RESEARCH ARTICLE

SEFF-Net: A Hybrid Feature Fusion Network for Accurate Segmentation of Breast Ultrasound Images

Tingli Su[1], **Rui Wan**[1], **Senmao Wang**[2,*] **and Yuting Bai**[1]

[1] School of Computer and Artificial Intelligence, Beijing Technology and Business University, Beijing 100048, China
[2] Taiyuan Central Hospital, Taiyuan 030024, China

Abstract

Breast ultrasound imaging plays a crucial role in early breast cancer screening and diagnosis due to its noninvasive nature and cost-effectiveness. However, accurate lesion segmentation remains challenging because of severe speckle noise, low contrast, and blurred tumor boundaries. To address these issues, this paper proposes SEFF-Net, a novel edge-aware feature fusion network with a U-shaped encoder–decoder architecture to capture multi-level semantic representations for breast ultrasound image segmentation task. To enhance boundary perception, a Self-learning Edge Enhancement Module is embedded in the shallow encoding stages, while a Spatial Feature Fusion Module is introduced to effectively integrate multi-scale features by leveraging spatial context, thereby achieving a better balance between low-level spatial details and high-level semantic information.To further alleviate the class imbalance between foreground and background regions and improve boundary learning, a novel joint loss function is designed by combining region-based consistency constraints with boundary-sensitive supervision. This optimization strategy reinforces contour awareness while maintaining overall segmentation accuracy. Experimental results demonstrate that SEFF-Net consistently outperforms state-of-the-art segmentation methods across multiple evaluation metrics, including Dice coefficient, IoU, and boundary-related measures. Overall, SEFF-Net provides an effective and reliable solution for accurate breast ultrasound image segmentation, showing promising potential for clinical computer-aided diagnosis systems.

Keywords: breast ultrasound segmentation, edge enhancement, feature fusion, hybrid loss strategy, image processing.

1 Introduction

Breast cancer is one of the most prevalent malignant tumors threatening women's health, with its incidence continuing to rise worldwide [1]. Early diagnosis plays a crucial role in improving patient survival rates and clinical outcomes. As a commonly used imaging modality in clinical practice, breast ultrasound has become an indispensable tool for breast cancer screening and diagnosis due to its noninvasive nature, low cost, and absence of radiation

Submitted: 10 January 2026
Accepted: 04 March 2026
Published: 06 March 2026

Vol. 3, **No.** 2, 2026.
10.62762/TETAI.2026.494190

***Corresponding author:**
✉Senmao Wang
wangjiaoyoutiao@163.com

Citation
Su, T., Wan, R., Wang, S., & Bai, Y. (2026). SEFF-Net: A Hybrid Feature Fusion Network for Accurate Segmentation of Breast Ultrasound Images. *ICCK Transactions on Emerging Topics in Artificial Intelligence*, 3(2), 128–141.

exposure [2]. However, ultrasound images inherently suffer from speckle noise, intensity inhomogeneity, complex tissue structures, and indistinct boundaries, which severely hinder accurate segmentation of lesion regions. Even for experienced radiologists, precisely delineating tumor contours remains challenging under conditions of heavy noise and weak boundaries, thereby increasing the risk of misdiagnosis [3].

In recent years, deep learning techniques have achieved remarkable progress in the field of medical image analysis. Convolutional neural networks (CNNs) have become the dominant approach for medical image segmentation, effectively extracting spatial features through local convolution operations. Representative encoder–decoder architectures, such as U-Net [4] and SegNet [5], have been widely applied to ultrasound image segmentation and have demonstrated promising performance. Nevertheless, due to the limited receptive field of CNNs, these methods still face difficulties in handling complex backgrounds, low contrast, and blurred boundaries commonly observed in ultrasound images [6]. To address these limitations, attention mechanisms and multi-scale feature fusion strategies have been introduced to enhance structural awareness. For instance, the application of attention-based U-Net in breast ultrasound imaging has been shown to significantly improve tumor segmentation accuracy [7].

Meanwhile, Transformer architectures, benefiting from their strong global modeling capability, have exhibited notable advantages in medical image segmentation. Hybrid CNN–Transformer models, such as HCT-Net [8], are capable of capturing long-range dependencies while preserving local details, achieving superior performance in breast ultrasound image segmentation tasks. However, relying solely on global attention remains insufficient for precise identification of weak boundary regions. Consequently, several studies have incorporated boundary-aware mechanisms. For example, BO-Net [9] employs a boundary-guided strategy to address low contrast and structural ambiguity in ultrasound images, while SMU-Net [10] further enhances sensitivity to tumor boundaries through saliency maps and morphological constraints.

Despite the improvements achieved by existing methods, several limitations persist: (1) most approaches introduce edge constraints only at the network output stage, resulting in insufficient interaction between boundary information and semantic features; (2) the complex multi-scale structures in ultrasound images make it difficult for single-scale features to simultaneously capture global context and local details; and (3) there is a lack of a unified framework capable of adaptively learning the importance of boundary regions during training while integrating semantic and boundary features across multiple scales. Therefore, effectively exploiting multi-scale features and enhancing discrimination of weak boundary regions remain critical challenges in breast ultrasound image segmentation.

To address these issues, this paper proposes a Self-learning Edge-enhanced and Feature Fusion Network (SEFF-Net) for high-precision segmentation of blurred lesion regions in breast ultrasound images, where a U-shaped encoder–decoder architecture is adopted to capture and fuse the multi-scale edge feature and semantic feature hidden in the images. The main contributions of this work can be summarized as follows:

1. An edge enhancement module is proposed to achieve adaptive enhancement of blurred boundary regions;
2. Deep aggregation and feature fusion modules are introduced to integrate edge features and deep semantic features, enabling effective global context modeling;
3. A hybrid loss function is designed to further improve boundary-aware segmentation performance.

The remainder of this paper is organized as follows. Section 2 (Related Work) reviews existing studies on medical image segmentation and breast ultrasound image analysis. Section 3 (Methodology) describes the proposed SEFF-Net framework, including the self-learning edge enhancement module, spatial feature fusion module, and the improved loss function. Section 4 (Experiments) presents the experimental datasets, evaluation metrics, and comparative results. Section 5 (Conclusion) summarizes the main findings and outlines future research directions.

2 Related Work

Medical image segmentation is a core task in computer-aided diagnosis systems and has been extensively studied across various imaging modalities, including CT, MRI, and ultrasound. With the rapid development of deep learning,

CNN-based segmentation methods have gradually become the mainstream approach. Among them, encoder–decoder architectures represented by U-Net effectively fuse low-level spatial details and high-level semantic information through skip connections, achieving remarkable success in a wide range of medical segmentation tasks [11]. Subsequently, numerous variants have been proposed by incorporating residual connections, dense connections, or attention mechanisms to further enhance feature representation capability and training stability [12]. Despite their effectiveness, these architectures are generally designed under modality-agnostic assumptions and often struggle to cope with the distinct imaging characteristics of breast ultrasound, where severe speckle noise, low contrast, and indistinct lesion boundaries coexist. As a result, conventional CNN-based methods frequently encounter difficulties in simultaneously maintaining region consistency and boundary accuracy in breast ultrasound segmentation [13].

To enhance the modeling of complex structures and long-range dependencies, Transformer-based architectures have recently been introduced into medical image segmentation [14]. Self-attention mechanisms enable effective global context modeling and have demonstrated advantages in handling objects with complex morphology or large-scale variations [15]. Motivated by this, several hybrid CNN–Transformer frameworks have been proposed, where CNNs capture local texture information while Transformers model global semantic relationships, achieving a balance between performance and computational cost [8]. However, in ultrasound imaging scenarios, although global semantic modeling improves contextual understanding, local boundary information is often weakened during feature propagation, leading to discontinuous or over-smoothed contours [16]. This limitation highlights that global modeling alone is insufficient for precise boundary delineation in breast ultrasound images.

Feature fusion mechanisms have therefore been widely investigated as a means of alleviating these challenges. Existing studies explore multi-scale, multi-level, and multi-source feature fusion strategies to integrate complementary information from different resolutions and semantic levels [17]. Multi-scale fusion approaches, such as feature pyramid structures and cross-layer skip connections, enable networks to capture both global lesion structure and fine-grained details [18]. Nevertheless, simple concatenation or element-wise summation treats features from different scales equally, neglecting their spatially varying importance and potentially introducing redundant or noisy information. Although adaptive fusion strategies based on attention mechanisms or learnable weighting have been proposed, many of them focus primarily on semantic enhancement and provide limited support for boundary-sensitive feature interaction, which remains critical for breast ultrasound images characterized by cluttered backgrounds and weak edges [19].

Beyond network architecture design, loss function optimization also plays a vital role in segmentation performance. Pixel-wise cross-entropy loss, while widely used, is sensitive to severe foreground–background class imbalance and tends to bias predictions toward background regions. Region-overlap-based losses such as Dice loss and IoU loss directly optimize spatial overlap and improve segmentation accuracy for small targets, but they offer limited constraints on boundary geometry. Focal loss partially alleviates this issue by emphasizing hard-to-classify samples, while boundary-aware and distance-based losses explicitly guide models toward contour alignment [20]. More recently, hybrid loss strategies that combine multiple complementary loss terms have emerged as an effective solution to jointly address region accuracy, boundary precision, and class imbalance [21]. Nevertheless, designing a loss function that can effectively cooperate with boundary-enhanced feature learning remains an open challenge in breast ultrasound image segmentation.

3 Methodology

In terms of network architecture, our model follows the U-Net paradigm due to its structural simplicity and proven applicability to medical ultrasound images, consisting of an encoder, skip connections, and a decoder. Overall, the input image is first fed into the encoder, where it passes through four encoding blocks to generate feature representations at four different hierarchical levels, as shown in the Figure 1. Among these features, the first three layers contain richer shallow representations with abundant edge information, while the deeper layers focus more on high-level semantic features. Therefore, the Self-learning Edge Enhancement Module (SEEM) is applied to the first three encoder layers to extract and enhance edge information, whereas the Aggregation Module (AGGM) is employed on the deeper layers

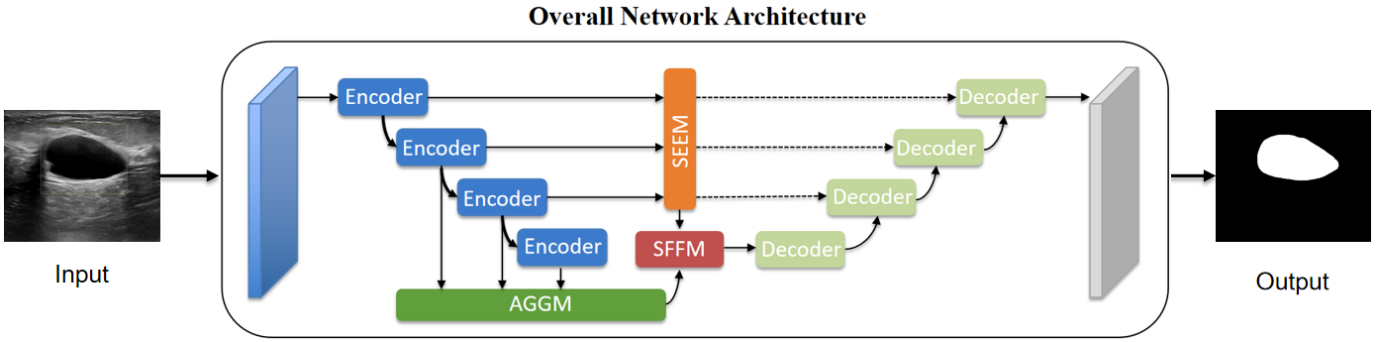

Figure 1. Overall network architecture.

to perform semantic feature aggregation. As a result, two groups of features with complementary properties are obtained. These features are then effectively fused by the Selective Feature Fusion Module (SFFM) and subsequently fed into the decoder for upsampling and final prediction.

3.1 Self-learning Edge Enhancement Module

The core idea of SEEM is to guide the network to automatically focus on edge-related features through a self-learning mechanism, thereby enhancing the precision of lesion boundary delineation. The detailed structure of SEEM is given in Figure 2. It is worth noting that edge information is predominantly concentrated in shallow feature maps; however, as feature extraction proceeds to deeper layers, such information tends to be gradually weakened. Directly utilizing shallow features without further processing may therefore limit the model's boundary segmentation capability. Consequently, an effective mechanism is required to further strengthen the edge characteristics extracted from multiple shallow layers, enabling them to be jointly leveraged with deep semantic features for accurate target representation.

The SEEM module primarily performs edge enhancement on the shallow features from the first three encoder layers, as these layers contain more prominent boundary cues. Given an input image, the Sobel operator is first applied to compute gradient responses in the horizontal and vertical directions:

$$G_x = I * S_x, \quad G_y = I * S_y \tag{1}$$

where $*$ denotes the two-dimensional convolution operation, and S_x and S_y represent the horizontal and vertical Sobel convolution kernels, respectively, defined as:

$$S_x = \begin{bmatrix} 1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \quad S_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \tag{2}$$

Here, S_x captures grayscale variations along the horizontal direction, emphasizing vertical edge structures, while S_y captures grayscale variations along the vertical direction, emphasizing horizontal edge structures. To further highlight high-frequency variation regions in the image and provide a foundational representation for subsequent edge enhancement, the Sobel edge map G_o of the input is computed as:

$$G_o = \sqrt{G_x^2 + G_y^2} \tag{3}$$

However, when applied to complex ultrasound images, the Sobel operator is prone to generating spurious or coarse edges, particularly when small convolution kernels are used, which may degrade edge extraction accuracy. To alleviate this issue, the SEEM module introduces multi-scale depthwise separable convolutions to process edge features, followed by feature fusion via summation. By applying convolution kernels with different receptive fields to the edge map, this strategy not only suppresses pseudo-edge interference but also enhances edge feature representation across multiple scales, resulting in more robust boundary perception.

Subsequently, the SEEM module incorporates a Squeeze-and-Excitation (SE) block to introduce channel-wise attention. This mechanism enables the network to adaptively reweight channels generated by multi-scale depthwise separable convolutions, emphasizing channels that are highly correlated with edge information while suppressing less informative ones. This process further strengthens the self-learning capability of the model, allowing it to automatically

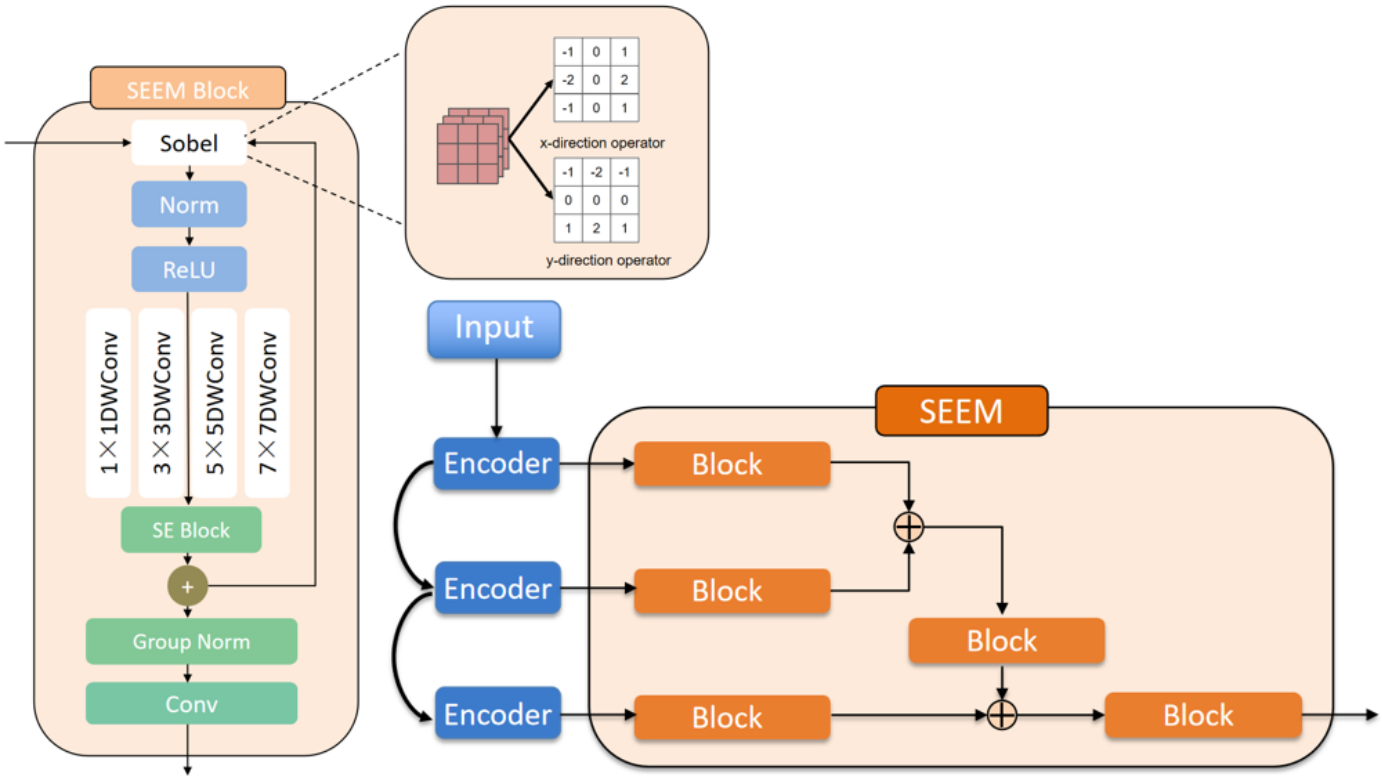

Figure 2. SEEM architecture.

optimize its focus on critical boundary features during training.

To ensure effective propagation of edge information across different scales, the SEEM module adopts a progressive enhancement strategy. Specifically, the output features of each layer are summed with the SEEM-processed features from the subsequent scale and then passed through the SEEM again. This design ensures that edge information from earlier layers is inherited and reinforced by deeper layers, enabling multi-scale edge features to be fully aggregated and utilized in the final output.

Overall, through the joint design of self-learning edge enhancement, channel attention, and multi-scale feature fusion, the SEEM enables the network to adaptively discover and reinforce critical boundary information, thereby significantly improving the accuracy and robustness of boundary segmentation in breast ultrasound images.

3.2 Feature Aggregation and Fusion Modules

To enhance feature representation capability for breast ultrasound image segmentation, we propose an AGGM as shown in the Figure 3. The core idea of AGGM is to effectively aggregate feature information across different scales, thereby improving feature fusion accuracy in the segmentation task. However, feature maps at different scales usually exhibit significant resolution discrepancies, and directly concatenating or summing them may lead to information loss or spatial inconsistency. To address this issue, AGGM performs progressive fusion of multi-scale features, generating richer and more efficient representations. Specifically, upsampling and convolution operations are employed to process features from different scales, ensuring spatial and semantic alignment.

First, AGGM upsamples low-resolution feature maps to match the spatial dimensions of higher-resolution features. Bilinear interpolation is adopted for upsampling to preserve spatial continuity. The upsampled feature maps are then processed using depthwise separable convolutions, which enable the extraction of more fine-grained representations while avoiding excessive information loss and maintaining computational efficiency. Through this design, informative features can be effectively extracted at each scale.

To further integrate multi-scale information, AGGM performs layer-wise feature fusion via concatenation and convolution, allowing each layer to inherit detailed information from adjacent layers. The aggregation

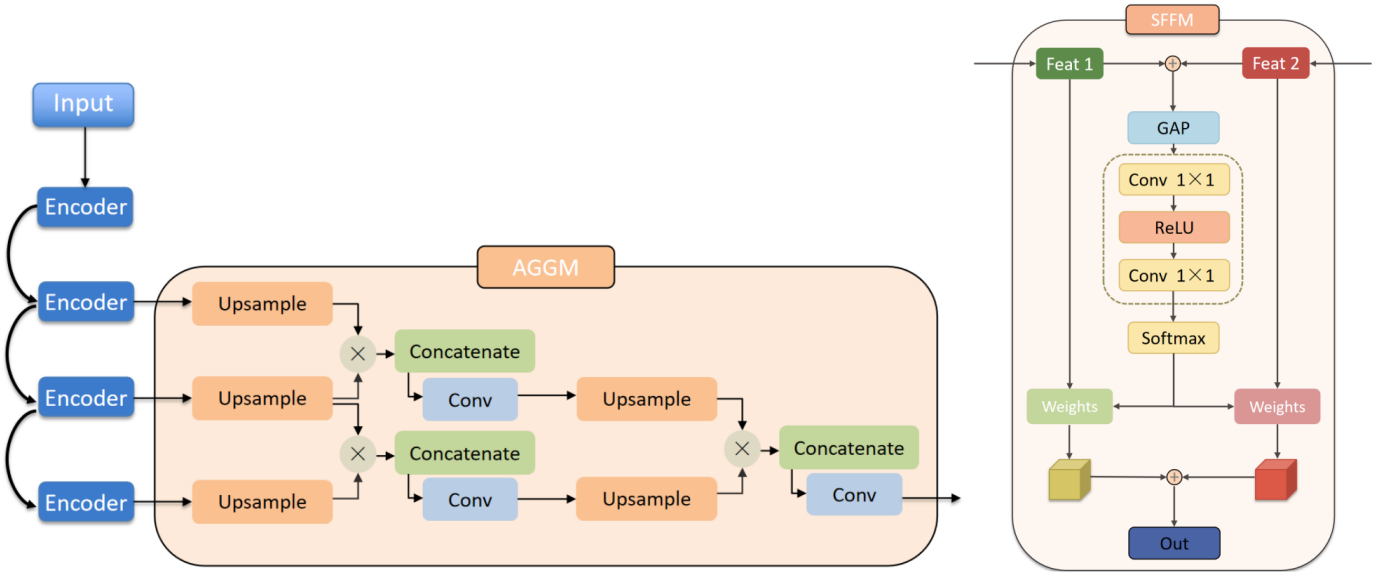

Figure 3. AGGM and SFFM architecture.

process is formulated as:

$$F_{agg}^{l} = \mathrm{Conv}(\mathrm{Concat}(F^{l}, F^{l+1})) \tag{4}$$

where $\mathrm{Concat}(\cdot)$ denotes feature concatenation along the channel dimension, and $\mathrm{Conv}(\cdot)$ represents a standard convolution layer used for cross-channel interaction and semantic refinement. F^{l} and F^{l+1} denote the feature maps from the l and $l+1$ layers, respectively. Through this progressive aggregation strategy, AGGM ensures effective propagation of boundary cues and fine-grained details across scales while enhancing the representational power of features at each level.

Meanwhile, to strengthen the integration of deep semantic features and edge features, we introduce the SFFM as shown in the Figure 3. The main function of SFFM is to dynamically adjust the importance of features from different scales by incorporating a spatial feature fusion mechanism guided by global contextual information. Essentially, SFFM performs scale-wise weighted fusion under the guidance of global spatial context, with pixel-level effectiveness in the spatial domain. By combining adaptive channel weighting with multi-scale feature fusion, SFFM significantly enhances the model's sensitivity to fine details and boundary information.

Specifically, SFFM first applies Global Average Pooling to the input feature maps, compressing the spatial dimensions (height and width) of each feature map into a single value. This operation aggregates spatial information to generate global contextual descriptors, enabling the model to capture the overall distribution of image features. Based on this global context, SFFM then employs convolution operations to generate a weight matrix that guides the fusion of multi-scale features.

Concretely, a 1×1 convolution is first used to reduce the channel dimension from dim to dim/r, where the reduction ratio r is set to 16 by default. This is followed by a ReLU activation to introduce nonlinearity. Another 1×1 convolution is subsequently applied to restore the channel dimension to dim $*$ height. This process adaptively assigns weights to features at different scales, allowing the model to dynamically select more discriminative representations according to task requirements. A Softmax operation is then applied to normalize the weights, ensuring that each scale receives an importance coefficient during fusion. As a result, informative features are emphasized while less relevant ones are suppressed.

Finally, SFFM stacks the edge features F_1 and deep semantic features F_2 and performs weighted fusion under spatial-context guidance, where the fusion coefficients α_1 and α_2 are learnable parameters dynamically generated from global contextual information, resulting in the fused output feature map:

$$F_{out} = \alpha_1 \cdot F_1 + \alpha_2 \cdot F_2 \tag{5}$$

This fusion strategy preserves the original spatial structure while introducing a global-context-driven scale selection mechanism, enabling more effective information integration across features with different resolutions.

3.3 Loss Function Improvement

In breast ultrasound image segmentation, most commonly used loss functions are based on region overlap, such as Dice loss and IoU loss. However, merely improving the overlap between predicted lesion areas and ground-truth regions is insufficient to guarantee accurate boundary localization. To address this limitation, we propose a hybrid loss that jointly optimizes region accuracy, edge consistency, and boundary alignment. The proposed loss consists of four complementary components: Focal Binary Cross-Entropy (Focal BCE), Tversky loss (a generalized form of Dice loss), gradient-based edge consistency constraint, and boundary-band Dice loss, where components gradient-based edge consistency constraint and boundary-band Dice loss are specifically designed to enhance boundary precision. In the following formulations, the predicted probability is denoted by p, and the ground-truth label is denoted by y.

3.3.1 Focal Binary Cross-Entropy Loss

To alleviate class imbalance across different lesion categories, we adopt Focal Loss, which enhances the model's ability to handle imbalanced data distributions. In this formulation, p_t denotes the predicted probability of the true class, while α_t and γ control class balancing and the focusing strength, respectively. The Focal BCE loss is defined as:

$$L_{focal} = -\alpha_t(1-p_t)^{\gamma}\left[y\log p + (1-y)\log(1-p)\right] \tag{6}$$

3.3.2 Tversky Loss

The Tversky loss is employed to impose asymmetric penalties on false positives and false negatives. Specifically, true positives (TP), false positives (FP), and false negatives (FN) are defined as:

$$TP = \sum py, \quad FP = \sum (1-y)p, \quad FN = \sum y(1-p) \tag{7}$$

where ε is a smoothing term (also referred to as a numerical stability term) used to prevent division by zero and improve training stability. The Tversky loss generalizes Dice loss by introducing weighting parameters α and β to control the relative importance of FP and FN.

$$L_{tversky} = 1 - \frac{TP+\varepsilon}{TP+\alpha FP+\beta FN+\varepsilon} \tag{8}$$

3.3.3 Sobel-gradient-based Edge Consistency Loss

To improve edge precision in the predicted probability maps, the Sobel operator $S(\cdot)$ is applied to each channel to extract gradient responses, yielding $S(p)$ and $S(y)$ for the prediction and ground truth, respectively. The L1 norm of their difference (i.e., the sum of pixel-wise absolute differences) is then computed as:

$$L_{edge} = \|S(p) - S(y)\|_1 \tag{9}$$

This loss term encourages the model to preserve sharp and consistent boundary contours while learning regional segmentation.

3.3.4 Boundary-band Dice Loss

To further focus optimization on boundary pixels, a narrow boundary band is generated by applying morphological dilation and erosion to the ground-truth mask. The boundary band is defined as:

$$\text{band} = \text{dilate}(y) - \text{erode}(y) \tag{10}$$

$$L_{\text{boundary}} = 1 - \frac{2\Sigma(p\cdot y\cdot \text{band})+\epsilon}{\Sigma(p\cdot \text{band})+\Sigma(y\cdot \text{band})+\epsilon} \tag{11}$$

where ϵ is a smoothing constant.

3.3.5 Final Hybrid Loss

The final hybrid loss is defined as a weighted sum of the four components:

$$L_{hybrid} = \lambda_{focal}L_{focal} + \lambda_{tversky}L_{tversky} + \lambda_{edge}L_{edge} + \lambda_{boundary}L_{boundary} \tag{12}$$

where the coefficients λ control the contribution of each sub-loss. By jointly optimizing pixel-level classification, region overlap, boundary accuracy, and local gradient consistency, the proposed hybrid loss yields more accurate segmentation results with clearer boundaries compared to single-loss formulations, as demonstrated by experimental results.

4 Experiments

4.1 Datasets

In this study, four publicly available breast ultrasound image datasets are employed, including BUSI, UDIAT, OMI, and BUET. Among them, the BUSI dataset is used for model training, while the UDIAT, OMI, and BUET datasets are utilized for generalization testing and performance evaluation.

4.1.1 BUSI Dataset

The BUSI dataset (Breast Ultrasound Images Dataset) was introduced by Al-Dhabyani et al. [11] in 2020 and collected at Baheya Hospital for Early Detection & Treatment of Breast Cancer, Cairo, Egypt. It comprises 780 breast ultrasound images with an average spatial resolution of approximately 500×500 pixels (PNG format), accompanied by corresponding ground-truth segmentation masks. The images are categorized into three classes: normal (133 images), benign (437 images), and malignant (210 images). In this study, the BUSI dataset is utilized for model training and validation to capture lesion boundary and tissue characteristics.

4.1.2 UDIAT Dataset

The UDIAT dataset is provided by the UDIAT Diagnostic Centre in Sabadell, Spain [22]. It consists of approximately 163 ultrasound images, each accompanied by precise lesion segmentation masks annotated by experienced radiologists. Similar to BUSI, the UDIAT dataset includes both benign and malignant masses and is widely regarded as a benchmark dataset for evaluating the generalization capability of breast ultrasound segmentation models.

4.1.3 OMI Dataset

The OMI dataset originates from a multi-center breast imaging database [23] and contains approximately 400 breast ultrasound images with corresponding segmentation masks. OMI generally refers to a collection of publicly available breast ultrasound images from multiple sources, encompassing both benign and malignant tumors. Due to variations in imaging devices, acquisition protocols, brightness, and noise levels, this dataset exhibits high diversity in lesion morphology and imaging conditions, making it well suited for assessing model robustness under heterogeneous ultrasound environments.

4.1.4 BUET Dataset

The BUET dataset was collected by the Bangladesh University of Engineering and Technology (BUET) during 2012–2013 [24] and serves as another public breast ultrasound image repository for computer-aided diagnosis (CAD) research. It contains approximately 260 B-mode ultrasound images acquired under varying conditions, including both benign and malignant lesions. Image resolutions and acquisition parameters are not standardized across cases. Due to its cross-device and cross-condition nature, the BUET dataset is commonly used to evaluate model generalization across different imaging systems.

In this study, the proposed model is trained and validated on the BUSI dataset, and subsequently tested independently on the UDIAT, OMI, and BUET datasets to evaluate its generalization performance under different data sources and ultrasound acquisition conditions. It should be noted that although the BUSI dataset contains normal cases, consistent with most breast ultrasound segmentation studies, only images containing lesions (i.e., benign and malignant cases) are used for training and testing. This is because the segmentation task aims to accurately extract lesion regions, and including normal images without target regions would lead to severe foreground–background imbalance, thereby interfering with effective boundary learning.

4.2 Evaluation Metrics

To comprehensively evaluate segmentation performance, eight widely used metrics are adopted in this study, including Intersection over Union (IoU), Mean Intersection over Union (mIoU), Overall Accuracy (OA), Cohen's Kappa coefficient, Precision, Dice Coefficient, 95% Hausdorff Distance (HD95), and Average Symmetric Surface Distance (ASSD).

IoU measures the overlap between the predicted region and the ground-truth region, where a higher value indicates better segmentation performance:

$$\text{IoU} = \frac{TP}{TP + FP + FN} \tag{13}$$

mIoU is defined as the average IoU across all classes and serves as a commonly used overall evaluation metric in semantic segmentation tasks:

$$\text{mIoU} = \frac{1}{N} \sum_{i=1}^{N} \frac{TP_i}{TP_i + FP_i + FN_i} \tag{14}$$

OA represents the proportion of correctly classified pixels among all pixels, reflecting overall pixel-level accuracy. Here, TP, TN, FP and FN denote the numbers of true positive, true negative, false positive, and false negative pixels, respectively:

$$\text{OA} = \frac{TP + TN}{TP + TN + FP + FN} \tag{15}$$

The Kappa coefficient measures the agreement between predictions and ground-truth labels while

accounting for random agreement. Its value ranges from $[-1, 1]$, with higher values indicating stronger consistency. Let P_o denote the observed agreement and P_e denote the expected agreement by chance:

$$\kappa = \frac{P_o - P_e}{1 - P_e} \tag{16}$$

Precision measures the proportion of correctly predicted lesion pixels among all pixels predicted as lesions:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{17}$$

The Dice coefficient is another widely used overlap-based metric, which is more sensitive to small targets compared to IoU:

$$\text{Dice} = \frac{2TP}{2TP + FP + FN} \tag{18}$$

HD95 evaluates the 95th percentile of the maximum distance between predicted and ground-truth boundaries, effectively reducing the influence of outliers and reflecting worst-case boundary deviation. Let S_p and S_g denote the sets of boundary points for the predicted segmentation and ground truth, respectively. For a predicted boundary point $x \in S_p$, $d(x, S_g)$ denotes the minimum Euclidean distance to the ground-truth boundary set, and similarly for $y \in S_g$:

$$\begin{aligned}\text{HD95}(S_p, S_g) = \max\big\{&\text{percentile95}\big(d(x, S_g) \mid x \in S_p\big),\\ &\text{percentile95}\big(d(y, S_p) \mid y \in S_g\big)\big\}\end{aligned} \tag{19}$$

ASSD measures the average symmetric distance between predicted and ground-truth boundaries, reflecting boundary consistency. Smaller values indicate better boundary alignment:

$$\text{ASSD}(S_p, S_g) = \frac{1}{|S_p| + |S_g|}\left(\sum_{x \in S_p} d(x, S_p) + \sum_{y \in S_g} d(y, S_g)\right) \tag{20}$$

4.3 Implementation Details

Data augmentation techniques, including random horizontal and vertical flipping as well as random rotation, are applied to augment malignant samples in the BUSI dataset [11]. During training, a unified set of hyperparameters is adopted across all experiments. To enhance generalization ability and robustness, the BUSI dataset is randomly split into training, validation, and test sets with a ratio of 8:2:2. All ultrasound images are resized to 256×256 pixels to ensure consistent input dimensions.

The model is trained directly with optimal weight selection. Specifically, the AdamW optimizer with a weight decay of 0.001 is employed. The initial learning rate, batch size, and number of training epochs are set to 1×10^{-3}, 4, and 1000, respectively. All experiments are conducted using PyTorch 2.0.0 under a Python 3.10 environment on an NVIDIA RTX 4090 GPU and an Intel® Xeon® Platinum 8352V CPU. Through empirical evaluation, the optimal weights for the hybrid loss are determined as $\lambda_{focal} = 0.25$, $\lambda_{tversky} = 0.6$, and $\lambda_{edge} = 0.05$, $\lambda_{boundary} = 0.1$.

4.4 Experimental Results

The experimental results comprehensively evaluate the proposed SEFF-Net through qualitative visualization, quantitative analysis, and comparison with state-of-the-art methods across multiple datasets. As illustrated in Figure 4, segmentation results on the BUSI dataset demonstrate that SEFF-Net performs consistently well under different lesion scales. For large lesions, the model accurately captures the overall lesion morphology, with predictions highly consistent with the ground truth in terms of both region coverage and contour alignment. Almost no under-segmentation or boundary discontinuities are observed, indicating strong global perception and region modeling capability.

For the more challenging scenario of small lesions, where target regions are limited in size, boundaries are blurred, and noise interference is severe, SEFF-Net still successfully localizes lesion regions and preserves boundary continuity in most cases. Only slight shape deviations or minor scale shrinkage are observed in a few samples. Overall, the qualitative results indicate that the proposed model exhibits stable and robust segmentation performance across lesions of different scales, with particular advantages in small lesion detection and boundary preservation.

Quantitative evaluation results are summarized in Tables 1 and 2. On the BUSI dataset, SEFF-Net achieves the best performance among all compared methods, attaining the highest mIoU (86.88%), OA (96.83%), and significantly improved boundary accuracy as reflected by the lowest ASSD (11.15). These results demonstrate that SEFF-Net can effectively learn complex lesion morphologies and boundary characteristics. In contrast, many existing methods suffer from performance degradation when lesion

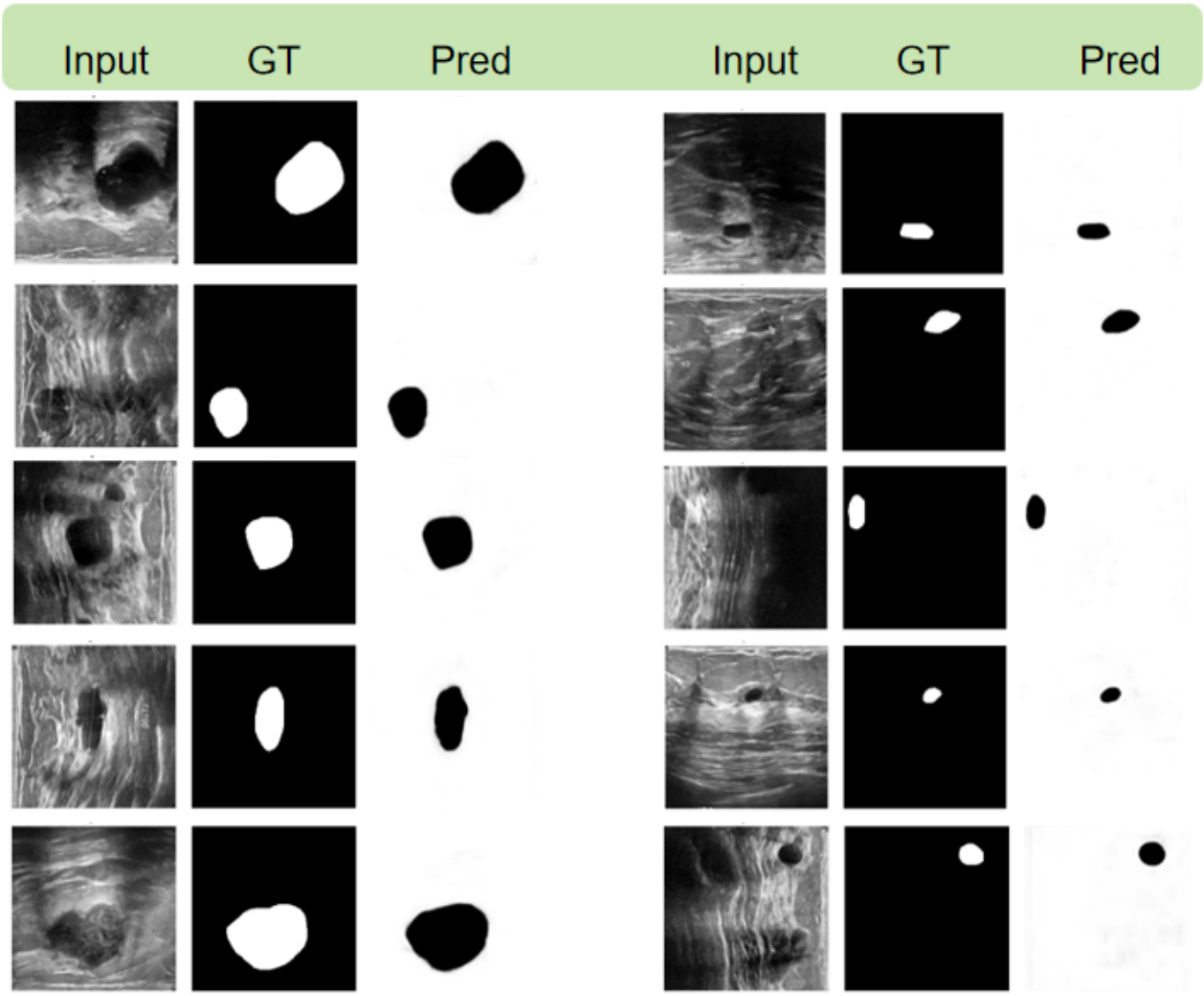

Figure 4. Segmentation results (large lesions on the left, small lesions on the right).

contours are blurred or when cancer-like artifact regions interfere with boundary localization, which is common in BUSI images.

Table 1. Performance of the model on different datasets.

	BUSI	UDIAT	OMI	BUET
mIoU	0.868	0.799	0.796	0.780
OA	0.968	0.970	0.955	0.921
Kappa	0.832	0.709	0.738	0.713
Precision	0.864	0.668	0.691	0.732
IoU	0.773	0.628	0.644	0.650
Dice	0.848	0.720	0.762	0.752
HD95	46.867	70.159	67.625	50.923
ASSD	11.157	19.345	21.015	14.854

In cross-dataset evaluation, SEFF-Net is trained solely on the BUSI dataset and directly tested on UDIAT, OMI, and BUET without any retraining or fine-tuning. Despite substantial differences in imaging devices, annotation standards, noise levels, and lesion morphologies, the proposed model consistently achieves high segmentation accuracy. Specifically,

Table 2. Performance of different models on the BUSI dataset.

	OA	Kappa	Precision	Recall	mIoU	ASSD
UNet	95.96	74.40	80.83	72.81	78.88	20.58
SegNet	95.86	73.67	80.41	71.93	78.38	24.07
TransUnet	95.62	72.58	78.02	72.16	77.64	33.07
MDA-Net	94.96	68.71	73.67	69.37	75.11	19.81
DualA-Net	94.02	65.32	65.6	71.93	72.91	35.12
EGEUnet	95.69	72.92	78.54	72.27	77.87	24.07
MGCC	95.84	74.35	78.28	75.06	78.03	23.89
EH-Former	95.85	73.79	79.46	73.30	77.29	24.29
ScribFormer	94.71	68.71	69.84	73.50	75.06	26.67
SEFF-Net	**96.83**	**83.24**	**86.47**	**85.62**	**86.88**	**11.15**

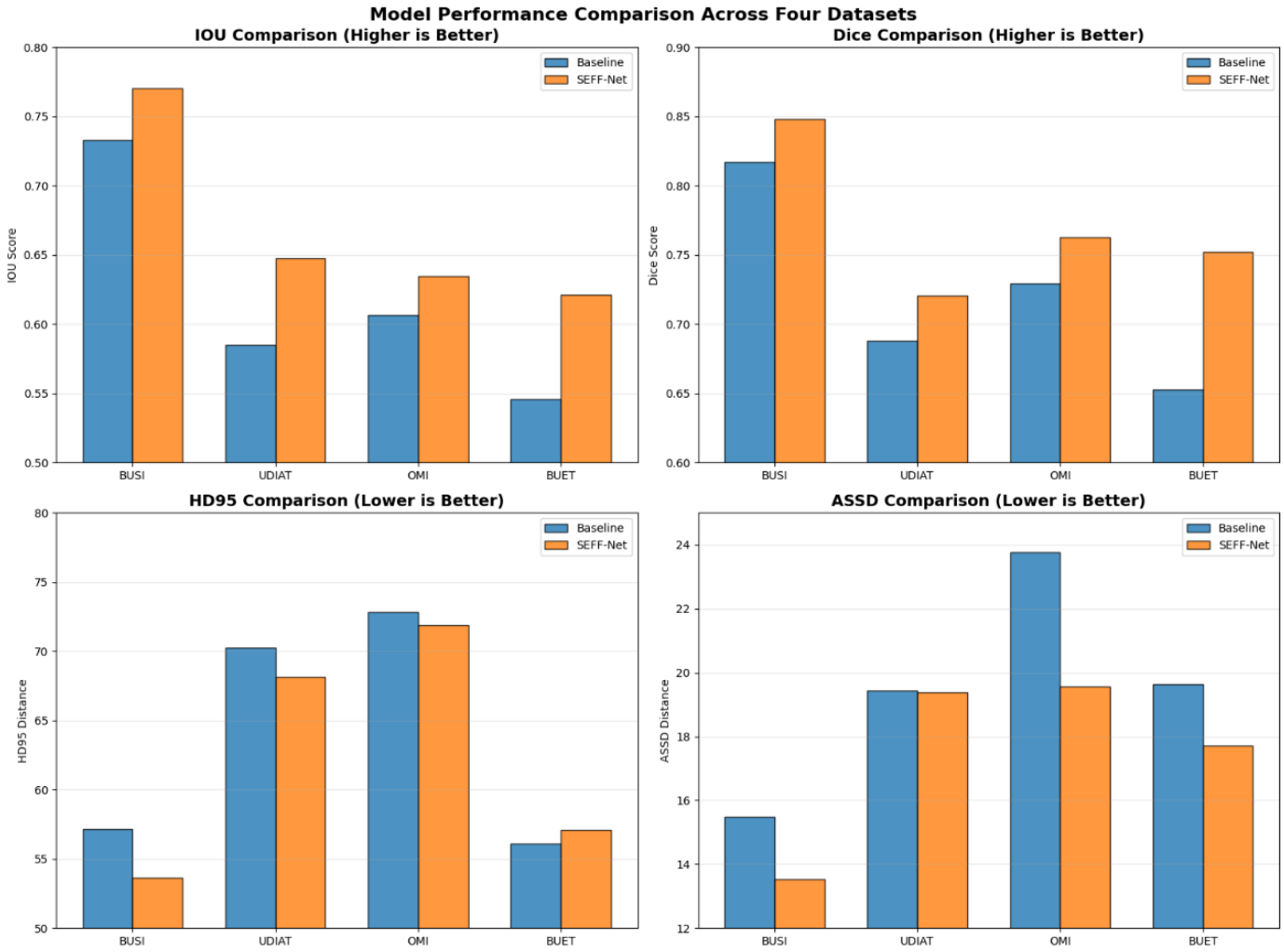

Figure 5. Performance comparison of the model on different datasets.

SEFF-Net obtains mIoU scores of 0.799 and 0.796 on the UDIAT and OMI datasets, respectively, indicating strong generalization capability across different imaging domains. Although the BUET dataset contains lower-quality images with more severe noise and irregular lesion shapes, SEFF-Net still maintains competitive performance, particularly in boundary-related metrics such as HD95 and ASSD, where the ASSD value is reduced to 14.854.

From the perspective of boundary accuracy, SEFF-Net consistently demonstrates superior contour localization ability across datasets. The BUSI dataset achieves the lowest HD95 and ASSD values, while slightly increased boundary errors are observed on UDIAT and OMI due to higher noise levels and annotation variability. Nevertheless, the overall trend confirms that SEFF-Net maintains stable boundary consistency and robustness under diverse imaging conditions.

In summary, the experimental results across all four datasets validate the robustness and effectiveness of SEFF-Net in breast ultrasound image segmentation. Benefiting from the self-learning edge enhancement mechanism of the SEEM module and the multi-scale feature fusion strategy of SFFM, the proposed model achieves superior performance in both qualitative visualization and quantitative evaluation. These results demonstrate strong cross-domain adaptability and highlight the potential clinical applicability of SEFF-Net in complex and unseen ultrasound imaging scenarios.

4.5 Ablation Studies

To evaluate the effectiveness of each component in SEFF-Net, comprehensive ablation experiments are conducted on the BUSI dataset, with additional reference to external dataset results. All experiments are performed under identical hardware conditions and deep learning frameworks, with standardized

training strategies, hyperparameters, and dataset splits. The model is decomposed into five configurations for analysis, and the results are summarized in Table 3.

Table 3. Comparison of ablation experiment performance.

	CMUNeXt	+Agg	+SEEM	+SFFM	+Loss
mIoU	0.844	0.849	0.858	0.861	0.868
OA	0.961	0.961	0.963	0.967	0.968
Kappa	0.798	0.804	0.817	0.821	0.832
precision	0.786	0.802	0.836	0.839	0.864
IOU	0.732	0.742	0.758	0.759	0.773
Dice	0.817	0.824	0.835	0.838	0.848
HD95	57.154	57.209	54.714	52.373	46.867
ASSD	15.484	14.329	14.781	14.028	11.157

Considering the limited training data and to prevent overfitting on the BUSI dataset, we select CMUNeXt [25] with a relatively small parameter count as the baseline model. Its encoder is composed of CMUNeXt blocks to extract deep features by effectively mixing spatial and channel information. Next, the AGGM module is introduced to aggregate features extracted from the last three encoder layers, forming a deep semantic feature block. Subsequently, the SEEM module is applied to enhance edge information from the first three encoder layers, producing an edge feature block. Finally, the SFFM module is employed to fuse deep semantic features and edge features in a weighted manner.

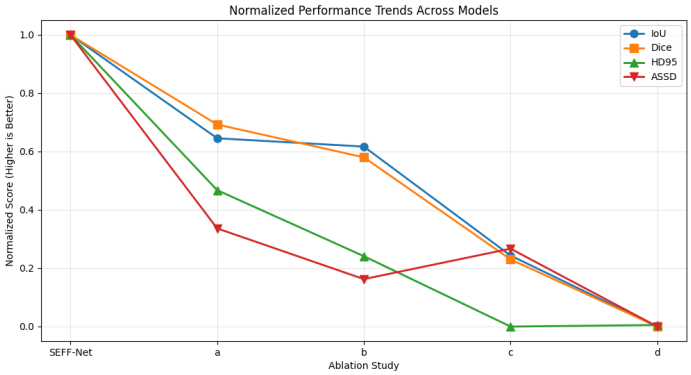

Figure 6. Comparison chart of ablation experiment indicators.

Eight evaluation metrics are used to assess performance differences among ablation settings. Before introducing SFFM, deep features and edge features are directly fused at a fixed ratio of 1:1. As shown in Figures 5 and 6, the AGGM slightly improves segmentation performance compared to the baseline, but yields limited improvement in boundary segmentation. After incorporating the edge enhancement module, the HD95 metric decreases significantly. With the introduction of SFFM, the model can automatically adapt the fusion ratio between deep semantic and edge features, resulting in substantial improvements in both segmentation accuracy and boundary quality. Furthermore, the improved loss function provides explicit supervision for boundary learning, leading to further performance gains. Specifically, the model achieves an IoU of 0.773 and a Dice score of 0.848, while ASSD and HD95 are reduced to 11.157 and 46.867, respectively.

5 Conclusion

To improve the accuracy of breast ultrasound image segmentation, we propose SEFF-Net, which effectively exploits both deep semantic and edge features in ultrasound images. An efficient AGGM is introduced to aggregate multi-scale deep semantic features, while the proposed SEEM module extracts informative edge features from multi-scale encoder outputs. Subsequently, the SFFM module adaptively integrates shallow edge features and rich deep semantic representations through learned weights. SEFF-Net demonstrates excellent segmentation performance on breast cancer ultrasound datasets, and ablation studies confirm the effectiveness of each proposed module. Furthermore, the proposed method achieves competitive performance on three additional datasets, indicating strong generalization ability.

SEFF-Net successfully achieves accurate segmentation for most breast ultrasound images, even when lesion boundaries are severely blurred. Although the current model already demonstrates strong accuracy, further optimization in terms of computational efficiency is required for deployment in clinical diagnostic equipment. Future work will focus on reducing model complexity and parameter count to accelerate inference speed, enabling SEFF-Net to better assist clinicians in real-time preliminary diagnosis.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported by the Taiyuan Bureau of Science and Technology through the Science, Technology and Innovation Program of the National Regional Medical Center under Grant 202243.

Conflicts of Interest

Senmao Wang is affiliated with the Taiyuan Central Hospital, Taiyuan 030024, China. The authors declare

that this affiliation had no influence on the study design, data collection, analysis, interpretation, or the decision to publish, and that no other competing interests exist.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

This study used solely publicly available, de-identified datasets with no human subject involvement or new data collection, no ethical approval or informed consent was required.

References

[1] Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians, 71*(3), 209-249. [CrossRef]

[2] Lee, C., Phillips, J., Sung, J., Lewin, J., & Newell, M. (2022). Breast Imaging Reporting and Data System: ACR BI-RADS breast imaging atlas. *Contrast enhanced mammography (CEM)(a supplement to ACR BI-RADS® Mammography 2013), 5th edn. American College of Radiology, Reston*.

[3] Wang, R., Wang, Z., Xiao, Y., Liu, X., Tan, G., & Liu, J. (2025). Application of deep learning on automated breast ultrasound: Current developments, challenges, and opportunities. *Meta-Radiology, 3*(2), 100138. [CrossRef]

[4] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Cham: Springer international publishing. [CrossRef]

[5] Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence, 39*(12), 2481-2495. [CrossRef]

[6] Huang, Z., Wang, X., Wei, Y., Huang, L., Shi, H., Liu, W., & Huang, T. S. (2020). CCNet: Criss-Cross Attention for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 45*(6), 6896-6908. [CrossRef]

[7] Al-Karawi, D., Al-Zaidi, S., Helael, K. A., Obeidat, N., Mouhsen, A. M., Ajam, T., ... & Ahmed, M. H. (2024). A review of artificial intelligence in breast imaging. *Tomography, 10*(5), 705-726. [CrossRef]

[8] He, Q., Yang, Q., & Xie, M. (2023). HCTNet: A hybrid CNN-transformer network for breast ultrasound image segmentation. *Computers in Biology and Medicine, 155*, 106629. [CrossRef]

[9] Zhang, M., Huang, A., Yang, D., & Xu, R. (2023). Boundary-oriented network for automatic breast tumor segmentation in ultrasound images. *Ultrasonic Imaging, 45*(2), 62-73. [CrossRef]

[10] Ning, Z., Zhong, S., Feng, Q., Chen, W., & Zhang, Y. (2021). SMU-Net: Saliency-guided morphology-aware U-Net for breast lesion segmentation in ultrasound image. *IEEE transactions on medical imaging, 41*(2), 476-490. [CrossRef]

[11] Al-Dhabyani, W., Gomaa, M., Khaled, H., & Fahmy, A. (2020). Dataset of breast ultrasound images. *Data in brief, 28*, 104863. [CrossRef]

[12] Peng, Y., Chen, D. Z., & Sonka, M. (2025, April). U-net v2: Rethinking the skip connections of u-net for medical image segmentation. In *2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI)* (pp. 1-5). IEEE. [CrossRef]

[13] Wu, R., Lu, X., Yao, Z., & Ma, Y. (2024). MFMSNet: a multi-frequency and multi-scale interactive CNN-transformer hybrid network for breast ultrasound image segmentation. *Computers in Biology and Medicine, 177*, 108616. [CrossRef]

[14] Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., ... & Zhou, Y. (2021). Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*.

[15] Sinha, A., & Dolz, J. (2020). Multi-scale self-guided attention for medical image segmentation. *IEEE journal of biomedical and health informatics, 25*(1), 121-130. [CrossRef]

[16] Zhuang, J. (2018). LadderNet: Multi-path networks based on U-Net for medical image segmentation. *arXiv preprint arXiv:1810.07810*.

[17] Kumar, A., Kim, J., Lyndon, D., Fulham, M., & Feng, D. (2016). An ensemble of fine-tuned convolutional neural networks for medical image classification. *IEEE journal of biomedical and health informatics, 21*(1), 31-40. [CrossRef]

[18] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017, July). Feature Pyramid Networks for Object Detection. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 936-944). IEEE. [CrossRef]

[19] Hu, J., Shen, L., Albanie, S., Sun, G., & Wu, E. (2019). Squeeze-and-Excitation Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 42*(8), 2011-2023. [CrossRef]

[20] Shelhamer, E., Long, J., & Darrell, T. (2017). Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(4), 640-651. [CrossRef]

[21] Dar, M. F., & Ganivada, A. (2025). Adaptive ensemble loss and multi-scale attention in breast ultrasound segmentation with UMA-Net. *Medical & Biological Engineering & Computing*, *63*(6), 1697-1713. [CrossRef]

[22] Salem, S., Mostafa, A., Ghalwash, Y. E., Mahmoud, M. N., Elnokrashy, A. F., & Mahmoud, A. M. (2023, November). Computer-Aided System for Breast Cancer Lesion Segmentation and Classification Using Ultrasound Images. In *International Conference on e-Health and Bioengineering* (pp. 297-305). Cham: Springer Nature Switzerland. [CrossRef]

[23] Agarwal, R., Diaz, O., Yap, M. H., Llado, X., & Marti, R. (2020). Deep learning for mass detection in full field digital mammograms. *Computers in biology and medicine*, *121*, 103774. [CrossRef]

[24] Hussain, M. S. (2019). *Breast Lesion Classification from Bi-modal Ultrasound Images by Convolutional Neural Network* (Doctoral dissertation, Bangladesh University of Engineering and Technology).

[25] Tang, F., Ding, J., Quan, Q., Wang, L., Ning, C., & Zhou, S. K. (2024, May). Cmunext: An efficient medical image segmentation network based on large kernel and skip fusion. In *2024 IEEE International Symposium on Biomedical Imaging* (*ISBI*) (pp. 1-5). IEEE. [CrossRef]

Tingli Su received her B.E. degree in Mechatronic Engineering and her Ph.D. degree in Control Science and Engineering from the Beijing Institute of Technology, China. From 2009 to 2012, she was a visiting student at the University of Bristol, where she conducted research on networked control systems. She is currently an Associate Professor at the Beijing Technology and Business University. Her research interests include multi-sensor fusion, data analytics, and time series-based state estimation. (Email: sutingli@btbu.edu.cn)

Rui Wan received the B.Eng. degree in Industrial Automation from East China Jiaotong University, China, in 2023. He is currently pursuing the M.Eng. degree in Control Engineering at Beijing Technology and Business University, Beijing, China. His research interests include medical image segmentation, medical image classification, deep learning, and related areas. (Email: 2330602073@st.btbu.edu.cn)

Senmao Wang obtained his Bachelor of Medicine in Clinical Medicine and Master of Medicine in Surgery from China Medical University. He is currently an attending physician and Secretary of the Surgical Teaching and Research Office at Taiyuan Central Hospital. His research interests include early diagnosis of breast cancer and the application of artificial intelligence in breast diseases. (Email: wangjiaoyoutiao@163.com)

Yuting Bai received the Ph.D. degree in control science and engineering from Beijing Institute of Technology, the M.S. degree in management science and engineering from Beijing Technology and Business University, and the B.S. degree in automation from Beijing Technology and Business University. He is now an associate professor in Beijing Technology and Business University. His research mainly covers information fusion, machine learning and decision-making method. (Email: baiyuting@btbu.edu.cn)