Boundary-Enhanced Attention Network for Breast Mass Segmentation

Rong Chen*, Stephen Karungaru*, Kenji Terada*, and Linhuang Wang†
* Tokushima University, Japan, E-mail: karungaru@tokushima-u.ac.jp
† The University of Tokyo, Japan, E-mail: wlh@g.ecc.u-tokyo.ac.jp

***Abstract*—Breast cancer is the most common malignancy in women, and mammography is widely used for early detection due to its clarity and low cost. The shape and boundary of breast masses are key features for distinguishing between benign and malignant tumors. However, accurate automatic segmentation remains challenging because the mass boundaries are often unclear, vary in shape, and have low contrast to surrounding tissues. To address these challenges, a Boundary-Enhanced Attention Network (BEANet) is proposed. This model is based on the U-Net architecture and uses a ResNet34 encoder to improve the extraction of deep features. It incorporates two attention mechanisms: Squeeze-and-Excitation (SE) blocks for channel attention, and Attention Gates (AG) in skip connections to enhance focus on important regions. Additionally, a boundary supervision branch is added to better detect and refine the edges of the masses. Experiments on the CBIS-DDSM dataset show that BEANet achieves a Dice Similarity Coefficient (DSC) of 88.39% and outperforms U-Net and other models in detecting small mass and segmenting unclear boundaries, highlighting its potential for accurate and robust breast mass segmentation.**

***Index Terms*—Breast mass segmentation, Mammography, Deep learning, Attention mechanism, Boundary supervision**

I. Introduction

Accurate segmentation of breast masses in mammograms is a fundamental task in computer-aided diagnosis (CAD), directly impacting mass detection, quantitative assessment, and treatment planning. However, automatic segmentation remains challenging due to irregular mass shapes, low mass-to-tissue contrast, and ambiguous boundaries. These issues require models capable of both precise localization and edge-aware prediction.

In recent years, deep learning—particularly convolutional neural networks (CNNs)—has achieved remarkable success in various computer vision tasks such as classification, detection, and segmentation [1], [2]. In the medical imaging domain, CNN-based methods have significantly advanced breast mass segmentation. Among them, U-Net [3] is widely used for its encoder–decoder structure with skip connections. However, its basic feature fusion limits effective integration of semantic and spatial information, reducing boundary precision. Improvements like CE-Net [4] and UNet++ [5] enhance contextual or semantic alignment, but often increase complexity.

This work was supported by JST SPRING, Grant Number JPMJSP2113.
Tokushima University, Japan
Corresponding author:
Stephen Karungaru
E-mail: karungaru@tokushima-u.ac.jp

Attention mechanisms have been introduced to enhance focus on diagnostically relevant features. Attention U-Net [6] incorporates attention gates (AG) in skip connections to suppress irrelevant background, while AUNet [7] extends attention modeling to decoder stages. However, these methods often lack explicit modeling of mass boundaries. Squeeze-and-Excitation (SE) blocks [8], though effective for channel-wise recalibration, also fail to capture fine-grained spatial structures crucial for edge delineation.

Recognizing the diagnostic importance of boundary information, several boundary-aware networks such as Shape Stream [9] and BA-Net [10] have been proposed. These models introduce edge-specific branches to enhance contour sensitivity. Nevertheless, most existing methods treat semantic feature extraction and boundary modeling as separate processes, leading to inefficient integration and suboptimal performance, particularly in cases involving blurry edges or tissue adhesion. Furthermore, the increased complexity of these models poses challenges for deployment in clinical environments where efficiency and robustness are critical.

To address these limitations, this study proposes a Boundary-Enhanced Attention Network (BEANet)-a lightweight and effective segmentation framework designed to jointly enhance semantic feature representation and boundary awareness. Built upon the U-Net architecture, BEANet incorporates a ResNet34 encoder [11] for deep feature extraction, integrates dual attention mechanisms, and introduces explicit boundary supervision. To further improve the localization performance, a region of interest (ROI) cropping strategy is employed to suppress background interference and enhance the detection of breast masses. The overall design enhances both the accuracy and robustness of mass delineation in mammograms, contributing to more reliable computer-aided diagnosis. The main contributions of this study are as follows:

- A ResNet34 encoder is employed to enhance multi-scale semantic feature extraction while maintaining model efficiency.
- A dual attention strategy is adopted by integrating SE blocks into the center and decoder stages, and embedding AG blocks in the skip connections.
- A dedicated boundary supervision branch is incorporated to explicitly model mass contours, enabling finer edge

979-8-3315-7206-8/25/$31.00

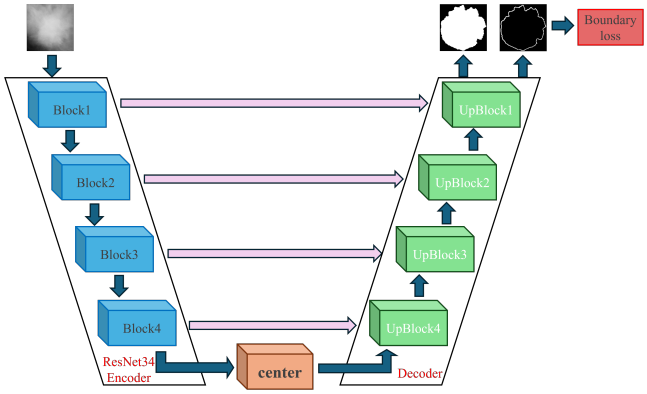

Fig. 1. Overview of the BEANet architecture, featuring a ResNet34 encoder, attention-enhanced center and decoder, Attention Gates (AG)-guided skip connections, and dual outputs for mass and boundary segmentation.

localization and enhancing boundary sensitivity.

II. Methodology

A. Overall Architecture

The proposed BEANet is designed for accurate breast mass segmentation in mammography. Built upon the U-Net framework, BEANet consists of three key components: a ResNet34 encoder, an attention-augmented center and decoder with SE blocks, and AG-guided skip connections. A dual-output decoder simultaneously predicts mass regions and boundary maps, supervised by a joint loss function to enhance both semantic segmentation and boundary localization. The overall architecture is illustrated in Fig. 1.

In the encoder, a pre-trained ResNet34 replaces standard U-Net blocks, enabling deeper semantic feature extraction. The decoder reconstructs spatial resolution using a series of UpBlocks, each receiving skip-connected encoder features to recover spatial detail. SE blocks are embedded in the center and decoder to perform channel recalibration, emphasizing discriminative features during reconstruction. AG blocks are used in skip connections to selectively transmit spatially relevant features, suppressing background noise and enhancing boundary localization. BEANet produces two outputs: a segmentation mask and a boundary map. The mask branch captures mass morphology. The boundary branch focuses on edge precision, guided by Edge loss to improve contour consistency. This dual-branch design enables the mask to learn semantic regions, while the boundary branch reinforces structural accuracy. Boundary supervision also acts as a regularization mechanism, mitigating annotation noise and improving robustness on blurred or adhesive mass edges.

In summary, BEANet balances semantic accuracy and boundary precision through an encoder-decoder backbone enhanced with attention and dual-task supervision, making it effective and efficient for real-world clinical applications.

B. Architectural Components

BEANet is built upon the classical U-Net architecture, adopting a symmetric encoder–decoder design with skip connections for high-precision pixel-wise segmentation. Compared to the original U-Net, BEANet enhances feature representation and structural modeling while maintaining computational efficiency.

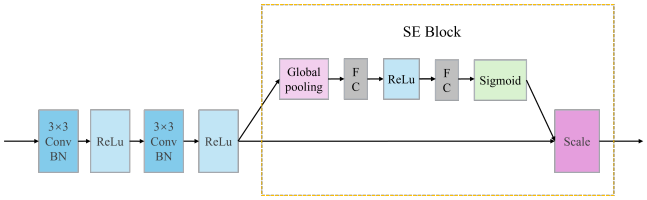

Fig. 2. Center layer of BEANet with a Squeeze-and-Excitation (SE) block for channel attention enhancement. The SE block recalibrates feature maps by modeling inter-channel dependencies, emphasizing informative channels while suppressing irrelevant ones.

1) ResNet34 Encoder: BEANet replaces the shallow convolutional encoder in U-Net with a pre-trained ResNet34. It begins with a 7×7 convolution (stride = 2), followed by four residual stages producing feature maps at $1/4$, $1/8$, $1/16$, and $1/32$ of the input resolution. These multi-scale features are connected to corresponding decoder stages via skip connections. Residual blocks with identity mappings facilitate gradient flow and improve feature stability.gradient problem in deep networks and help stabilize feature propagation.

2) Center Layer with SE block: At the center, an SE block recalibrates feature responses by modeling inter-channel dependencies. As shown in Fig. 2, global average pooling generates a compact descriptor, which passes through two fully connected layers and a sigmoid activation to produce channel attention weights. These weights are used to emphasize informative channels and suppress irrelevant ones.

3) Decoder with Attention-Guided Skip Connections: The decoder mirrors the encoder and consists of UpBlocks that progressively restore spatial resolution. Each UpBlock includes transposed convolution, skip connections, and convolutional refinement, as shown in Fig. 3. To further enhance feature representation, BEANet integrates two attention mechanisms within each UpBlock: (a) SE Block (Channel Attention): Each UpBlock includes an SE block (same as in the center) to refine channel-level feature importance. (b) AG Block (Spatial Attention): AG blocks are applied to the skip connections, where they combine encoder and decoder features to generate spatial attention maps. This mechanism emphasizes mass-relevant regions while suppressing background noise, improving spatial precision.

4) Boundary Supervision Output Branch: To enhance contour delineation, BEANet includes a parallel boundary prediction branch. This branch outputs a boundary map aligned with the input resolution using convolution and upsampling layers. Ground-truth boundary maps are generated by applying the Sobel operator within the ROI extracted from the mask. Compared to morphological gradients, the Sobel operator provides better edge continuity and directional sensitivity, producing sharper binary boundary labels.

During training, the predicted boundary map is supervised

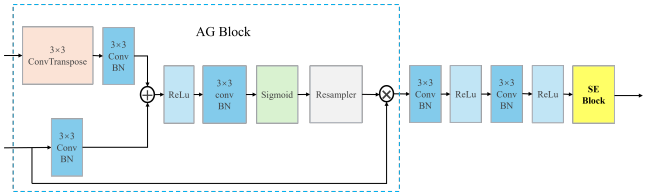

Fig. 3. Structure of an UpBlock in the decoder. Each block includes upsampling, convolution, and dual attention: an SE block that recalibrates channel responses to emphasize informative features, and an AG that guides spatial focus toward mass-relevant regions.

using these binary edges via a weighted binary cross-entropy loss, termed Edge Loss. This loss emphasizes edge pixels to address class imbalance and guide the network in learning fine-grained boundary structures, ultimately improving segmentation accuracy in cases with blurred or adhesive mass borders.

C. Loss Function

To handle the challenges of class imbalance and blurry boundaries in breast mass segmentation, BEANet adopts a dual-branch structure with a composite loss function that jointly supervises region and edge predictions.

Dice loss [12] is employed to optimize the spatial overlap between the predicted masks and the ground truth, effectively addressing foreground-background imbalance. It is applied to both the mass mask and boundary map outputs:

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2\sum y_{\text{true}} y_{\text{pred}}}{\sum y_{\text{true}}^2 + \sum y_{\text{pred}}^2 + \epsilon}, \tag{1}$$

where ϵ prevents division by zero. This loss promotes region-level accuracy and smooth boundary predictions.

Since boundary pixels typically constitute only a small portion of the image, a significant class imbalance exists between edge and non-edge pixels. To address this, a weighted binary cross-entropy loss, referred to as Edge loss, is introduced to emphasize learning in the edge regions:

$$\begin{aligned}\mathcal{L}_{\text{Edge}} = & -\beta \sum_{j \in y_+} \log P(y_{\text{pred},j} = 1) \\ & -(1-\beta) \sum_{j \in y_-} \log P(y_{\text{pred},j} = 0),\end{aligned} \tag{2}$$

where y_+ and y_- represent the sets of boundary and non-boundary pixels, respectively, and β is the proportion of non-boundary pixels. Probabilities $P(y_{\text{pred},j})$ are computed using a sigmoid activation function. This Edge loss helps compensate for class imbalance and increases the model's sensitivity to fine boundary structures.

Inspired by Hatamizadeh et al. [13], We integrate region-level and edge-level supervision into a unified training objective. The final composite loss function is defined as:

$$\begin{aligned}\mathcal{L}_{\text{total}} = & \lambda_1 \mathcal{L}_{\text{Dice}}(y_{\text{pred}}, y_{\text{true}}) \\ & + \lambda_2 \mathcal{L}_{\text{Dice}}(s_{\text{pred}}, s_{\text{true}}) \\ & + \lambda_3 \mathcal{L}_{\text{Edge}}(s_{\text{pred}}, s_{\text{true}}),\end{aligned} \tag{3}$$

where $y_{\text{pred}}, y_{\text{true}}$ are the predicted and ground-truth segmentation masks, and $s_{\text{pred}}, s_{\text{true}}$ are the predicted and ground-truth boundary maps. The parameters $\lambda_1, \lambda_2, \lambda_3$ control the relative contributions of each term. In this study, we used $\lambda_1 = 1.0, \lambda_2 = 0.5, \lambda_3 = 0.1$. These weights were empirically determined to balance region accuracy and boundary precision, and the resulting performance remained robust to small perturbations.

III. Experimental Results and Discussion

A. Dataset and Preprocessing

We evaluate BEANet on two publicly available mammography datasets: CBIS-DDSM [14] and INbreast [15]. CBIS-DDSM (Curated Breast Imaging Subset of DDSM) is a standardized subset of the DDSM database, providing high-resolution mammograms with expert-annotated mass masks. We used all 1,592 mammograms annotated for mass segmentation, with 1,231 for training and 361 for testing. INbreast is a full-field digital mammography dataset with precise pixel-level annotations. To assess cross-dataset generalization, 107 images with confirmed mass labels were selected from INbreast for evaluation.

A unified preprocessing pipeline was applied to all images. Background pixels were removed, and ROIs were cropped using the smallest bounding boxes enclosing each mass. To enhance local contrast, particularly for low-intensity or blurry masses, CLAHE (Contrast Limited Adaptive Histogram Equalization) was applied [16]. All ROIs and corresponding masks were resized to 256×256 pixels (px), normalized to the [0, 1] range, and further standardized using Z-score normalization. Binary masks were processed to ensure foreground pixels were labeled as 1 and background as 0. For boundary supervision, edge maps were generated by applying the Sobel operator [17] to the mass masks, capturing gradient-based edges with directional sensitivity. These maps were used as auxiliary labels for the boundary prediction branch.

To improve model robustness, standard data augmentation strategies—including horizontal/vertical flipping, random rotation, and zooming—were applied during training [18]. This preprocessing strategy ensures consistent input quality and supports the model in capturing both semantic content and fine-grained structural boundaries critical for accurate mass segmentation.

B. Implementation Details

All experiments were conducted on a workstation equipped with an Intel Core i7-14700F CPU, an NVIDIA GeForce RTX 4070 Ti SUPER GPU (16 GB), and 64 GB RAM. The software environment was based on Windows 10, using Python 3.10 and PyTorch 2.1 for model implementation and training.

The Adam optimizer was employed with an initial learning rate of 1×10^{-4}, which was adaptively adjusted using the ReduceLROnPlateau scheduler based on validation loss. The batch size was set to 16. Automatic mixed precision (AMP) was enabled to accelerate training and reduce GPU memory usage. Training was performed for up to 150 epochs, with early

stopping applied based on validation loss (patience = 20). The model checkpoint with the highest validation Dice Similarity Coefficient (DSC) [19] was saved. To ensure fair comparison and reproducibility, all experiments used fixed random seeds and identical hyperparameter settings across trials.

C. Evaluation Metrics

This study adopts multiple evaluation metrics to measure the effectiveness of image segmentation, including the DSC, Intersection over Union (IoU) [20], and the Hausdorff distance [21]. These indicators were utilized to comprehensively assess the segmentation performance of the proposed approach.

Let TP denote true positive pixels, TN denote true negatives, FP denote false positives, FN denote false negatives.

The DSC measures the overlap between predicted and ground-truth regions. A higher DSC indicates better agreement:

$$DSC = \frac{2 \times TP}{2 \times TP + FP + FN}. \tag{4}$$

IoU evaluates the ratio between the intersection and union of the predicted and ground-truth regions:

$$IoU = \frac{TP}{TP + FP + FN}. \tag{5}$$

The Hausdorff distance quantifies boundary alignment by computing the maximum distance between the boundaries of the predicted and ground-truth masks. A lower Hausdorff distance reflects closer alignment and improved segmentation precision. Let T denote the ground truth region of breast mass manually annotated by a physician on images and $t \in T$ be a point within this region. Similarly, let P denote the predicted mass region obtained by the BEANet network, and let $p \in P$ be a point within the predicted region. The function $d(t, p)$ represents the Euclidean distance between two points. The Hausdorff distance is defined as:

$$Hausdorff(T, P) = \max \Big\{ \sup_{t \in T} \inf_{p \in P} d(t, p), \sup_{p \in P} \inf_{t \in T} d(t, p) \Big\}, \tag{6}$$

where, sup represents the upper bound and inf represents the lower bound. In this study, all images were resized to 256×256 px prior to evaluation, and Hausdorff distances were calculated in px units unless otherwise specified.

D. Results and Discussion

To comprehensively evaluate the segmentation performance of the proposed BEANet, we conducted comparative experiments on two public mammography datasets: CBIS-DDSM and INbreast. The comparison includes four representative methods: U-Net, Attention U-Net, AUNet, and BA-Net. Table I summarizes the segmentation results on the CBIS-DDSM dataset, evaluated using DSC, IoU, and Hausdorff distance.

As shown in Table I, BEANet achieves the best performance across all three evaluation metrics. Specifically, BEANet attains a DSC of 88.39%, representing an improvement of 6.13% over the conventional U-Net. In addition, the Hausdorff distance is significantly reduced to 2.90, more than 50% lower than the 6.09 obtained by U-Net, clearly demonstrating BEANet's superior capability in boundary precision. These performance gains are attributed to the synergistic effect of multi-scale attention mechanisms (SE and AG blocks) and boundary-aware supervision. The channel and spatial attention modules enhance the discriminability and robustness of feature representations, while the boundary prediction branch explicitly guides the model to focus on structural contours, particularly in regions with blurred or adhesive mass edges. To qualitatively assess segmentation performance, representative examples are illustrated in Fig. 4, including the original image, ROI, ground truth, and predicted masks from U-Net, Attention U-Net, AUNet, BA-Net, and BEANet. It is evident that BEANet produces more accurate and complete mass contours, with superior boundary clarity. Even in challenging cases involving indistinct edges or tissue adhesion, BEANet maintains high structural consistency and localization accuracy.

In summary, BEANet effectively balances model efficiency and segmentation precision by integrating attention mechanisms with boundary-aware design, enabling joint modeling of semantic regions and structural edges. The proposed model demonstrates clear advantages in breast mass segmentation.

To assess the cross-dataset generalization of BEANet, we conducted an experiment in which the model trained exclusively on the CBIS-DDSM dataset was directly applied to

TABLE I
PERFORMANCE COMPARISON ON CBIS-DDSM DATASET

Method	DSC (%)	IoU (%)	Hausdorff Distance (px)
U-Net	82.26	71.52	6.09
Attention U-Net	84.42	74.51	5.37
AUNet	85.03	73.98	4.21
BA-Net	86.98	76.68	3.86
BEANet(Ours)	**88.39**	**79.22**	**2.90**

* Bold indicates the best performance.

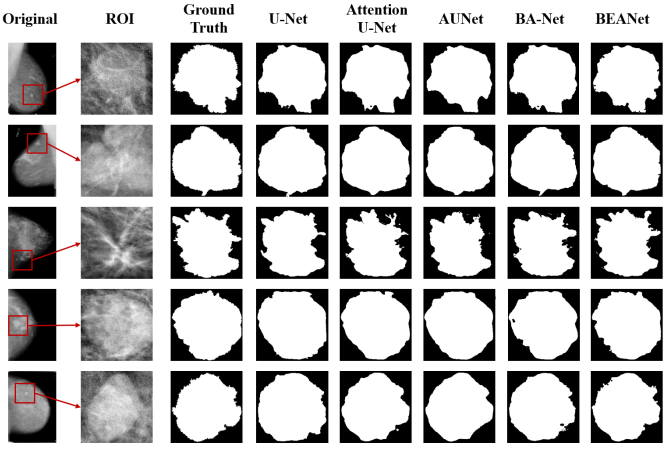

Fig. 4. Visual comparison of breast mass segmentation results on the CBIS-DDSM dataset. Showing the original image, ROI image, ground truth, and outputs of U-Net, Attention U-Net, AUNet, BA-Net, and BEANet

TABLE II
PERFORMANCE COMPARISON ON INBREAST DATASET

Method	DSC (%)	IoU (%)	Hausdorff Distance (px)
U-Net	73.03	58.93	15.17
Attention U-Net	76.47	63.54	12.01
AUNet	77.09	64.39	11.35
BA-Net	78.94	66.95	9.81
BEANet(Ours)	**79.78**	**66.36**	**8.44**

* Bold indicates the best performance.

the INbreast dataset without any fine-tuning. Both datasets underwent identical preprocessing to ensure consistency. This setting simulates a practical clinical scenario where a trained model is deployed on data from a different institution or imaging protocol. The evaluation on INbreast reflects the model's robustness to domain shifts, including variations in contrast, image quality, and annotation style. The corresponding results are summarized in Table II.

Although the INbreast dataset differs from CBIS-DDSM in terms of image source, contrast, mass morphology, and annotation style, BEANet achieves strong cross-dataset performance without fine-tuning. Specifically, it achieves a DSC of 79.78% and a Hausdorff distance of 8.44, outperforming U-Net (73.03%, 15.17) and Attention U-Net (76.47%, 12.01), highlighting its superior generalization and boundary modeling ability.

This stable transfer performance demonstrates the robustness and generalization capability of BEANet in handling cross-domain mammographic images. Despite the differences in imaging style, boundary annotation, and anatomical structures between the two datasets, BEANet effectively captures mass regions while maintaining boundary clarity. This robust performance can be attributed to BEANet's architecture, which integrates boundary-aware supervision and dual attention mechanisms. The boundary branch encourages structural consistency across domains, while SE and AG blocks enhance the model's sensitivity to key mass features, allowing it to maintain segmentation quality under distributional shifts. Fig. 5 presents visual comparisons on INbreast samples. The figure displays the original image, ROI image, ground truth, and segmentation results produced by U-Net, Attention U-Net, AUNet, BA-Net, and BEANet. BEANet delivers more complete contours, clearer boundaries, and stronger noise suppression than competing methods, confirming its effectiveness in cross-domain segmentation scenarios.

E. Ablation Study

To evaluate the contribution of each key component in BEANet, we conducted five ablation experiments, summarized in Table III. The first setting uses the original U-Net as a baseline for performance comparison. In the second, the encoder is replaced with a pre-trained ResNet34 to evaluate the impact of a deeper backbone on semantic feature extraction. The third variant introduces SE blocks into the center and decoder to enhance channel feature recalibration. The fourth

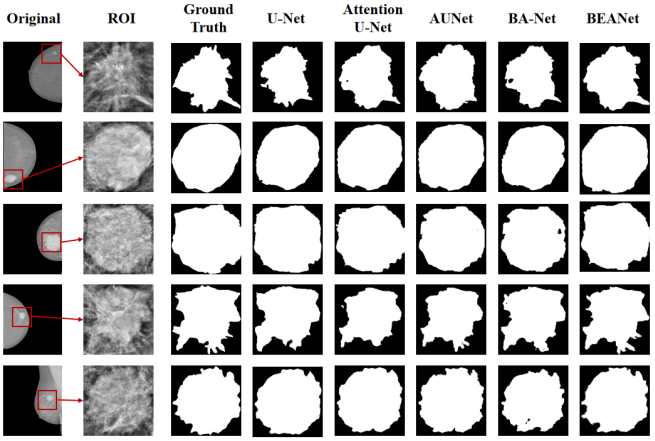

Fig. 5. Visual comparison of breast mass segmentation results on the INbreast dataset. Showing the original image, ROI image, ground truth, and outputs of U-Net, Attention U-Net, AUNet, BA-Net, and BEANet

TABLE III
STEPWISE ABLATION STUDY OF BEANET ON THE CBIS-DDSM DATASET

No.	Configuration	DSC (%)	IoU (%)	Hausdorff Distance (px)
1	Baseline (U-Net)	82.26	71.52	6.09
2	Group 1 + ResNet34 Encoder	83.62	71.90	5.58
3	Group 2 + SE blocks	84.50	74.63	5.43
4	Group 3 + AG blocks	86.19	75.75	3.99
5	**Group 4 + Boundary supervision (BEANet)**	**88.39**	**79.22**	**2.90**

* Bold indicates the best performance. SE = Squeeze-and-Excitation, AG = Attention Gate.

adds AG to the skip connections for spatial attention and background suppression, aiming to improve boundary localization. Finally, the fifth configuration incorporates the boundary-aware supervision branch into the full network to assess its effect on structural modeling and cross-domain generalization. This stepwise evaluation isolates the contribution of each module and validates the overall architectural design of BEANet.

As shown in Table III, each component of BEANet contributes positively to overall performance. The original U-Net serves as the baseline (Group 1), achieving a DSC of 82.26%, IoU of 71.52%, and a Hausdorff distance of 6.09. In Group 2, replacing the encoder with ResNet34 improves semantic feature extraction, leading to a DSC of 83.62% and a Hausdorff distance of 5.58. In Group 3, the addition of SE blocks further improves performance (DSC: 84.50%, IoU: 74.63%), indicating the effectiveness of channel attention. Group 4 incorporates AG blocks into the skip connections, resulting in noticeable boundary refinement and a Hausdorff distance reduction to 3.99. Finally, Group 5 introduces a boundary supervision branch, which yields the best results (DSC: 88.39%, IoU: 79.22%, Hausdorff distance: 2.90), confirming its role in enhancing structural consistency.

This ablation study quantitatively validates the design of

BEANet and highlights the incremental benefits of each component, particularly in improving segmentation accuracy and boundary delineation.

IV. Conclusions

In this study, we proposed a BEANet for accurate breast mass segmentation. By incorporating a ResNet34 encoder, SE blocks, AG blocks, and a boundary supervision branch, the model significantly improves both segmentation accuracy and boundary localization capability. Extensive experiments on the CBIS-DDSM and INbreast datasets demonstrate that BEANet achieves superior performance and strong generalization, especially in challenging cases with blurred boundaries and cross-domain variations.

In future work, we plan to extend BEANet to support 3D medical image segmentation tasks, such as volumetric MRI or DCE-MRI analysis. Moreover, we aim to explore its applicability to multi-organ segmentation and clinical decision support, further enhancing its potential in diverse medical imaging scenarios.

References

[1] G. Litjens, T. Kooi, B. E. Bejnordi, *et al.*, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[2] Z. Rezaei, "A review on image-based approaches for breast cancer detection, segmentation, and classification," *Expert Systems with Applications*, vol. 182, p. 115 204, 2021.

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, Springer, 2015, pp. 234–241.

[4] Z. Gu, J. Cheng, H. Fu, *et al.*, "Ce-net: Context encoder network for 2d medical image segmentation," *IEEE transactions on medical imaging*, vol. 38, no. 10, pp. 2281–2292, 2019.

[5] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Proc. Int. Workshop Deep Learn. Med. Image Anal. (DLMIA)*, Springer, 2018, pp. 3–11.

[6] O. Oktay, J. Schlemper, L. L. Folgoc, *et al.*, "Attention u-net: Learning where to look for the pancreas," *arXiv preprint*, vol. arXiv:1804.03999, 2018.

[7] H. Sun, C. Li, B. Liu, *et al.*, "Aunet: Attention-guided dense-upsampling networks for breast mass segmentation in whole mammograms," *Phys. Med. Biol.*, vol. 65, no. 5, p. 055 005, 2020.

[8] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 7132–7141.

[9] N. Zou, Z. Xiang, Y. Chen, S. Chen, and C. Qiao, "Boundary-aware cnn for semantic segmentation," *IEEE Access*, vol. 7, pp. 114 520–114 528, 2019.

[10] R. Wang, S. Chen, C. Ji, J. Fan, and Y. Li, "Boundary-aware context neural network for medical image segmentation," *Medical image analysis*, vol. 78, p. 102 395, 2022.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.

[12] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*, Ieee, 2016, pp. 565–571.

[13] A. Hatamizadeh, D. Terzopoulos, and A. Myronenko, "End-to-end boundary aware networks for medical image segmentation," in *Proc. Int. Workshop Mach. Learn. Med. Imaging (MLMI)*, Springer, 2019, pp. 187–194.

[14] R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy, and D. L. Rubin, "A curated mammography data set for use in computer-aided detection and diagnosis research," *Scientific data*, vol. 4, no. 1, pp. 1–9, 2017.

[15] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "Inbreast: Toward a full-field digital mammographic database," *Academic radiology*, vol. 19, no. 2, pp. 236–248, 2012.

[16] A. M. Reza, "Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement," *Journal of VLSI signal processing systems for signal, image and video technology*, vol. 38, pp. 35–44, 2004.

[17] W. Gao, X. Zhang, L. Yang, and H. Liu, "An improved sobel edge detection," in *2010 3rd International conference on computer science and information technology*, IEEE, vol. 5, 2010, pp. 67–71.

[18] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of big data*, vol. 6, no. 1, pp. 1–48, 2019.

[19] A. A. Taha and A. Hanbury, "Metrics for evaluating 3d medical image segmentation: Analysis, selection, and tool," *BMC medical imaging*, vol. 15, pp. 1–28, 2015.

[20] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *Int. J. Comput. Vis.*, vol. 88, pp. 303–338, 2010.

[21] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the hausdorff distance," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 15, no. 9, pp. 850–863, 1993.