ITC 3/55
Information Technology and Control
Vol. 55 / No. 3/ 2026
pp. 1073-1099
DOI 10.5755/j01.itc.55.3.44982

SDFS-Net: A Breast Ultrasound Segmentation Network combining Static–Dynamic Receptive Field and Frequency-Selective Modeling

Received 2026/05/09 | Accepted after revision 2026/07/20

HOW TO CITE: Zhao, H., Liu, Z. (2026) SDFS-Net: A Breast Ultrasound Segmentation Network combining Static–Dynamic Receptive Field and Frequency-Selective Modeling. *Information Technology and Control*, 55(3), 1071-1097. https://doi.org/10.5755/j01.itc.55.3.44982

SDFS-Net: A Breast Ultrasound Segmentation Network combining Static–Dynamic Receptive Field and Frequency-Selective Modeling

Hong Zhao, Ze Liu*

School of Computer Science and Artificial Intelligence, Lanzhou University of Technology, Lanzhou, 730050, Gansu, China; e-mails: zhaoh@lut.edu.cn (Zhao); 15755269725@163.com (Liu)

Corresponding author: 15755269725@163.com

Breast ultrasound (BUS) plays a critical role in breast cancer diagnosis. However, the large-scale variability of breast tumors poses challenges to receptive field modeling in segmentation methods. Existing approaches are often constrained in their receptive field design, making it challenging to coordinate dynamic and static receptive fields effectively. Moreover, benign and malignant breast tumors exhibit substantial differences in morphological characteristics and structural complexity. By leveraging frequency-domain analysis, this paper finds that these differences are clearly reflected in their spectral distributions. Nevertheless, most methods neglect such spectral heterogeneity. To address these challenges, this paper proposes a novel breast ultrasound segmentation network, termed SDFS-Net. The network adopts the Static-Dynamic Receptive Field (SDRF) block as its visual backbone, which integrates static and dynamic receptive fields through a dual-branch design for complementary and collaborative modeling. Additionally, we design a Frequency-Selective Modeling (FSM) module to selectively model the spectral information of different types of breast tumors. On three breast ultrasound datasets (PACE 2025, BUI, and BUS_UC), SDFS-Net achieves Dice scores of 84.04%, 87.15%, and 92.43%, outperforming the recent state-of-the-art method H-vmunet by 6.91%, 0.29%, and 0.52%, respectively, validating the effectiveness of static-dynamic receptive field collaboration and frequency-selective modeling.

KEYWORDS: Static and dynamic receptive fields, frequency-domain modeling, breast tumors, semantic segmentation, medical image analysis

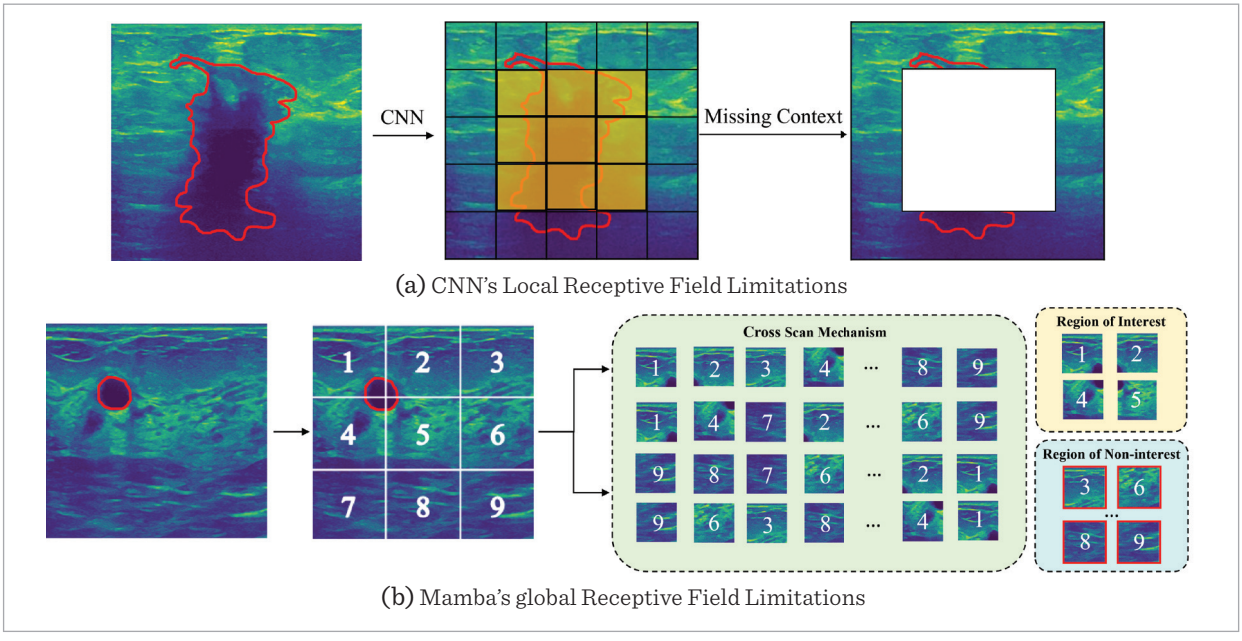
1. Introduction

Breast cancer remains a significant challenge in global public health. According to the latest statistical analysis of breast cancer incidence across 185 countries [23], breast cancer has become one of the leading causes of cancer-related mortality among women, and its high incidence and mortality rates continue to pose a serious threat to women's health. In this context, population-wide regular screening has been widely recognized as an important strategy for breast cancer prevention and control. Currently, clinical breast screening and diagnosis mainly rely on medical imaging techniques [3, 24, 42]. Among them, ultrasound (US) [44] is widely used owing to its safety, ease of access, and low cost. Despite the importance of regular screening, the diagnostic process heavily depends on radiologists' subjective experience, which not only increases labour and time costs but also increases the risk of misdiagnosis, particularly under high clinical workloads.

To this end, developing automated, high-accuracy medical image segmentation methods has become essential for improving the efficiency of breast cancer diagnosis. In recent years, driven by advances in deep learning, convolutional neural network (CNN)-based methods [14, 32, 55] have been extensively applied in medical image segmentation. Among these, U-Net [40] has become a foundational framework due to its classic encoder-decoder architecture, inspiring a series of improved U-shaped network variants [5, 9, 16, 61]. Nevertheless, CNN-based methods face challenges in modeling semantic relationships across regions. Compared with CNNs, vision transformer-based methods [10, 13, 60] employ self-attention mechanisms to capture global image context. However, their inherent quadratic computational complexity results in substantial overhead in dense prediction scenarios [39]. Recently, state space models (SSMs) [11, 12, 25] have attracted increasing attention. Mamba introduces an input-dependent selective mechanism and captures long-range dependencies while maintaining linear computational complexity. To adapt to vision

Figure 1

Limitations of static receptive fields in CNN- and Mamba-based methods under different tumor scales. The red contours denote the ground-truth boundaries of breast tumors. For better visualization, the original ultrasound images are shown in pseudo-color. (a) For large tumors, CNNs are limited by local receptive fields. They fail to model long-range dependencies and lose contextual information. (b) Mamba introduces a global receptive field via a cross-scan mechanism. When dealing with small-scale tumors, its global modeling strategy tends to incorporate substantial redundant information from non-interest regions, thereby interfering with accurate foreground modeling.



tasks, VMamba [28] proposes a cross-scan mechanism for efficiently modeling global image context. Subsequently, improved scanning strategies [17, 45, 56] have been proposed, further facilitating the application of Mamba-based architectures in medical image segmentation tasks [35, 41, 54, 64].

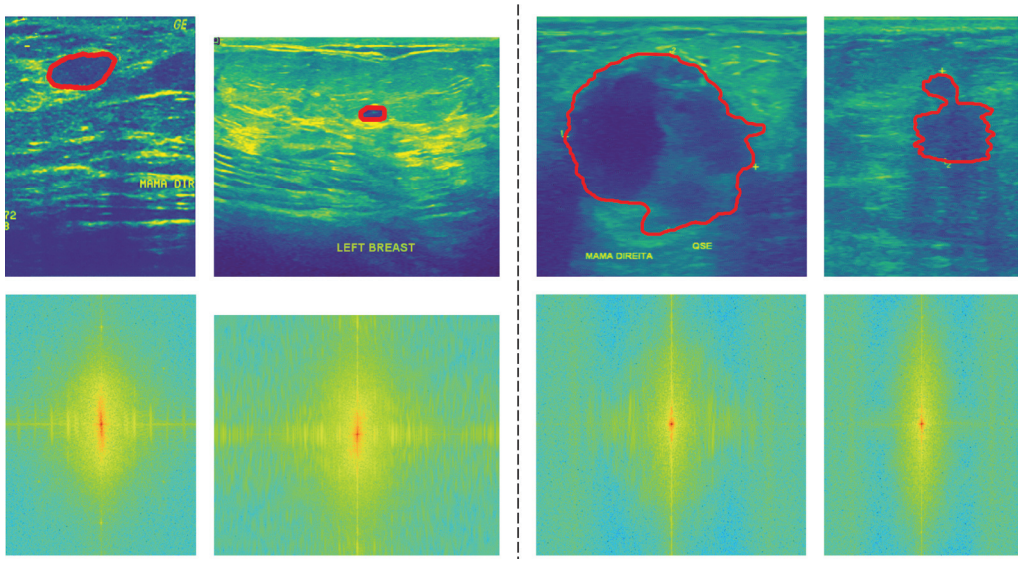
It is worth noting that, in breast ultrasound segmentation tasks, breast tumors exhibit pronounced inter-subject variability in scale. However, existing methods predominantly rely on static receptive field strategies, where the receptive field range is predefined by convolution kernels or network architectures and remains unchanged across different samples, limiting their ability to adaptively model tumor scale variations. As illustrated in Figure 1(a), local static receptive field–based methods, represented by CNNs, primarily focus on local regions. Limited by their receptive fields, these methods struggle to perceive relevant information from distant spatial locations, leading to insufficient contextual understanding of the overall lesion structure and the surrounding background. Furthermore, as shown in Figure 1(b), global static receptive field–based methods, exemplified by VMamba [28], are more susceptible to interference from redundant information and background noise during modeling, thereby weakening their ability to focus on key foreground regions effectively.

To account for the scale variability of breast tumors, several studies [2, 48, 57] have attempted to incorporate dynamic receptive field mechanisms to enhance model adaptability, enabling the receptive field range to be adaptively adjusted according to the content of the input features. However, their modeling strategies often rely excessively on the outcomes of dynamic algorithms, failing to fully exploit the foundational and dominant role of static receptive fields in providing stable feature representations. Overall, existing methods still suffer from limitations in receptive field design and struggle to effectively coordinate the complementary strengths of dynamic and static receptive fields. Consequently, accurately and efficiently segmenting breast tumors with substantial scale variations remains challenging.

From a pathological perspective, benign and malignant breast tumors exhibit substantial differences in growth patterns and tissue structural complexity.

Figure 2

Visualization of benign and malignant breast tumors in the spatial and frequency domains. The first row shows ultrasound images, and the second row presents the corresponding frequency spectra. The red contours indicate the ground-truth boundaries of breast tumors. (a) For benign tumors, spectral energy is mainly concentrated in the central frequency region, resulting in a significantly high-intensity (red) low-frequency response. (b) Malignant tumors exhibit a weakened central low-frequency dominance, with spectral energy spreading toward the mid- to high-frequency regions, resulting in enhanced blue-green responses at the spectral edges.



(a) Benign tumor (b) Malignant tumor

These spatial-domain variations in morphology and structure are reflected in their spectral distribution characteristics. As illustrated in Figure 2(a), benign tumors typically exhibit relatively regular boundaries and homogeneous internal structures, resulting in spectral energy predominantly concentrated in low-frequency components. In contrast, Figure 2(b) shows that malignant tumors often exhibit irregular growth patterns, spiculated margins, and complex internal texture variations, leading to more pronounced high-frequency responses in the frequency domain. These observations indicate that different breast tumor types exhibit distinct representations in the frequency domain, offering new perspectives for frequency-aware modeling in breast tumor segmentation. However, existing frequency-domain methods [4, 31, 51, 52] pay limited attention to such differences. Most of them treat spectral information as uniformly distributed and adopt a single-frequency modeling strategy, thereby failing to capture the distinct spectral distributions of benign and malignant tumors. This limitation ultimately constrains further improvements in segmentation accuracy.

To address these challenges, we propose SDFS-Net for breast ultrasound segmentation. Specifically, we design a Static-Dynamic Receptive Field (SDRF) block that integrates static and dynamic receptive fields, enabling complementary and collaborative modeling between different receptive field mechanisms. In addition, we propose a Frequency-Selective Modeling (FSM) module to selectively extract spectral information from lesion regions, thereby enhancing the representation of discriminative frequency components across different types of breast tumors.

The main contributions of this paper are summarised below:

- To address the scale variation in breast tumors and the morphological differences between benign and malignant lesions, we propose a BUS segmentation network, SDFS-Net, with the SDRF block as its visual backbone. This block integrates complementary static and dynamic receptive field algorithms, enabling targeted modeling of lesions with varying scales.
- We design the FSM module to selectively model lesion-related frequency-band information based on spectral differences between benign and malignant tumors, thereby enhancing the representation of frequency-domain discriminative features and improving segmentation accuracy for tumors with significant morphological and structural differences.
- The introduction of FRLoss guides the model to focus on subtle feature differences, refines the delineation of complex boundary regions, and reduces false-positive predictions.
- Extensive experiments on three public breast ultrasound datasets demonstrate that SDFS-Net consistently outperforms state-of-the-art methods, achieving superior segmentation performance.

2. Related Works

2.1. Medical Image Segmentation Algorithms with Dynamic and Static Receptive Fields

In recent years, the field of medical image segmentation has made significant progress in receptive field modeling. Early CNN-based methods, such as FCN [29] and UNet [40], mainly rely on static receptive fields, extracting local features through convolution kernels or pooling operations with fixed sizes. Building upon this paradigm, subsequent studies have introduced various improvements to static receptive field frameworks. C-Net [7] employs cascaded bidirectional attention and residual refinement to suppress interference from visually similar background regions effectively. AMS-PAN [33] combines attention mechanisms with multi-scale pyramid structures and demonstrates exemplary performance in breast ultrasound (BUS) image segmentation.

In addition, segmentation approaches built on Transformer and Mamba architectures have achieved competitive performance by modeling global, static receptive fields [30, 34]. TransUNet [8] leverages Transformers as a powerful encoder to capture long-range dependencies and combines a U-Net-style decoder to recover local details. State-space model (SSM)-based segmentation methods, such as VM-UNet [41], and H-vmunet [49], efficiently construct global context through selective scanning mechanisms.

However, as shown in Table 1, the above methods based on static receptive fields suffer from inherent limitations when handling breast tumors with large-

Table 1
Comparison of representative medical image segmentation methods in terms of static and dynamic receptive fields.

Static or Dynamic	Methods	Year	Limitation
Static	UNet [40]	2015	The receptive field is fixed. It cannot balance the global contour of large tumors and the delicate structures of small tumors.
	TransUNet [8]	2021	The static global receptive field lacks spatial adaptivity. Global semantics may dominate and weaken the modeling of small lesions and fine details.
	VM-UNet [41]	2024	The receptive field based on global sequence scanning lacks spatial selectivity. It is easily affected by large-scale background interference, which reduces focus on foreground targets.
	H-vmunet [49]	2025	Both global and local receptive fields are constructed using predefined structures, lacking a dynamic receptive field adjustment mechanism for lesion-scale variations.
Dynamic	D-LKA Net [2]	2024	Stacked local operations still form the dynamic receptive field. The effective modeling range is limited.
	BFNet [57]	2025	The flexibility of the dynamic receptive field depends on the accuracy of offset learning. The receptive range is still constrained to local regions.
	HMDA [48]	2025	Offset modeling of sampling points relies on the discriminability of local features. When lesion boundaries are blurred, receptive field adjustment may fail.

scale variations, due to their fixed receptive fields. Against this background, dynamic modeling methods that adaptively adjust receptive fields based on foreground regions have attracted increasing attention. These methods flexibly adjust the shape, size, or feature-aggregation weights of receptive fields, enabling more accurate characterization of lesion structures. For example, BFNet [57] introduces a flexible feature aggregation module and achieves cross-level spatial alignment using deformable convolution. D-LKA Net [2] proposes Deformable Large Kernel Attention, aggregating global context while alleviating the locality limitation of standard convolutions. HMDA [48] presents a multi-scale, spatially adaptive deformable attention mechanism to enable content-driven dynamic receptive field modeling.

Nevertheless, research on dynamic receptive field modeling in existing methods remains relatively limited, and collaborative modeling of dynamic and static receptive fields within a unified framework is even rarer. Some existing approaches introduce dynamic receptive fields via deformable convolutions or adaptive attention mechanisms. However, they often treat dynamic mechanisms as independent modules or embed them into networks in a stage-wise manner, lacking a systematic design that exploits the complementary relationship between dynamic and static receptive fields. Therefore, achieving collaborative modeling of dynamic and static receptive fields within a unified framework is of great importance for improving segmentation performance on lesions with scale variations.

2.2. Application of Frequency-Domain Modeling in Medical Image Segmentation

Existing medical image segmentation methods predominantly employ spatial-domain attention mechanisms to enhance the feature representation of foreground regions. By contrast, frequency-domain modeling provides a way to represent features in the Fourier space. It can effectively highlight key pathological characteristics in noisy medical images, such as tumor texture details and tissue structural variations, leading to a more concentrated distribution of discriminative information in the frequency domain. As a result, frequency-domain modeling has been widely applied in medical image segmentation.

Depending on how frequency-domain information is introduced into the model and the stage at which it is processed, existing frequency-domain-based methods can be broadly categorized into two groups. The first group primarily focuses on spatial-domain feature learning and incorporates frequency-domain transformations to decompose, model, or enhance features, treating frequency information as a complementary component of spatial representations. For example, MFMSNet [51] designs a Multi-frequency Transformer block that simulates the separation and interaction of high- and low-frequency features in the spatial domain. HCMNet [52] proposes a Wavelet Feature Enhancement module that integrates low- and high-frequency features decomposed via wavelet transforms into the spatial domain to alleviate detail loss.

In contrast, the second category of methods explicitly maps features into the frequency domain and directly performs key modeling, selection, or fusion operations in the spectral space. Representative methods include EM-Net [6], MCF-Net [22], and Polyp-Mamba [62]. Specifically, EM-Net [6] proposes an Efficient Frequency-domain Learning (EFL) layer, which employs the Fast Fourier Transform (FFT) to map features into the spectral space and adaptively learns the importance of different frequency components through learnable weights. Similarly, MCF-Net [22] introduces a frequency-domain adaptive learning module, which performs spectral decomposition and dynamic fusion after applying FFT to the input features, thereby integrating local and global information in the frequency domain. Polyp-Mamba [62] designs a Hybrid Multi-frequency Fusion Module (HMFM), which utilizes the Discrete Cosine Transform (DCT) for frequency-domain transformation and subsequently employs multi-frequency channel attention to comprehensively extract broadband information from polyp images, effectively alleviating the issue of blurred polyp boundaries.

Although the aforementioned methods incorporate frequency-domain mechanisms to improve segmentation performance, in breast tumor segmentation, the pronounced differences in morphological structures between benign and malignant tumors are further reflected in distinct spectral distributions. Existing methods do not fully account for the spectral distribution differences caused by tumor properties, thereby limiting the discriminative power of frequency-domain information for delineating lesion boundaries and complex structures.

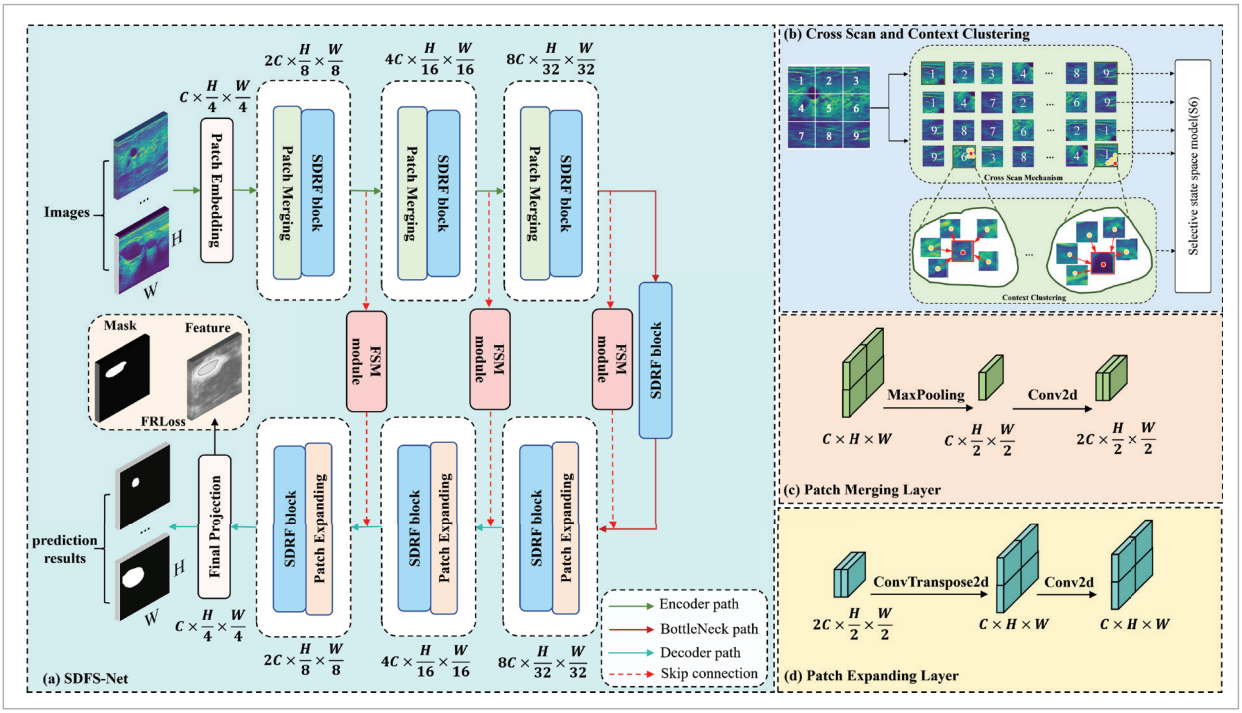
3. Methodology

3.1. Model Overview

SDFS-Net adopts a U-shaped encoder–decoder architecture, as shown in Figure 3(a), comprising a Patch Embedding Layer, an Encoder, a Bottleneck, a Decoder, a Final Projection Layer, and Skip Connections. Both the encoder and the decoder contain three stages. At each stage, feature-scale transformation is performed via a patch-merging layer (Figure 3(c)) or a patch-expanding layer (Figure 3(d)), with the proposed Static–Dynamic Receptive Field (SDRF) block embedded. As illustrated in Figure 3(b), the SDRF block adopts a dual-branch design that integrates static receptive field modeling methods (CNN, Mamba) and a dynamic receptive field algorithm (Context Clustering) within a unified framework. This block is also incorporated into the bottleneck layer. Different from the configurations used in the encoder and decoder, its dynamic receptive field branch employs two consecutive C25 dynamic clustering operations (as shown in Figure 5) to enhance the contextual modeling of deep semantic features. Serving as the fundamental visual processing unit throughout the entire segmentation network, the SDRF block continuously maintains the complementarity and collaboration between static and dynamic receptive fields along the main pathway. In addition, a Frequency-Selective Modeling (FSM) module is embedded in the skip connections, which adaptively extracts key frequency-band features of breast tumors, enhances the transmission of lesion boundary and texture information, and provides more discriminative semantic features for feature reconstruction in the decoder stage. Subsequently, FRLoss is introduced between the feature maps output by the final projection layer and the ground-truth masks, guiding the model to focus on difficult regions, such as complex boundaries, fine structures, and regions with highly similar appearances at the feature level, thereby improving segmentation accuracy. Finally, the decoder outputs are mapped to the corresponding category space through the final projection layer to obtain the segmentation results.

Figure 3

The overall architecture of SDFS-Net. (a) The encoder–decoder framework of SDFS-Net. It consists of the core SDRF block, the FSM module, and FRLoss. (b) In the dynamic context modeling branch of the SDRF block, the outputs of cross-scanning and the context clustering (CC) algorithm are fed into the selective state space model (S6). (c) The patch merging layer, which performs progressive feature downsampling in the encoder. (d) The patch expanding layer, which upsamples features and restores spatial resolution in the decoder.



3.2. Static-Dynamic Receptive Field (SDRF) Block

To address the limitations of existing methods in collaborative modeling of dynamic and static receptive fields for breast tumor segmentation with large-scale variations, we propose the SDRF block and adopt it as the core visual backbone of SDFS-Net. As shown in Figure 4(a), this block comprises two branches: a dynamic context modeling branch and a local static branch, aiming to achieve complementary and collaborative modeling of dynamic and static receptive fields.

Specifically, the input feature $X \in R^{H\times W\times C}$ is first normalized by Layer Normalization (LN), and then evenly split into $X_1 \in R^{H\times W\times C/2}$ and $X_2 \in R^{H\times W\times C/2}$ along the channel dimension, which are fed into the two branches for parallel modeling, respectively.

Algorithm 1 Context Clustering Process

Input:

Feature map F; Number of cluster centers t; Neighborhood size for initialization k; Learnable parameters α, β.

Output:

Updated feature map F'.

Step 1: Feature Projection & Initialization

1. P = Reshape(F); P_s = Linear(P); // where Linear denotes a linear layer with bias, mapping C to C/8 dimensions.

2. $\{c_j\}_{j=1}^{t}$ = InitCenters(P_s,t,k); // where InitCenters denotes uniformly sampling t initial centers and initializing each center by averaging its k-nearest neighboring features.

Step 2: Similarity Calculation & Clustering

3. S = CosineSimilarity(P_s,{c_j}); $Assign_i$=$argmax_j S_{i,j}$; // Assign each point to the most similar center.

Step 3: Cluster-wise Feature Aggregation & Update

4. **for** each cluster m **do**

P_m, s = Gather(P_s, S, m) // where Gather denotes partitioning the input feature map into non-overlapping spatial blocks for local clustering.

P'_m = Linear(P_m); v_c = GetCenterFeature(P'_m) // where GetCenterFeature denotes extracting cluster center representations by applying adaptive average pooling.

w_i = Getweight(α, β,s); T = Sum(w_i,1); // where Getweight denotes computing the i-th pixel weight based on the learnable parameters α and β and the similarity scores s.

g = GetAggregationFeatures(T, w_i, p_i, s, v_c); // where GetAggregationFeatures denotes performing similarity-weighted aggregation of pixel features within each cluster.

p'_i = Sum(FC(w_i * g), p_i); // Update features; where FC denotes a fully connected layer with learnable parameters.

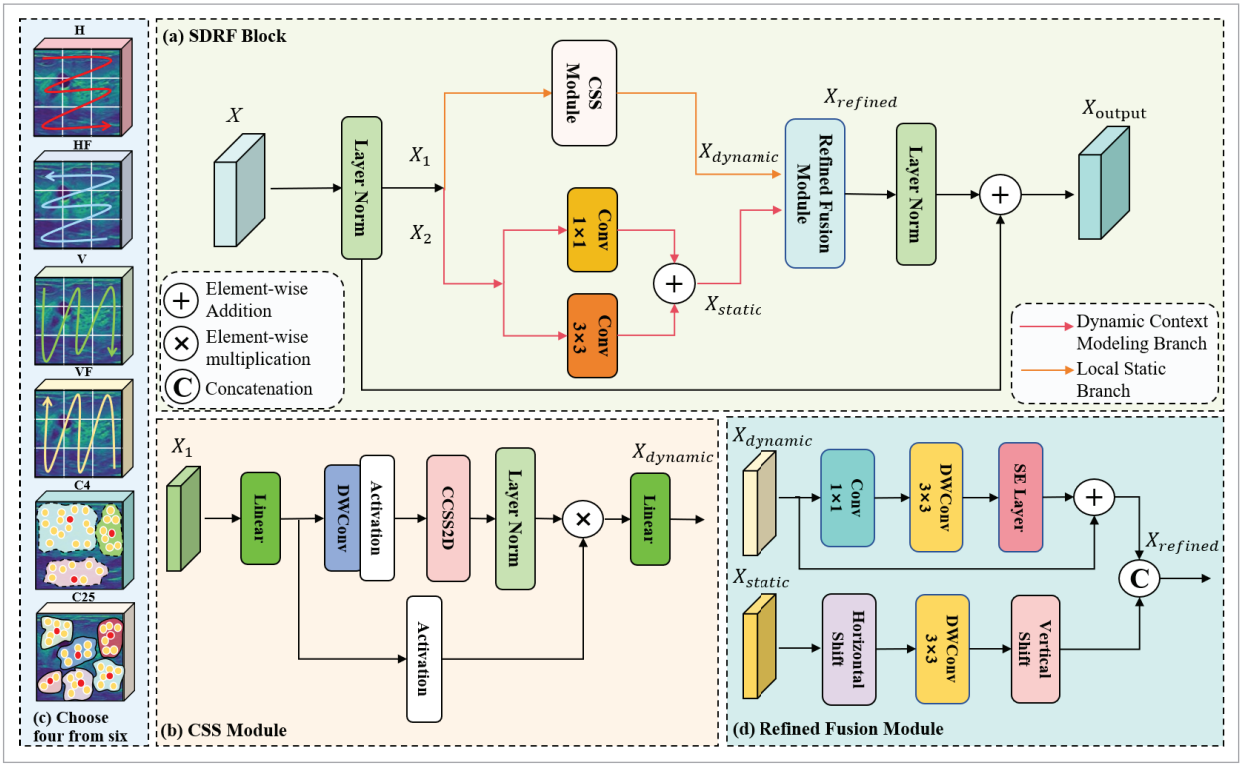
end for

Step 4: Reassemble Output

5. F'=Scatter({p'_i}); // where Scatter denotes reconstructing the full feature map.

Figure 4

Internal details of the Static-Dynamic Receptive Field (SDRF) block. (a) The SDRF block, which serves as the core visual backbone of SDFS-Net. (b) The Context State Space (CSS) module is embedded in the dynamic context modeling branch. (c) In the Context Clustering 2D Selective Scan (CCSS2D), the CSS module combines four cross-scan modes (H, HF, V, and VF) with two context clustering strategies (C4 and C25), forming six candidate modeling strategies. Based on the network stage, four optimal combinations are adaptively selected to achieve efficient dynamic context modeling. (d) The refined fusion module.

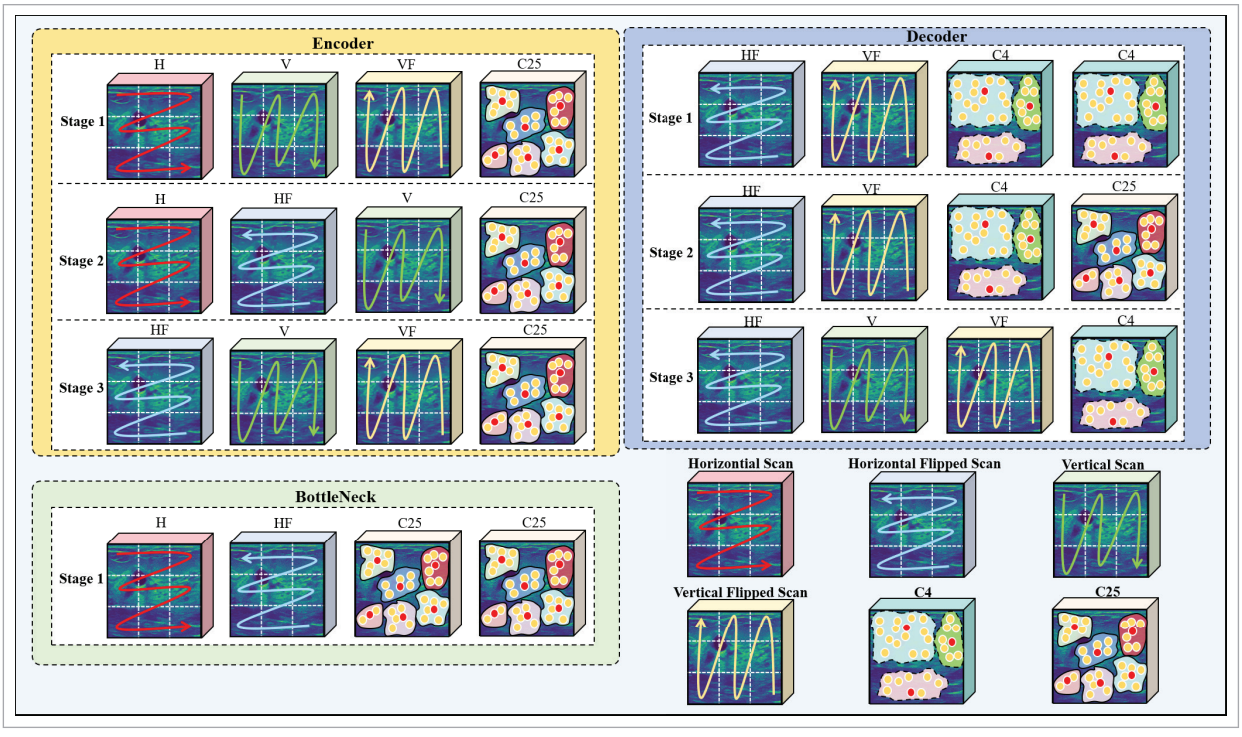


3.2.1. Dynamic Context Modeling Branch

In the dynamic context modeling branch, the feature X_1 is fed into the Context State Space (CSS) module, as illustrated in Figure 4(b). This module is designed based on the VMamba architecture, where the cross-scan mechanism serves as the foundation for sequence modeling, and a dynamic receptive field algorithm, Context Clustering (CC), is introduced. The CC algorithm achieves adaptive dynamic receptive field modeling through feature clustering: First, the input features are represented as a set of data points, and t initial cluster centres $\{c_j\}_{j=1}^{k}$ are obtained through adaptive pooling. For an input feature map with a spatial size of $H\times W$, each initial cluster centre corresponds to a local neighbourhood containing k feature points, defined as $k=\frac{H}{\sqrt{t}}\times\frac{W}{\sqrt{t}}$. The number of cluster centres, t, directly affects the modeling range of the dynamic receptive field, while k is jointly determined by the spatial resolution of the input feature map and the number of cluster centres t. Following the setting of the representative context clustering method CCViM, we adopt $t=4$ and $t=25$, corresponding to the large receptive field clustering variant C4 and the fine-grained receptive field clustering variant C25, respectively. Specifically, C4 aggregates large-range features through four cluster centres, constructing a dynamic receptive field with a broader spatial perception, whereas C25 employs a larger number of cluster centres to capture fine-grained local contextual information, thereby facilitating refined dynamic receptive field modeling. Subsequently, we calculate the semantic similarity between feature points and each cluster centre, and assign the data points in the image to different clusters, allowing each cluster to form a dynamic receptive field region whose spatial range can be adaptively adjusted according to the input content. During this process, feature points within the same cluster aggregate features based on shared global information within the cluster, thereby suppressing redundant information, enhancing feature consistency, and modeling the contextual relationships of scale-varying breast tumors. Algorithm 1 provides the pseudocode description of the CC algorithm.

Figure 5

The dynamic receptive field configuration strategies at different stages of the encoder, bottleneck, and decoder. By adopting stage-specific scanning modes and varying clustering granularities, this design enables adaptive modeling of breast tumors at different scales.



Notably, in the Context Clustering 2D Selective Scan (CCSS2D) of the CSS module, we construct a stage-wise differentiated dynamic receptive field configuration strategy by selecting four optimal combinations from four cross-scan patterns and two context clustering variants (as shown in Figure 4(c)).

Specifically, as illustrated in Figure 5, the encoder stage employs a high-density, strongly localized dynamic receptive field. By using the C25 clustering variant combined with horizontal (H) and vertical (V) scanning patterns, the receptive field is constrained to a local range while establishing global contextual dependencies, enabling the aggregation of semantically consistent feature points to capture fine-grained pathological characteristics. At the bottleneck stage, two consecutive C25 dynamic clustering operations are further cascaded to strengthen contextual modeling of deep semantic features. At the decoder stage, a coarse-grained dynamic receptive field with a broader perceptual range and global integration capability is constructed. By adopting C4 clustering combined with horizontal flip (HF) and vertical flip (VF) scanning patterns, contextual information is integrated over a larger spatial extent, enabling accurate reconstruction of the overall lesion morphology and spatial location. This stage-wise dynamic receptive field configuration strategy enables the model to fully exploit the advantages of dynamic receptive field modeling at different stages of feature extraction and reconstruction, thereby better addressing the high heterogeneity of breast tumors in terms of scale and structure.

The dynamic context modeling branch models the global context of lesions based on cross-scan operations and further incorporates the CC algorithm. By leveraging dynamic receptive fields, it adaptively captures local pathological details of tumors, thereby enabling adaptive modeling of lesion regions at different scales. Finally, global and local information is fused within this branch, producing the output feature $X_{dynamic}$.

3.2.2. Local Static Branch and Refined Fusion Module

We design a local static branch in the SDRF block in parallel with the dynamic context modeling branch to fully exploit the stability advantages of CNN-based static receptive fields in structural modeling. As shown in Figure 4(a), this branch adopts a parallel dual-kernel convolution design, where the input feature X_2 is processed by convolution kernels of different scales, enabling efficient and stable extraction of local structural information within fixed receptive fields and providing structural complementarity for the feature modeling process. The output feature of this branch is denoted as X_{static}.

After obtaining the output features $X_{dynamic}$ and X_{static} from the two branches, respectively, we feed them into a refined fusion module for refinement and fusion. As illustrated in Figure 4(d), for the feature $X_{dynamic}$, we further constrain its stability and discriminability in both spatial structure and channel responses. Specifically, a 1×1 convolution is first applied for channel mapping and reorganization; subsequently, a depthwise separable convolution is introduced to enable lightweight modeling and enhance spatial continuity; finally, a Squeeze-and-Excitation (SE) module is incorporated to adaptively recalibrate channel-wise responses via channel attention, thereby strengthening the responses of relevant feature channels. For the output feature X_{static} from the local static branch, we introduce a direction-aware window-shifting operation. By rearranging features along the horizontal and vertical directions, features within fixed receptive fields can interact more sufficiently across multiple directions, thereby strengthening structural feature representation under static receptive fields. Subsequently, features generated by both branches are merged through channel-wise concatenation to form $X_{refined}$. Finally, this feature is processed by LN again and added to the input feature X via a residual connection, yielding the output feature $X_{output} \in R^{H\times W\times C}$ of the SDRF block. The process can be formally described as follows:

$$X_1, X_2 \in R^{H\times W\times \frac{C}{2}} = Chunk(LN(X), 2) \tag{1}$$

$$X_{dynamic} = CSS(X_1) \tag{2}$$

$$X_{static} = Conv_{1\times 1}(X_2) \bigoplus Conv_{3\times 3}(X_2) \tag{3}$$

$$X_{refined} = Cat\left(SE\left(DWConv_{3\times 3}\left(Conv_{1\times 1}\left(X_{dynamic}\right)\right)\right) + X_{dynamic}, Shift(X_{static})\right) \tag{4}$$

$$X_{output} \in R^{H\times W\times C} = LN\left(X_{refined}\right) + X\ , \quad (5)$$

where Chunk denotes the tensor splitting operation, CSS denotes the Context State Space (CSS) module, SE denotes the Squeeze-and-Excitation (SE) module, and Shift denotes the window shifting operation.

Through a parallel dual-branch architecture, the SDRF block integrates static receptive field modeling methods (Mamba and CNN) with the dynamic algorithm (CC), achieving complementary and collaborative modeling of dynamic and static receptive fields between the dynamic context modeling branch and the local static branch, thereby making the model more suitable for breast tumor segmentation scenarios with significant scale variations.

3.3. Frequency-Selective Modeling (FSM) Module

To further analyze differences in spectral distribution characteristics across breast tumor types, we conduct a quantitative statistical analysis of frequency-domain features of benign and malignant breast tumors using the large-scale PACE 2025 dataset. This dataset contains 1,941 benign case images and 956 malignant case images. Based on the normalized radial distance from the spectrum centre to each frequency component, threshold ratios of 0.25, 0.30, and 0.35 are used to divide the spectrum into low- and high-frequency bands, enabling a systematic evaluation of spectral distribution differences across different high–low-frequency partitioning strategies. As shown in Table 2, across all threshold settings, malignant tumors exhibit a higher proportion of high-frequency energy and a larger high-to-low frequency energy ratio, whereas benign tumors show a significantly higher proportion of low-frequency energy. Further statistical tests demonstrate that these spectral distribution differences are significant ($p < 0.05$), confirming that differences in growth patterns and tissue structural complexity between benign and malignant tumors result in markedly distinct spectral distributions.

Although all three selected high- and low-frequency partition thresholds sufficiently demonstrate the significant spectral distribution differences between benign and malignant breast tumors, varying threshold settings result in distinct feature learning outcomes. Therefore, determining an optimal frequency partition threshold for efficiently extracting frequency-domain discrepancies remains challenging. Motivated by this observation, we expect the model to adaptively learn the spectral differences between benign and malignant tumors, thereby obtaining more discriminative frequency-domain representations.

Table 2

Statistical comparison of frequency-domain features between benign and malignant breast ultrasound images on the PACE 2025 dataset under different high–low frequency partition thresholds. Mean ± Std are reported as descriptive statistics over all samples in each class.

Threshold	Feature	Benign (%)	Malignant (%)	p-value*
0.25	low_energy_ratio	58.78 ± 6.93	56.60 ± 7.03	< 1e-9
	high_energy_ratio	41.22 ± 6.93	43.40 ± 7.03	
	high_low_ratio	72.55 ± 21.00	79.55 ± 23.55	
0.3	low_energy_ratio	65.74 ± 6.34	63.78 ± 6.63	< 1e-9
	high_energy_ratio	34.26 ± 6.34	36.22 ± 6.63	
	high_low_ratio	53.57 ± 15.36	58.58 ± 17.43	
0.35	low_energy_ratio	71.68 ± 5.58	70.05 ± 5.99	< 1e-9
	high_energy_ratio	28.32 ± 5.58	29.95 ± 5.99	
	high_low_ratio	40.37 ± 11.30	43.85 ± 12.97	

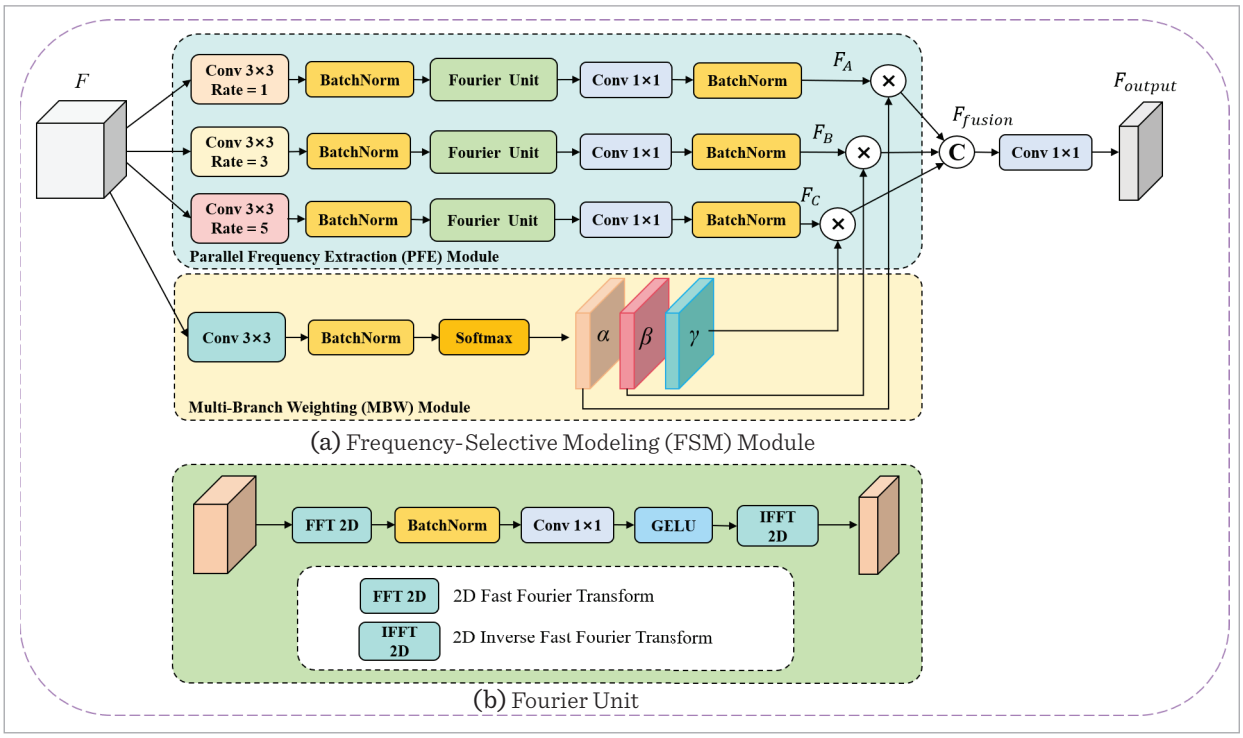
**Statistical significance between benign and malignant samples was evaluated using the Mann–Whitney U test, with $p < 0.05$ considered statistically significant.*

To this end, as shown in Figure 6(a), we propose a Frequency-Selective Modeling (FSM) module that selectively enhances discriminative frequency-domain information based on the spectral distribution of the input tumor features. The proposed module is embedded in the network's skip connections to strengthen the transmission and representation of semantic features highly relevant to breast tumors.

The FSM module consists of a Parallel Frequency Extraction (PFE) module and a Multi-Branch Weighting (MBW) module. Specifically, the input feature F is processed by the PFE module. Considering the significant scale variations of breast tumors, we construct three convolutional branches with different dilation rates for multi-scale feature extraction. Following the multi-scale dilated convolution design widely adopted in existing medical image analysis tasks [38, 59], the dilation rates of the three branches are set to 1, 3, and 5, respectively, to capture local, intermediate-scale, and large-scale tissue structures, thereby enhancing the model's ability to represent scale-varying lesions. By sparse sampling with different dilation rates, frequency-domain modeling can capture discriminative spectral features across multiple structural levels. Meanwhile, dilated convolutions introduce different sparse sampling rates, allowing subsequent modules to exhibit preferences toward different frequency bands. Based on this, as shown in Figure 6(b), the Fourier Unit in each branch selectively enhances frequency components that are relevant to the lesion. In this unit, features are first mapped to the frequency domain via the fast Fourier transform (FFT), followed by local modeling in the frequency domain using a 1×1 convolution. This convolution layer adaptively learns the weighted relationships among different frequency components, thereby strengthening frequency bands of interest. Finally, the features are transformed back to the spatial domain through the inverse FFT (IFFT) and further processed by convolution, generating the output features F_A, F_B, and F_C from the three branches, respectively. The above process can be formulated as:

Figure 6

Structure of the Frequency-Selective Modeling (FSM) module. (a) The FSM module consists of the Parallel Frequency Extraction (PFE) and Multi-Branch Weighting (MBW) modules. (b) The Fourier Unit is the core part of the FSM module.



$$F_A = BN(Conv_{1\times1}(Fourier(BN(Conv_{rate=1}(F))))) \quad (6)$$

$$F_B = BN(Conv_{1\times1}(Fourier(BN(Conv_{rate=3}(F))))) \quad (7)$$

$$F_C = BN(Conv_{1\times1}(Fourier(BN(Conv_{rate=5}(F))))), \quad (8)$$

where $Conv_{rate=1}$, $Conv_{rate=3}$ denote dilated convolutions with dilation rates of 1, 3, and 5, respectively, BN denotes batch normalization, and $Fourier$ denotes the Fourier Unit.

In addition, to effectively integrate the morphological characteristics of breast tumors at different scales, we design a Multi-Branch Weighting (MBW) module to perform adaptive weighted fusion of multi-branch features. As shown in Figure 6(a), the input feature F is sequentially fed into a standard convolution layer and a Softmax function to generate a set of learnable spatial weight maps, α, β and γ. Subsequently, α, β and γ are multiplied element-wise with the output features F_A, F_B, and F_C from the three branches, respectively, achieving pixel-level modulation of branch responses. The weighted outputs are concatenated across channels to generate the fused feature map F_{fusion}. Through this spatially adaptive weighting strategy, the MBW module enables dynamic, scale-aware feature selection and effectively coordinates multi-branch representations to accommodate the large-scale variations in breast tumors. Finally, F_{fusion} is fed into a convolution layer for further integration, producing the output feature F_{output} of the FSM module. The above process can be expressed as:

$$\alpha, \beta, \gamma = Softmax(\, BN(Conv_{3\times3}(F))) \quad (9)$$

$$F_{fusion} = Cat(\alpha \odot F_A, \beta \odot F_B, \gamma \odot F_C) \quad (10)$$

$$F_{output} = Conv_{1\times1}(F_{fusion}). \quad (11)$$

In this module, frequency-domain information can be selectively enhanced and fused at different spatial scales and structural levels, enabling the model to use frequency-domain differences across tumor types as a basis for segmentation, thereby improving its ability to segment tumors with distinct spatial morphologies.

3.4. Loss Function

To enhance boundary delineation of breast tumors and reduce false positive predictions, this paper introduces the Feature-guided Region-aware Loss [26] (FRLoss), which has shown strong performance in MRI segmentation tasks, into ultrasound image segmentation, on the basis of jointly adopting the Binary Cross-Entropy (BCE) loss [36] function and the Dice loss [27] function, so as to enhance feature consistency and improve the model's ability to distinguish subtle differences within tumor regions.

3.4.1. BCE Loss

BCE loss focuses on the classification accuracy of individual pixels, constraining the foreground prediction probability at each pixel location, making it suitable for medical image segmentation tasks. The mathematical form of BCE loss is:

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)], \quad (12)$$

where N denotes the total number of pixels, y_i and $\hat{y}_i$ represents the ground-truth label and the predicted foreground probability of the i-th pixel, respectively.

3.4.2. Dice Loss

Dice loss, by maximizing overlap between predicted and ground-truth regions, effectively mitigates the impact of foreground-background class imbalance in medical image segmentation. Its mathematical form is defined as:

$$L_{Dice} = 1 - \frac{2\sum_{i=1}^{N} y_i \hat{y}_i + 1}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \hat{y}_i + 1}, \quad (13)$$

where N denotes the total number of pixels, y_i and $\hat{y}_i$ represents the ground-truth label and the predicted foreground probability of the i-th pixel, respectively.

3.4.3. FRLoss

FRLoss uses pixel-level ground truth annotations to construct multiple optimization objectives in the feature space with the foreground centre as a reference, respectively constraining the feature compactness of foreground regions, the discriminability of boundary-confused regions, and the separability between hard negative samples and foreground fea-

tures, thereby guiding the model to focus on subtle feature differences and improving its robustness in complex BUS segmentation tasks. This loss function consists of three sub-loss terms: the Positive Compactness Loss, the Boundary-Aware Loss, and the Hard Negative Pixels Mining Loss. By combining the above loss terms, FRLoss can be expressed as:

$$L_{FR} = \lambda_1 L_{pos} + \lambda_2 L_{boundary} + \lambda_3 L_{negative}, \quad (14)$$

where λ_1, λ_2, and λ_3 are the weighting factors for the three sub-loss terms, respectively set to 1, 0.5, and 0.3. Algorithm 2 presents the pseudocode of the FRLoss.

Algorithm 2 Feature-guided Region-aware Loss (FRLoss)

Input:

Predicted feature map $F \in R^{H \times W \times C}$; Ground-truth mask $Y \in \{0,1\}^{H \times W}$; Positive pixel set Ω_{pos}; Negative pixel set Ω_{neg}; Hyperparameters: $\lambda_1, \lambda_2, \lambda_3$, top-k;

Output:

Feature-guided Region-aware Loss L_{FR};

Step 1: Foreground Prototype Computation

1. Compute foreground feature center: f_p = Mean(f_i), for all pixel $i \in \Omega_{pos}, f_i \in F$;

Step 2: Positive Compactness Loss (L_{pos})

2. Initialize L_{pos}=0;
3. For each pixel $i \in \Omega_{pos}$:
   Compute cosine similarity s_i = sim(f_i, f_p);
   Accumulate L_{pos} += (1 - s_i);

$L_{pos} = L_{pos} / |\Omega_{pos}|$

Step 3: Boundary-aware Loss ($L_{boundary}$)

4. Generate boundary region: M = dilate(Ω_{pos})$\cap\Omega_{neg}$
5. Initialize $L_{boundary}$ = 0;
6. For each pixel $i \in M$:
   Compute similarity s_i' = sim(f_i, f_p);
   Accumulate $L_{boundary}$ += ReLU(s_i');
7. $L_{boundary} = L_{boundary} / |M|$;

Step 4: Hard Negative Mining Loss(L_{neg})

8. Compute similarity scores: S = {sim(f_i, f_p) | $i \in \Omega_{neg}$};
9. Select hard negatives: N = dilate(top-k(S))$\cap\Omega_{neg}$;
10. Initialize L_{neg} = 0;
11. For each pixel $i \in N$:
12. Accumulate L_{neg} += ReLU($sim(f_i, f_p)$);
13. $L_{neg} = L_{neg} / |N|$;

Step 5: Loss Aggregation

14. $L_{FR} = \lambda_1 L_{pos} + \lambda_2 L_{boundary} + \lambda_3 L_{neg}$

3.4.4. Overall Loss Function

The overall loss function is composed of FRLoss, BCE loss, and Dice loss. Its formulation is given as follows:

$$L_{total} = L_{Bce} + L_{Dice} + \lambda L_{FR}, \quad (15)$$

where λ denotes the weight of FRLoss and is set to 1 by default.

4. Experiments and Results

4.1. Datasets

To comprehensively evaluate the proposed model, we conducted experiments across three publicly available BUS image datasets: PACE 2025, BUI, and BUS_UC. Table 3 provides the composition of each dataset, with all datasets split into training and testing sets at a 7:3 ratio.

Table 3

Overview of Breast Ultrasound Datasets. N/A indicates that no normal breast ultrasound images are available in the corresponding dataset.

Dataset	Normal	Benign	Malignant	Total
PACE 2025	266	1941	956	3163
BUI	N/A	109	123	232
BUS_UC	N/A	358	453	811

The PACE 2025 dataset is derived from the first phase of the PRECISE-Africa Breast Ultrasound Segmentation and Classification Challenge 2025. The dataset contains 3,163 breast ultrasound images, including 1,941 benign, 956 malignant, and 266 normal cases. All images are provided with finely annotated ground-truth masks. During the data preprocessing stage, 47 low-quality or invalid samples were removed. After removing invalid data, the training set contains 2,178 images, and the test set contains 938 images.

The BUI dataset [1] was collected from Shahid Beheshti University of Medical Sciences in Tehran, Iran. This dataset comprises 232 breast tumor ultrasound cases, including 123 malignant and 109 be-

nign samples. The training set contains 162 images, and the test set contains 70 images, ensuring the rationality of model training and evaluation.

The BUS_UC dataset contains a total of 811 breast ultrasound images, collected jointly by SonoSkills and Hitachi Medical Systems Europe. The dataset includes ultrasound images of 358 benign tumors and 453 malignant tumors, with all images providing finely annotated ground-truth masks as in previous studies [19]. The training set contains 566 images, and the test set contains 245 images.

4.2. Implementation Details

All experiments are implemented based on the PyTorch deep learning framework and conducted on a Linux operating system (kernel 5.4.0). Python 3.9.23, CUDA 11.7, and PyTorch 1.13.0 are used for model training and inference. In all experiments, the batch size is set to 8. The number of worker processes is set to 8 to improve data loading efficiency. The AdamW optimizer is used for model training, with an initial learning rate of 1e-3. The first-order and second-order momentum parameters are set to 0.9 and 0.999, respectively. To suppress overfitting and improve generalization, the weight decay is set to 1e-2. The learning rate scheduler adopts CosineAnnealingLR to dynamically adjust the learning rate using cosine annealing. The maximum number of iterations in each annealing cycle is set to 50. The learning rate decays from its initial value to a minimum of 1e-5 within each cycle. The total number of training epochs is set to 300. All experiments are conducted on a single NVIDIA A100-PCIE-40GB GPU.

4.3. Evaluation Metrics

To evaluate the segmentation performance of SDFS-Net, we first use a confusion matrix to analyze pixel-level predictions. As shown in Table 4, the predicted values are compared with the ground-truth labels to classify each pixel into one of four categories: true positive (TP), true negative (TN), false positive (FP), or false negative (FN).

Table 4
Constituents of the Confusion Matrix (TP, TN, FP, FN).

	Ground Truth: Positive	Ground Truth: Negative
Predicted: Positive	TP	FP
Predicted: Negative	FN	TN

We selected five core metrics for a comprehensive evaluation: for segmentation accuracy, we use the mean Intersection over Union (mIoU) and Dice similarity coefficient (DSC) to measure consistency between the segmented region and the ground-truth mask. For classification performance, accuracy (Acc), sensitivity (Sen), and specificity (Spe) are used to evaluate the model's pixel-level classification performance from different perspectives. The formulas for these metrics are as follows:

$$Acc = \frac{TP+TN}{TP+FP+TN+FN} \tag{16}$$

$$Dice = \frac{2TP}{2TP+FP+FN} \tag{17}$$

$$IoU = \frac{TP}{TP+FP+FN} \tag{18}$$

$$Sen = \frac{TP}{TP+FN} \tag{19}$$

$$Spe = \frac{TN}{TN+FP}, \tag{20}$$

where TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative pixel predictions, respectively.

4.4. Comparison with State-of-the-Art Models

To validate the effectiveness of the proposed SDFS-Net, we conducted systematic comparisons with a series of classical and state-of-the-art segmentation models, including UNet [40], UNet++ [61], ResUNet++ [20], TransUNet [8], VMUNet [41], VMUNetv2 [58], Mamba-UNet [47], MLLAUNet [21], H-vmunet [49, 50], CCViM [63, 64], and HCMNet [52, 53]. These models cover three mainstream architectures, namely CNN-, Transformer-, and Mamba-based methods, ensuring a comprehensive comparison. All experiments were conducted in the same software and hardware environment, with identical dataset splits, to eliminate confounding factors and ensure fair comparisons.

4.4.1. Results on PACE 2025

Table 5 compares the performance of the proposed SDFS-Net with various advanced segmentation models on the PACE 2025 dataset. The results show that SDFS-Net achieves the best overall performance, with mIoU, DSC, and Acc reaching 72.48%, 84.04%, and 97.56%, respectively. Compared with Mamba-based methods proposed in 2025, including H-vmunet [49], CCViM [64], and HCMNet [52], our model improves mIoU by 9.67%, 5.57%, and 1.43%, respectively. These findings confirm that our model outperforms all baseline approaches.

Figure 7 presents qualitative comparisons of segmentation results from various methods on the PACE 2025 dataset. TransUNet [8] and H-vmunet [49] suffer from under-segmentation, where some lesion areas are misclassified as background. VMUNet [41] and Mamba-UNet [47] show instability at the le-

Figure 7

Visualization of breast tumor segmentation results from different methods on the PACE 2025 dataset. The first column shows the input ultrasound images, the second column presents the corresponding ground truth (GT), and the remaining columns show the segmentation results from the proposed SDFS-Net and various advanced methods. Both GT and prediction results are displayed as binary masks, where white and black regions represent tumor and background areas, respectively.

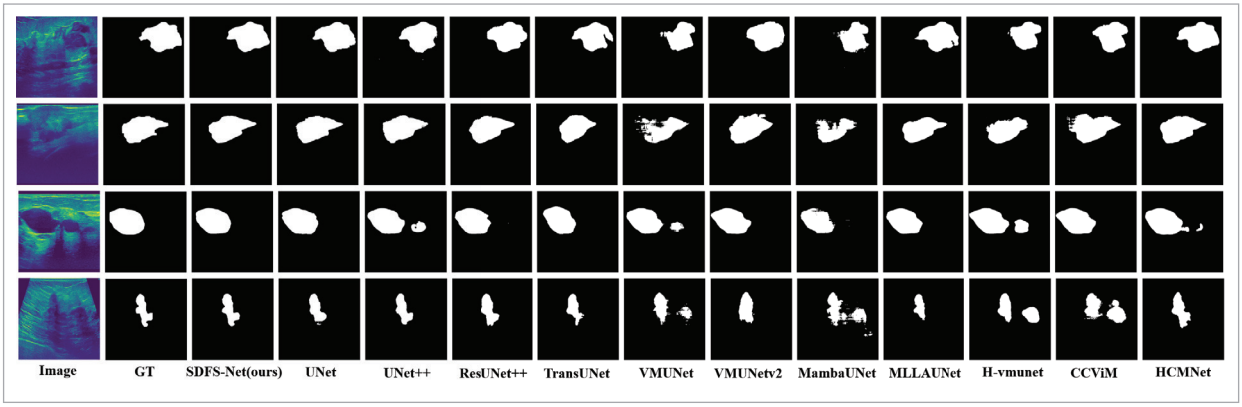


Table 5

Quantitative comparison results on the PACE 2025 dataset. The symbol ↑ indicates the larger the better. The best results are in bold.

Methods	Year	mIoU (%)↑	DSC (%)↑	Acc (%)↑	Spe (%)↑	Sen (%)↑
UNet [40]	2015	65.52 ± 1.08	79.17 ± 0.79	96.87 ± 0.11	98.83 ± 0.10	74.37 ± 1.29
UNet++ [61]	2018	63.05 ± 1.40	77.33 ± 1.05	96.53 ± 0.19	98.49 ± 0.22	73.96 ± 1.08
ResUNet++ [20]	2019	65.78 ± 1.56	79.35 ± 1.14	96.96 ± 0.14	99.02 ± 0.08	73.17 ± 1.81
TransUNet [8]	2021	70.47 ± 1.25	82.67 ± 0.86	97.37 ± 0.10	99.12 ± 0.06	77.51 ± 1.83
VMUNet [41]	2024	60.85 ± 2.36	75.64 ± 1.81	96.32 ± 0.29	98.46 ± 0.23	71.57 ± 1.95
VMUNetv2 [58]	2024	65.54 ± 3.11	79.14 ± 2.24	96.80 ± 0.32	98.60 ± 0.41	76.08 ± 4.67
Mamba-UNet [47]	2024	62.47 ± 1.86	76.88 ± 1.41	96.73 ± 0.17	99.21 ± 0.06	68.12 ± 1.86
MLLAUNet [21]	2024	69.58 ± 2.59	82.03 ± 1.82	97.38 ± 0.22	99.31 ± 0.06	75.13 ± 2.84
H-vmunet [49]	2025	62.81 ± 2.47	77.13 ± 1.85	96.49 ± 0.25	98.44 ± 0.16	74.05 ± 2.86
CCViM [64]	2025	66.91 ± 1.80	80.16 ± 1.28	96.97 ± 0.16	98.72 ± 0.12	76.80 ± 2.48
HCMNet [52]	2025	71.05 ± 1.24	83.07 ± 0.84	97.41 ± 0.09	98.94 ± 0.17	79.74 ± 2.49
SDFS-Net (ours)	-	72.48 ± 0.85	84.04 ± 0.57	97.56 ± 0.11	99.04 ± 0.13	80.47 ± 0.20

sion boundaries. Compared with these methods, our SDFS-Net maintains lesion contour integrity and achieves more precise segmentation results in these complex scenarios. The fourth row presents a more challenging case, where the lesion is a highly irregular malignant breast tumor, and the overall segmentation area is small. For this sample, most comparison methods fail to provide reliable segmentation results. In contrast, SDFS-Net demonstrates more stable performance in modeling the lesion's edge details and better restores the lesion's true contours.

4.4.2. Results on BUI

As shown in Table 6, on the smaller BUI dataset, SDFS-Net also achieves the best performance, with mIoU and DSC of 77.24% and 87.15%, respectively. Compared with the latest Mamba-based methods proposed in 2025, SDFS-Net shows clear

Figure 8

Visualization of breast tumor segmentation results from different methods on the BUI dataset. The first column shows the input ultrasound images, the second column presents the corresponding ground truth (GT), and the remaining columns show the segmentation results from the proposed SDFS-Net and various advanced methods. Both GT and prediction results are displayed as binary masks, where white and black regions represent tumor and background areas, respectively.

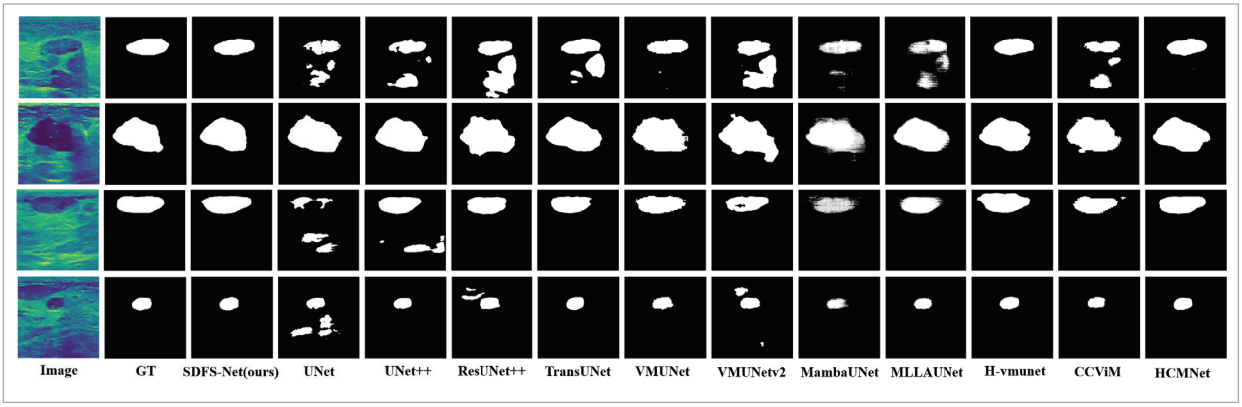


Table 6

Quantitative comparison results on the BUI dataset. The symbol ↑ indicates the larger the better. The best results are in bold.

Methods	Year	mIoU (%)↑	DSC (%)↑	Acc (%)↑	Spe (%)↑	Sen (%)↑
UNet [40]	2015	62.94 ± 0.82	77.26 ± 0.62	91.88 ± 0.52	94.57 ± 1.50	79.12 ± 4.43
UNet++ [61]	2018	67.94 ± 0.30	80.91 ± 0.21	93.39 ± 0.29	96.16 ± 0.88	80.28 ± 2.48
ResUNet++ [20]	2019	63.07 ± 3.29	77.30 ± 2.51	91.55 ± 1.26	93.55 ± 1.59	82.10 ± 1.03
TransUNet [8]	2021	76.62 ± 1.85	86.75 ± 1.18	95.47 ± 0.38	97.67 ± 0.50	85.07 ± 2.53
VMUNet [41]	2024	68.47 ± 1.82	81.27 ± 1.29	93.17 ± 0.73	94.94 ± 1.31	84.77 ± 2.07
VMUNetv2 [58]	2024	57.22 ± 2.13	72.77 ± 1.73	89.62 ± 1.23	91.80 ± 2.06	79.28 ± 2.88
Mamba-UNet [47]	2024	64.24 ± 1.30	78.22 ± 0.97	92.67 ± 0.10	96.25 ± 0.57	75.67 ± 3.24
MLLAUNet [21]	2024	70.34 ± 2.13	82.57 ± 1.46	93.83 ± 0.68	96.01 ± 1.03	83.52 ± 0.95
H-vmunet [49]	2025	76.77 ± 0.71	86.86 ± 0.46	95.37 ± 0.15	96.99 ± 0.23	87.72 ± 1.14
CCViM [64]	2025	61.99 ± 2.81	76.50 ± 2.16	91.11 ± 1.65	93.00 ± 3.19	82.14 ± 5.86
HCMNet [52]	2025	76.00 ± 2.22	86.35 ± 1.45	95.03 ± 0.73	96.14 ± 1.34	89.75 ± 2.22
SDFS-Net (ours)	-	77.24 ± 1.56	87.15 ± 0.99	95.54 ± 0.33	97.40 ± 0.16	86.73 ± 1.35

advantages in both mIoU and DSC. Specifically, it outperforms H-vmunet [49] by 0.47% and 0.29%, and achieves much larger gains over CCViM [64], with improvements of 15.25% and 10.65%, respectively. In addition, even in small-sample scenarios where CNN-based methods usually perform better, SDFS-Net still shows clear superiority over representative CNN models. Compared with ResUNet++ [20], the improvements reach 14.17% and 9.85%. These results indicate that SDFS-Net can maintain high segmentation accuracy even under limited data conditions.

Figure 8 shows the segmentation results of different methods on the BUI dataset.

The first row presents benign tumor ultrasound images, where the tumor regions are highly similar to normal breast tissue. This similarity leads other comparison methods to produce false-positive predictions. In contrast, SDFS-Net effectively suppresses such errors by introducing FRLoss. In addition, for the malignant lesions shown in the second and fourth rows, Mamba-based methods such as VMUNet [41], VMUNetv2 [58], and MLLAUNet [21] fail to delineate tumor boundaries and overall shapes accurately. By contrast, SDFS-Net employs an FSM module to model key frequency-domain features effectively. It enables the model better to capture morphological and structural differences across breast tumor types, leading to more accurate and stable lesion segmentation results.

4.4.3 Results on BUS_UC

As shown in Table 7, the comparison results on the BUS_UC dataset indicate that SDFS-Net also achieves the best segmentation performance. It attains 85.93% mIoU and 92.43% DSC, ranking first among all compared methods. These results consistently demonstrate that the proposed SDFS-Net achieves higher segmentation accuracy than methods based on different network architectures and shows stronger robustness across datasets of different scales in breast ultrasound image segmentation tasks.

Figure 9 shows examples of breast tumor segmentation with large-scale variations and apparent shape differences. Most competing methods suffer from over-segmentation or missing foreground regions. In contrast, SDFS-Net produces more complete and accurate lesion segmentation results.

Together with the previous visualization results, these findings verify that the proposed SDFS-Net, through the complementary collaboration of static

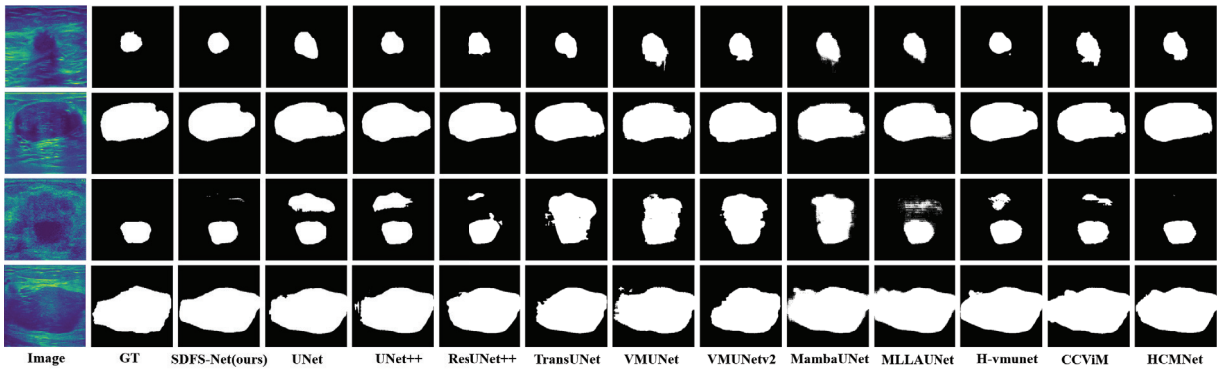
Table 7

Quantitative comparison results on the BUS_UC dataset. The symbol ↑ indicates the larger the better. The best results are in bold.

Methods	Year	mIoU (%)↑	DSC (%)↑	Acc (%)↑	Spe (%)↑	Sen (%)↑
UNet [40]	2015	79.32 ± 0.58	88.47 ± 0.36	96.97 ± 0.05	98.20 ± 0.22	88.82 ± 1.81
UNet++ [61]	2018	81.83 ± 0.56	90.00 ± 0.34	97.36 ± 0.08	98.33 ± 0.22	90.90 ± 1.51
ResUNet++ [20]	2019	80.60 ± 0.74	89.26 ± 0.45	97.09 ± 0.20	97.84 ± 0.54	92.12 ± 2.04
TransUNet [8]	2021	82.48 ± 1.09	90.40 ± 0.66	97.54 ± 0.16	98.94 ± 0.35	88.30 ± 2.25
VMUNet [41]	2024	79.64 ± 0.83	88.67 ± 0.51	97.07 ± 0.10	98.48 ± 0.14	87.68 ± 1.65
VMUNetv2 [58]	2024	78.12 ± 0.93	87.71 ± 0.59	96.77 ± 0.21	98.08 ± 0.42	88.06 ± 1.23
Mamba-UNet [47]	2024	81.76 ± 0.56	89.96 ± 0.34	97.46 ± 0.08	99.08 ± 0.09	86.76 ± 0.68
MLLAUNet [21]	2024	83.17 ± 0.86	90.81 ± 0.51	97.69 ± 0.11	99.25 ± 0.05	87.31 ± 1.11
H-vmunet [49]	2025	85.04 ± 0.59	91.91 ± 0.35	97.89 ± 0.09	98.82 ± 0.15	91.71 ± 0.71
CCViM [64]	2025	84.15 ± 0.35	91.39 ± 0.20	97.75 ± 0.05	98.77 ± 0.03	91.01 ± 0.23
HCMNet [52]	2025	85.08 ± 0.83	91.94 ± 0.48	97.89 ± 0.13	98.77 ± 0.14	92.04 ± 0.69
SDFS-Net (ours)	-	85.93 ± 0.58	92.43 ± 0.33	98.03 ± 0.11	98.99 ± 0.28	91.66 ± 1.20

Figure 9

Visualization of breast tumor segmentation results from different methods on the BUS_UC dataset. The first column shows the input ultrasound images, the second column presents the corresponding ground truth (GT), and the remaining columns show the segmentation results from the proposed SDFS-Net and various advanced methods. Both GT and prediction results are displayed as binary masks, where white and black regions represent tumor and background areas, respectively.



and dynamic receptive fields and the selective modeling of frequency-domain features, can effectively handle complex segmentation scenarios with large-scale variations and significant shape differences, achieving accurate and efficient breast tumor segmentation.

4.5. Ablation Studies

4.5.1. Ablation of Model Components

In this section, we conduct ablation studies on the PACE 2025 dataset to evaluate the contribution of each component in SDFS-Net. Specifically, we analyze the effects of the Static–Dynamic Receptive Field (SDRF) block, the Frequency-Selective Modeling (FSM) module, and the Feature-Guided Region-Aware Loss (FRLoss) on overall segmentation performance. Baseline models are constructed by removing the above components from the full model. The ablation results are reported in Table 8. The baseline model achieves 58.43% mIoU and 73.74% DSC on this dataset.

We then gradually add each component to the baseline model. As shown in Table 8, introducing the SDRF block leads to a clear performance improvement. Compared with the baseline, mIoU and DSC increase by 5.94% and 4.57%, respectively. As the visual backbone of the model, the SDRF block enables cooperation and complementarity between different receptive fields. This design improves the modeling and adaptation capabilities for lesions with large-scale variations, thereby improving segmentation accuracy. On this basis, we further introduce the FSM module. Benefiting from its selective enhancement of frequency-domain information, the transmission and representation of key semantic features in the skip connections are strengthened. As a result, mIoU and DSC are further improved by 3.65% and 2.65%, respectively. Finally, FRLoss is introduced to form the complete SDFS-Net. This loss function guides the model to perceive subtle feature differences and effectively suppresses false positive predictions. The final performance reaches 72.48% mIoU and 84.04% DSC. These results demonstrate that the proposed components make complementary contributions to achieving accurate breast tumor segmentation.

Furthermore, we observed that introducing FRLoss into the baseline model improved the DSC by 2.41%, which further verifies its effectiveness in the breast ultrasound segmentation task. Although this loss function was originally proposed for MRI segmentation, it can likewise effectively alleviate challenges in ultrasound segmentation, such as ambiguous boundaries and low tissue contrast, through learning and optimization of difficult-to-segment regions, thereby demonstrating its excellent cross-modality applicability.

To further validate the role of each module in lesion localization and segmentation, we use Grad-CAM to visualize the segmentation results from differ-

Table 8

Ablation studies on the component contributions of SDFS-Net on the PACE 2025 dataset. The symbol ↑ indicates the larger the better.The best results are in bold. The baseline refers to the backbone network without SDRF, FSM and FRLoss. The symbol "—" denotes that the corresponding module is not used.

Setting	SDRF	FSM	FRLoss	mIoU(%)↑	DSC(%)↑	Acc(%)↑	Spe(%)↑	Sen(%)↑
Baseline	—	—	—	58.43 ± 1.88	73.74 ± 1.49	96.02 ± 0.23	98.29 ± 0.20	69.89 ± 1.91
+ SDRF	✓	—	—	64.37 ± 1.15	78.31 ± 0.85	96.66 ± 0.06	98.51 ± 0.25	75.47 ± 3.06
+ FSM	—	✓	—	67.73 ± 1.17	80.75 ± 0.83	97.08 ± 0.16	98.86 ± 0.19	76.57 ± 0.44
+ FRLoss	—	—	✓	60.84 ± 2.25	75.63 ± 1.73	96.37 ± 0.26	98.61 ± 0.29	70.57 ± 2.66
+ SDRF + FSM	✓	✓	—	68.02 ± 0.77	80.96 ± 0.54	97.05 ± 0.15	98.68 ± 0.26	78.33 ± 1.22
+ SDRF + FRLoss	✓	—	✓	64.83 ± 2.99	78.62 ± 2.20	96.81 ± 0.31	98.82 ± 0.18	73.59 ± 2.90
+ FSM + FRLoss	—	✓	✓	69.70 ± 1.16	82.14 ± 0.80	97.32 ± 0.14	99.08 ± 0.13	77.10 ± 0.22
+ SDRF + FSM + FRLoss	✓	✓	✓	72.48 ± 0.85	84.04 ± 0.57	97.56 ± 0.11	99.04 ± 0.13	80.47 ± 0.20

ent model components on the PACE 2025 dataset. As shown in Figure 10, after introducing the SDRF block into the baseline model, the model can more effectively capture the diversity of breast tumors in terms of shape and scale by leveraging complementary modeling across different receptive fields, leading to more accurate lesion localization. Moreover, the FSM module efficiently captures key frequency-band information about lesions, enabling the model to focus on lesion-related regions. FRLoss effectively strengthens the model's ability to delineate tumor boundaries. By combining these modules, SDFS-Net not only achieves precise foreground localization but also shows more complete, continuous, and consistent high-activation responses.

4.5.2. Ablation Study of SDRF Blocks

As the core visual backbone of the model, the SDRF block plays a crucial role in receptive field modeling within SDFS-Net. To analyze its specific contribution at different stages of the network, we conducted ablation experiments on SDRF blocks in the encoder, decoder, and bottleneck. Specifically, while keeping the rest of the network structure unchanged, we removed the SDRF blocks at the aforementioned locations and quantitatively analyzed segmentation performance on the PACE 2025 dataset. The experimental results are shown in Table 9.

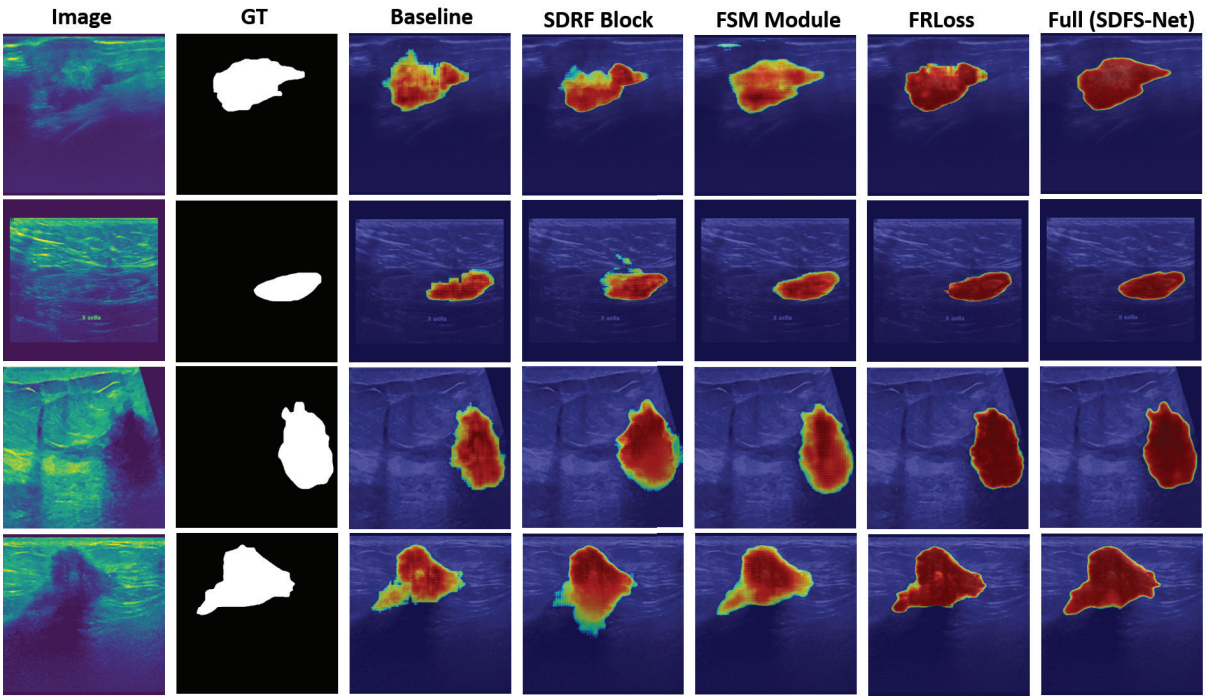
The experimental results show that the SDRF blocks, as the core visual backbone of SDFS-Net, play an important role at all network stages, while

Table 9

Ablation studies on the contribution of SDRF blocks in the encoder, decoder, and bottleneck of SDFS-Net. The symbol ↑ indicates the larger the better. The best results are in bold.

Module	mIoU (%)↑	DSC (%)↑	Acc (%)↑	Spe (%)↑	Sen (%)↑
Encoder w/o SDRF blocks	69.81 ± 2.27	82.20 ± 1.56	97.36 ± 0.22	99.19 ± 0.17	76.28 ± 2.07
Decoder w/o SDRF blocks	71.34 ± 1.41	83.26 ± 0.96	97.50 ± 0.13	99.19 ± 0.02	78.00 ± 1.44
BottleNeck w/o SDRF blocks	70.84 ± 0.91	82.93 ± 0.62	97.38 ± 0.10	98.92 ± 0.21	79.64 ± 2.00
Ours	72.48 ± 0.85	84.04 ± 0.57	97.56 ± 0.11	99.04 ± 0.13	80.47 ± 0.20

Figure 10
Comparison of Grad-CAM heatmaps produced by different model components on the PACE 2025 dataset.



their contributions vary across stages. When all SDRF blocks in the encoder are removed, the model's performance drops noticeably. The mIoU and DSC decrease by 2.67% and 1.84%, respectively. This result indicates that during the feature encoding stage, the SDRF block effectively extracts lesion features from scale-variant breast tumors by collaboratively modeling static and dynamic receptive fields. Furthermore, removing the SDRF blocks from the decoder stage leads to a severe performance degradation, with mIoU and DSC decreasing by 1.14% and 0.78%, respectively. By integrating static and dynamic receptive fields, the SDRF blocks facilitate feature reconstruction and semantic refinement in the decoder and improve robustness to lesion-scale variations, leading to higher segmentation accuracy.

4.5.3. Comparison of FSM Module with Spatial-Domain Attention Mechanisms

The FSM module selectively enhances frequency-domain information and suppresses irrelevant spectral noise. In fact, most existing methods still enhance features in the spatial domain. To validate the effectiveness of the proposed FSM module, we conduct comparison experiments with several representative spatial-domain attention mechanisms, including Self-Attention (SA) [43], Deformable Large Kernel Attention (D-LKA) [2], Hierarchical Attention Module (HAM) [18], Channel-Spatial Attention (CSA) [37], Coordinate Attention (CA) [15], and Efficient Channel Attention (ECA) [46]. To ensure fair comparison, all comparison modules are replaced at the same network position and trained and evaluated on the PACE 2025 dataset.

As shown in Table 10, compared with the baseline model, our FSM module improves mIoU and DSC by 9.30% and 7.01%, respectively. Additionally, compared to the recently proposed attention mechanisms D-LKA [2] and HAM [18], the FSM module improves DSC by 10.40% and 1.30%, respectively. These results suggest that spatial-domain attention mechanisms struggle to consistently focus on foreground regions when handling breast tumors with significant morphological and structural differenc-

Table 10
Performance comparison of the FSM module with spatial-domain attention mechanisms on the PACE 2025 dataset. The symbol ↑ indicates the larger the better. The best results are in bold.

Module	mIoU (%)↑	DSC (%)↑	Acc (%)↑	Spe (%)↑	Sen (%)↑
Baseline	58.43 ± 1.88	73.74 ± 1.49	96.02 ± 0.23	98.29 ± 0.20	69.89 ± 1.91
Baseline + FSM module	67.73 ± 1.17	80.75 ± 0.83	97.08 ± 0.16	98.86 ± 0.19	76.57 ± 0.44
Baseline + SA [43]	54.56 ± 1.78	70.58 ± 1.50	95.64 ± 0.19	98.24 ± 0.53	65.65 ± 4.94
Baseline + D-LKA [2]	54.37 ± 3.99	70.35 ± 3.30	95.45 ± 0.49	97.86 ± 0.29	67.64 ± 3.74
Baseline + HAM [18]	65.93 ± 1.87	79.45 ± 1.35	96.80 ± 0.17	98.48 ± 0.06	77.48 ± 2.51
Baseline + CSA [37]	65.54 ± 2.40	79.16 ± 1.75	96.93 ± 0.16	99.00 ± 0.18	73.10 ± 4.02
Baseline + CA [15]	65.32 ± 1.89	79.01 ± 1.38	96.77 ± 0.21	98.55 ± 0.13	76.17 ± 1.33
Baseline + ECA [46]	62.22 ± 2.05	76.69 ± 1.55	96.44 ± 0.34	98.45 ± 0.43	73.22 ± 0.82

Table 11
Comparison of computational complexity across methods for a 1×3×256×256 input image.

Methods	Year	FLOPs (G)	Parameters (M)	Inference time (ms)
UNet [40]	2015	54.73	31.03	243.73
UNet++ [61]	2018	34.90	9.16	272.30
ResUNet++ [20]	2019	70.99	14.48	548.96
TransUNet [8]	2021	32.23	93.23	361.55
VMUNet [41]	2024	4.11	22.03	303.20
VMUNetv2 [58]	2024	4.40	17.90	384.10
Mamba-UNet [47]	2024	4.60	15.47	280.40
MLLAUNet [21]	2024	7.69	28.50	456.44
H-vmunet [49]	2025	0.74	6.43	706.74
CCViM [64]	2025	4.56	23.61	501.61
HCMNet [52]	2025	6.10	26.97	417.58
SDFS-Net (ours)	-	9.42	51.76	376.97

es, resulting in reduced segmentation performance. In contrast, the proposed FSM module demonstrates higher segmentation accuracy and robustness.

Additionally, introducing FSM results in a computational cost of 7.16 G FLOPs. Although this is higher than lightweight attention mechanisms such as CA (2.73 G) and ECA (2.73 G), FSM improves DSC by 1.74% and 4.06% compared with CA and ECA, respectively, achieving a better performance–complexity trade-off under an acceptable computational cost.

4.6. Efficiency Analysis

In this section, we evaluate the computational complexity of the proposed SDFS-Net in terms of FLOPs, number of parameters, and inference time, and compare it with other methods, as shown in Ta-

ble 11. SDFS-Net requires 9.42 G FLOPs and 51.76 M parameters, and achieves an average inference time of 376.97 ms for a single BUS image with a resolution of 1 × 3 × 256 × 256. These results demonstrate a favourable trade-off between segmentation accuracy and computational cost. Specifically, compared with H-vmunet [49] and CCViM [64], the inference time of SDFS-Net is reduced by 329.77 ms and 124.64 ms, respectively. Compared with HCMNet [52], SDFS-Net also achieves a 40.61 ms reduction. In addition, CNN- and Transformer-based models, such as UNet [40], ResUNet++ [20], and TransUNet [8], often incur higher FLOPs or longer inference times, limiting their applicability in resource-constrained scenarios. In contrast, our model achieves state-of-the-art performance while maintaining a favorable accuracy–complexity trade-off.

Some lightweight models (e.g., H-vmunet and Mamba-UNet) have lower computational costs but relatively limited segmentation performance, making them more suitable for deployment scenarios with constrained computational resources or high real-time requirements. In contrast, in clinical application scenarios with more sufficient computational resources and higher requirements for segmentation accuracy, SDFS-Net achieves the best segmentation performance, demonstrating a superior performance–efficiency trade-off.

5. Conclusion

In this paper, we propose a novel BUS segmentation method, SDFS-Net, to address the challenges posed by lesion-scale variability and the pronounced structural differences between benign and malignant breast tumors. SDFS-Net adopts the SDRF block as its core visual backbone. This block jointly models static and dynamic receptive field information, thereby improving the robustness of breast tumor segmentation to scale variations. In addition, frequency-domain analysis shows that differences in morphology and tissue structure between benign and malignant tumors lead to distinct spectral distributions. Based on this observation, we design the FSM module to selectively model frequency-domain features. It enhances key frequency components in lesion regions and enables the model to better capture tumors with large spatial and structural variations. Moreover, the introduced FRLoss guides the model to focus on subtle feature differences. It improves boundary delineation and effectively reduces false-positive predictions.

Experiments on three BUS datasets (PACE 2025, BUI, and BUS_UC) show that SDFS-Net consistently outperforms existing mainstream methods in overall segmentation performance. It demonstrates its effectiveness and superiority for BUS segmentation. In addition, ablation studies clearly verify the contributions of the proposed SDRF block, FSM module, and FRLoss to performance improvement. The SDRF block has a significant impact on segmentation performance across different network stages, validating the effectiveness of the receptive field collaborative modeling strategy for multi-scale breast tumor segmentation. On the other hand, the FSM module improves segmentation performance for lesions with significant structural differences by adaptively capturing frequency-domain features across different tumor types. This result confirms the feasibility of frequency–domain–based modeling for accurate breast tumor segmentation.

The current work originates from clinical observations in ultrasound imaging, where we identified the differences in spectral distributions and scale variability between benign and malignant breast tumors. However, we have not yet further incorporated more explicit clinical priors and related clinical knowledge, particularly under clinical scenarios with limited computational resources, high real-time requirements, and the context of large-scale multi-centre datasets. In future work, we will conduct an in-depth investigation of clinical literature related to breast tumors to further explore clinical facts regarding tumor characteristics across different scales and frequency domains, and incorporate these findings as prior knowledge into the design and optimization of the proposed model to improve its real-time inference efficiency. Furthermore, we will seek further clinical pilot studies and collaborations to promote the model's diagnostic support in real-world clinical scenarios.

Funding

This work was supported by the National Natural Science Foundation of China under Grant 62166025.

References

1. Ardakani, A. A., Mohammadi, A., Mirza-Aghazadeh-Attari, M., Acharya, U. R. An Open-Access Breast Lesion Ultrasound Image Database: Applicable in Artificial Intelligence Studies. Computers in Biology and Medicine, 2023, 152, 106438. https://doi.org/10.1016/j.compbiomed.2022.106438
2. Azad, R., Niggemeier, L., Hüttemann, M., Kazerouni, A., Aghdam, E. K., Velichko, Y., Bagci, U., Merhof, D. Beyond Self-Attention: Deformable Large Kernel Attention for Medical Image Segmentation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2024, 1287-1297. https://doi.org/10.1109/WACV57701.2024.00132
3. Boone, J. M., Kwan, A. L., Yang, K., Burkett, G. W., Lindfors, K. K., Nelson, T. R. Computed Tomography for Imaging the Breast. Journal of Mammary Gland Biology and Neoplasia, 2006, 11(2), 103-111. https://doi.org/10.1007/s10911-006-9017-1
4. Cai, G., Chen, H., Li, Y., Sun, J., Chen, Z., Geng, Q. WD-FFU-Mamba: A Wavelet-Guided Dual-Attention Feature Fusion Mamba for Breast Tumor Segmentation in Ultrasound Images. arXiv preprint arXiv:2512.17278, 2025, https://doi.org/10.48550/arXiv.2512.17278
5. Cai, S., Tian, Y., Lui, H., Zeng, H., Wu, Y., Chen, G. Dense-UNet: A Novel Multiphoton in Vivo Cellular Image Segmentation Model Based on a Convolutional Neural Network. Quantitative Imaging in Medicine and Surgery, 2020, 10(6), 1275. https://doi.org/10.21037/qims-19-1090
6. Chang, A., Zeng, J., Huang, R., Ni, D. EM-Net: Efficient Channel and Frequency Learning with Mamba for 3d Medical Image Segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention, 2024, 266-275. https://doi.org/10.1007/978-3-031-72114-4_26
7. Chen, G., Dai, Y., Zhang, J. C-Net: Cascaded Convolutional Neural Network with Global Guidance and Refinement Residuals for Breast Ultrasound Images Segmentation. Computer Methods and Programs in Biomedicine, 2022, 225, 107086. https://doi.org/10.1016/j.cmpb.2022.107086
8. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A. L., Zhou, Y. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. arXiv Preprint arXiv:2102.04306, 2021, https://doi.org/10.48550/arXiv.2102.04306
9. Diakogiannis, F. I., Waldner, F., Caccetta, P., Wu, C. ResUNet-A: A Deep Learning Framework for Semantic Segmentation of Remotely Sensed Data. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 162, 94-114. https://doi.org/10.1016/j.isprsjprs.2020.01.013
10. Dosovitskiy, A. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv Preprint arXiv:2010.11929, 2020, https://doi.org/10.48550/arXiv.2010.11929
11. Gu, A. and Dao, T. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. arXiv Preprint arXiv:2312.00752, 2023, https://doi.org/10.48550/arXiv.2312.00752
12. Gu, A., Goel, K., Ré, C. Efficiently Modeling Long Sequences with Structured State Spaces. arXiv Preprint arXiv:2111.00396, 2021, https://doi.org/10.48550/arXiv.2111.00396
13. Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y. A Survey on Vision Transformer. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(1), 87-110. https://doi.org/10.1109/TPAMI.2022.3152247
14. Haq, I. U., Ali, H., Li, Y., Liu, Z. MAR-GAN: Multi Attention Residual Generative Adversarial Network for Tumor Segmentation in Breast Ultrasounds. Biomedical Signal Processing and Control, 2025, 100, 107171. https://doi.org/10.1016/j.bspc.2024.107171
15. Hou, Q., Zhou, D., Feng, J. Coordinate Attention for Efficient Mobile Network Design. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, 13713-13722. https://doi.org/10.1109/CVPR46437.2021.01350
16. Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., Han, X., Chen, Y.-W., Wu, J. Unet 3+: A Full-Scale Connected Unet for Medical Image Segmentation. In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020, 1055-1059. https://doi.org/10.1109/ICASSP40776.2020.9053405
17. Huang, T., Pei, X., You, S., Wang, F., Qian, C., Xu, C. LocalMamba: Visual State Space Model with Windowed Selective Scan. In European Conference on Computer Vision (ECCV), 2024, 12-22. https://doi.org/10.1007/978-3-031-91979-4_2

18. Huang, Z., Wang, M., Hou, M., Yu, Z., Wang, S., Li, X., Wang, J., Yan, Y., Liu, Y. BACFormer: A Robust Boundary-Aware Transformer for Medical Image Segmentation. Knowledge-Based Systems, 2025, 115209. https://doi.org/10.1016/j.knosys.2025.115209

19. Iqbal, A. and Sharif, M. Memory-Efficient Transformer Network with Feature Fusion for Breast Tumor Segmentation and Classification Task. Engineering Applications of Artificial Intelligence, 2024, 127, 107292. https://doi.org/10.1016/j.engappai.2023.107292

20. Jha, D., Smedsrud, P. H., Riegler, M. A., Johansen, D., De Lange, T., Halvorsen, P., Johansen, H. D. ResUNet++: An Advanced Architecture for Medical Image Segmentation. In 2019 IEEE International Symposium on Multimedia (ISM), 2019, 225-230. https://doi.org/10.1109/ISM46123.2019.00049

21. Jiang, Y., Li, Z., Chen, X., Xie, H., Cai, J. MLLA-UNet: Mamba-Like Linear Attention in an Efficient U-Shape Model for Medical Image Segmentation. arXiv Preprint arXiv:2410.23738, 2024, https://doi.org/10.48550/arXiv.2410.23738

22. Jin, B., Chen, C., Li, X., Wang, X., Lin, L., Chen, J. MCF-Net: Mamba-Based Channel-Frequency Dual Fusion Network for CBCT Dental Image Segmentation. Biomedical Signal Processing and Control, 2026, 115, 109372. https://doi.org/10.1016/j.bspc.2025.109372

23. Kim, J., Harper, A., McCormack, V., Sung, H., Houssami, N., Morgan, E., Mutebi, M., Garvey, G., Soerjomataram, I., Fidler-Benaoudia, M. M. Global Patterns and Trends in Breast Cancer Incidence and Mortality across 185 Countries. Nature Medicine, 2025, 1-9. https://doi.org/10.1038/s41591-025-03502-3

24. Kuhl, C. MRI of Breast Tumors. European radiology, 2000, 10(1), 46-58. https://doi.org/10.1007/s003300050006

25. Li, H., Pan, H., Liu, X., Ren, J., Du, Z., Cao, J. GLVMamba: A Global-Local Visual State Space Model for Remote Sensing Image Segmentation. IEEE Transactions on Geoscience and Remote Sensing, 2025, https://doi.org/10.1109/TGRS.2025.3572127

26. Li, H., Song, W., Qin, P., Yuan, X., Chen, Z. HCMA-UNet: A Hybrid CNN-Mamba UNet with Axial Self-Attention for Efficient Breast Cancer Segmentation. In 2025 IEEE International Conference on Multimedia and Expo (ICME), 2025, 1-6. https://doi.org/10.1109/ICME59968.2025.11209819

27. Liu, B., Dolz, J., Galdran, A., Kobbi, R., Ayed, I. B. Do We Really Need Dice? The Hidden Region-Size Biases of Segmentation Losses. Medical Image Analysis, 2024, 91, 103015. https://doi.org/10.1016/j.media.2023.103015

28. Liu, Y., Tian, Y., Zhao, Y., Yu, H., Xie, L., Wang, Y., Ye, Q., Jiao, J., Liu, Y. VMamba: Visual State Space Model. Advances in Neural Information Processing Systems, 2024, 37, 103031-103063. https://doi.org/10.52202/079017-3273

29. Long, J., Shelhamer, E., Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, 3431-3440. https://doi.org/10.1109/CVPR.2015.7298965

30. Lu, J., Ding, H., Zhang, S., Huo, G. M-Net: Mri Brain Tumor Sequential Segmentation Network Via Mesh-Cast. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2025, 20116-20125. https://doi.org/10.1109/ICCV51701.2025.01871

31. Luo, X., Wang, Y., Ou-Yang, L. LGFFM: A Localized and Globalized Frequency Fusion Model for Ultrasound Image Segmentation. IEEE Transactions on Medical Imaging, 2026, 45(2), 515-527. https://doi.org/10.1109/TMI.2025.3600327

32. Luo, Y., Huang, Q., Li, X. Segmentation Information with Attention Integration for Classification of Breast Tumor in Ultrasound Image. Pattern Recognition, 2022, 124, 108427. https://doi.org/10.1016/j.patcog.2021.108427

33. Lyu, Y., Xu, Y., Jiang, X., Liu, J., Zhao, X., Zhu, X. AMS-PAN: Breast Ultrasound Image Segmentation Model Combining Attention Mechanism and Multi-Scale Features. Biomedical Signal Processing and Control, 2023, 81, 104425. https://doi.org/10.1016/j.bspc.2022.104425

34. Ma, Z., Qi, Y., Xu, C., Zhao, W., Lou, M., Wang, Y., Ma, Y. ATFE-Net: Axial Transformer and Feature Enhancement-Based CNN for Ultrasound Breast Mass Segmentation. Computers in Biology and Medicine, 2023, 153, 106533. https://doi.org/10.1016/j.compbiomed.2022.106533

35. Maklachur Rahman, M., Jung, S. K., Hammond, T. MambaLiteUNet: Cross-Gated Adaptive Feature Fusion for Robust Skin Lesion Segmentation. arXiv e-prints arXiv: 2604.20286, 2026, https://doi.org/10.48550/arXiv.2604.20286

36. Mao, A., Mohri, M., Zhong, Y. Cross-Entropy Loss Functions: Theoretical Analysis and Applications. arXiv Preprint arXiv:2304.07288, 2023, https://doi.org/10.48550/arXiv.2304.07288

37. Mou, L., Zhao, Y., Chen, L., Cheng, J., Gu, Z., Hao, H., Qi, H., Zheng, Y., Frangi, A., Liu, J. CS-Net: Channel and Spatial Attention Network for Curvilinear Structure Segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2019, 721-730. https://doi.org/10.1007/978-3-030-32239-7_80

38. Qiu, Y., Chai, S., Zhu, E., Zhang, N., Zhang, G., Zhao, X., Cui, L., Farhan, I. M. Deep Multi-Scale Dilated Convolution Network for Coronary Artery Segmentation. Biomedical Signal Processing and Control, 2024, 92, 106021. https://doi.org/10.1016/j.bspc.2024.106021

39. Ranftl, R., Bochkovskiy, A., Koltun, V. Vision Transformers for Dense Prediction. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021, 12179-12188. https://doi.org/10.1109/ICCV48922.2021.01196

40. Ronneberger, O., Fischer, P., Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

41. Ruan, J., Li, J., Xiang, S. VM-UNet: Vision Mamba UNet for Medical Image Segmentation. ACM Transactions on Multimedia Computing, Communications and Applications, 2025, https://doi.org/10.1145/3767748

42. Sehgal, C. M., Weinstein, S. P., Arger, P. H., Conant, E. F. A Review of Breast Ultrasound. Journal of Mammary Gland Biology and Neoplasia, 2006, 11(2), 113-123. https://doi.org/10.1007/s10911-006-9018-0

43. Shaw, P., Uszkoreit, J., Vaswani, A. Self-Attention with Relative Position Representations. arXiv Preprint arXiv:1803.02155, 2018, https://doi.org/10.18653/v1/N18-2074

44. Tan, T., Platel, B., Huisman, H., Sánchez, C. I., Mus, R., Karssemeijer, N. Computer-Aided Lesion Diagnosis in Automated 3-D Breast Ultrasound Using Coronal Spiculation. IEEE Transactions on Medical Imaging, 2012, 31(5), 1034-1042. https://doi.org/10.1109/TMI.2012.2184549

45. Tang, L., Xiao, H., Jiang, P.-T., Zhang, H., Chen, J., Li, B. Scalable Visual State Space Model with Fractal Scanning. arXiv Preprint arXiv:2405.14480, 2024, https://doi.org/10.48550/arXiv.2405.14480

46. Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, 11534-11542. https://doi.org/10.1109/CVPR42600.2020.01155

47. Wang, Z., Zheng, J.-Q., Zhang, Y., Cui, G., Li, L. Mamba-UNet: UNet-Like Pure Visual Mamba for Medical Image Segmentation. arXiv Preprint arXiv:2402.05079, 2024, https://doi.org/10.48550/arXiv.2402.05079

48. Wu, M., Liu, T., Dai, X., Ye, C., Wu, J., Funahashi, S., Yan, T. HMDA: A Hybrid Model with Multi-Scale Deformable Attention for Medical Image Segmentation. IEEE Journal of Biomedical and Health Informatics, 2024, https://doi.org/10.1109/JBHI.2024.3469230

49. Wu, R., Liu, Y., Liang, P., Chang, Q. H-vmunet: High-Order Vision Mamba Unet for Medical Image Segmentation. Neurocomputing, 2025, 624, 129447. https://doi.org/10.1016/j.neucom.2025.129447

50. Wu, R., Liu, Y., Liang, P., Chang, Q. H-vmunet: Official Pytorch Implementation. GitHub repository. Available: https://github.com/wurenkai/H-vmunet.

51. Wu, R., Lu, X., Yao, Z., Ma, Y. MFMSNet: A Multi-Frequency and Multi-Scale Interactive CNN-Transformer Hybrid Network for Breast Ultrasound Image Segmentation. Computers in Biology and Medicine, 2024, 177, 108616. https://doi.org/10.1016/j.compbiomed.2024.108616

52. Xiong, Y., Shu, X., Liu, Q., Yuan, D. HCMNet: A Hybrid CNN-Mamba Network for Breast Ultrasound Segmentation for Consumer Assisted Diagnosis. IEEE Transactions on Consumer Electronics, 2025, https://doi.org/10.1109/TCE.2025.3593784

53. Xiong, Y., Shu, X., Liu, Q., Yuan, D. HCMNet: Official Pytorch Implementation. GitHub repository. Available: https://github.com/XYQ1517/HCMNet.

54. Xu, J., Li, J., Cui, F., Zhang, Z., Yang, J., Jin, S., Jin, Q., Meng, Y. Eccomamba: Enhanced Cross-Hierarchical Continuity Orthogonal Mamba for Medical Image Segmentation. In Proceedings of the AAAI Conference on Artificial Intelligence, 2026, 11287-11295. https://doi.org/10.1609/aaai.v40i13.38109

55. Xue, C., Zhu, L., Fu, H., Hu, X., Li, X., Zhang, H., Heng, P.-A. Global Guidance Network for Breast Lesion Segmentation in Ultrasound Images. Medical Image Analysis, 2021, 70, 101989. https://doi.org/10.1016/j.media.2021.101989

56. Yang, Z. and Farsiu, S. Directional Connectivity-Based Segmentation of Medical Images. In Proceedings of

the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023, 11525-11535. https://doi.org/10.1109/CVPR52729.2023.01109

57. Yue, G., Wu, S., Li, G., Zhao, C., Hao, Y., Zhou, T., Zhao, B. Boundary-Guided Feature-Aligned Network for Colorectal Polyp Segmentation. IEEE Transactions on Circuits and Systems for Video Technology, 2025, https://doi.org/10.1109/TCSVT.2025.3542969
58. Zhang, M., Yu, Y., Jin, S., Gu, L., Ling, T., Tao, X. VM-UNET-V2: Rethinking Vision Mamba UNet for Medical Image Segmentation. In International Symposium on Bioinformatics Research and Applications, 2024, 335-346. https://doi.org/10.1007/978-981-97-5128-0_27
59. Zhang, Q. A Novel Resnet101 Model Based on Dense Dilated Convolution for Image Classification. SN Applied Sciences, 2022, 4(1), 9. https://doi.org/10.1007/s42452-021-04897-7
60. Zhou, D., Kang, B., Jin, X., Yang, L., Lian, X., Jiang, Z., Hou, Q., Feng, J. DeepViT: Towards Deeper Vision Transformer. arXiv Preprint arXiv:2103.11886, 2021, https://doi.org/10.48550/arXiv.2103.11886
61. Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., Liang, J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In International Workshop on Deep Learning in Medical Image Analysis, 2018, 3-11. https://doi.org/10.1007/978-3-030-00889-5_1
62. Zhu, X., Wang, W., Zhang, C., Wang, H. Polyp-Mamba: A Hybrid Multi-Frequency Perception Gated Selection Network for Polyp Segmentation. Information Fusion, 2025, 115, 102759. https://doi.org/10.1016/j.inffus.2024.102759
63. Zhu, Y., Zhang, D., Lin, Y., Feng, Y., Tang, J. CCViM: Official Pytorch Implementation. GitHub repository. Available: https://github.com/zymissy/CCViM.
64. Zhu, Y., Zhang, D., Lin, Y., Feng, Y., Tang, J. Merging Context Clustering with Visual State Space Models for Medical Image Segmentation. IEEE Transactions on Medical Imaging, 2025, https://doi.org/10.1109/TMI.2025.3525673

This article is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 (CC BY 4.0) License (http://creativecommons.org/licenses/by/4.0/).