Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

ElaD-Net: An Elastic Semantic Decoupling Network for Lesion Segmentation in Breast Ultrasound Images

Lijuan Xu[1,2], **Kai Wang**[1,2], **Fuqiang Yu**[*1,2], **Fenghua Tong**[1,2], **Mengran Li**[3] and **Dawei Zhao**[*1,2]
[1]Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center (National Supercomputer Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences), Jinan, China
[2]Shandong Provincial Key Laboratory of Industrial Network and Information System Security, Shandong Fundamental Research Center for Computer Science, Jinan, China
[3]School of Intelligent Systems Engineering, Sun Yat-sen University
{yufq, zhaodw}@sdas.org

Abstract

Breast diseases pose a significant threat to women's health. Automatic lesion segmentation in breast ultrasound images (BUSI) plays a crucial role in fast diagnosis. While various enhanced U-Net-based models have achieved success in multi-scale feature analysis and handling blurred boundaries, two key challenges persist that could guide the improvement of BUSI segmentation networks: 1) significant fluctuations in pixel intensity dis tribution similarity between the lesion and surrounding tissues, and 2) inconsistent transmission of spatial detail due to multi-scale lesion sampling. These issues highlight the necessity of semantic elasticity understanding and consistency control. To this end, we propose ElaD-Net, an Elastic Semantic Decoupling Network for lesion segmentation in BUSI. This network uses the pre-trained EfficientNet-B2 for multi-scale encoding of BUSI. The decoding stage features two key modules: Elastic Semantic Decoupling (ESD) and Spatial Semantic Reconstruction (SSR). ESD learns and decouples multi-frequency semantics in multi-scale channels with a self-calibration mechanism, enabling dynamic adjustment of receptive depth to resist similarity fluctuations. SSR further optimizes ESD outputs via feature branching, compression, and excitation to ensure spatial semantic consistency, thereby separately reconstructing edge and body.

1 Introduction

Breast cancer is a leading cause of cancer-related deaths in women [Wilcock and Webster, 2021]. Automated breast ultrasound image (BUSI) segmentation is a key tool to assist early diagnosis and treatment planning. Convolutional neural networks (CNNs) [O'Shea and Nash, 2015] are widely used for medical image segmentation, but traditional CNNs struggle with pixel-level classification. To address this, Ronneberger et al. [Ronneberger *et al.*, 2015] proposed U-Net, an encoder-decoder architecture with skip connections that preserves spatial details and integrates multi-level features.

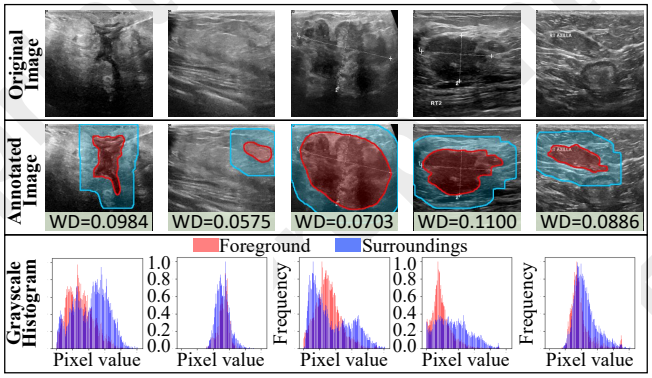

Figure 1: Foreground-surroundings similarity quantification.

In addition, several U-Net enhancements, such as **U²-Net** [Qin *et al.*, 2020], AAU-Net [Chen *et al.*, 2023a], and Swin-Unet [Cao *et al.*, 2022], integrate residual structures, multi-scale analysis, and attention mechanisms to capture intricate image details. However, these methods mainly focus on texture [Geirhos *et al.*, 2019a], making it difficult to capture global shape information in BUSI with blurred boundaries. Thus, strategies like multi-scale fusion and deep supervision, as seen in UNet++ [Zhou *et al.*, 2018], have been proposed. Yet, these approaches still struggle with the unique challenges of BUSI, hindering precise lesion segmentation.

1) Significant similarity fluctuations in pixel intensity distributions of foreground and surroundings. A major challenge in BUSI segmentation is the blurred edges of lesions. To elucidate the nature of this blur, we quantitatively analyze the similarity between the foreground (lesion) and surroundings. Figure 1 presents results from five random samples, where the lesion (in red) and surrounding tissue (in blue) exhibit similar morphological features (semantics). To quantify this, we calculate the Wasserstein Distance (WD) between their pixel intensity distributions, shown in grayscale

[*] Corresponding Authors

Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

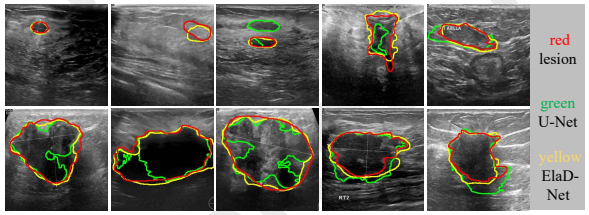

Figure 2: Spatial detail segmentation: U-Net vs. ElaD-Net.

histograms. Sample 2 demonstrates significantly higher semantic similarity than others (e.g., samples 1, 3, and 4) according to the proportion of overlapping areas. Additionally, even for the same lesion (e.g., sample 1), different edge positions show noticeable variations in similarity. We term this fluctuation in semantic similarity, a core challenge in addressing blurred boundaries often overlooked in existing studies.

2) Inconsistency in spatial details transmission caused by sampling of multi-scale lesions. U-Net, while foundational, relies heavily on downsampling, leading to the loss of key features [Rahman *et al.*, 2024] and difficulty in recovering critical details. As shown in Figure 2, edge details are often captured roughly, with regions frequently missed. This issue arises from the challenge of maintaining spatial semantic consistency during transmission, as varying lesion sizes in BUSI complicate dynamic adaptation. Misalignment between edge and internal structure sampling disrupts semantic consistency, preventing accurate restoration of spatial details.

The core challenge lies in ensuring the elastic and consistent information preservation during semantic decoupling. Our main goal is to answer how to adaptively decouple semantics despite fluctuations in intensity similarity between the lesion and surroundings, while maintaining consistent spatial semantics across multi-scale lesions during sampling.

To this end, we propose an Elastic semantic Decoupled Network (ElaD-Net) for lesion segmentation in BUSI. ElaD-Net utilizes EfficientNet-B2 for downsampling BUSI, with mobile inverted bottleneck convolutions to suppress noise and extract key features. Each upsampling step incorporates two modules: Elastic Semantic Decoupling (ESD) and Spatial Semantic Reconstruction (SSR), addressing similarity fluctuations and spatial detail inconsistencies. ESD creates multi-scale channels, performs self-calibrated decoupling through multiple branches within each channel, and fuses correlated features to generate the overall map. SSR decomposes this map into edge and body semantic feature maps, which are optimized separately and reconstruct the feature maps for predicting the edge and body. Losses for body, edge, and overall features are balanced during experimentation.

Our main contributions are: 1) An elastic semantic decoupling module resists fluctuations in the pixel intensity distribution similarity between foreground and surroundings, enhancing semantic understanding in complex cases. 2) Spatial semantic reconstruction refines the feature maps generated by ESD, maintaining consistency in spatial detail transmission during sampling. 3) Extensive experiments on three real BUSI datasets demonstrate superior segmentation performance and robustness of the proposed method ElaD-Net.

2 Related Work

Feature Enhancement for BUSI Segmentation. U-Net-based methods have achieved notable success in medical image segmentation [Dong *et al.*, 2023; Ning *et al.*, 2022; Lin *et al.*, 2022]. However, in BUSI, the foreground and surroundings have similar distributions, leading to blurred boundaries. To this end, some studies have introduced attention mechanisms and multi-scale feature fusion strategies. For example, Attention U-Net [Oktay *et al.*, 2018] enhances focus on important regions through self-attention mechanisms. ESKnet [Chen *et al.*, 2024a] designs enhanced selective kernel convolutions that integrate multiple feature map regions and adaptively recalibrate their weights from both channel and spatial dimensions, mitigating interference from less relevant areas. Moreover, to effectively suppressing irrelevant features in BUSI, Punn et al. [Punn and Agarwal, 2022] introduce residual initial convolutions (RIC) and cross-spatial attention (CSA) blocks into U-net. Liu et al. [Liu *et al.*, 2024a] propose CMFF-Net, which combines Transformer's global context information with CNN-extracted local spatial information via cross-attention feature fusion. It enhances the feature representation. Lee et al. [Lee *et al.*, 2020] introduce channel attention modules to improve U-Net's performance.

Information Loss Alleviating. Information loss during upsampling and downsampling is common in U-Net-based models for BUSI segmentation. To address this, some studies propose non-stride or dilated convolutions to retain spatial details. For example, DeepLab [Chen *et al.*, 2015; Chen *et al.*, 2018a; Chen *et al.*, 2018b] addresses information loss with dilated convolutions and multi-scale context aggregation. In addition, some studies focus on designing reconstruction modules to recover the edge information lost during downsampling. Chen et al. [Chen *et al.*, 2024b] use adaptive low-pass filter (ALPF) generators and adaptive high-pass filter (AHPF) generators to reduce position shifts caused by upsampling and recover high-frequency edge details lost during downsampling. BATFormer [Lin *et al.*, 2023]ensures accurate edge information learning at different scales by providing supervision signals. Moreover, SegNet [Badrinarayanan *et al.*, 2017] proposes max-pooling indices, where the decoder uses indices computed by the encoder to upsample low-resolution feature maps, achieving precise pixel-level recovery. Also, it effectively mitigates information loss.

However, the issues of fluctuations in the similarity of intensity distribution between the foreground and surroundings, as well as the inconsistencies in multi-scale spatial detail transmission, remain unresolved, making flexible semantic understanding and segmentation of BUSI challenging.

3 Method

3.1 Semantic Encoding

The similar morphological features between lesions and surrounding tissues pose challenges for sampling (semantic encoding). EfficientNet [Tan and Le, 2019] adopts a compound scaling strategy to adjust depth, width, and resolution, enabling efficient multi-scale feature extraction and noise suppression. EfficientNet-B2, using Mobile Inverted Bottleneck

Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

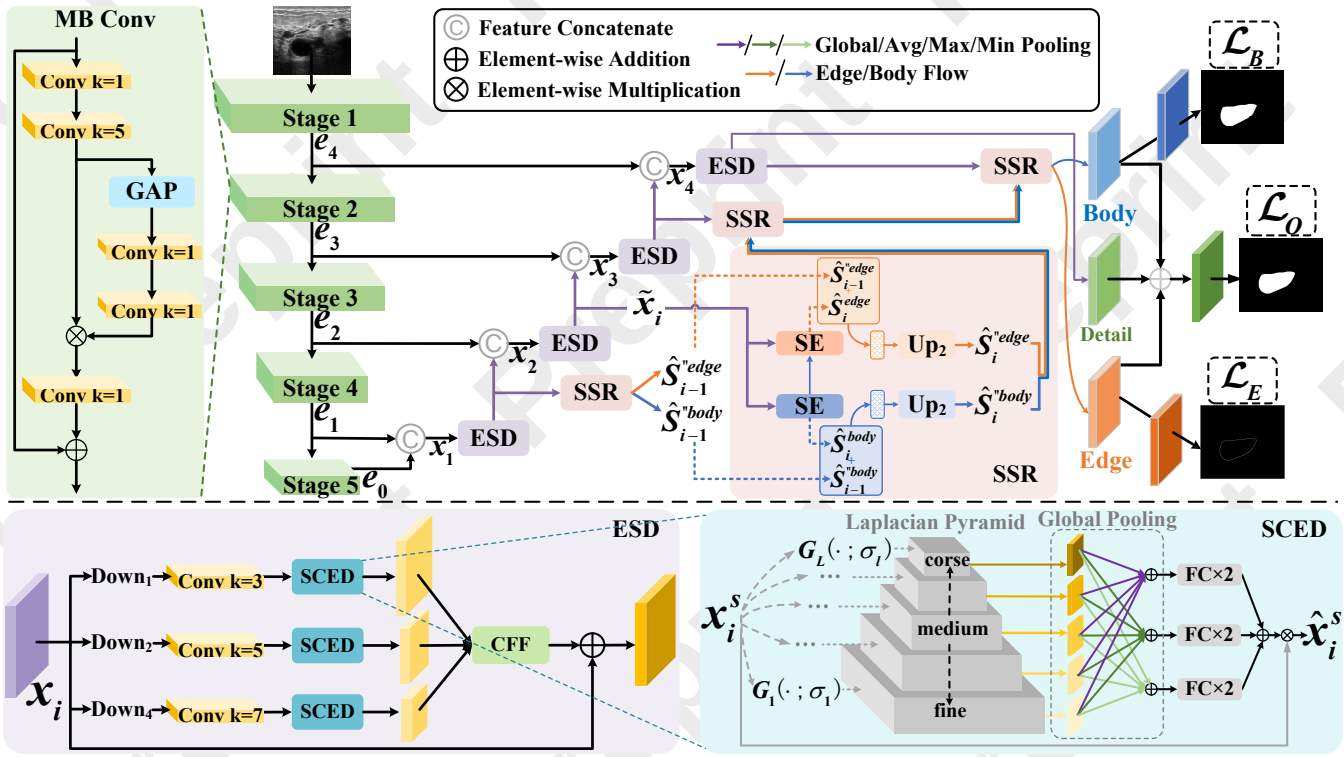

Figure 3: The overall architecture of ElaD-Net.

Stege i	Conv	Channels	Size
Stege 1	Stem(7×7 Conv, Stride=2), Block1	16	1/2
Stege 2	Block 2	24	1/4
Stege 3	Block 3	48	1/8
Stege 4	Block 4	88	1/16
Stege 5	Block 5, Block 6, Block 7	1408	1/32

Table 1: Detailed parmeters setting for each stage of encoder.

Convolution (MBConv) as its core building block, is chosen for our encoder design, as shown in the upper-left of Figure 3.

To address the cold start problem on BUSI and capture general image features, we pre-trained EfficientNet-B2 on ImageNet [Russakovsky *et al.*, 2015]. The encoder details at each downsampling stage are shown in Table 1. Outputs e_i from each stage are passed to the corresponding decoder via skip connections and mixed with decoupled semantics from the previous stage to preserve spatial details, as follows:

$$x_i = [\mathrm{Up}_2(e_0) : e_1] \text{ if } i = 1,\ [\mathrm{Up}_2(\tilde{x}_{i-1}) : e_i] \text{ if } i > 1, \quad (1)$$

where $[a : b]$ is concatenation, and $\mathrm{Up}_u(\cdot)$ denotes upsampling by a factor of u through a transpose convolution.

3.2 Elastic Semantic Decoupling (ESD)

The pixel intensity similarity between lesions and surrounding tissues in BUSI varies significantly, known as the FP problem (Section 1). Since fixed decoupling criteria are impractical for such mixed contexts, we expand branching in multi-scale channels and introduce elasticity into the decoupling process. *Inspired by the multi-frequency segmentation strategy under multi-scale attention* from [Nam *et al.*, 2024], we further extend this idea by designing the Elastic Semantic Decoupling module, which introduces adaptive semantic disentanglement and dynamic frequency calibration for robust handling of foreground–background similarity fluctuations.

Multi-Scale Semantic Extraction

Breast lesions vary in scale, requiring dynamic adjustment of the proportion of key information extracted. We employ three convolutional layers with different kernel sizes, termed multi-scale attention [Nam *et al.*, 2024], for semantic extraction:

$$x_i^s = \mathrm{Conv2D}_k(\mathrm{Down}_s(x_i)) \in \mathbb{R}^{C_k \times H/s \times W/s}, \quad (2)$$

where $x_i \in \mathbb{R}^{C \times H \times W}$ denotes the feature map at the i-th decoder layer and $k \in \{3, 5, 7\}$ the kernel size. $Down_s(\cdot)$ indicates downsampling by a factor of $s \in \{1, 2, 4\}$, corresponding to no/2×/4× downsampling.

Self-Calibrating Elastic Decoupling (SCED)

Multiscale extraction allows the model to dynamically adjust its receptive field. If we treat the domain of similarity fluctuations as image depth, the next step is to endow the model with the ability to adjust this receptive depth. Features of different semantics can be emphasized by their corresponding frequencies [Azad *et al.*, 2021; Geirhos *et al.*, 2019b]. We believe that adjusting the focus of the model's decoupled semantics by utilizing frequency at different depths will help achieve the desired outcome.

We use the Laplacian pyramid mechanism [Lai *et al.*, 2017] to map convolutional feature maps to the frequency

Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

domain, dynamically adjusting their weights based on frequency component importance. Low-frequency captures overall structure, while high-frequency captures edges and textures. Specifically, we approximate the Laplacian with the Difference of Gaussian (DoG). Starting from x_i^s, we extract "$L+1$" levels (branches) of Gaussian representations with incrementally increasing variances $\varphi_1, \varphi_2, \ldots, \varphi_L$:

$$\mathrm{G}_l(x_i^s) = \mathrm{G}(\cdot;\varphi_l) * x_i^s, \tag{3}$$

$$\mathrm{G}(j_1, j_2; \varphi) = \frac{1}{\varphi\sqrt{2\pi}} e^{-\frac{{j_1}^2+{j_1}^2}{2\varphi^2}}. \tag{4}$$

Above, $\mathrm{G}(\cdot)$ is the Gaussian filter, with * denoting convolution, and j_1, j_2 as spatial positions in the feature space.

The Laplacian Pyramid LP is constructed by calculating differences between adjacent Gaussian representations:

$$LP_{i,l}^s = \begin{cases} \mathrm{G}_l(x_i^s) - \mathrm{G}_{l+1}(x_i^s), & \text{if } 1 \le l < L \\ \mathrm{G}_l(x_i^s), & \text{if } l = L \end{cases}, \tag{5}$$

where $LP_{i,l}^s$ is the output of the l-th layer LP from the i-th decoder layer and scale s. For each spectral feature map $LP_{i,l}^s$, we apply average pooling to mitigate noise, max pooling to emphasize locally salient features, and min pooling to capture dark areas or negative outliers. This multi-pooling strategy enhances discriminative power for elastic decoupling:

$$Z_{p,l}^{(s,i)} = \begin{cases} \text{GlobalAveragePooling}(LP_{i,l}^s), & \text{if } p = \text{avg} \\ \text{GlobalMaxPooling}(LP_{i,l}^s), & \text{if } p = \max \\ -\text{GlobalMaxPooling}(-LP_{i,l}^s), & \text{if } p = \min \end{cases}, \tag{6}$$

where $Z_{p,l}^{(s,i)} \in \mathbb{R}^{C_s \times 1 \times 1}$. We then average these pooling results across all pyramid levels to yield combined features:

$$Z_p^{(s,i)} = \frac{1}{L} \sum_{l=1}^{L} Z_{p,l}^{(s,i)}. \tag{7}$$

To dynamically adjust the importance of frequency components, we compute an attention matrix from the pooling results to calibrate the original semantics. Specifically, the pooled features are first passed through two fully connected layers for self-calibration: FC1 reduces the dimension to C/r with reduction rate r, while FC2 restores it to C. Here, δ denotes the ReLU activation function. The attention matrix M_i^s is then generated by adding and activating with the Sigmoid function $\sigma()$, enabling debiasing in the fluctuation (depth)-based elastic semantic decoupling, as follows:

$$\hat{x}_i^s = x_i^s \otimes M_i^s \in \mathbb{R}^{C_s \times H \times W}, \tag{8}$$

$$M_i^s = \sigma\left(\textstyle\sum_{p \in \{\text{avg,max,min}\}} \text{FC2}(\delta(\text{FC1}(Z_p^{(s,i)})))\right) \in \mathbb{R}^{C_s \times 1 \times 1}. \tag{9}$$

Correlation Feature Fusion (CFF)

To aggregate the correlated foreground details from multi-scale semantics while distinguishing the surroundings, we developed an improved multi-scale feature fusion mechanism inspired by [Nam *et al.*, 2024], generating attention maps for both foreground and surroundings. We established two independent attention leaves for $\hat{x}_i^s$:

$$F_i^s = \sigma(\text{Conv2D}_1(\hat{x}_i^s)) \otimes \hat{x}_i^s \in \mathbb{R}^{C_s \times H \times W}, \tag{10}$$

$$B_i^s = (1 - F_i^s) \otimes \hat{x}_i^s \in \mathbb{R}^{C_s \times H \times W}. \tag{11}$$

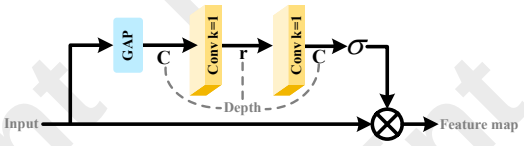

Figure 4: Squeeze-and-Excitation (SE) layer.

Squeeze-and-Excitation (SE) layer, illustrated in Figure 4, is designed here to suppress irrelevant features and highlight important ones within each leaf. SE utilizes Global Average Pooling (GAP) to convert the input into a vector, with dimension C representing the leaf's size. Two convolution layers are then applied: one reduces the size from C to r, and the other restores it to C. Afterward, the Sigmoid function generates quantized values, which are multiplied element-wise with original input to suppress or enhance features. The process is defined as follows:

$$\hat{x}_{\text{i,f}}^s = \text{SE}(F_i^s) = \sigma(Cov2D_1^2(GAP(F_i^s))) \otimes F_i^s \in \mathbb{R}^{C \times H \times W}, \tag{12}$$

$$\hat{x}_{\text{i,b}}^s = \text{SE}(B_i^s) = \sigma(Cov2D_1^2(GAP(B_i^s))) \otimes B_i^s \in \mathbb{R}^{C \times H \times W}. \tag{13}$$

The feature maps are integrated and upsampled to maintain dimensional consistency across different scale channels.

$$\bar{x}_i^s = \text{Conv2D}_3(\hat{x}_{\text{i,f}}^s + \hat{x}_{\text{i,b}}^s) \in \mathbb{R}^{C \times H \times W}, \tag{14}$$

$$\tilde{x}_i^s = \text{Up}_s(\bar{x}_i^s) \in \mathbb{R}^{C \times H \times W}, \tag{15}$$

Finally, we fuse the spatially refined feature maps from all scales with learnable weights W_s:

$$\tilde{x}_i = \sum_{s=1}^{3} W_s \tilde{x}_i^s. \tag{16}$$

3.3 Spatial Semantic Reconfiguration (SSR)

To address spatial detail inconsistency (Section 1), we propose to enhance the transmission process by decomposing it into three key steps: decomposition, optimization, and reconstruction, thereby recovering the lost spatial details.

Decomposition. Thanks to ESD's semantic decoupling, we can easily separate the body component s_i^{body} and the edge component s_i^{edge} from the stage feature map$\tilde{x}_i$, as follows:

$$s_i^{\text{body}} = \text{AvgPool}(\tilde{x}_i), \quad s_i^{\text{edge}} = \tilde{x}_i - s_i^{\text{body}}, \tag{17}$$

where AvgPool is a 5×5 average pooling with $stride = 1, padding = 2$. This facilitates the independent transmission of the two components to reduce spatial detail loss.

Optimization. To mitigate body component attenuation during the cross-scale transmission, we introduce a transmission optimization mechanism to enhance spatial semantics:

$$\hat{s}_i^{\text{body}} = \text{SE}(s_i^{\text{body}}), \tag{18}$$

$$\hat{s}_i'^{\text{body}} = \text{Up}_2(\hat{s}_{i-1}'^{\text{body}} + \hat{s}_i^{\text{body}}). \tag{19}$$

Here, SE is employed to amplify key features while suppressing irrelevant ones, and integrate legacy information from the previous level to restore finer details, ultimately achieving a

Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

finely tuned body component $\hat{s}_i^{\prime\prime\text{body}}$. The same approach is applied to tune the edge component $\hat{s}_i^{\prime\prime\text{edge}}$:

$$\hat{s}_i^{\text{edge}} = \text{SE}(s_i^{\text{edge}}), \tag{20}$$

$$\hat{s}_i^{\prime\prime\text{edge}} = \text{Up}_2(\hat{s}_{i-1}^{\prime\prime\text{edge}} + \hat{s}_i^{\text{edge}}). \tag{21}$$

Reconstruction. We reintegrate the two components with the final semantic feature map $\tilde{x}_i$ decoupled from SSR, forming a cohesive whole that effectively recovers the spatial details lost during downsampling, as follows:

$$p_{\text{o}} = \text{ConV2D}_1(\hat{s}_i^{\prime\prime\text{edge}} + \hat{s}_i^{\prime\prime\text{body}} + \tilde{x}_i) \tag{22}$$

Both $\hat{s}_i^{\prime\prime\text{edge}}$ and $\hat{s}_i^{\prime\prime\text{body}}$ are passed through a 1x1 convolution to obtain the predicted edge probability map p_{e} and the body probability map p_{b}, respectively.

3.4 Loss Function

To enable more comprehensive segmentation, we compute the loss for three segmentation results p_{e}, p_{b} and p_{o}:

$$\mathcal{L}_{\text{E}} = \mathcal{L}_{\text{Dice}}(p_{\text{e}}, l_{\text{e}}), \tag{23}$$

$$\mathcal{L}_{\text{B}} = \mathcal{L}_{\text{wBCE}}(p_{\text{b}}, l_{\text{b}}) + \mathcal{L}_{\text{wIoU}}(p_{\text{b}}, l_{\text{b}}), \tag{24}$$

$$\mathcal{L}_{\text{O}} = \mathcal{L}_{\text{wBCE}}(p_{\text{o}}, l_{\text{o}}) + \mathcal{L}_{\text{wIoU}}(p_{\text{o}}, l_{o}). \tag{25}$$

Above, l_{e}, l_{b}, and l_{o} are the ground truth. Considering the class imbalance in edge information, the Dice loss [Milletari *et al.*, 2016] $\mathcal{L}_{\text{Dice}}$ is utilized to ensure the accuracy of edge detection. The weighted binary cross-entropy loss $\mathcal{L}_{\text{wBCE}}$ and the weighted intersection over union (IoU) loss $\mathcal{L}_{\text{wIoU}}$ are combined to enhance the model's ability to learn different classes and complex shapes. Finally, we calculate the total loss by assigning weights, i.e., α, β, μ, to the aforementioned components: $\mathcal{L}_{\text{F}} = \alpha\mathcal{L}_{\text{O}} + \beta\mathcal{L}_{\text{E}} + \mu\mathcal{L}_{\text{B}}$.

4 Experiments

4.1 Experimental Setup

Datasets Experiments are conducted on three publicly BUSI datasets (Table 2): the BUSI-E9 [Al-Dhabyani *et al.*, 2020] dataset from Bahia Hospital, Dataset B [Yap *et al.*, 2020] acquired using the Siemens ACUSON Sequoia C512 system, and the STU [Zhuang *et al.*, 2019] dataset from the Imaging Department of Shantou University Medical Center.

Evaluation Metrics We adopt five common segmentation metrics to evaluate performance: Jaccard Index (IoU), Precision, Recall, Specificity, and Dice Coefficient (DC).

Experimental Details We employ a four-fold cross-validation approach in experiments. We use the Adamax optimizer with an initial learning rate of 0.002 to train our model. The hyperparameters for loss calculation, α, β, and μ, are set to 0.4, 0.3, and 0.3, respectively. The epoch is set to 300, and the batch size is 6. The experimental environment includes PyTorch 1.13, Python 3.8, and an NVIDIA RTX 4090 GPU.

4.2 Ablation Study

Parameter Ablation. To evaluate the impact of different loss weight configurations on segmentation performance, we designed a series of ablation experiments testing various $\alpha : \beta : \mu$. Experimental results shown in Table 3 indicate that when the weights are set to $0.4 : 0.3 : 0.3$, the segmentation accuracy reaches the optimal state on BUSI-E9 and Dataset B because the edge loss and body loss are properly balanced.

Dataset	Benign	Malignant	Normal	Total	External-validation
BUSI-E9	437	210	133	780	False
Dataset B	110	53	No	163	True
STU	Unknow	Unknow	No	42	True

Table 2: Sample distribution of the three public BUSI datasets.

Dataset	α, β, μ	IoU	Pre	Rec	Spec	DC
BUSI-E9	0.2, 0.4, 0.4	72.91	80.62	85.10	97.66	81.18
	0.4, 0.3, 0.3	**73.86**	**82.58**	84.56	**97.78**	**81.96**
	0.6, 0.2, 0.2	73.59	82.07	**85.70**	97.66	81.87
Dataset B	0.2, 0.4, 0.4	75.81	83.23	89.06	99.29	84.48
	0.4, 0.3, 0.3	**77.25**	83.85	89.56	**99.36**	**85.38**
	0.6, 0.2, 0.2	77.16	**84.07**	**89.66**	99.10	85.17

Table 3: Ablation study of different Loss.

Method	IoU	Pre	Rec	Spec	DC
U-Net	60.70	71.88	76.30	96.18	70.10
Efficient UNet	71.11	81.02	82.69	98.78	80.31
Efficient Unet+ESD	72.77	81.37	84.56	**98.88**	81.31
Efficient U-Net+ESD+SSR	**73.86**	**82.58**	**84.56**	97.78	**81.96**

Table 4: Ablation study of network components on BUSI-E9.

Evaluation of Module Effectiveness. 1) To systematically verify the impact of different modules on model performance, we conducted ablation studies (Table 4) on the BUSI-E9 dataset using four-fold cross-validation. Replacing traditional U-Net with EfficientNet-B2 led to significant improvements across all metrics, with the Dice coefficient rising from 70.1 to 80.31. As ESD and SSR were progressively introduced, segmentation performance further improved, owing to FP problem resistance from multi-scale feature extraction, self-calibrated elastic semantic decoupling, and spatial semantic consistency from SSR. 2) To assess the impact of edge and body reconstruction within SSR, we test three variants: w/o edge (no edge branch), w/o body (no body branch), and w/ body-edge (both branches). Results in Table 5 show that the w/ body-edge variant yields the best segmentation, with the Jaccard index rising from 72.87 (w/o edge) and 73.49 (w/o body) to 73.86. Adding either branch improves performance, though not as much as using both.

Variants	IoU	Pre	Rec	Spec	DC
w/o edge	72.87	81.22	84.59	97.71	80.96
w/o body	73.49	81.44	**85.92**	97.58	81.72
w/ body-edge	**73.86**	**82.58**	84.56	**97.78**	**81.96**

Table 5: Ablation study of body-edge on BUSI-E9.

Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

Method	BUSI-E9					Dataset B				
	IoU	Pre	Rec	Spec	DC	IoU	Pre	Rec	Spec	DC
U-Net	60.70±2.36	71.88±2.41	76.30±2.48	96.18±0.55	70.10±2.20	58.44±4.26	70.27±6.11	75.32±2.85	98.44±0.40	68.20±4.23
Att U-Net	57.09±1.22	78.78±4.67	66.97±4.08	96.87±0.83	67.99±1.18	59.93±4.53	70.40±6.05	76.15±4.21	98.43±0.33	69.30±4.07
U-Net++	61.38±1.73	79.68±3.07	71.44±2.77	97.04±0.54	71.58±2.09	61.19±5.86	68.32±5.73	79.64±3.84	98.44±0.41	69.77±5.30
SegNet	67.31±1.87	76.09±2.00	79.85±1.03	96.99±0.53	75.64±1.80	62.83±2.20	71.72±1.70	80.15±3.90	98.59±0.30	72.16±1.52
BASNet	69.49±2.30*	78.25±3.07	82.28±1.72*	97.21±0.62	77.75±2.51*	68.27±4.09	77.05±4.70	82.83±4.27*	98.81±0.44	77.17±3.22
AAU-Net	68.82±0.44	79.61±1.07	81.10±0.52	97.57±0.24	77.51±0.68	69.10±2.98*	78.83±2.40*	82.22±3.84	98.82±0.35	78.14±2.41*
ESK-net	70.20±2.28	79.57±1.65	82.41±2.84	97.47±0.35	78.71±2.37	71.65±2.39	81.01±3.91	82.66±1.40	99.01±0.35	79.92±2.21
NU-Net	70.35±1.54	79.56±1.17	82.46±1.02	97.48±0.49	78.62±1.38	72.03±0.82	81.49±0.44	84.13±1.73	98.96±0.17	80.80±0.57
Rolling-Unet	67.31±2.44	81.77±2.01	77.20±1.36	**98.75±0.44**	76.65±1.42	64.81±3.31	79.43±2.92	74.47±2.45	**99.39±0.26**	73.84±2.77
Ours	**73.86±2.17**	**82.58±1.54**	**84.56±2.02**	97.78±0.62	**81.96±2.01**	**77.25±2.98**	**83.85±4.29**	**89.56±0.87**	99.36±0.35	**85.38±2.59**

Table 6: Segmentation results (Mean ± Std) of different methods on BUSI-E9 and Dataset B.

Method	Benign Lesions					Malignant Lesions				
	IoU	Pre	Rec	Spec	DC	IoU	Pre	Rec	Spec	DC
U-Net	61.53±3.98	74.97±2.80	73.97±5.81	97.72±0.59	70.49±3.23	51.11±2.62	64.96±2.55	68.86±4.27	93.63±1.28	63.47±2.38
Att U-Net	65.03±2.05	75.24±1.68	79.44±2.84	97.68±0.62	73.30±2.00	51.12±2.35	61.62±0.97	72.57±2.17	93.12±1.00	62.95±2.14
U-Net++	68.25±2.75	75.93±3.66	81.58±1.09	97.74±0.62	75.56±2.79	54.03±3.03	65.50±2.94	73.43±2.10	93.73±1.31	65.52±2.75
SegNet	67.89±3.31	76.96±3.11	79.57±2.21	97.98±0.46	75.47±2.91	54.89±1.78	63.79±2.65	77.25±4.02*	94.00±1.14	65.90±1.97
BASNet	70.55±2.38	78.00±2.83	82.25±1.17	98.13±0.54	77.78±2.29	58.94±1.71	68.68±2.15	76.30±3.36	94.79±0.90	69.27±1.51
AAU-Net	73.33±2.09*	82.70±2.90*	83.14±0.87*	98.39±0.47*	80.88±2.06*	60.60±1.70*	72.62±3.13*	76.13±5.66	95.11±1.27*	71.54±1.74*
ESK-net	72.73±2.12	81.50±2.62	82.69±0.40	98.29±0.40	80.17±1.79	59.63±1.57	71.52±3.40	74.71±3.21	95.24±1.31	70.43±1.32
NU-Net	74.34±2.83	82.91±2.42	85.56±3.59	98.43±0.40	81.43±2.85	61.37±0.96	72.88±1.90	77.41±2.99	95.15±1.14	72.15±0.70
Rolling-Unet	67.86±2.67	81.24±2.33	80.26±1.75	98.44±0.47	77.86±2.31	56.80±2.17	73.96±2.42	71.59±1.96	**95.86±1.43**	68.63±1.62
Ours	**77.22±2.41**	**84.89±2.56**	**87.38±1.93**	**98.53±0.55**	**84.56±2.20**	**66.94±1.18**	**75.26±3.85**	**80.24±4.20**	94.87±1.72	**75.07±1.05**

Table 7: Comparison of segmentation results for benign and malignant lesions on the BUSI dataset.

4.3 Segmentation Results

To evaluate the effectiveness of ElaD-Net, we select several state-of-the-art methods as baselines, including U-Net [Ronneberger *et al.*, 2015], Att U-Net [Oktay *et al.*, 2018], U-Net++ [Zhou *et al.*, 2018], SegNet [Badrinarayanan *et al.*, 2017], BasNet [Qin *et al.*, 2021], AAU-Net [Chen *et al.*, 2023a], NU-Net [Chen *et al.*, 2023b], ESKnet [Chen *et al.*, 2024a] and Rolling-Unet [Liu *et al.*, 2024b].

Comparison of Segmentation Results. The experimental results are summarized in Table 6. **Bold** and underline represent the **best** and second-best performance results, respectively. Asterisks ($* : p < 0.05$) indicates a significant difference via a paired t-test. U-Net and Att U-Net perform the worst, since they fall short in precisely capturing the lesion features and distinguishing interference in BUSI. U-Net++ leverages deep supervision by utilizing intermediate layer outputs for auxiliary loss calculations, yet it struggles to maintain edge information. SegNet mitigates information loss from downsampling via max-pooling indices but still lacks precision in segmenting fine structures in BUSI. Other methods (e.g., BasNet, AAU-Net, NU-Net, and ESKnet) have improved segmentation performance but often rely on specific contextual information or local feature extraction, limiting their adaptability for variably shaped lesions. In contrast, ElaD-Net exhibits significant advantages across multiple aspects. Its IoU value improves by 4.9% on BUSI and 7.2% on Dataset B compared to the second-best method, NU-Net. it achieves elastic semantic decoupling through the ESD module, allowing it to more accurately capture lesion features and differentiate from surrounding environmental interference. Additionally, the SSR model optimizes the transmission of body and edge information, addressing the issue of spatial detail loss caused by downsampling.

Qualitative Comparison. Figure 5 illustrates a visual comparison of different segmentation methods, featuring examples with varying lesion scales and foreground-surrounds intensity distribution similarities. It is clear that our method (last column) produces segmentation results that are closer to the ground truth (second column). Other methods, particularly U-Net, AttU-Net, U-Net++, SegNet, and ESKNet, perform poorly in complex scenarios (rows 2, 3, 5, and 6). This further highlights the robustness of our method.

4.4 Robustness Analysis

Robustness on Benign and Malignant Lesions. Benign lesions have regular shapes and clear boundaries, while malignant lesions are irregular, with fuzzy boundaries and complex intensity distributions, demanding higher segmentation capabilities from the model. As shown in Table 7, our method achieved the best results in both benign and malignant lesion segmentation. The IoU for benign and malignant lesions were 77.22 and 66.94, respectively, demonstrating the robustness of the method in handling the complex morphology of malignant lesions. In contrast, methods like U-Net and Att U-Net struggled with malignant lesions due to their reliance on sim-

Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

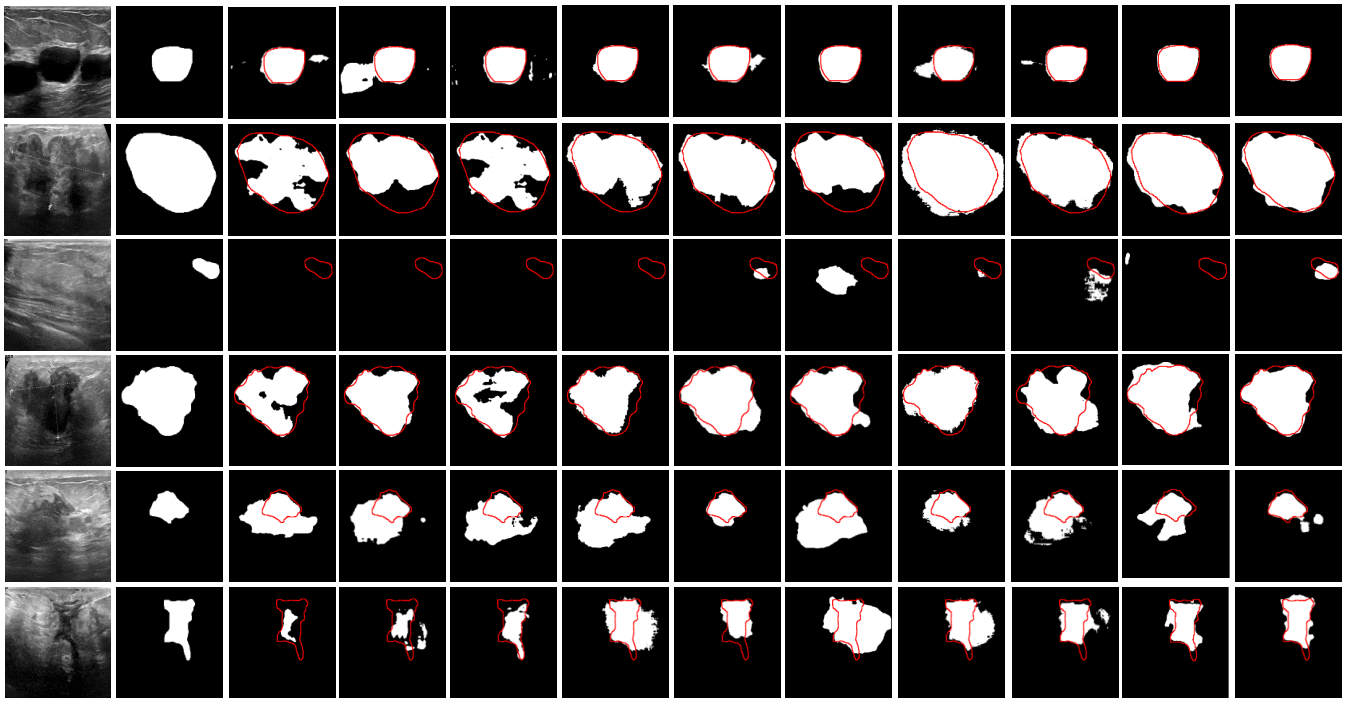

Figure 5: Qualitative comparison of segmentation results on BUSI-E9.

Method	Dataset B on BUSI-E9					STU on Dataset B				
	IoU	Pre	Rec	Spec	DC	IoU	Pre	Rec	Spec	DC
U-Net	42.72±4.65	63.44±5.55	54.48±6.86	98.17±0.70	53.31±4.31	58.90±3.75	66.27±5.46	86.88±1.60	94.54±0.74	71.41±3.67
Att U-Net	38.48±4.19	47.72±5.37	58.38±4.06	97.38±0.71	47.36±5.57	52.65±2.29	59.26±2.86	86.35±1.29	93.41±0.41	65.19±2.73
U-Net++	46.05±3.54	54.14±5.96	61.94±4.52	97.84±0.77	54.89±4.21	59.18±4.21	64.86±5.36	89.67±1.59	94.33±0.83	70.70±4.19
SegNet	42.56±8.30	63.39±10.94	51.59±9.73	98.29±0.87	51.61±9.02	62.70±3.09	66.57±3.05	91.36±0.38	95.04±0.49	73.50±3.62
BASNet	50.86±4.67	69.27±6.62	60.63±3.74	98.66±0.78	59.60±4.87	72.19±3.11*	77.87±3.71*	92.01±1.42	96.30±0.29*	82.12±2.92*
AAU-Net	51.27±5.52	79.71±1.73*	56.13±6.60	98.91±0.61*	61.34±5.65	68.99±3.29	74.91±3.18	92.12±0.75*	95.94±0.71	80.23±2.60
ESK-net	58.36±2.62	**81.21±3.71**	66.60±2.42	98.97±0.71	67.92±3.48	72.63±2.77	79.23±3.39	92.01±0.81	96.57±0.53	82.72±1.70
NU-Net	57.60±2.77	80.50±4.59	64.91±2.51	**99.01±0.61**	67.56±2.88	74.06±0.82	79.87±1.10	**92.38±1.07**	96.80±0.24	84.10±0.76
Rolling-Unet	54.23±1.26	61.29±2.9	75.96±5.6	96.98±0.86	63.33±1.34	67.80±4.27	**87.70±2.76**	77.24±6.74	**98.63±0.05**	79.04±3.98
Ours	**64.79±0.46**	68.71±0.95	**89.34±0.95**	97.64±0.19	**74.15±0.43**	**77.26±1.66**	84.34±2.73	86.72±0.87	97.76±0.37	**84.57±2.03**

Table 8: Comparison of segmentation results for cross-dataset validation.

pler contextual information, resulting in poor segmentation with missed detections and false positives.

Cross-Dataset Validation. To evaluate the model's generalization, we conducted cross-dataset validation (Table 8). Our method outperformed all others, achieving a Dice coefficient of 74.15, Precision of 89.34 on Dataset B (using a model trained on BUSI-E9). On the STU dataset, it achieved a Dice of 84.57 and Specificity of 97.76, demonstrating strong robustness. In contrast, methods like Att U-Net and U-Net++ showed significant performance degradation, highlighting our method's superior generalization.

5 Conclusion

This study presents ElaD-Net, a novel model for BUSI segmentation that tackles complex backgrounds, similarity fluctuations, and spatial detail loss. Built on a pre-trained EfficientNet-B2 encoder, ElaD-Net integrates two key modules: Elastic Semantic Decoupling (ESD) and Spatial Semantic Reconstruction (SSR). ESD enables multi-scale extraction, self-calibration, and feature fusion to improve robustness against FP issues. SSR decouples body and edge features, then optimizes and reconstructs them to preserve spatial consistency. We balance losses across body, edge, and full outputs to enhance accuracy. Extensive experiments on three real BUSI datasets validate each module's effectiveness, showing superior performance and robustness. Future work will focus on accelerating segmentation for clinical use.

Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

Acknowledgments

This work was supported in part by the National Key R&D Program of China under Grant 2023YFB3107303, in part by the National Natural Science Foundation of China (62402253 and 62301289), in part by the project ZR2024MF050, ZR2024QF109 and ZR2022QF041 supported by Shandong Provincial Natural Science Foundation, in part by the Taishan Scholars Program under Grant tsqn202211210, in part by the Pairing Plan Project of the School of Computer Science and Technology of Qilu University of Technology (Shandong Academy of Sciences) under Grant 2024JDJH12, in part by the "20 New Universities" Project of Jinan City under Grant 202333023 and Grant 202333045, and in part by the Young Innovation Team of Colleges and Universities in Shandong Province 2024KJH044.

References

[Al-Dhabyani *et al.*, 2020] Walid Al-Dhabyani, Mohammed Gomaa, Hussien Khaled, and Aly Fahmy. Dataset of breast ultrasound images. *Data in Brief*, 28:104863, 2020.

[Azad *et al.*, 2021] Reza Azad, Afshin Bozorgpour, Maryam Asadi-Aghbolaghi, Dorit Merhof, and Sergio Escalera. Deep frequency re-calibration u-net for medical image segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pages 3267–3276, 2021.

[Badrinarayanan *et al.*, 2017] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12):2481–2495, 2017.

[Cao *et al.*, 2022] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *Proceedings of the European Conference on Computer Vision Workshops*, volume 13803, pages 205–218, 2022.

[Chen *et al.*, 2015] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. In *Proceedings of the International Conference on Learning Representations*, 2015.

[Chen *et al.*, 2018a] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):834–848, 2018.

[Chen *et al.*, 2018b] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European Conference on Computer Vision*, volume 11211, pages 833–851, 2018.

[Chen *et al.*, 2023a] Gongping Chen, Lei Li, Yu Dai, Jianxun Zhang, and Moi Hoon Yap. Aau-net: An adaptive attention u-net for breast lesions segmentation in ultrasound images. *IEEE Transactions on Medical Imaging*, 42(5):1289–1300, 2023.

[Chen *et al.*, 2023b] Gongping Chen, Lei Li, Jianxun Zhang, and Yu Dai. Rethinking the unpretentious u-net for medical ultrasound image segmentation. *Pattern Recognition*, 142:109728, 2023.

[Chen *et al.*, 2024a] Gongping Chen, Lu Zhou, Jianxun Zhang, Xiaotao Yin, Liang Cui, and Yu Dai. Esknet: An enhanced adaptive selection kernel convolution for ultrasound breast tumors segmentation. *Expert Systems with Applications*, 246:123265, 2024.

[Chen *et al.*, 2024b] Linwei Chen, Ying Fu, Lin Gu, Chenggang Yan, Tatsuya Harada, and Gao Huang. Frequency-aware feature fusion for dense image prediction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):10763–10780, 2024.

[Dong *et al.*, 2023] Bo Dong, Pichao Wang, and Fan Wang. Head-free lightweight semantic segmentation with linear transformer. In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence*, pages 516–524, 2023.

[Geirhos *et al.*, 2019a] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A. Wichmann, and Wieland Brendel. Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness. In *Proceedings of the International Conference on Learning Representations*, 2019.

[Geirhos *et al.*, 2019b] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A. Wichmann, and Wieland Brendel. Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness. In *Proceedings of the 7th International Conference on Learning Representations*, 2019.

[Lai *et al.*, 2017] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, pages 5835–5843. IEEE Computer Society, 2017.

[Lee *et al.*, 2020] Haeyun Lee, Jinhyoung Park, and Jae Youn Hwang. Channel attention module with multiscale grid average pooling for breast cancer segmentation in an ultrasound image. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 67(7):1344–1353, 2020.

[Lin *et al.*, 2022] Ailiang Lin, Bingzhi Chen, Jiayu Xu, Zheng Zhang, Guangming Lu, and David Zhang. Ds-transunet: Dual swin transformer u-net for medical image segmentation. *IEEE Transactions on Instrumentation and Measurement*, 71:1–15, 2022.

[Lin *et al.*, 2023] Xian Lin, Li Yu, Kwang-Ting Cheng, and Zengqiang Yan. Batformer: Towards boundary-aware

Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

lightweight transformer for efficient medical image segmentation. *IEEE Journal of Biomedical and Health Informatics*, 27(7):3501–3512, 2023.

[Liu *et al.*, 2024a] Guoqi Liu, Yanan Zhou, Jiajia Wang, Zongyu Chen, Dong Liu, and Baofang Chang. A cross-attention and multilevel feature fusion network for breast lesion segmentation in ultrasound images. *IEEE Transactions on Instrumentation and Measurement*, 73:1–13, 2024.

[Liu *et al.*, 2024b] Yutong Liu, Haijiang Zhu, Mengting Liu, Huaiyuan Yu, Zihan Chen, and Jie Gao. Rolling-unet: Revitalizing mlp's ability to efficiently extract long-distance dependencies for medical image segmentation. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence*, pages 3819–3827, 2024.

[Milletari *et al.*, 2016] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *Proceedings of the Fourth International Conference on 3D Vision*, pages 565–571, 2016.

[Nam *et al.*, 2024] Ju-Hyeon Nam, Nur Suriza Syazwany, Su Jung Kim, and Sang-Chul Lee. Modality-agnostic domain generalizable medical image segmentation by multi-frequency in multi-scale attention. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11480–11491, 2024.

[Ning *et al.*, 2022] Zhenyuan Ning, Shengzhou Zhong, Qianjin Feng, Wufan Chen, and Yu Zhang. Smu-net: Saliency-guided morphology-aware u-net for breast lesion segmentation in ultrasound image. *IEEE Transactions on Medical Imaging*, 41(2):476–490, 2022.

[Oktay *et al.*, 2018] Ozan Oktay, Jo Schlemper, Loïc Le Folgoc, Matthew C. H. Lee, Mattias P. Heinrich, Kazunari Misawa, Kensaku Mori, Steven G. McDonagh, Nils Y. Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention u-net: Learning where to look for the pancreas. *CoRR*, abs/1804.03999, 2018.

[O'Shea and Nash, 2015] Keiron O'Shea and Ryan Nash. An introduction to convolutional neural networks. *CoRR*, abs/1511.08458, 2015.

[Punn and Agarwal, 2022] Narinder Singh Punn and Sonali Agarwal. Rca-iunet: a residual cross-spatial attention-guided inception u-net model for tumor segmentation in breast ultrasound imaging. *Machine Vision and Applications*, 33(2):27, 2022.

[Qin *et al.*, 2020] Xuebin Qin, Zichen Zhang, Chenyang Huang, Masood Dehghan, Osmar R. Zaïane, and Martin Jägersand. U^2-net: Going deeper with nested u-structure for salient object detection. *Pattern Recognition*, 106:107404, 2020.

[Qin *et al.*, 2021] Xuebin Qin, Deng-Ping Fan, Chenyang Huang, Cyril Diagne, Zichen Zhang, Adrià Cabeza Sant'Anna, Albert Suàrez, Martin Jägersand, and Ling Shao. Boundary-aware segmentation network for mobile and web applications. *CoRR*, abs/2101.04704, 2021.

[Rahman *et al.*, 2024] Md Mostafijur Rahman, Mustafa Munir, and Radu Marculescu. EMCAD: efficient multi-scale convolutional attention decoding for medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11769–11779, 2024.

[Ronneberger *et al.*, 2015] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, volume 9351, pages 234–241, 2015.

[Russakovsky *et al.*, 2015] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael S. Bernstein, Alexander C. Berg, and Li Fei-Fei. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015.

[Tan and Le, 2019] Mingxing Tan and Quoc V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pages 6105–6114, 2019.

[Wilcock and Webster, 2021] Paul Wilcock and Rachel M Webster. The breast cancer drug market. *Nature Reviews Drug Discovery*, 20(5):339–340, 2021.

[Yap *et al.*, 2020] Moi Hoon Yap, Manu Goyal, Fatima Osman, Robert Martí, Erika R. E. Denton, Arne Juette, and Reyer Zwiggelaar. Breast ultrasound region of interest detection and lesion localisation. *Artificial Intelligence in Medicine*, 107:101880, 2020.

[Zhou *et al.*, 2018] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Proceedings of the International Workshop on Deep Learning in Medical Image Analysis Deep Learning*, volume 11045, pages 3–11, 2018.

[Zhuang *et al.*, 2019] Zhemin Zhuang, Nan Li, Alex Noel Joseph Raj, Vijayalakshmi G. V. Mahesh, and Shunmin Qiu. An rdau-net model for lesion segmentation in breast ultrasound images. *PLoS ONE*, 14, 2019.