentation	High Dice; modality-independent performance
[25]	Deep-supervised CNN (kidney)	Multi-scale inputs, pyramid pooling and deep supervision for kidney ultrasound segmentation. Limitation: not designed for multi-organ or MAS segmentation settings.	Kidney ultrasound	Dice: 95.86%; robust under typical ultrasound noise
[26]	BACANet	Lightweight liver ultrasound segmentation with convolutional attention, boundary supervision and large kernels. Limitation: small structures and fine details can be difficult to capture.	Liver ultrasound	Lightweight with high segmentation accuracy
[27]	Residual attention U-Net	Residual blocks with attention fusion to improve breast tumor ultrasound segmentation and training stability. Limitation: limited discussion of failure cases under strong ultrasound artifacts.	Breast ultrasound	Improved accuracy and gradient stability
[28]	Multi-modality segmentation framework	Fusion framework that separates single-modality and multi-modality information for segmentation. Limitation: high-level semantic alignment may reduce modality-specific details.	Multi-modality medical images	Improved over single-modality approaches
[29]	Kolmogorov-Arnold Networks (KAN)	Two-phase encoder-decoder combining convolutional feature extraction with tokenized KAN for complex pattern modeling. Limitation: limited evidence for real-time or general-purpose deployment beyond validation studies.	Medical image data	IoU: 78.69%; F1-score: 87.22%
[30]	MCAFT	Transformer-based segmentation with frequency-domain analysis via Discrete Wavelet Transform for edge/structure decomposition. Limitation: computational overhead can be high for high-resolution images.	Synapse abdominal multi-organ data	Dice: 83.87%
[31]	PMFSNet	Simplified U-Net and self-attention to balance global/local features with reduced redundancy. Limitation: reduced capacity may limit very fine anatomical detail capture.	Multi-organ and multi-task segmentation (7 datasets)	Strong performance across all evaluated datasets
[32]	SegFormerB0-Pix2Pix (two-stage)	Two-stage pipeline with SegFormerB0 for coarse masks and Pix2Pix refinement for fine boundaries. Limitation: region-of-interest focus may limit generalization to broader MAS settings.	Spleen ultrasound	Mean Dice: 96.82%; Mean IoU: 94.17%

For the multi-anatomical segmentation problem considered in this work, the model is configured with $K = 6$ classes:

$$\mathcal{K} = \{\text{background, capsule, central echo complex, medulla, cortex, spleen}\}.$$

(1)

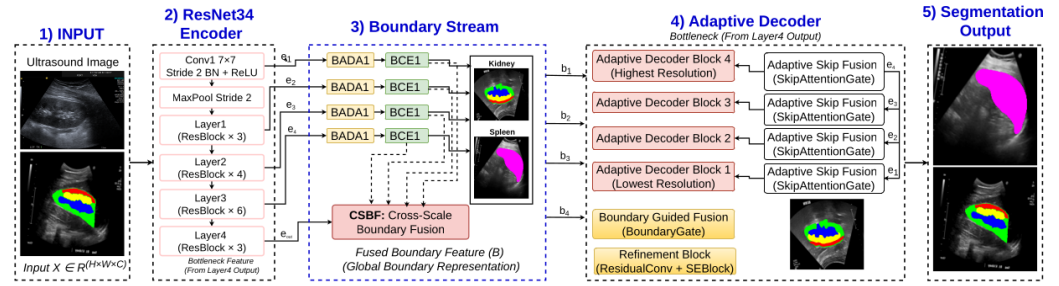


Figure 1. Overview of the proposed BADF-Net architecture. A ResNet-34 encoder produces skip features e_1 – e_4 , each refined by a BADA–BCE module and aggregated by Cross-Scale Boundary Fusion (CSBF) into a shared boundary feature B , which guides a four-stage adaptive decoder to produce the final segmentation. The internal structure of BADA/BCE and of CSBF/the decoder block is shown in Figures 2 and 3, respectively.

Let the input image batch be

$$\mathbf{X} \in \mathbb{R}^{N \times C_{in} \times H \times W}, \quad (2)$$

where N is the batch size, C_{in} denotes the number of input channels, and H and W denote image height and width, respectively. The implemented pretrained ResNet–34 receives three-channel input, i.e., $C_{in} = 3$.

3.1. Overall Architecture and ResNet–34 Encoder

BADF-Net employs an ImageNet-pretrained ResNet–34 as the feature encoder. The original ResNet–34 stem and residual stages are retained, allowing the network to exploit pretrained hierarchical representations while progressively reducing spatial resolution.

The first encoder feature is obtained as

$$\mathbf{E}_1 = \delta[\text{BN}(\text{Conv}_{7 \times 7, s=2}(\mathbf{X}))], \quad (3)$$

followed by max pooling and the four standard ResNet–34 residual stages:

$$\begin{aligned} \mathbf{E}_2 &= \text{Layer}_1(\text{MaxPool}(\mathbf{E}_1)), \\ \mathbf{E}_3 &= \text{Layer}_2(\mathbf{E}_2), \\ \mathbf{E}_4 &= \text{Layer}_3(\mathbf{E}_3), \\ \mathbf{E}_5 &= \text{Layer}_4(\mathbf{E}_4), \end{aligned} \quad (4)$$

where $\delta(\cdot)$ denotes ReLU and $\text{BN}(\cdot)$ denotes batch normalization.

The resulting encoder feature dimensions are

$$\begin{aligned} \mathbf{E}_1 &\in \mathbb{R}^{N \times 64 \times H/2 \times W/2}, \\ \mathbf{E}_2 &\in \mathbb{R}^{N \times 64 \times H/4 \times W/4}, \\ \mathbf{E}_3 &\in \mathbb{R}^{N \times 128 \times H/8 \times W/8}, \\ \mathbf{E}_4 &\in \mathbb{R}^{N \times 256 \times H/16 \times W/16}, \\ \mathbf{E}_5 &\in \mathbb{R}^{N \times 512 \times H/32 \times W/32}. \end{aligned} \quad (5)$$

Here, $\mathbf{E}_1$ – $\mathbf{E}_4$ serve as hierarchical skip features, whereas $\mathbf{E}_5$ represents the bottleneck. BADA and BCE are applied only to $\mathbf{E}_1$ – $\mathbf{E}_4$; the bottleneck is transferred directly to the decoder.

For convenience, the Conv–BN–ReLU operation used throughout the proposed modules is denoted as

$$\mathcal{C}(\mathbf{X}) = \delta[\text{BN}(\text{Conv}_{3 \times 3}(\mathbf{X}))]. \quad (6)$$

3.2. Boundary-Aware Dynamic Attention and Boundary Confidence Estimation

To strengthen anatomically relevant structures before decoder fusion, BADF-Net applies a BADA module independently to each encoder feature $\mathbf{E}_l$, where $l \in \{1, 2, 3, 4\}$. BADA combines efficient boundary-oriented feature extraction with channel and spatial attention, followed by residual and SE-based refinement.

Boundary-aware feature extraction.

The encoder feature is first processed by a depthwise-separable convolution:

$$\text{DSC}(\mathbf{E}_l) = \delta[\text{BN}(\text{PWConv}_{1 \times 1}(\text{DWConv}_{3 \times 3}(\mathbf{E}_l)))], \quad (7)$$

where $\text{DWConv}(\cdot)$ denotes depthwise convolution and $\text{PWConv}(\cdot)$ denotes pointwise convolution. A subsequent Conv–BN–ReLU operation produces

$$\mathbf{U}_l = \mathcal{C}(\text{DSC}(\mathbf{E}_l)). \quad (8)$$

Channel attention.

Both global average pooling and global max pooling are applied to $\mathbf{U}_l$. The two descriptors share the same channel transformation:

$$g(\mathbf{z}) = \mathbf{W}_2^c \delta(\mathbf{W}_1^c \mathbf{z}), \quad (9)$$

and the resulting channel-attention weights are

$$\mathbf{A}_l^c = \sigma[g(\text{GAP}(\mathbf{U}_l)) + g(\text{GMP}(\mathbf{U}_l))], \quad (10)$$

where $\sigma(\cdot)$ is the sigmoid function. The channel-refined representation is

$$\mathbf{U}_l^c = \mathbf{U}_l \odot \mathbf{A}_l^c. \quad (11)$$

Spatial attention.

Spatial attention is computed using channel-wise average and maximum responses:

$$\mathbf{A}_l^s = \sigma[\text{Conv}_{7 \times 7}([\text{Mean}_c(\mathbf{U}_l^c); \text{Max}_c(\mathbf{U}_l^c)])]. \quad (12)$$

The resulting attended boundary feature is

$$\mathbf{Q}_l = \mathbf{U}_l^c \odot \mathbf{A}_l^s. \quad (13)$$

The attended representation is added to the original encoder feature, and the result is further refined:

$$\tilde{\mathbf{E}}_l = \text{SE}[\mathcal{C}(\mathbf{E}_l + \mathbf{Q}_l)]. \quad (14)$$

The SE operation performs global channel recalibration. For a feature $\mathbf{F} \in \mathbb{R}^{N \times C \times H_f \times W_f}$, the channel descriptor is

$$z_c = \frac{1}{H_f W_f} \sum_{i=1}^{H_f} \sum_{j=1}^{W_f} F_{c,i,j}, \quad (15)$$

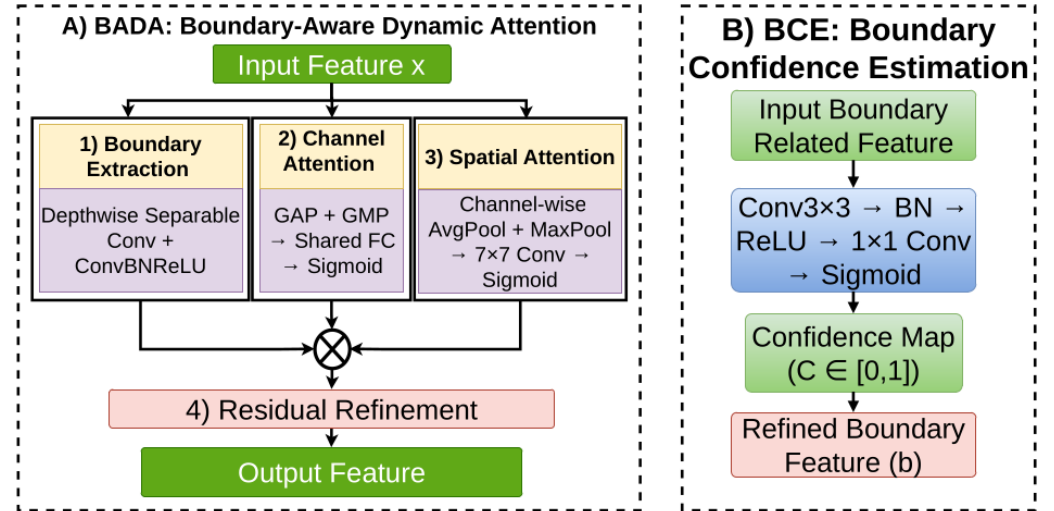


Figure 2. Internal structure of the boundary-refinement modules. (A) BADA fuses boundary-extraction, channel-attention and spatial-attention branches, followed by residual refinement. (B) BCE converts the BADA output into a per-pixel confidence map that weights the feature before it is passed to CSBF.

and the channel weights are

$$\mathbf{s} = \sigma(\mathbf{W}_2^{se} \delta(\mathbf{W}_1^{se} \mathbf{z})), \quad (16)$$

where the reduction ratio is $r = 16$. The output is

$$\text{SE}(\mathbf{F}) = \mathbf{F} \odot \mathbf{s}. \quad (17)$$

Boundary confidence estimation.

Because the reliability of enhanced features may vary spatially, the BADA-refined feature $\tilde{\mathbf{E}}_l$ is processed by BCE:

$$\mathbf{C}_l = \sigma \left[\text{Conv}_{1 \times 1} \left(\delta \left[\text{BN} \left(\text{Conv}_{3 \times 3} (\tilde{\mathbf{E}}_l) \right) \right] \right) \right], \quad (18)$$

where $\mathbf{C}_l \in [0, 1]^{N \times 1 \times H_l \times W_l}$ is the confidence map.

The BCE output is then

$$\mathbf{B}_l = \tilde{\mathbf{E}}_l \odot \mathbf{C}_l. \quad (19)$$

Thus, the four boundary-refined features have channel dimensions

$$(C_{B_1}, C_{B_2}, C_{B_3}, C_{B_4}) = (64, 64, 128, 256). \quad (20)$$

Here, $\odot$ denotes element-wise multiplication with appropriate broadcasting, $\mathbf{Q}_l$ is the BADA-attended feature, $\tilde{\mathbf{E}}_l$ is the final BADA-refined feature, $\mathbf{C}_l$ is the confidence map and $\mathbf{B}_l$ is the confidence-refined feature supplied to CSBF.

3.3. Cross-Scale Boundary Fusion

The four BCE outputs contain complementary boundary information at different spatial scales. CSBF combines these representations into a single shared boundary feature that guides every decoder stage.

First, each feature is projected to 256 channels:

$$\mathbf{P}_l = \delta[\text{BN}(\text{Conv}_{1 \times 1}(\mathbf{B}_l))], \quad l = 1, \dots, 4. \quad (21)$$

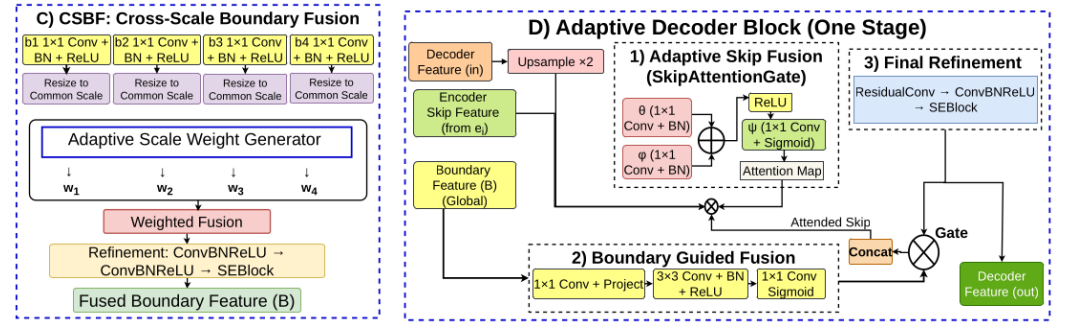


Figure 3. Internal structure of the fusion and decoding modules. (C) CSBF projects, resizes and adaptively weights the four BCE outputs into a single fused boundary feature B . (D) Each adaptive decoder block applies attention-gated skip fusion, boundary-guided gating from B , and residual refinement to reconstruct the decoder feature at that stage.

The fourth scale determines the target spatial resolution, $H/16 \times W/16$. Therefore,

$$\bar{P}_l = \mathcal{I}(P_l; H/16, W/16), \quad l = 1, 2, 3, \quad (22)$$

while

$$\bar{P}_4 = P_4. \quad (23)$$

Here, $\mathcal{I}(\cdot)$ denotes bilinear interpolation.

To estimate the relative contribution of each scale, global average pooling is performed:

$$g_l = \text{GAP}(\bar{P}_l). \quad (24)$$

The pooled descriptors are concatenated:

$$\mathbf{g} = [g_1; g_2; g_3; g_4]. \quad (25)$$

Adaptive scale weights are then generated by

$$\mathbf{w} = \text{Softmax}[\mathbf{W}_2^w \delta(\mathbf{W}_1^w \mathbf{g} + \mathbf{b}_1^w) + \mathbf{b}_2^w], \quad (26)$$

where

$$\mathbf{w} = [w_1, w_2, w_3, w_4], \quad \sum_{l=1}^4 w_l = 1. \quad (27)$$

The dynamically fused feature becomes

$$\mathbf{F}_B = \sum_{l=1}^4 w_l \bar{P}_l. \quad (28)$$

Finally, two Conv–BN–ReLU blocks and an SE block refine the fused feature:

$$\mathbf{B} = \text{SE}[\mathcal{C}_2(\mathcal{C}_1(\mathbf{F}_B))], \quad (29)$$

where

$$\mathbf{B} \in \mathbb{R}^{N \times 256 \times H/16 \times W/16}. \quad (30)$$

The resulting $\mathbf{B}$ represents a shared multi-scale boundary-guidance feature and is supplied to all four decoder stages.

3.4. Adaptive Boundary-Guided Decoder

The decoder reconstructs spatial resolution through four adaptive decoder blocks. Each block performs bilinear upsampling, attention-based skip fusion, boundary-guided modulation and residual refinement.

Adaptive skip fusion.

Let $\mathbf{D}$ denote an upsampled decoder feature and $\mathbf{E}$ the corresponding encoder skip feature. The skip-attention gate first transforms the two features:

$$\begin{aligned}\theta &= \text{BN}\left(\text{Conv}_{1 \times 1}^{\theta}(\mathbf{E})\right), \\ \phi &= \text{BN}\left(\text{Conv}_{1 \times 1}^{\phi}(\mathbf{D})\right).\end{aligned}\quad (31)$$

The spatial skip-attention coefficient is

$$\alpha = \sigma\left[\text{Conv}_{1 \times 1}^{\psi}(\delta(\theta + \phi))\right]. \quad (32)$$

The attended encoder feature is

$$\mathbf{E}^a = \mathbf{E} \odot \alpha. \quad (33)$$

The attended skip and decoder feature are concatenated and refined:

$$\mathbf{S} = \text{SE}[\mathcal{C}_2(\mathcal{C}_1([\mathbf{E}^a; \mathbf{D}]))], \quad (34)$$

where $[\cdot; \cdot]$ denotes channel-wise concatenation.

Boundary-guided refinement.

The shared CSBF feature $\mathbf{B}$ is resized to the current decoder resolution when required and projected to the decoder channel dimension:

$$\mathbf{B}^p = \delta[\text{BN}(\text{Conv}_{1 \times 1}(\mathcal{I}(\mathbf{B})))] \quad (35)$$

A boundary gate is generated as

$$\mathbf{G} = \sigma[\text{Conv}_{1 \times 1}(\delta[\text{BN}(\text{Conv}_{3 \times 3}(\mathbf{B}^p))])]. \quad (36)$$

The decoder representation is enhanced through residual multiplicative gating:

$$\mathbf{S}^g = \mathbf{S} \odot \mathbf{G} + \mathbf{S}. \quad (37)$$

A residual convolution and SE block are then applied:

$$\mathbf{F}_{bg} = \text{SE}[\text{ResConv}(\mathbf{S}^g)]. \quad (38)$$

For an input $\mathbf{X}$, the custom residual convolution is

$$\text{ResConv}(\mathbf{X}) = \mathcal{C}_2(\mathcal{C}_1(\mathbf{X})) + \text{Conv}_{1 \times 1}(\mathbf{X}). \quad (39)$$

Each adaptive decoder block applies one additional residual refinement:

$$\mathbf{D}_{\text{out}} = \text{SE}\left[\mathcal{C}\left(\text{ResConv}(\mathbf{F}_{bg})\right)\right]. \quad (40)$$

Before skip fusion, the incoming decoder feature is bilinearly upsampled by a factor of two. Therefore, the four-stage decoder is

$$\begin{aligned} \mathbf{D}_4 &= \text{ADB}_4(\mathbf{E}_5, \mathbf{E}_4, \mathbf{B}), \\ \mathbf{D}_3 &= \text{ADB}_3(\mathbf{D}_4, \mathbf{E}_3, \mathbf{B}), \\ \mathbf{D}_2 &= \text{ADB}_2(\mathbf{D}_3, \mathbf{E}_2, \mathbf{B}), \\ \mathbf{D}_1 &= \text{ADB}_1(\mathbf{D}_2, \mathbf{E}_1, \mathbf{B}), \end{aligned} \quad (41)$$

with

$$\begin{aligned} \mathbf{D}_4 &\in \mathbb{R}^{N \times 256 \times H/16 \times W/16}, \\ \mathbf{D}_3 &\in \mathbb{R}^{N \times 128 \times H/8 \times W/8}, \\ \mathbf{D}_2 &\in \mathbb{R}^{N \times 64 \times H/4 \times W/4}, \\ \mathbf{D}_1 &\in \mathbb{R}^{N \times 64 \times H/2 \times W/2}. \end{aligned} \quad (42)$$

The corresponding encoder, decoder and output channel configurations are

$$(256, 512, 256), \quad (128, 256, 128), \quad (64, 128, 64), \quad (64, 64, 64), \quad (43)$$

for decoder stages 4 to 1, respectively.

3.5. Segmentation Head and End-to-End Prediction

The final decoder representation

$$\mathbf{D}_1 \in \mathbb{R}^{N \times 64 \times H/2 \times W/2}$$

is passed to the segmentation head.

A Conv–BN–ReLU block first produces

$$\mathbf{H}_s = \mathcal{C}(\mathbf{D}_1), \quad (44)$$

after which spatial dropout with probability $p = 0.1$ is applied:

$$\mathbf{H}_d = \text{Dropout2D}_{0.1}(\mathbf{H}_s). \quad (45)$$

A 1×1 convolution produces K class logits:

$$\mathbf{Z}_{1/2} = \text{Conv}_{1 \times 1}^K(\mathbf{H}_d), \quad (46)$$

where

$$\mathbf{Z}_{1/2} \in \mathbb{R}^{N \times K \times H/2 \times W/2}. \quad (47)$$

The logits are finally bilinearly resized directly to the original input resolution:

$$\mathbf{Z} = \mathcal{I}(\mathbf{Z}_{1/2}; H, W) \quad (48)$$

with

$$\mathbf{Z} \in \mathbb{R}^{N \times K \times H \times W}. \quad (49)$$

The network returns the full-resolution logits $\mathbf{Z}$. For multi-class inference, class probabilities may subsequently be computed outside the network using

$$P_{n,k,i,j} = \frac{\exp(Z_{n,k,i,j})}{\sum_{q=1}^K \exp(Z_{n,q,i,j})}, \quad (50)$$

and the corresponding segmentation label is

$$\hat{Y}_{n,i,j} = \arg \max_k Z_{n,k,i,j}. \quad (51)$$

Here, $\mathbf{Z}$ denotes the output logit tensor, $P_{n,k,i,j}$ denotes the optional softmax probability of class k , and $\hat{Y}_{n,i,j}$ denotes the predicted class label. The softmax operation is used for interpretation of the logits and is not part of the network's internal forward output.

4. Experiments

This section describes the experimental setup used to evaluate BADF-Net. It includes the dataset, preprocessing, training settings, evaluation metrics and baseline methods. All settings follow the methodology in Section 3 to ensure a fair and reproducible evaluation. Table 3 summarizes the main hyperparameters.

Table 3. Hyperparameter settings used for training proposed BADF-Net for multi-anatomical ultrasound segmentation.

Hyperparameter	Value
Input image size	256×256
Number of classes	6
Batch size	4
Optimizer	AdamW
Initial learning rate	1×10^{-4}
Weight decay	1×10^{-4}
Maximum epochs	200
Loss function	Composite Loss ($0.5 \times \text{Dice} + 0.5 \times \text{Focal}$)
Focal loss α	0.25
Focal loss γ	2.0
Dice smoothing factor	1.0
Data augmentation	Enabled
Random seed	42
Hardware	CUDA (GPU)

4.1. Dataset Description and Preparation

Experiments were conducted on a curated multi-anatomical ultrasound dataset constructed by merging two publicly available datasets: a kidney ultrasound dataset [33] containing capsule, central echo complex, medulla and cortex annotations and a spleen ultrasound dataset [32] with pixel-level labels. The merged dataset enables multi-class segmentation under heterogeneous anatomical and imaging conditions, reflecting realistic clinical variability. Table 5 shows the summary of the ultrasound MAS dataset used in our experiments.

To create a unified MAS segmentation dataset, annotations from the kidney and spleen ultrasound datasets were mapped into a common label space, as shown in Table 4. The kidney dataset contributes four anatomical structures (capsule, central echo complex, medulla and cortex), while the spleen dataset contributes the spleen class. Background labels from both datasets were mapped to a shared background class (Class ID 0). Since the

Table 4. Annotation harmonization of the merged multi-anatomical structure (MAS) ultrasound dataset.

Dataset Source	Original Annotation	Unified Class ID
Kidney Dataset	Background	0
Kidney Dataset	Capsule	1
Kidney Dataset	Central Echo Complex (CEC)	2
Kidney Dataset	Medulla	3
Kidney Dataset	Cortex	4
Spleen Dataset	Background	0
Spleen Dataset	Spleen	5

Table 5. Class presence counts in the merged MAS dataset. The numbers show how many images contain each labeled structure.

Tissue Type	Capsule	Central Echo Complex	Medulla	Cortex	Spleen
No. of images with label	534	534	534	534	450
Labeled Color	red	green	blue	yellow	purple

two datasets contain distinct anatomical structures rather than duplicate annotations of the same organ, no anatomical class was annotated by multiple sources. This harmonization enables joint training and evaluation within a consistent six-class segmentation framework.

Label standardization:

We unify the labels from both datasets into one shared label set with $K = 6$ classes. The class order is background, capsule, central echo complex, medulla, cortex and spleen. We convert each mask into a single-channel label map. We also convert masks to one-hot format during training.

All images were resized to 256×256 pixels and normalized to the range $[0, 1]$. We use $K = 6$ classes, including background. Ground-truth masks are converted to one-hot encoding to match the softmax output.

Data Split Strategy:

We divided the dataset into training (80%), validation (10%) and (10%) test sets. The validation set monitored training for early stopping, while the test set remained completely untouched during training. All reported results come from this independent test set. When patient identifiers were available, we ensured no patient appeared in multiple splits. This prevents data leakage and gives a realistic estimate of clinical performance [12].

Hyperparameter Selection:

Key hyperparameters were chosen based on validation performance, an ablation study on the loss function formulation (Table 9) and established practices. We used ResNet-34 as it offered the best balance between depth and efficiency in preliminary tests. For the loss function, we evaluated ten formulations – including Dice-only, Focal-only, Tversky-only, standard cross-entropy and several composite weightings – and selected an equal-weighted combination of Dice and focal loss (0.5 Dice / 0.5 Focal), which achieved the highest overall test Dice score (0.8174) among all variants tested, outperforming both single-term losses and other composite weightings such as the three-term 0.5 Dice + 0.3 Focal + 0.2 Tversky combination (0.7792) and the equal-weighted Dice + Tversky combination using a recall-favoring Tversky parameterization ($\alpha = 0.3, \beta = 0.7$) (0.7633). The focal parameters were set to $\alpha = 0.25$ and $\gamma = 2.0$, following common usage for emphasizing hard, minority-class examples [22]. The Dice smoothing factor was set to 1.0 to avoid division instability on classes with few or no foreground pixels. We used the AdamW optimizer with an initial

learning rate of 1×10^{-4} and weight decay of 1×10^{-4} , which provided stable convergence across all compared methods [13,15].

4.2. Implementation Details and Training Protocol

We implemented BADF-Net in PyTorch. All experiments were conducted on a high-performance workstation equipped with an NVIDIA GeForce RTX 4090 GPU with 24 GB of memory (CUDA-enabled). The reported inference time corresponds to the average forward-pass time per image on the independent test set using an input resolution of 256×256 and a batch size of 1 under evaluation mode. The measured average processing time was 0.016 s per image. The reported timing excludes model loading and preprocessing overhead.

We initialize the ResNet-34 encoder with ImageNet-pretrained weights. All decoder layers, BADA modules and SE blocks are trained from scratch. We train using the AdamW optimizer with an initial learning rate of 1×10^{-4} , weight decay of 1×10^{-4} and a batch size of 4 for a maximum of 200 epochs, using 256×256 input images across all 6 segmentation classes (Table 3). The model is optimized using the composite Dice-Focal loss ($0.5 \times \text{Dice} + 0.5 \times \text{Focal}$, with focal $\alpha = 0.25$, $\gamma = 2.0$ and Dice smoothing factor 1.0), selected via the ablation study in Table 9. A fixed random seed (42) is used throughout training for reproducibility.

We use early stopping based on validation loss to reduce overfitting. We apply data augmentation during training, including random rotation, horizontal flip and intensity scaling.

To ensure reproducibility, the same preprocessing steps, augmentation strategies and hyperparameters were applied to all baseline models. Code and trained weights are available on GitHub for verification. This allows other researchers to verify our results and build upon this work.

4.3. Reproducibility checklist

To keep the comparison fair, we keep the same settings for all methods.

- Fixed train, validation and test splits.
- Same preprocessing and input size (256×256) for all methods.
- Same augmentations for all methods (rotation, flip and intensity scaling).
- Same optimizer (AdamW), learning rate (1×10^{-4}), weight decay (1×10^{-4}), batch size (4) and epochs (200).
- Same loss setup for all methods (Dice and focal loss, weighted equally at $\lambda = 0.5$, with focal parameters $\alpha = 0.25$ and $\gamma = 2.0$).
- Fixed random seed (42) for reproducibility across all runs.
- Code and weights: <https://github.com/rashed200613/BADF-Net>.

4.4. Loss Function

To address class imbalance and boundary precision in multi-class segmentation, the training objective combines Dice loss and focal loss. Let $\hat{Y}_{i,j,k}$ and $Y_{i,j,k}$ represent the predicted probability and ground-truth label for class k at pixel (i, j) , in the corresponding order and let K denote the total number of classes.

The multi-class soft Dice loss is defined as

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{1}{K} \sum_{k=1}^K \frac{2 \sum_{i,j} \hat{Y}_{i,j,k} Y_{i,j,k} + \epsilon}{\sum_{i,j} \hat{Y}_{i,j,k} + \sum_{i,j} Y_{i,j,k} + \epsilon}, \quad (52)$$

where $\epsilon = 1.0$ is a smoothing constant added for numerical stability.

The focal loss is given by

$$\mathcal{L}_{\text{Focal}} = \frac{1}{HW} \sum_{i,j} \alpha (1 - p_t)^\gamma \text{CE}(i, j), \quad (53)$$

where $\text{CE}(i, j)$ denotes the standard pixel-wise cross-entropy between the prediction and the ground-truth label at pixel (i, j) , $p_t = \exp(-\text{CE}(i, j))$ is the model's estimated probability for the correct class, γ controls the focusing strength on hard, misclassified pixels and α balances the contribution of the modulating factor. In our implementation, $\alpha = 0.25$ and $\gamma = 2.0$.

The final training loss is computed as a weighted combination of the two terms:

$$\mathcal{L} = \lambda \mathcal{L}_{\text{Dice}} + (1 - \lambda) \mathcal{L}_{\text{Focal}}, \quad (54)$$

with $\lambda = 0.5$ in all experiments, giving equal weight to region-level overlap (Dice) and pixel-level, hard-example-focused (Focal) supervision.

4.5. Comparative Baselines

The proposed BADF-Net was evaluated against other established segmentation models reported in the literature, including but not limited to Att-UNet, CMUNeXt, UDABUNet, UNet, UNet++, ResUNet++, UDBRNet, Residual U-Net, MISSFormer, EMCAD and MultiResUNet. All baseline models were trained and evaluated under identical conditions, including dataset splits, input resolution, loss functions and optimization settings, to ensure balanced comparison.

5. Results

This section reports the quantitative and qualitative performance of the proposed BADF-Net for multi-anatomical structure (MAS) segmentation in ultrasound (US) images. All results are computed on the independent test set using the same preprocessing, training configuration and composite Dice-Focal objective defined in the Experiments section. We first formalize the evaluation metrics (consistent with the confusion-matrix formulation), then present overall and class-wise quantitative comparisons against competitive baselines, followed by qualitative visual assessment, confusion-matrix analysis, ablation of each component of architecture and loss-function ablation study that validates the selected training objective. Performance is evaluated using standard metrics that measure both overlap accuracy (Dice, IoU) and boundary precision (Hausdorff Distance, ASD). All results are reported with mean values across the test set.

5.1. Evaluation Metrics

We evaluate segmentation performance using standard overlap, pixel-wise and boundary metrics. The metrics are defined below.

- Pixel Accuracy (PA) is the ratio of correctly classified pixels to the total number of pixels.
- Mean Pixel Accuracy (MPA) is the mean of per-class pixel accuracies.
- Precision is the ratio of correctly predicted positive pixels to all predicted positive pixels.
- Recall is the ratio of correctly predicted positive pixels to all ground-truth positive pixels.
- Dice Similarity Coefficient (Dice) measures region overlap between prediction and ground truth.

- HD95 is the 95th percentile Hausdorff distance between the predicted boundary and the ground-truth boundary.
- Average Symmetric Surface Distance (ASD) is the mean symmetric distance between the predicted boundary and the ground-truth boundary.

All presented metrics are computed on the test set and aggregated across classes unless otherwise specified.

5.2. Overall Quantitative Results

Table 6 summarizes the overall quantitative comparison between the proposed BADF-Net and multiple established segmentation baselines under identical training and evaluation conditions. The proposed approach achieves the highest mean Dice score (0.8174), substantially outperforming the next-best baseline, Att-UNet (0.7049), by more than 11 percentage points. BADF-Net also achieves the highest Pixel Accuracy (PA, 0.9703) and the highest Mean Pixel Accuracy (MPA, 0.8521), indicating improved overlap accuracy and markedly better class-balanced pixel performance across anatomical structures, with the MPA gap over the strongest baseline (CMUNeXt, 0.7546) exceeding 9 percentage points.

The proposed model also yields the highest precision (0.7744), recall (0.7759) and F1-score (0.7741) among all compared methods, demonstrating stronger and more balanced sensitivity in capturing foreground regions, which is particularly important in ultrasound, where faint boundaries often lead to under-segmentation.

For boundary-based metrics, BADF-Net achieves the lowest HD95 (11.49) among all baselines, indicating more accurate contour localization under challenging ultrasound conditions. Its ASD (2.51) is close to the best-observed value (CMUNeXt, 2.49) and clearly outperforms the remaining baselines, several of which exceed an ASD of 3.0. Despite this strong overlap and boundary performance, BADF-Net is also the fastest model at inference, with an average processing time of 0.016 s per image, faster than every baseline reported, including much lighter-performing models such as ResUNet++ (0.029 s) and UNet (0.030 s).

While Att-UNet achieves the lowest training loss (0.0822), it yields substantially lower Dice and MPA scores on our test set. This confirms that a lower training loss does not necessarily translate into better class-balanced segmentation under ultrasound artifacts; BADF-Net attains a higher training loss (0.1407) yet delivers markedly stronger overlap, boundary and pixel-wise accuracy. Similarly, CMUNeXt shows comparable Dice performance among the baselines (0.7005) and the best ASD overall, but BADF-Net still achieves higher Dice, MPA, precision, recall and F1-score, along with faster inference, which is consistent with the boundary-aware and channel-recalibration design of our architecture. Overall, the combination of highest Dice, PA, MPA, precision, recall and F1-score, together with the lowest HD95 and fastest inference time, indicates that BADF-Net achieves the best balance of accuracy, boundary fidelity and computational efficiency among all evaluated methods.

Practical Significance of Improvements:

The absolute Dice improvement of BADF-Net over the closest baseline (0.8174 vs. 0.7049 for Att-UNet) is substantial rather than marginal, representing an improvement of over 11% points and it reflects meaningful clinical progress. In ultrasound segmentation, even small improvements in challenging regions like kidney medulla and cortex are valuable [25]. These areas are clinically important but difficult to segment due to similar textures and weak boundaries [7,8]. Our model's consistent gains across multiple metrics (Dice, PA, MPA, precision, recall, F1-score and boundary distances) and across anatomical classes suggest robust improvement rather than marginal optimization. The improved boundary metrics (HD95: 11.49, ASD: 2.51) are particularly relevant for clinical measure-

Table 6. Comparative performance analysis of BADF-Net and other state-of-the-art (SOTA) models for multi-anatomical structure ultrasound segmentation. ↑ and ↓ denote higher-is-better and lower-is-better metrics, respectively. Abbreviations: HD95 – 95th percentile Hausdorff Distance; ASD – Average Symmetric Distance; PA – Pixel Accuracy; MPA – Mean Pixel Accuracy; Avg. – Average.

Method	Dice Score ↑	HD95↓	ASD↓	Avg. Time (s)↓	Loss↓	PA↑	MPA↑	Precision↑	Recall↑	F1-Score↑
Att-UNet	0.7049	12.02	2.63	0.053	0.0822	0.9688	0.7514	0.7565	0.7514	0.7537
CMUNeXt	0.7005	12.61	2.49	0.031	0.0856	0.9689	0.7546	0.7489	0.7546	0.7477
UDBA-UNet	0.6991	14.50	3.04	0.032	0.1059	0.9674	0.7436	0.7383	0.7436	0.7368
UNet	0.6953	13.80	2.98	0.030	0.1533	0.9673	0.7407	0.7454	0.7407	0.7403
UNet++	0.6927	15.51	3.46	0.042	0.1791	0.9665	0.7261	0.7490	0.7261	0.7354
ResUNet++	0.6924	11.98	2.63	0.029	0.1007	0.9665	0.7240	0.7451	0.7240	0.7327
UDBRNet	0.6017	13.18	3.11	0.040	0.2142	0.9590	0.6923	0.7104	0.6923	0.6976
Residual U-Net	0.5731	15.95	3.89	0.048	0.2082	0.9571	0.6470	0.7299	0.6470	0.6846
MultiResUNet	0.4761	26.27	10.04	0.036	0.3171	0.7892	0.5369	0.5375	0.5669	0.5603
MISSFormer	0.7028	13.02	2.61	0.047	0.1105	0.9589	0.6922	0.7103	0.6922	0.6976
EMCAD	0.6170	13.15	3.09	0.030	0.1420	0.9492	0.6368	0.6344	0.6368	0.6302
Proposed	0.8174	11.49	2.51	0.016	0.1407	0.9703	0.8521	0.7744	0.7759	0.7741

Table 7. Per-tissue Dice score comparison of BADF-Net and other state-of-the-art (SOTA) models for ultrasound image segmentation. ↑ denotes higher is better.

Method	Capsule ↑	Central Echo Complex ↑	Medulla ↑	Cortex ↑	Spleen ↑	Average ↑
Att-UNet	0.6402	0.8077	0.5799	0.5351	0.9615	0.7049
CMUNeXt	0.6092	0.8161	0.5952	0.5131	0.9689	0.7005
UNet	0.6065	0.8063	0.6039	0.4983	0.9616	0.6953
UDBA-UNet	0.6109	0.8172	0.5999	0.4970	0.9706	0.6991
UNet++	0.6194	0.7833	0.5836	0.5093	0.9679	0.6927
ResUNet++	0.6145	0.8119	0.5849	0.4932	0.9574	0.6924
UDBRNet	0.5213	0.7124	0.5023	0.4515	0.8209	0.6017
Residual U-Net	0.5750	0.7173	0.5379	0.4990	0.5365	0.5731
MultiResUNet	0.3849	0.6374	0.4343	0.3056	0.6181	0.4761
MISSFormer	0.6522	0.7653	0.5773	0.5936	0.9258	0.7028
EMCAD	0.5711	0.7125	0.5376	0.4907	0.7729	0.6170
Proposed	0.7404	0.8405	0.7525	0.7751	0.9786	0.8174

ments where precise contours matter [10,11] and the gains are achieved together with the fastest inference time among all evaluated methods (0.016 s per image), indicating that the improvement in accuracy does not come at the cost of computational efficiency.

5.3. Per-Tissue Dice Score Analysis

Table 7 reports per-tissue Dice scores for Capsule, Central Echo Complex, Medulla, Cortex and Spleen. The proposed BADF-Net achieves the highest average Dice score (0.8174) and outperforms all baselines on every individual tissue. The largest gains appear in Cortex (0.7751 vs. 0.5351 for the best baseline, Att-UNet) and Medulla (0.7525 vs. 0.6039 for the best baseline, UNet) – the two structures typically hardest to segment due to low contrast and close proximity of kidney sub-regions. Consistent gains are also seen in Capsule (0.7404) and Central Echo Complex (0.8405), while Spleen remains the easiest structure overall (0.9786). These gains, concentrated most heavily on the low-contrast, boundary-ambiguous tissues, are consistent with BADF-Net’s Boundary-Aware Dynamic Attention (BADA) module, Boundary Confidence Estimation (BCE) and SE-based channel recalibration, which jointly emphasize boundary-relevant regions and discriminative channels before decoder fusion.

5.4. Qualitative Results

Figure 4 and Figure 5 provide representative qualitative comparisons between the proposed BADF-Net and baseline models. Traditional encoder-decoder models (e.g., UNet and UNet++) generally recover coarse organ shapes but often produce blurred boundaries and missing sub-regions under speckle noise. Residual and multi-resolution variants (e.g., Residual U-Net and MultiResUNet) may partially preserve structure but frequently exhibit boundary leakage and fragmented predictions, particularly around ambiguous

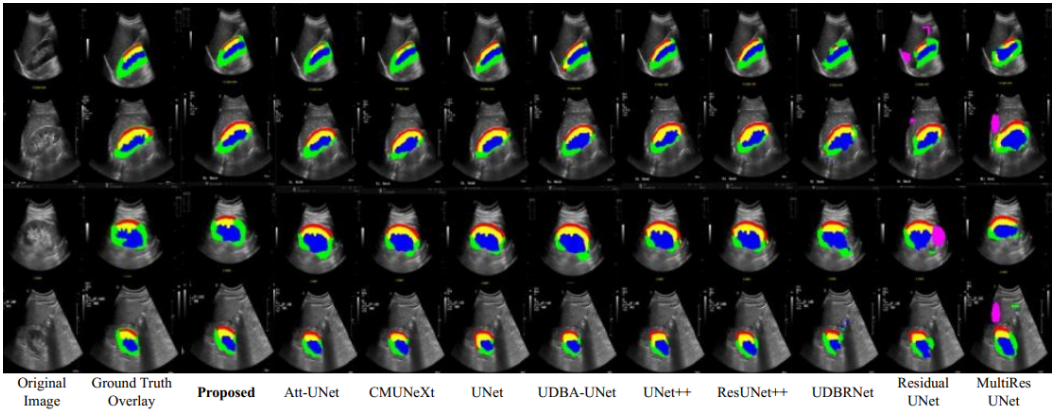


Figure 4. Visual comparison of MAS segmentation results on the kidney dataset using BADF-Net and Other SOTA models.

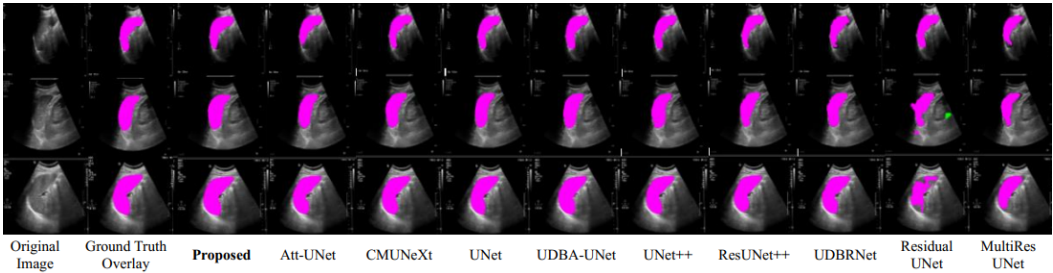


Figure 5. Visual comparison of MAS segmentation results on the spleen dataset using BADF-Net and Other SOTA models

tissue transitions. In contrast, the proposed BADF-Net produces smoother and more anatomically consistent boundaries and better separation between adjacent kidney tissues. These improvements visually support the quantitative findings and reflect the effects of the Boundary-Aware Dynamic Attention (BADA) module and Boundary Confidence Estimation (BCE) in suppressing irrelevant background activations while preserving fine boundary detail and the SE block in strengthening informative channels for fine-structure discrimination.

5.5. Confusion Matrix and Metric Visualization

Figure 6 shows the confusion matrix of the proposed BADF-Net, where dominant diagonal entries indicate strong per-class recognition. Minor confusions are primarily observed between anatomically adjacent regions, which is expected due to weak boundaries and texture similarity in US imaging.

To further investigate the per-structure behavior underlying the aggregate Dice score reported in Table 6, we analyzed the distribution of Dice scores across the five anatomical classes on the test set, as shown in Figure 7.

5.6. Ablation Study

We ran two ablation studies to check whether our architectural design and loss function choices actually help. The first study removes or changes one architectural component at a time. The second study compares different loss function combinations.

Table 8 shows the architectural ablation results. The largest performance drops come from the confidence/boundary fusion components: removing the decoder Boundary Gate reduces Dice to 0.7530 (−0.0644) and removing Boundary Confidence Estimation gives a nearly identical drop to 0.7533 (−0.0641), indicating that boundary-aware gating and confidence weighting are the most critical elements for accurate boundary recovery. Re-

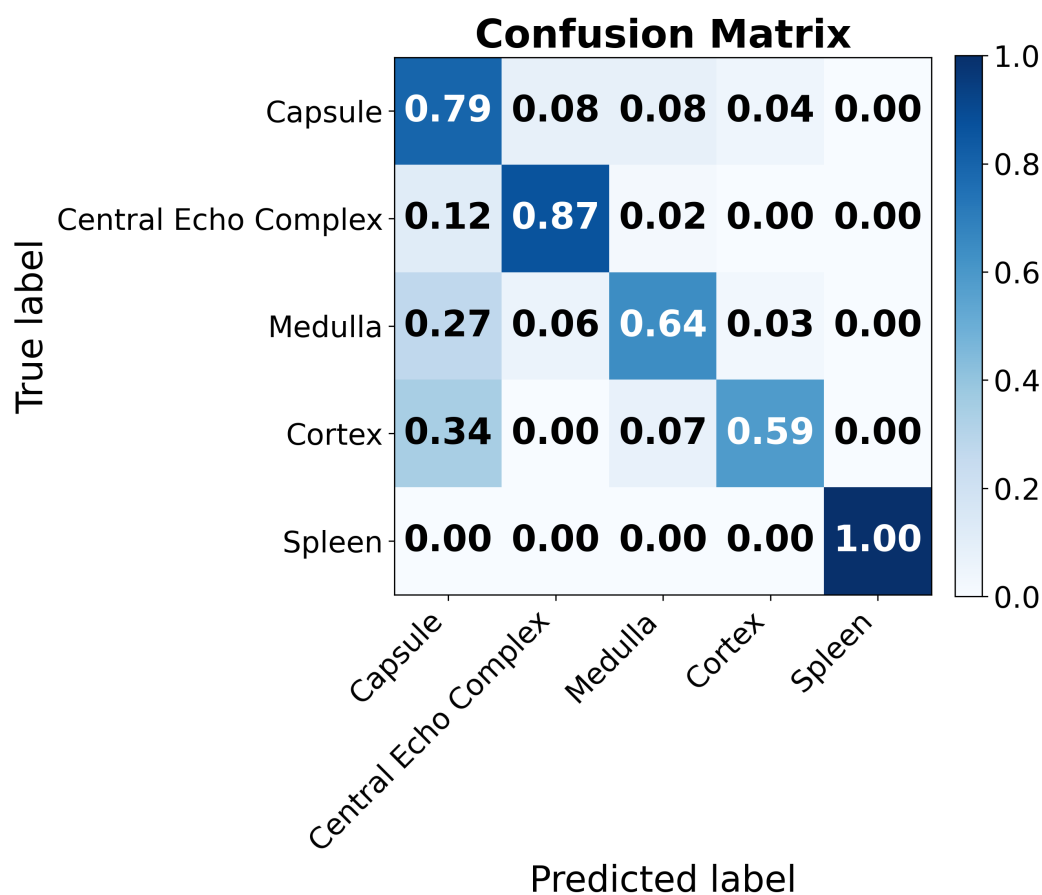


Figure 6. Confusion matrix for MAS segmentation using our proposed BADF–Net model



Figure 7. Box plots of per-class Dice scores for the proposed model on the test set, showing the median (orange), mean (green triangle), interquartile range and outliers (circles) for each anatomical structure.

placing the CSBF module with a simple concat + 1×1 conv also causes a substantial

drop (0.7725) and even using fixed equal scale weights within CSBF reduces performance (0.7870), confirming that the learned, dynamic weighting of confidence and boundary cues is responsible for most of this gain.

Within the BADA module, removing it entirely lowers Dice to 0.7683 (-0.0491), confirming that dynamically fusing local and dilated context before decoder fusion also plays a meaningful role in boundary recovery and background suppression, though its individual contribution is smaller than that of the confidence/boundary fusion components. Replacing BADA's attention-gated skip fusion with plain concatenation likewise reduces performance (0.7806), showing that the learned attention map, rather than simple feature concatenation, accounts for part of this gain; using only channel or only spatial attention within BADA yields intermediate results (0.7752 and 0.7853, respectively), suggesting the two attention types are complementary.

Removing all SE blocks causes a smaller but still consistent drop (0.7885), indicating that channel-wise recalibration in the encoder contributes to, but is not solely responsible for, the overall performance gain. Using a lighter ResNet18 backbone instead of ResNet34 reduces performance more substantially (0.7636), suggesting that encoder depth matters for capturing the multi-scale anatomical context needed for this task. Removing the decoder's residual refinement also degrades performance (0.7815), highlighting its role in final boundary sharpening. Overall, every component we tested contributes to the final result, with the confidence/boundary fusion mechanism proving most critical; the full BADF-Net, combining CSBF, BADA, SE-based recalibration, residual refinement and the ResNet34 backbone, performs best with a Dice score of 0.8174.

Table 9 shows the loss function results. Using just one loss term, Dice, Focal, Tversky, or cross-entropy, performs worse than our combined loss. This means no single loss captures everything we need: good region overlap and attention to hard pixels. Adding a third Tversky term, either in an equal three-way combination or paired with Dice instead of Focal, does not beat our two-term loss. This suggests Tversky's precision-recall balance does not add much once Focal loss is already handling hard boundary pixels. Combining Dice with cross-entropy instead of Focal also gives a lower score (0.7836), which supports our choice of Focal loss for down-weighting easy background pixels. Our final loss, an equal combination of Dice and Focal loss ($0.5 \mathcal{L}_{\text{Dice}} + 0.5 \mathcal{L}_{\text{Focal}}$), gives the best overall Dice score of 0.8174.

These two studies together show that our architecture and loss function both matter and combining them gives the most accurate and reliable segmentation across different anatomical structures.

5.7. Clinical Validation with Expert Comparison

To assess clinical relevance beyond agreement with the original dataset's ground truth, we evaluated BADF-Net against an independent, blinded expert radiologist annotation on a separate 50-case subset. BADF-Net achieved a Dice agreement of 0.78 with the radiologist's independent annotations and 88% of the 50 predicted segmentations were judged clinically acceptable for measurement purposes. This independent agreement score is somewhat lower than the 0.8174 overall Dice reported against the original dataset ground truth in Table 6, which is expected: the radiologist's masks reflect an entirely independent annotation process, unconstrained by the original dataset's specific labeling conventions and some divergence between two independent expert annotators is a normal feature of manual ultrasound segmentation rather than a model failure. The 88% clinical-acceptability rate indicates that, even where pixel-level Dice agreement is imperfect, the large majority of BADF-Net's predictions were still judged usable for the quantitative measurements (e.g., organ dimensions) that motivate this work.

Table 8. Ablation study of BADF-Net components on the test set. Dice denotes the Dice similarity coefficient (higher is better); Δ denotes the change in Dice relative to the full model. CSBF: Cross-Scale Boundary Fusion; BCE: Boundary Confidence Estimation; BADA: Boundary-Aware Dynamic Attention; SE: Squeeze-and-Excitation; DW-sep: depthwise-separable convolution. **BADF-Net** denotes our proposed model.

Group	Variant	Dice $\uparrow$	Δ
Conf./ Boundary fusion	Remove decoder Boundary Gate	0.7530	−0.0644
	Remove Boundary Confidence Estimation (BCE)	0.7533	−0.0641
	CSBF replaced with concat + 1×1 conv	0.7725	−0.0449
	CSBF with fixed equal scale weights	0.7870	−0.0304
	BADA: no channel/spatial attention	0.7635	−0.0539
BADA (internal)	Boundary extraction: standard conv (no DW-sep)	0.7653	−0.0521
	Removed Boundary-Aware Dynamic Attention (BADA)	0.7683	−0.0491
	BADA: channel attention only	0.7752	−0.0422
	Skip fusion: plain concat (no attention gate)	0.7806	−0.0368
Backbone/ encoder	BADA: spatial attention only	0.7853	−0.0321
	Lighter backbone: ResNet18 instead of ResNet34	0.7636	−0.0538
	Remove all SE blocks	0.7885	−0.0289
Decoder	Remove residual refinement in decoder	0.7815	−0.0359
Proposed	Full BADF-Net	0.8174	—

Table 9. Ablation study on the loss function formulation. All variants are evaluated on the test set using the overall Dice score.

Description	Test Dice (Overall)
0.5 Dice + 0.3 Focal + 0.2 Tversky	0.7792
Dice only	0.7571
Focal only	0.7868
Tversky only ($\alpha = 0.7, \beta = 0.3$)	0.7525
Standard Cross-Entropy only	0.7898
Dice + CE (0.5 / 0.5)	0.7836
Dice + Tversky (0.5 / 0.5), Focal dropped	0.7411
Composite with equal weights (0.33/0.33/0.33)	0.7886
Composite, Tversky $\alpha = 0.3/\beta = 0.7$ (favor recall)	0.7633
Dice + Focal (0.5 / 0.5) (Proposed)	0.8174

6. Discussion

We discuss why BADF-Net improves MAS segmentation under ultrasound-specific artifacts, its practical clinical relevance and remaining limitations that motivate future work. Table 10 presents a comparison of various FCN architectures for MAS segmentation in US imaging.

6.1. Effect of the Proposed Architectural Design

The performance gains of BADF-Net are consistent with the boundary-aware, channel-recalibrated design introduced in Section 3. The ResNet-34 encoder improves representation robustness through residual learning, facilitating stable gradient propagation and multi-scale feature extraction for anatomically diverse regions. However, US skip features can include irrelevant clutter and speckle-driven activations. The Boundary-Aware Dynamic Attention (BADA) module mitigates this by dynamically fusing a standard convolution branch with a dilated convolution branch into a single boundary-aware attention

Table 10. Comparison table demonstrating the effectiveness of BADF-Net against SOTA models on same dataset.

Ref.	Method	Dice (%)
[17]	Att-UNet	70.49
[18]	CMUNeXt	70.05
[34]	UDBA-UNet	69.91
[22]	UNet	69.53
[35]	UNet++	69.27
[23]	ResUNet++	69.24
[36]	UDBRNet	60.17
[21]	Residual U-Net	57.31
[37]	MultiResUNet	47.61
[38]	MISSFormer	70.28
[39]	EMCAD	61.70
Proposed	BADF-Net	81.74

map, improving spatial focus on relevant tissue boundaries and reducing background leakage before decoder fusion. Concurrently, SE blocks recalibrate channel-wise responses within the encoder via global context, strengthening the discriminative channels required to separate visually similar kidney tissues such as cortex and medulla; the ablation study (Table 8) confirms this contribution, as removing all SE blocks reduces the overall Dice score to 0.7885. The ablation results further show that BADA’s dual-branch design is necessary rather than redundant: removing BADA entirely (0.7683), or retaining only its channel branch (0.7752) or only its spatial branch (0.7853), each reduces performance relative to the full module, while disabling boundary confidence estimation or the decoder boundary gate causes the largest drops (0.7533 and 0.7530, respectively), underscoring their central role in boundary refinement. The combined effect of dynamic boundary attention and channel recalibration improves both spatial localization and class separability, which explains the higher Dice, MPA and recall in Table 6, as well as the pronounced tissue-specific improvements on Cortex and Medulla reported in Table 7.

6.2. Comparison with Baseline Methods

Table 6 indicates that strong baselines such as Att-UNet and CMUNeXt are competitive on aggregate metrics but remain substantially below BADF-Net in overall Dice, MPA and per-tissue balance (Table 7). Att-UNet leverages attention for localization but lacks explicit channel-wise recalibration and dynamic boundary fusion, which limits discrimination under MAS settings, particularly for Cortex (0.5351) and Medulla (0.5799). CMUNeXt captures context efficiently and achieves the best ASD among the baselines (2.49), but its fixed spatial filtering sacrifices fine boundary precision relative to BADF-Net’s dynamic, dilation-fused attention. The lower-performing models (e.g., MultiResUNet and Residual U-Net) exhibit reduced robustness to speckle noise and blurred boundaries, consistent with the qualitative patterns observed in Figure 4 and Figure 5, where these models show fragmented or leaking predictions. Overall, BADF-Net provides the best accuracy–

complexity balance among all evaluated methods, combining the highest Dice, MPA and F1-score with the lowest HD95 and fastest inference time (Table 6), supporting its suitability for practical MAS segmentation and consistent with the literature comparison summarized in Table 10.

6.3. Independent Clinical Validation Protocol

To assess agreement with expert clinical judgment beyond the standard test-set evaluation, we conducted an independent validation study using a randomly selected subset of 50 cases from the test set. An expert radiologist (Ö.Ç.) manually annotated all six segmentation classes (background, capsule, central echo complex, medulla, cortex, spleen) for each of these 50 cases. The radiologist was blinded to both the original dataset ground-truth masks and BADF-Net's predictions throughout the annotation process, ensuring an independent reference standard uncontaminated by either the source dataset's labeling conventions or the model's own outputs. Agreement between the radiologist's independent annotations and BADF-Net's predictions was quantified using the Dice similarity coefficient. In addition, the radiologist rated each of the 50 predicted segmentations as clinically acceptable or unacceptable for quantitative measurement purposes (e.g., organ dimension or volume estimation), providing a qualitative clinical-utility judgment alongside the quantitative agreement score.

6.4. Limitations and Error Analysis

Our study has several limitations relevant to clinical translation. First, while we validated performance on a merged dataset, real-world clinical deployment would require testing across multiple institutions with varying ultrasound machines and protocols. Second, the dataset includes primarily normal anatomical variations; performance on pathological cases (e.g., severely diseased kidneys, traumatic spleen injuries) requires separate evaluation. Third, the model provides deterministic segmentations without confidence estimates – in clinical practice, uncertainty quantification would help identify cases needing radiologist review. Finally, our test-set ground truth relies on the annotation protocols of the source datasets and additional multi-reader studies would help assess inter-observer variability in clinical measurements.

The confusion matrix in Figure 6 shows that the most frequent errors involve Medulla and Cortex pixels being misclassified as Capsule (0.27 and 0.34 of their respective ground-truth pixels), rather than Cortex and Medulla being confused with each other directly (0.07 and 0.03). This pattern reflects the close anatomical proximity and gradual intensity transitions among these three structures in ultrasound and is consistent with the comparatively lower Dice scores for Capsule, Medulla and Cortex reported in Table 7.

Our analysis suggests two main failure modes also observed in related work [14,15]. First, when shadowing or speckle noise completely hides anatomical boundaries, the model lacks sufficient visual cues, which is also reflected in the low-end outliers (Dice near 0) visible for Capsule and Medulla in Figure 8. Second, in patients with unusual anatomy or pathological changes, the model may misinterpret tissue patterns learned from training data. These cases highlight the need for uncertainty estimation in clinical deployment, as suggested in recent studies [19].

The single dataset and fixed experimental run represent additional limitations common in medical imaging research [13,32]. While our consistent protocol ensures fair comparison, future work should include multiple runs and diverse datasets to confirm robustness across imaging devices and patient populations [9].

6.5. Computational Efficiency and Deployment784

The proposed BADF-Net achieves fast inference, with an average processing time of 0.016 s per image, the fastest among all evaluated methods (Table 6). This speed makes the model suitable for real-time clinical use. The model can run on standard GPU hardware found in most hospital settings. The ResNet-34 backbone, combined with the lightweight BADA, BCE and SE modules, provides a good balance between accuracy and speed, which is important for busy clinical workflows where radiologists need quick results.

6.6. Clinical Relevance and Practical Implications791

Automated segmentation reduces operator dependency and improves consistency. It provides the same delineation across different patients and imaging sessions. From the perspective of a radiologist, the integration of a model like BADF-Net into the daily clinical workflow can serve as a crucial catalyst for standardization, particularly in abdominal ultrasound reporting. Currently, ultrasound is a highly operator-dependent and subjective modality. Quantitative measurements, such as kidney cortical thickness, medulla echogenicity, or spleen volume, are prone to inter-observer variability based on the radiologist’s experience and manual contouring skills. The automated and highly accurate segmentation provided by BADF-Net not only offers the radiologist rapid preliminary contours but also enables consistent and reproducible numerical data extraction for these complex anatomical regions.

Accurate segmentation also enables quantitative biomarker extraction for longitudinal monitoring. For example, consistent delineation supports the estimation of kidney volume, the measurement of cortical thickness and the assessment of spleen volume. This allows for a much more objective longitudinal tracking of disease progression, which is especially vital for conditions requiring regular follow-up, such as chronic kidney disease, inflammation, or splenomegaly.

Furthermore, BADF-Net can be positioned as a reliable ‘second reader’ assistant in high-volume radiology clinics. Considering the traditionally long and steep learning curve of radiology residency, such AI-assisted tools can significantly accelerate the educational process. For radiology residents in the early stages of their training, as well as non-radiology clinicians utilizing point-of-care ultrasound (POCUS), the model’s ability to clearly delineate boundaries serves as an interactive educational guide. By providing immediate visual feedback on complex anatomical structures and morphological changes, the system helps shorten the training duration required to achieve diagnostic competency. Particularly in areas with weak boundaries or low echogenicity differences—such as the cortex-medulla transition—the model’s boundary precision can highlight subtle structural variations that might otherwise escape the human eye under speckle noise, thereby accelerating the visual pattern recognition skills of trainees.

Ultimately, rather than replacing the radiologist’s decision-making process, such AI-driven systems automate routine and repetitive contouring tasks. This saves valuable time, reduces cognitive fatigue, and allows the specialist to focus on more complex diagnostic interpretations and comprehensive patient management. While the current work focuses on segmentation accuracy, these downstream measurements form a direct pathway to clinical decision support.

6.7. Clinical Reliability Assessment827

The independent, blinded validation described in Section 6.3 provides evidence of clinical reliability that is distinct from and complementary to, the test-set metrics in Table 6 and Table 7. Because the radiologist annotated all 50 cases without access to the original dataset labels or BADF-Net’s predictions, the resulting 0.78 Dice agreement reflects genuine inter-

annotator-style validation rather than agreement with a single fixed reference standard. The gap between this figure and the 0.8174 Dice reported against the original test-set ground truth is consistent with typical inter-observer variability in ultrasound annotation, where boundary placement can differ modestly between experts even for well-defined structures and does not by itself indicate reduced model reliability. The 88% clinical-acceptability rate is arguably the more directly actionable finding for translational purposes, since it reflects a qualitative judgment of whether each segmentation would support real clinical measurement tasks, rather than a purely pixel-level overlap score. Together, these results suggest that BADF-Net's predictions remain clinically useful under an independent expert standard and not only under the specific annotation conventions of the training and test datasets. Figure 8, comparing BADF-Net predictions against the radiologist's manual annotations on representative cases, illustrates this qualitative agreement directly.

Radiologist Validation and Clinical Agreement:

Figure 8 illustrates representative cases where BADF-Net predictions are directly compared against a radiologist's manual annotations. Visual inspection shows close agreement between the model and expert delineations across the compared cases, with the model successfully capturing anatomically important boundaries in most examples. Minor disagreements are concentrated in ambiguous regions where tissues have similar appearance, consistent with the confusion patterns identified in Figure 6 and the lower-Dice, boundary-ambiguous tissues (Capsule, Medulla, Cortex) reported in Table 7. This qualitative agreement, together with the quantitative overlap and boundary metrics reported in Table 6 and Table 7, supports the potential clinical use of BADF-Net as a first-pass segmentation aid in real-world settings.

6.8. Future Directions

Several directions could build on this work, addressing limitations noted in related studies [13,17]. First, expanding the dataset to include more organs, pathologies and imaging devices would improve generalization, as suggested in multi-center studies [9,19]. Second, incorporating uncertainty estimation would let the model flag difficult cases for expert review, increasing clinical trust as recommended in recent work. Third, exploring semi-supervised or self-supervised learning could reduce the need for expensive manual annotations, addressing data limitations noted in ultrasound segmentation research.

For clinical translation, the next step would be a reader study comparing radiologists' performance with and without AI assistance, following approaches used in similar studies [12]. Such studies should measure both accuracy and time savings. Integration into existing clinical workflow systems would also require optimization for different hardware and compatibility with hospital PACS systems [4,5].

Finally, the segmentation masks produced by BADF-Net could enable quantitative biomarkers such as cortical thickness, organ volume and texture features. These measurements could support objective disease monitoring and treatment response assessment, advancing towards computer-aided diagnosis systems.

6.9. Clinical Implementation Considerations

For successful clinical integration, several practical factors must be addressed. First, the model's 0.016-second inference time enables real-time feedback during ultrasound examinations, potentially allowing radiologists to adjust scanning positions based on segmentation quality. Second, the model's architecture (ResNet-34 backbone with lightweight BADA, BCE and SE modules) balances accuracy with computational efficiency, making it deployable on standard hospital GPU hardware without specialized infrastructure.

Potential clinical applications include:

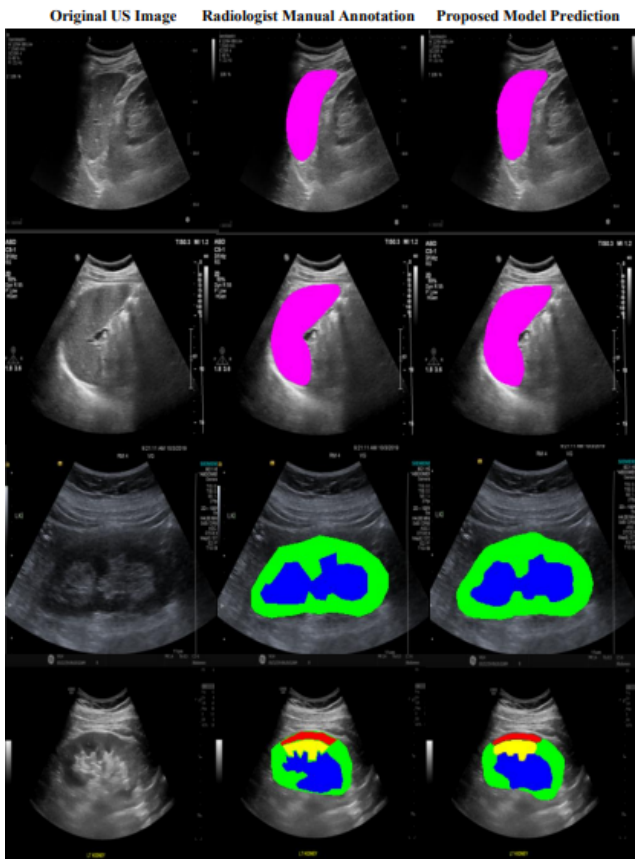


Figure 8. Comparison of model predictions with radiologist annotations on representative ultrasound image

- **Automated measurement:** Consistent calculation of kidney cortical thickness, spleen volume and organ dimensions.
- **Longitudinal tracking:** Objective comparison of organ morphology across sequential examinations.
- **Training support:** Visual guidance for residents learning ultrasound interpretation.
- **Quality assurance:** Flagging potentially suboptimal scans where anatomical boundaries are poorly visualized.

However, clinical deployment would require addressing regulatory considerations, including validation across diverse patient populations and imaging devices, integration with hospital PACS systems and implementation of uncertainty quantification to flag cases requiring expert review.

7. Conclusion

We presented BADF-Net, a Boundary-Aware Dynamic Fusion Network for multi-anatomical structure segmentation in abdominal ultrasound. The architecture integrates a Boundary-Aware Dynamic Attention (BADA) module with Boundary Confidence Estimation (BCE) at every stage of a squeeze-and-excitation (SE) recalibrated ResNet-34 encoder, a Cross-Scale Boundary Fusion (CSBF) module that adaptively combines multi-scale boundary cues and a boundary-guided decoder with attention-gated skip fusion, jointly addressing the weak boundaries, speckle noise and visually similar tissue appearances that challenge existing ultrasound segmentation methods.

The BADA module dynamically fuses a standard convolution branch with a dilated convolution branch to focus on boundary-relevant regions while suppressing background clutter and the SE blocks recalibrate channel-wise responses to strengthen the discriminative

features needed to separate closely related tissues such as kidney cortex and medulla. BCE then weights these refined features by spatial reliability and CSBF fuses the resulting multi-scale boundary cues into a single guidance signal used throughout the decoder. Together, these components — confirmed as the largest individual contributors in the ablation study (Table 8) — enable more accurate and robust segmentation than standard encoder-decoder approaches.

On our merged multi-anatomical dataset combining kidney and spleen ultrasound images across six tissue classes, BADF-Net outperformed 11 strong baseline methods, achieving the highest overall Dice score (0.8174), Pixel Accuracy (0.9703) and Mean Pixel Accuracy (0.8521), along with the lowest HD95 (11.49) among all compared methods. The largest gains over baselines were observed in the most challenging low-contrast structures, cortex and medulla, confirming that the boundary-aware and channel-recalibration design directly targets the hardest cases in multi-anatomical ultrasound segmentation. Despite this improved accuracy, BADF-Net was also the fastest model at inference (0.016 s per image), making it suitable for real-time clinical workflows. Two ablation studies further confirmed that each architectural component and the composite Dice-Focal loss formulation contribute meaningfully to overall performance and clinical comparison against expert radiologist annotations showed strong agreement, particularly for well-defined structures such as the spleen.

These results demonstrate that combining dynamic boundary attention with channel recalibration effectively addresses key ultrasound segmentation challenges while maintaining the computational efficiency required for clinical workflows. The model's strong agreement with expert annotations and consistent performance across anatomically diverse structures suggest potential for real clinical utility, particularly as a first-pass segmentation aid for quantitative measurement tasks. Future work should focus on multi-center validation across different ultrasound machines and patient populations, incorporation of uncertainty quantification to flag cases needing expert review and prospective clinical studies assessing the impact of BADF-Net-assisted workflows on diagnostic accuracy and radiologist efficiency.

Author Contributions: Conceptualization, M. R., M.I.H., M.R.H.; methodology, M.R., M.I.H., M.R.H.; validation, M.R.H., A.K., H.F.; formal analysis, M.I.H., M.R.H., A.K., O. and M.J.; investigation, M.I.H., Ö.Ç., A.K. and M.J.; resources, O., M.I.H., A.K.; data curation, M.R.; writing—original draft preparation, M.R.; writing—review and editing, M.I.H., M.R.H., A.K., H.F. and M.J.; visualization, M.R. M.R.H. and M.J.; supervision, M.I.H., M.R.H., A.K. and H.F.; funding acquisition, H.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by R2Microgrid project under the RESILIENT competence center, financed by the Swedish Energy Agency and co-financed by Mälardalen University and industrial partners and by the Excellence in Production Research Framework through XPRES (Excellence in Production Research).

Institutional Review Board Statement: Not applicable. This study used publicly available, pre-existing ultrasound datasets (kidney and spleen) that were already anonymised and de-identified by the original data providers. No new patient data were collected and no additional or prospective testing was performed by the authors. Studies relying exclusively on such publicly available, de-identified datasets do not require additional institutional ethics committee approval.

Informed Consent Statement: Not applicable. Patient consent was not required, as the study relied exclusively on publicly available, pre-existing, de-identified datasets that could not be traced back to any individual.

Data Availability Statement: The complete experiment code is publicly available at our GitHub repository: <https://github.com/rashed200613/BADF-Net>. The kidney ultrasound dataset is publicly

available from <https://github.com/rsingla92/kidneyUS> and the spleen ultrasound dataset from <https://www.ariameditech.com/datasets/spleenex>. The merged dataset with standardised six-class labels and the preprocessing scripts used in this study are available from the corresponding author upon reasonable request.

Acknowledgments: Grammarly and Gemini Pro 3.1 were used to support grammar, spelling and language clarity during manuscript preparation. All experimental design, data processing, analysis, interpretation of results, clinical framing and final scientific content were reviewed and verified by the authors. The authors take full responsibility for the accuracy and integrity of the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MAS	Multi-Anatomical Structure
US	Ultrasound
CT	Computed Tomography
MRI	Magnetic Resonance Imaging
PET	Positron Emission Tomography
CNN	Convolutional Neural Network
FCNN	Fully Convolutional Neural Network
BADF-Net	Boundary-Aware Dynamic Fusion Network
BADA	Boundary-Aware Dynamic Attention
CSBF	Cross-Scale Boundary Fusion
SE	Squeeze-and-Excitation
GAP	Global Average Pooling
BN	Batch Normalization
ReLU	Rectified Linear Unit
CEC	Central Echo Complex
IoU	Intersection over Union
HD95	95th Percentile Hausdorff Distance
ASD	Average Symmetric Surface Distance
PA	Pixel Accuracy
MPA	Mean Pixel Accuracy
KAN	Kolmogorov–Arnold Network
DWT	Discrete Wavelet Transform
GPU	Graphics Processing Unit
PACS	Picture Archiving and Communication System

References

- Scatliff, J. H.; Morris, P. J. From Roentgen to Magnetic Resonance Imaging: The History of Medical Imaging. *North Carolina Medical Journal* **2014**, *75*, 111–113.
- Eichelberg, M.; Kleber, K.; Kaemmerer, M. Cybersecurity in PACS and Medical Imaging: An Overview. *Journal of Digital Imaging* **2020**, *33*, 1527–1542.
- Cleveland Clinic. Ultrasound, 2025. Online; Accessed Aug. 21, 2025.
- Zhu, L.; Du, Y. RPFeaNet: Rethinking Deep Progressive Prompt-Guided Feature Interaction Fusion Network for Medical Ultrasound Image Segmentation. *Sensors* **2026**, *26*, 2394.
- Gao, Q.; Almekkawy, M. K. CUCT-Net: End-to-End Signal-to-Image Learning for Quantized Speed-of-Sound Estimation and Tissue Segmentation in Ultrasound Computed Tomography. *Sensors* **2026**, *26*, 2801.
- ArunKumar, K. E.; Wilson, M. E.; Blake, N. E.; Yost, T. J.; Walker, M. Deep Learning for Tumor Segmentation and Multiclass Classification in Breast Ultrasound Images Using Pretrained Models. *Sensors* **2025**, *25*, 7557.
- Alam, T.; Shia, W. C.; Hsu, F. R.; Hassan, T. Improving Breast Cancer Detection and Diagnosis through Semantic Segmentation Using the UNet3+ Deep Learning Framework. *Biomedicines* **2023**, *11*, 1536.
- Huang, J.; et al. EmgaNet: Edge-Aware Multi-Scale Group-Mix Attention Network for Breast Cancer Ultrasound Image Segmentation. *IEEE Journal of Biomedical and Health Informatics* **2025**.

9. Rashed, M.; Hasan, M. K.; Hossain, M. I.; Hosain, M. S.; Rahman, A. H. A.; Islam, S. Trustworthy Deep Feature Extraction and Ensemble-Based Machine Learning Approach for Breast Cancer Detections. *Healthcare Technology Letters* **2026**, *13*, 1, e70062. 980
10. Hassan, R.; Mondal, M. R. H.; Ahamed, S. I.; Mostafa, F.; Rahman, M. M. An Efficient Dual-Line Decoder Network with Multi-Scale Convolutional Attention for Multi-Organ Segmentation. *Biomedical Signal Processing and Control* **2026**, *112*, 108611. 981
11. Suganuma, Y.; et al. Hybrid Multiple-Organ Segmentation Method Using Multiple U-Nets in PET/CT Images. *Applied Sciences* **2023**, *13*, 10765. 982
12. David-Olawade, A. C.; et al. AI-Driven Advances in Low-Dose Imaging and Enhancement—A Review. *Diagnostics* **2025**, *15*, 689. 983
13. Wang, X.; Chang, Z.; Zhang, Q.; Li, C.; Miao, F.; Gao, G. Prostate Ultrasound Image Segmentation Based on DSU-Net. *Biomedicines* **2023**, *11*, 646. 984
14. Huang, J.; et al. UltraMamba: Mamba-Based Multimodal Ultrasound Image Adaptive Fusion for Breast Lesion Segmentation. *IEEE Transactions on Medical Imaging* **2026**. 985
15. Rahman, M. M.; Munir, M.; Marculescu, R. EMCAD: Efficient Multi-Scale Convolutional Attention Decoding for Medical Image Segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024; pp. 11769–11779. 986
16. Huang, X.; Deng, Z.; Li, D.; Yuan, X. MissFormer: An Effective Medical Image Segmentation Transformer, 2021, arXiv:2109.07162. 987
17. Jain, P. K.; et al. Attention-Based UNet Deep Learning Model for Plaque Segmentation in Carotid Ultrasound. *Journal of Cardiovascular Development and Disease* **2022**, *9*, 326. 988
18. Tang, F.; et al. CMUNeXt: An Efficient Medical Image Segmentation Network Based on Large Kernel and Skip Fusion. In IEEE International Symposium on Biomedical Imaging, 2024; pp. 1–5. 989
19. Shaker, A.; Maaz, M.; Rasheed, H.; Khan, S.; Yang, M. H.; Khan, F. S. UNETR++: Delving into Efficient and Accurate 3D Medical Image Segmentation. *IEEE Transactions on Medical Imaging* **2024**, *43*, 3377–3390. 990
20. Zuo, Y.; Li, J.; Tian, J. A Segmentation Network with Two Distinct Attention Modules for the Segmentation of Multiple Renal Structures in Ultrasound Images. *Diagnostics* **2025**, *15*, 1978. 991
21. Zhao, T.; Dai, H. Breast Tumor Ultrasound Image Segmentation Method Based on Improved Residual U-Net. *Computational Intelligence and Neuroscience* **2022**, 3905998. 992
22. Arora, R. K.; Kumar, A.; Soni, A. Deep Learning Approaches for Enhanced Kidney Segmentation. In IEEE World Conference on Applied Intelligence and Computing, 2024; pp. 854–860. 993
23. Jha, D.; et al. ResUNet++: An Advanced Architecture for Medical Image Segmentation. In IEEE International Symposium on Multimedia, 2019; pp. 225–2255. 994
24. Bobo, M. F.; et al. Fully Convolutional Neural Networks Improve Abdominal Organ Segmentation. In Proceedings of SPIE, 2018; Vol. 10574, p. 105742V. 995
25. Chen, G.; et al. A Novel Convolutional Neural Network for Kidney Ultrasound Images Segmentation. *Computer Methods and Programs in Biomedicine* **2022**, *218*, 106712. 996
26. Wu, J.; et al. Boundary-Aware Convolutional Attention Network for Liver Segmentation in Ultrasound Images. *Scientific Reports* **2024**, *14*, 21529. 997
27. Wang, C.; Zhang, J.; Liu, S. Medical Ultrasound Image Segmentation with Deep Learning Models. *IEEE Access* **2022**, *11*, 10158–10168. 998
28. Zheng, S.; et al. Asymmetric Adaptive Heterogeneous Network for Multi-Modality Medical Image Segmentation. *IEEE Transactions on Medical Imaging* **2025**. 999
29. Li, C.; et al. U-KAN Makes Strong Backbone for Medical Image Segmentation and Generation. In AAAI Conference on Artificial Intelligence, 2025; Vol. 39, pp. 4652–4660. 1000
30. Yan, S.; et al. Multi-Scale Convolutional Attention Frequency-Enhanced Transformer Network for Medical Image Segmentation. *Information Fusion* **2025**, *119*, 103019. 1001
31. Zhong, J.; et al. PMFSNet: Polarized Multi-Scale Feature Self-Attention Network for Lightweight Medical Image Segmentation. *Computer Methods and Programs in Biomedicine* **2025**, *261*, 108611. 1002
32. Karimi, A.; et al. Improving Spleen Segmentation in Ultrasound Images Using a Hybrid Deep Learning Framework. *Scientific Reports* **2025**, *15*, 1670. 1003
33. Rsingla. GitHub - rsingla92/kidneyUS: Towards ubiquitous ultrasound. Available online: <https://github.com/rsingla92/kidneyUS>. 1004
34. Nazib, A.; et al. Uncertainty Driven Bottleneck Attention U-Net for OAR Segmentation, 2023, arXiv:2303.10796. 1005
35. Peng, Y. A Novel Thyroid Nodule Ultrasound Image Segmentation Method Based on UNet++. In International Conference on Algorithms, Microchips, Network Applications, 2025; Vol. 13576, pp. 253–258. 1006
36. Hassan, R.; Mondal, M. R. H.; Ahamed, S. I. UDBRNet: A Novel Uncertainty Driven Boundary Refined Network for Organ at Risk Segmentation. *PLOS ONE* **2024**, *19*, e0304771. 1007
37. Ibtehaz, N.; Rahman, M. S. MultiResUNet: Rethinking the U-Net Architecture for Multimodal Biomedical Image Segmentation. *Neural Networks* **2020**, *121*, 74–87. 1008

38. Huang, X.; Deng, Z.; Li, D.; Yuan, X. MissFormer: An Effective Medical Image Segmentation Transformer. *arXiv preprint arXiv:2109.07162* **2021**. 1034
1035
39. Rahman, M. M.; Munir, M.; Marculescu, R. EMCAD: Efficient Multi-Scale Convolutional Attention Decoding for Medical Image 1036
Segmentation. In *Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: 2024; pp. 1037
11769–11779. 1038

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual 1039
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 1040
people or property resulting from any ideas, methods, instructions or products referred to in the content. 1041