electronics

MDPI

Article

WMCAF-Net: A Wavelet-Enhanced Mamba and Cross-Scale Adaptive Fusion Network for Breast Ultrasound Image Segmentation

Xicheng Fu [1], Hongwei Deng [1,2,3,*], Jingshi Lei [1], Yi Lei [1], Lei Xu [1] and Qiangfei Wang [1]

[1] College of Computer Science and Technology, Hengyang Normal University, Hengyang 421002, China; seaking_007@163.com (X.F.); jingshilei616@163.com (J.L.); yilei112000@163.com (Y.L.); xl200208281553@163.com (L.X.); wqf19118703774@163.com (Q.W.)
[2] Hunan Engineering Research Center of Cyberspace Security Technology and Applications, Hengyang Normal University, Hengyang 421002, China
[3] Hunan Provincial Key Laboratory of Intelligent Information Processing and Application, Hengyang Normal University, Hengyang 421002, China
* Correspondence: hongweid51@hynu.edu.cn

Abstract

Breast ultrasound image segmentation remains challenging because of low contrast, speckle noise, and blurred lesion boundaries. Although Mamba can model long-range dependencies efficiently, directly processing noisy ultrasound features may introduce unreliable contextual responses. Moreover, conventional skip connections can transmit redundant background information and hinder accurate boundary recovery. To address these issues, we propose WMCAF-Net, a Wavelet-Enhanced Mamba and Cross-Scale Adaptive Fusion Network. The Wave-Mamba module decomposes features into low- and high-frequency components using wavelet transform. Low-frequency components are processed by a visual state-space module to capture stable global structures, while high-frequency components are enhanced by directional convolutions to strengthen boundary and texture representation. The Cross-Scale Adaptive Fusion Module improves encoder–decoder feature fusion through channel-response recalibration and adaptive spatial selection. On BUSI, WMCAF-Net achieves 75.47% IoU and 83.47% Dice, surpassing the strongest competing methods by 2.08 and 1.97 percentage points, respectively. Under five-fold cross-validation on Dataset B, it achieves 72.89% IoU and 82.27% Dice, surpassing the strongest competing methods by 1.53 and 2.57 percentage points, respectively. Furthermore, bidirectional cross-dataset evaluation was conducted to analyze the segmentation performance of WMCAF-Net when transferred across different ultrasound datasets.

Keywords: breast ultrasound image; Mamba; wavelet transform; channel–spatial attention

Check for updates

Academic Editor: George A. Tsihrintzis

Received: 17 August 2026
Revised: 26 October 2026
Accepted: 5 November 2026
Published: 8 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Breast tumors are among the most common diseases threatening women's health, and early screening and accurate assessment are of great importance for clinical diagnosis, treatment planning, and prognostic analysis [1,2]. Breast ultrasound (BUS) imaging has been widely used for breast disease screening and auxiliary diagnosis because of its non-ionizing nature, real-time imaging capability, relatively low cost, and good applicability to dense breast tissues [3]. However, owing to the intrinsic imaging mechanism of ultrasound, BUS images often suffer from low contrast, speckle noise, acoustic shadow artifacts, and blurred lesion boundaries [4–6]. In addition, lesion regions and surrounding tissues often

exhibit similar gray-level and texture distributions, while different lesions may vary considerably in size, shape, and boundary regularity, which can lead to under-segmentation, over-segmentation, or boundary deviation [6,7]. Therefore, BUS image segmentation models should be capable of capturing global semantic information while remaining sensitive to local boundaries and fine-grained texture details.

In recent years, deep learning-based methods have achieved remarkable progress in medical image segmentation [8]. Among them, CNN-based encoder–decoder architectures represented by U-Net have been widely adopted. By extracting multi-scale features and using skip connections to fuse shallow spatial details with deep semantic information, U-Net has shown favorable performance in medical image segmentation tasks [9]. Subsequently, improved variants such as UNet++ and ResUNet have further enhanced feature propagation, gradient flow, and representation ability [10,11]. CNN-based methods have strong local feature modeling capability and are helpful for capturing lesion edges, textures, and local morphological patterns. However, convolution operations are mainly constrained by local receptive fields, which limits their ability to model long-range dependencies and global contextual information [12–14]. To enhance global modeling capability, Transformer architectures have been introduced into medical image segmentation tasks, where self-attention mechanisms are used to establish long-range feature dependencies. Representative methods such as TransUNet and TransFuse attempt to combine the local feature extraction ability of CNNs with the global modeling advantage of Transformers, thereby achieving complementary modeling of local details and global semantic information [15,16].

Compared with the computationally expensive self-attention mechanism in Transformers, Mamba-based architectures have emerged as a promising alternative for medical image segmentation due to their linear computational complexity and ability to model long-range dependencies [17–20]. However, directly applying Mamba to BUS features may introduce unreliable contextual responses due to noise and background interference, potentially affecting lesion structure representation and boundary localization [21]. Wavelet transform provides a frequency-aware approach by decomposing features into low-frequency structural components and high-frequency detail components [22]. While low-frequency components facilitate global structural modeling, high-frequency components preserve directional information relevant to lesion boundaries and textures. This complementary representation motivates the integration of wavelet decomposition with Mamba for more effective global and local feature modeling.

However, enhancing global and local representations only at the feature extraction stage is still insufficient for accurate recovery in the decoding stage. In encoder–decoder architectures, skip connections are commonly used to transfer shallow spatial details to the decoder, compensating for information loss caused by downsampling [9,23]. Nevertheless, conventional skip connections usually adopt direct concatenation or addition, without fully considering the differences between encoder and decoder features in spatial distribution and channel response. For BUS images, shallow features often contain speckle noise, acoustic shadow artifacts, and redundant background textures. If these features are directly passed to the decoder without effective selection, they may weaken lesion-related responses and lead to inaccurate boundary reconstruction. Therefore, how to adaptively calibrate feature discrepancies between the encoder and decoder during cross-scale feature fusion, highlight lesion-relevant information, and suppress redundant background interference remains an important problem in BUS image segmentation.

Motivated by the above analysis, we propose WMCAF-Net, a breast ultrasound image segmentation network that integrates wavelet-enhanced Mamba with cross-scale adaptive feature fusion. Specifically, we design a Wave-Mamba module that employs wavelet decomposition to separate low-frequency structural components from high-frequency detail

components. The low-frequency features are processed by a visual state-space module (VSS) to capture long-range dependencies and enhance global structural representations, while the high-frequency subbands are processed by horizontal–vertical cross convolution (HVC) and diagonal cross convolution (DCC) to strengthen directional boundary and texture representations. Furthermore, we introduce a Cross-Scale Adaptive Fusion Module (CAFM) into the skip connections, which combines multi-pooling channel attention and dynamic spatial attention to recalibrate channel responses and adaptively select spatial information from encoder and decoder features. By integrating frequency-specific feature modeling with adaptive cross-scale fusion, WMCAF-Net aims to improve lesion structure representation and boundary reconstruction in BUS images.

The main contributions of this work are summarized as follows:

- We propose WMCAF-Net, a wavelet-guided visual Mamba architecture that integrates CNNs, Mamba, and wavelet transform for effective BUS image segmentation.
- We design a Wave-Mamba module that performs global modeling on low-frequency components and directional feature enhancement on high-frequency components, improving the comprehensive representation of global semantics and local details.
- We develop a CAFM module that optimizes encoder–decoder feature fusion through multi-pooling-based channel recalibration and dynamic spatial attention, reducing information redundancy and improving detail recovery.

2. Related Works

2.1. CNN-Based Breast Ultrasound Image Segmentation

Numerous convolutional neural network (CNN)-driven methods have been proposed for breast ultrasound (BUS) image segmentation. Among these, U-Net [9] is widely regarded as a foundational architecture, as its symmetric encoder–decoder design combined with skip connections supports multi-scale feature fusion while maintaining spatial information.

Based on the U-Net architecture, numerous extensions have been explored to cope with severe noise contamination and heterogeneous lesion appearances in BUS images. For instance, UNeXt [24] enhances computational efficiency by incorporating lightweight multi-layer perceptron (MLP) modules without compromising segmentation accuracy, making it better suited for modeling complex ultrasound features. Rolling-UNet [25] focuses on segmentation refinement by iteratively updating prediction results, which improves robustness to irregular lesion shapes and indistinct boundaries. CMUNeXt [26] emphasizes contextual modeling and multi-scale feature interaction, facilitating more discriminative feature extraction under complex background conditions. In addition, DCSAU-Net [27] adopts dense connectivity and context-aware designs to encourage feature reuse and enlarge the effective receptive field, thereby enhancing segmentation stability in noisy ultrasound environments.

Although these methods differ in architectural design and feature fusion strategies, their reliance on convolution-based operations may limit their ability to capture long-range dependencies and global contextual information.

2.2. Mamba-Based Breast Ultrasound Image Segmentation

Owing to the limited ability of convolutional neural networks (CNNs) to capture long-range dependencies, Vision Transformers (ViTs) have gained increasing attention in medical image segmentation by enabling explicit global context modeling through self-attention [12,28]. This development has led to a variety of hybrid CNN–Transformer architectures. Representative examples include TransUNet [15], which embeds self-attention modules into the U-Net encoder to strengthen global representation learning, and Trans-

Fuse [16], which adopts a parallel fusion paradigm to jointly exploit CNN and Transformer branches for local detail modeling and global semantic perception, thereby maintaining a favorable trade-off between segmentation performance and computational cost.

Nevertheless, self-attention requires exhaustive pairwise interactions over global features, resulting in substantial computational cost and memory consumption when processing high-resolution medical images [29]. Such limitations restrict the practical deployment of Transformer-based models in real-world clinical environments. To better reconcile global modeling capability with computational efficiency, the Mamba [17] architecture based on state space models (State Space Model, SSM) has recently attracted increasing attention. By leveraging the Visual State Space [30] (VSS) mechanism, Mamba enables efficient modeling of long-range dependencies while maintaining linear computational complexity, making it well suited for global context modeling in high-resolution medical image segmentation [31,32].

Building upon this paradigm, several improved methods have been proposed. VMC-UNet [33] integrates Mamba into the U-Net architecture by embedding visual state space modules within residual connections, enabling long-range dependency modeling on convolutional features while preserving local details and enhancing global contextual representations. However, this method primarily focuses on the integration of spatial feature modeling and state space modules, rather than explicitly exploiting frequency-domain information. FMamba [34] introduces a frequency enhancement module into the deep feature extraction stages, where parallel VSS and frequency enhancement branches are employed to jointly strengthen global feature representations. This approach further demonstrates the complementary roles of frequency-domain features and state space modeling. However, for breast ultrasound images, different frequency components play distinct roles in representing global structures, lesion boundaries, and local textures. Therefore, how to further exploit these differences for more targeted feature modeling remains an important research direction. DD-Mamba [35] adopts a different strategy by constructing a dual-domain collaborative modeling framework that integrates spatial-domain and frequency-domain representations, employing visual state space modules in both domains to capture global dependencies. In addition, this method applies the Fast Fourier Transform (FFT) to map features into the frequency domain, where high- and low-frequency information is processed separately using high-pass and low-pass filters together with learnable complex-valued weights to enhance the representation of global structures and local details. This approach demonstrates that differentiated processing of frequency components can facilitate complementary representations of global structures and local details, providing inspiration for further exploring wavelet-based frequency decomposition and feature modeling.

In summary, existing studies have provided important foundations for medical image segmentation through global dependency modeling, frequency-domain feature enhancement, and spatial-frequency collaborative modeling. These studies demonstrate that combining frequency-domain information with state space modeling can facilitate complementary representations of global structures and local details. Inspired by these studies, we further investigate wavelet-based frequency decomposition and differentiated feature modeling. Unlike approaches that primarily rely on parallel frequency-domain enhancement or Fourier-based dual-domain collaborative modeling, our method employs wavelet transform to decompose features into low-frequency structural components and high-frequency detail components with different directional characteristics, followed by targeted feature modeling according to their respective properties. Furthermore, a cross-scale adaptive feature fusion mechanism is incorporated to improve feature interactions between the encoder and decoder, thereby enhancing lesion structure representation and boundary recovery in complex breast ultrasound images.

2.3. Wavelet Transform for Medical Image Processing

Wavelet transform (WT) [36] is a classical multi-resolution analysis technique that decomposes image features into components at different scales and frequency bands, enabling simultaneous modeling of global structural information and local detail features within a unified framework. In general, low-frequency components primarily capture overall contours and regional structural information, while high-frequency components emphasize edges, textures, and fine-grained details [37]. In medical image segmentation, low-frequency components capture global structural characteristics, while high-frequency components capture fine-grained patterns and localized details. As a result, incorporating wavelet transform into deep learning models has become an important research direction for improving segmentation performance in recent years [38–40].

In the field of breast ultrasound image segmentation, several studies have incorporated wavelet transforms into network architectures as frequency-domain priors or feature enhancement mechanisms. HCMNet [41] introduces a wavelet-based feature extraction module to explicitly decompose features into different frequency components and integrates low-frequency structural information with high-frequency details during encoding, thereby alleviating the influence of noise interference on lesion structure representation. This approach highlights the role of wavelet decomposition in facilitating complementary representations of structural information and fine-grained details, and provides inspiration for further exploring frequency-specific feature modeling. FABRF-Net [42] adopts a frequency-aware feature modeling strategy, employing wavelet decomposition to extract multi-scale frequency-domain features and coordinating boundary and region branches to enhance lesion boundary details and regional consistency. This approach demonstrates that integrating frequency-domain information with boundary and region modeling can improve lesion contour delineation. Building upon this perspective, further exploration of the directional structural information contained in different high-frequency wavelet components may contribute to more targeted modeling of local boundaries and fine-grained textures. UWT-Net [43] combines wavelet transforms with convolutional operations through wavelet convolution (WTConv), leveraging frequency features at different scales to expand the effective receptive field. By integrating wavelet convolution with standard convolution, it further enhances structural feature representations. This approach offers an alternative strategy for combining wavelet transforms with convolutional operations and further demonstrates the potential of frequency-domain features for multi-scale information extraction.

In summary, existing studies have explored the application of frequency-domain information to medical image segmentation through wavelet-based feature fusion, frequency-aware collaborative boundary and region modeling, and wavelet convolution [44,45]. Inspired by these studies, we further focus on the directional structural characteristics of different high-frequency components obtained through wavelet decomposition. Unlike approaches that primarily enhance feature representations through frequency-domain feature fusion, collaborative boundary and region modeling, or multi-scale wavelet convolution, our method emphasizes targeted enhancement of horizontal, vertical, and diagonal directional information within high-frequency wavelet components. This is further complemented by global modeling of low-frequency structural features to improve local detail representation.

3. Method

3.1. Architecture Overview

As illustrated in Figure 1, the proposed WMCAF-Net adopts a symmetric encoder–decoder architecture similar to U-Net. The entire network consists of five hierarchical stages, with skip connections employed to enable cross-level feature fusion between the encoder and the decoder. During downsampling, max pooling operations are applied, while bilinear

interpolation is used in the upsampling stage to ensure smooth feature recovery. The numbers of channels across the five stages are set to {32, 64, 128, 256, 512}, respectively.

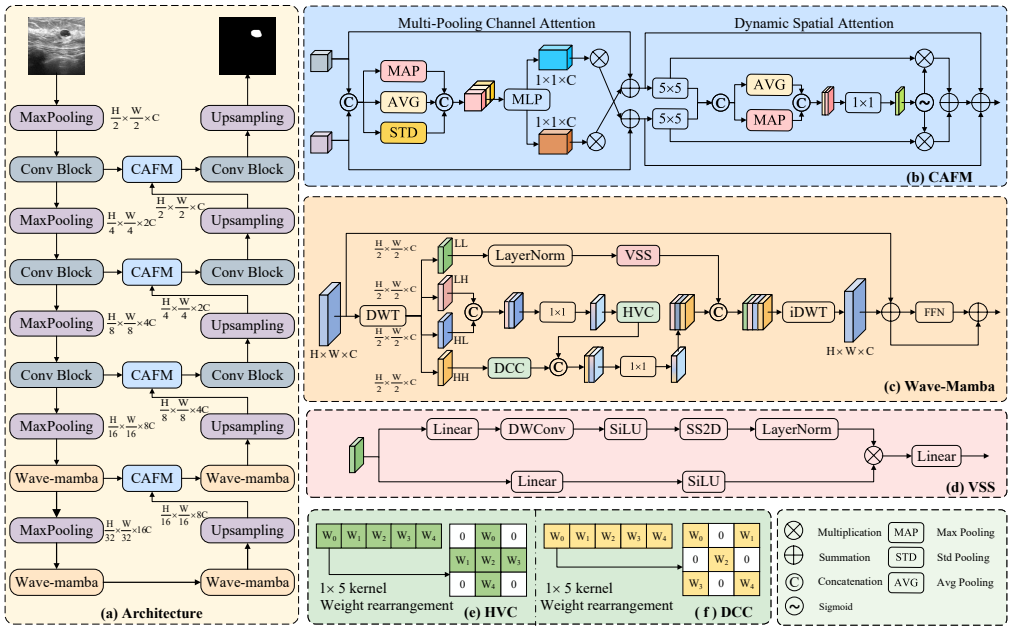


Figure 1. (**a**) Overall architecture of WMCAF-Net. (**b**) CAFM: Cross-Scale Adaptive Fusion Module, composed of multi-pooling channel attention and dynamic spatial attention. (**c**) Wave-Mamba: DWT-based low-frequency and high-frequency decomposition, VSS global modeling on low-frequency features, and directional enhancement on high-frequency details, followed by IDWT reconstruction. (**d**) VSS block for efficient long-range dependency modeling. (**e**) HVC: horizontal–vertical cross convolution for enhancing horizontal frequency and vertical high-frequency structural features. (**f**) DCC: diagonal cross convolution for enhancing diagonal high-frequency structural features.

Specifically, the architecture is organized as follows:

1. The first three stages adopt standard convolutional blocks to capture shallow semantic representations and fine-grained local features.
2. The fourth and fifth stages incorporate the proposed wavelet-enhanced Mamba (Wave-Mamba) Block. This block decompose features into low-frequency and high-frequency components via discrete wavelet transform (DWT). The low-frequency components are used for global dependency modeling, while the high-frequency components are enhanced using horizontal–vertical convolutions and diagonal convolutions to better capture directional structural details and local boundary information. In the decoder, features at each stage are upsampled and fused with the corresponding encoder features through skip connections.
3. To enhance the representational capacity of skip connections, a Cross-Scale Adaptive Fusion Module (CAFM) is introduced. This module employs a joint spatial channel attention mechanism to dynamically recalibrate the relative importance of encoder and decoder features within skip connections, thereby strengthening the discriminability of the fused representations. Through progressive decoding and feature reconstruction across multiple scales, high-resolution segmentation predictions are ultimately obtained.

3.2. Wave-Mamba Block Module

The proposed Wave-Mamba module integrates wavelet decomposition, visual state-space modeling, and directional convolutions to jointly capture global structural information

and local details. Specifically, low-frequency components are processed by the visual state-space module (VSS) for long-range dependency modeling, while high-frequency subbands are enhanced through directional convolutions to strengthen boundary and texture representations.

As illustrated in Figure 1c, given an input feature map X, a discrete wavelet transform (DWT) is first applied to decompose it into four subbands, including one low-frequency subband X_{ll} and three high-frequency subbands X_{lh}, X_{hl}, and X_{hh}. The low-frequency subband X_{ll} mainly encodes global structural and semantic information. Therefore, in the low-frequency branch, X_{ll} is first normalized by LayerNorm and then fed into the selective state space module (VSS), as shown in Figure 1d, to obtain the low-frequency enhanced feature F_L, which effectively captures long-range dependencies, strengthens the representation of global morphology and contextual relationships, and suppresses the interference of noise and artifacts on global feature modeling. These processes can be represented by Equations (1) and (2).

$$X_{ll}, X_{lh}, X_{hl}, X_{hh} = \mathrm{DWT}(X), \tag{1}$$

$$F_L = \mathrm{VSS}(\mathrm{LayerNorm}(X_{ll})). \tag{2}$$

where X_{ll}, X_{lh}, X_{hl}, and X_{hh} denote the low-frequency subband and the high-frequency subbands in the horizontal, vertical, and diagonal directions obtained by the discrete wavelet transform (DWT), $\mathrm{LayerNorm}(\cdot)$ denotes the layer normalization operation, $\mathrm{VSS}(\cdot)$ denotes the selective state space module, and F_L denotes the low-frequency enhanced feature.

In the high-frequency branch, the horizontal high-frequency subband X_{lh} and the vertical high-frequency subband X_{hl} are first concatenated along the channel dimension and then compressed using a 1×1 convolution to generate a joint high-frequency feature map X_{HV}, which encodes complementary horizontal and vertical directional information. The resulting X_{HV} is then fed into the horizontal–vertical cross convolution (HVC) in Figure 1e to produce the direction-enhanced feature F_{HV}. Specifically, the HVC is initially parameterized using a 1×5 convolution kernel, which is subsequently rearranged to form a sparse 3×3 kernel with a cross-shaped structure, thereby explicitly strengthening horizontal and vertical responses and enabling effective modeling of directional local details. Meanwhile, the diagonal high-frequency subband X_{hh} is processed by the diagonal cross convolution (DCC) Figure 1f branch to obtain the diagonal high-frequency feature F_{DC}. Similarly, the DCC also adopts a 1×5 kernel and rearranges its weights to generate a diagonally sparse 3×3 kernel, thereby enhancing diagonal directional responses. Subsequently, F_{HV} and F_{DC} are concatenated along the channel dimension and fused via a 1×1 convolution to yield the aggregated high-frequency feature map F_H, which integrates multi-directional high-frequency information. During backpropagation, the sparse 3×3 kernels are constructed from the five learnable parameters through differentiable tensor concatenation and reshaping, allowing gradients to be automatically propagated to the corresponding parameters by standard automatic differentiation. The fixed zero entries are non-trainable; therefore, no additional gradient-transfer mechanism is required. These processes can be represented by Equations (3) to (6).

$$X_{HV} = \mathrm{Conv}_{1\times1}(\mathrm{Concat}(X_{lh}, X_{hl})), \tag{3}$$

$$F_{HV} = \mathrm{HVC}(X_{HV}), \tag{4}$$

$$F_{DC} = \mathrm{DCC}(X_{hh}), \tag{5}$$

$$F_H = \mathrm{Conv}_{1\times1}(\mathrm{Concat}(F_{HV}, F_{DC})). \tag{6}$$

where X_{HV} is the fused horizontal–vertical high-frequency feature; HVC$(\cdot)$ denotes the horizontal–vertical cross convolution block; DCC$(\cdot)$ denotes the diagonal cross convolution block; and F_H is the fused multi-directional high-frequency feature.

After high-frequency and low-frequency feature modeling, the low-frequency feature F_L and the high-frequency feature F_H are concatenated in the wavelet domain and transformed back to the spatial domain through an inverse discrete wavelet transform (IDWT), resulting in a spatial feature map Y that combines global structural information with fine-grained local details. To preserve the original input information and stabilize training, a residual connection is established between Y and the input feature map X. The fused features are subsequently normalized using LayerNorm to improve numerical stability and adaptability across different feature scales, and are finally refined by a feed-forward network (FFN$(\cdot)$) to produce the output $\hat{Y}$ of the Wave-Mamba module. These processes can be represented by Equations (7) and (8).

$$Y = \text{IDWT}(\text{Concat}(F_L, F_H)), \tag{7}$$

$$\hat{Y} = \text{FFN}(\text{LayerNorm}(X + Y)). \tag{8}$$

where F_L denotes the low-frequency enhanced feature, F_H is the fused multi-directional high-frequency feature, and IDWT$(\cdot)$ is the inverse discrete wavelet transform.

3.3. Cross-Scale Adaptive Fusion Module

As illustrated in Figure 1b, the proposed Cross-Scale Adaptive Fusion Module (CAFM) consists of multi-pooling channel attention and dynamic spatial attention to facilitate adaptive fusion of encoder and decoder features. Specifically, multi-pooling channel attention utilizes complementary channel-wise statistics to recalibrate feature responses across the two branches, while dynamic spatial attention selectively emphasizes informative spatial regions during feature aggregation. Through the coordinated channel and spatial recalibration, CAFM enhances cross-scale feature interaction and supports effective detail reconstruction in the decoder.

3.3.1. Multi-Pooling Channel Attention

In this module, feature maps X_1 and X_2 are derived from the encoder and decoder, respectively, and are jointly fed into the Multi-Pooling Channel Attention module. First, X_1 and X_2 are concatenated in the channel direction to form the combined feature representation X'. Subsequently, global average pooling, global max pooling, and global standard deviation pooling are applied to X' to extract complementary channel-wise statistics. Specifically, average pooling captures the overall channel response, max pooling emphasizes salient activations, and standard deviation pooling reflects the spatial dispersion of responses, enabling a more comprehensive evaluation of channel importance.

The resulting statistical descriptors are concatenated to form the channel-wise descriptor Y, which is subsequently processed by a multi-layer perceptron (MLP$(\cdot)$) yielding an intermediate channel weight representation Y'. The output Y' is further split into two sets of channel-wise weights, W_1 and W_2, which serve as channel correction coefficients for the two feature branches. The resulting weight tensor has a shape of $[B, 2C, 1, 1]$. Finally, channel interaction correction is performed in a cross-branch manner: W_1 is used to reweight X_2 and fed back to X_1, while W_2 is applied to reweight X_1 and fed back to X_2. This cross-branch channel interaction mechanism effectively models the complementary relationship between encoder and decoder features, suppresses redundant responses, and enhances informative features, thereby improving the quality of subsequent feature fusion. The complete formulation of this procedure can be expressed as follows:

$$X' = \text{Concat}(X_1, X_2) \tag{9}$$

$$Y = \text{Concat}(\text{Avg}(X'), \text{Max}(X'), \text{Std}(X')) \tag{10}$$

$$Y' = \text{MLP}(Y) \tag{11}$$

$$W_1, W_2 = \text{Split}(Y') \tag{12}$$

$$X_1 = X_1 \times W_2 + X_1 \tag{13}$$

$$X_2 = X_2 \times W_1 + X_2 \tag{14}$$

where X_1 and X_2 are feature maps from the encoder and decoder, respectively; $\text{Avg}(\cdot)$, $\text{Max}(\cdot)$, and $\text{Std}(\cdot)$ denote global average pooling, global max pooling, and global standard deviation pooling, respectively; Y denotes the channel-wise statistical descriptor; W_1 and W_2 are channel weights used for cross-branch channel calibration; and $\times$ denotes element-wise multiplication.

3.3.2. Dynamic Spatial Attention

The input of the Dynamic Spatial Attention (DSA) module is derived from the output features X_1 and X_2 of the Multi-Pooling Channel Attention module. First, 5×5 convolution is applied to each feature branch to perform preliminary local context modeling, yielding the convolved feature representations X_1' and X_2'. Subsequently, X_1' and X_2' are merged via channel-wise concatenation, producing an integrated feature representation Y that combines spatial information from both the encoder and decoder.

To obtain adaptive spatial weights, average pooling and max pooling are separately applied to Y to extract complementary spatial descriptors. These descriptors are then concatenated and mapped through a convolution followed by a sigmoid activation to a generate spatial attention map, denoted as W, which characterizes the importance of different spatial locations. Finally, W is applied to X_1' and X_2', respectively, via element-wise modulation, and the modulated features are aggregated to produce the final output of the DSA module. This process enhances responses in salient regions while suppressing irrelevant spatial interference, thereby improving spatial consistency and effective information transmission during cross-scale feature fusion, which is beneficial for subsequent feature recovery in the decoding stage. The complete formulation of this procedure can be expressed as follows:

$$X_1' = \text{Conv}(X_1), \tag{15}$$

$$X_2' = \text{Conv}(X_2), \tag{16}$$

$$Y = \text{Concat}(X_1', X_2'), \tag{17}$$

$$W = \text{Sigmoid}(\text{Conv}(\text{Concat}(\text{Mean}_c(Y), \text{Max}_c(Y)))) \tag{18}$$

$$O = X_1' \times W + X_2' \times W. \tag{19}$$

where X_1 and X_2 denote the two input feature maps obtained from the Multi-Pooling Channel Attention module; $\text{Mean}(\cdot)$ and $\text{Max}(\cdot)$ denote average pooling and max pooling along the channel dimension, respectively; W denotes the spatial attention weight map; $\times$ denotes element-wise multiplication; and O denotes the final fused output feature map of the DSA module.

4. Experiments and Results

In this section, we first provide a description of the two publicly available datasets used in our experiments. Subsequently, the hyperparameter settings employed during the training process are presented in detail. We then describe the quantitative evaluation metrics used to assess and compare different methods. Finally, we present quantitative results for each dataset and conduct qualitative analyses.

4.1. Datasets

We evaluate the segmentation performance of the proposed method on two publicly available datasets: BUSI and Dataset B. For the BUSI dataset, the 647 benign and malignant images are randomly divided at the image level into training, validation, and test sets with a ratio of 7:2:1 using a fixed random seed of 41. The resulting data partition is kept unchanged across all repeated experiments to ensure consistent evaluation. Considering the relatively small sample size of Dataset B, five-fold cross-validation is adopted instead of a single random split. Specifically, Dataset B is divided into five mutually exclusive folds, with each fold used once as the independent test set while the remaining data are used for model training and validation. All compared methods use the same five-fold partitions to ensure a fair comparison.

BUSI: The BUSI dataset [46] consists of breast ultrasound images acquired in 2018 using LOGIQ E9 and LOGIQ E9 Agile scanners. The images were collected from Baheya Hospital in Cairo, Egypt, and include data from 600 female patients aged between 25 and 75 years. The dataset contains a total of 780 images with an average resolution of approximately 500 × 500 pixels, including 133 normal images, 437 benign images, and 210 malignant images. For the segmentation task, only the 647 images corresponding to benign and malignant cases are used in this study. The dataset is publicly available at https://www.kaggle.com/datasets/sabahesaraki/breast-ultrasound-images-dataset (accessed on 4 November 2026).

Dataset B: Dataset B [47] comprises 163 breast ultrasound images acquired using a Siemens ACUSON Sequoia C512 scanner, including 109 benign images and 54 malignant images. The dataset is publicly available at https://github.com/openmedlab/Awesome-Medical-Dataset/blob/main/resources/Breast_Ultrasound_DatasetB.md (accessed on 4 November 2026).

4.2. Implementation Details

All experiments in this study were conducted on a workstation equipped with a 12th Gen Intel Core i5-12600KF CPU and an NVIDIA GeForce RTX 4090D GPU with 24 GB of VRAM. The software environment was based on Python 3.9 and the PyTorch 1.13 deep learning framework. The Adam optimizer was used for network training. During the training process, the model checkpoint with the lowest validation loss was selected as the optimal model and used for final evaluation. For experiments on the BUSI dataset, the data partition generated using the fixed splitting seed of 41 is kept unchanged across all repeated experiments. Each experimental configuration is independently trained three times using training random seeds of 41, 42, and 43, respectively. The quantitative results are reported as the mean ± standard deviation over these three independent runs. For Dataset B, five-fold cross-validation is adopted, and the final results are reported as the mean ± standard deviation across the five test folds.

All input images and corresponding segmentation masks are resized to 256 × 256 pixels using direct resizing. The original aspect ratio is not preserved, and no cropping or padding is performed. Bilinear interpolation is used for image resizing, and nearest-neighbor interpolation is used for mask resizing. The input images are normalized using the Normalize operation in Albumentations. Data augmentation is applied only to the training set, including random 90° rotation, random horizontal flipping, and random vertical flipping. The same geometric transformations are applied to the images and their corresponding masks. For the validation and test sets, only resizing and normalization are performed. The model is trained for 300 epochs using the following hyperparameter configuration. The initial learning rate is fixed at 0.0003, with a weight decay of 0.00001 and Adam coefficients of $\beta_1 = 0.9$ and $\beta_2 = 0.999$, while the batch size is set to 8. Moreover, a

composite loss function, denoted as L_{BCEDice}, is employed by jointly combining the Dice loss (L_{Dice}) and the binary cross-entropy loss (L_{BCE}). The overall loss function is defined as:

$$L_{\text{BCEDice}}(y, \hat{y}) = 0.5 \times L_{\text{BCE}}(y, \hat{y}) + L_{\text{Dice}}(y, \hat{y}), \tag{20}$$

where y denotes the ground truth label and $\hat{y}$ represents the model prediction.

4.3. Evaluation Metrics

To quantitatively evaluate segmentation performance, five commonly used metrics are adopted, including Accuracy (Acc), Intersection over Union (IoU), Dice Similarity Coefficient (Dice), Sensitivity (SE), and Precision (PC). Their mathematical formulations are provided as follows:

$$IoU = \frac{TP}{TP + FP + FN} \tag{21}$$

$$Dice = \frac{2TP}{2TP + FP + FN} \tag{22}$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{23}$$

$$SE = \frac{TP}{TP + FN} \tag{24}$$

$$PC = \frac{TP}{TP + FP} \tag{25}$$

where TP, TN, FP, and FN denote the numbers of true positives, true negatives, false positives, and false negatives, respectively. All evaluation metrics are calculated at the individual-image level and then averaged over the corresponding test set or test fold.

4.4. Experimental Results

To comprehensively evaluate the proposed WMCAF-Net, we conduct comparative experiments with eleven representative segmentation methods selected based on architectural representativeness, methodological diversity, and relevance to breast ultrasound image segmentation. Specifically, the selected methods include CNN-based architectures UNet [9], UNeXt [24], Rolling-U-Net [25], CMUNeXt [26], and DCSAU-Net [27]; the hybrid CNN–Transformer architecture TransFuse [16]; Mamba-based architectures VMC-UNet [33], FMamba [34], and DD-Mamba [35]; and methods incorporating wavelet transforms or frequency-domain feature modeling, UWT-Net [43] and HCMNet [41]. These methods cover diverse technical approaches, including local feature extraction, global contextual modeling, and frequency-domain feature enhancement, providing a representative basis for evaluating the segmentation performance of WMCAF-Net. All baseline models are implemented using publicly available open-source code and trained and evaluated under consistent experimental settings to ensure a fair comparison.

4.4.1. Quantitative Evaluation

Table 1 reports the experimental results of WMCAF-Net on the BUSI dataset. WMCAF-Net achieves superior performance across multiple evaluation metrics, attaining an IoU of 75.47%, a Dice score of 83.47%, and an accuracy of 96.80%, which are the highest among all compared methods. In particular, compared with FMamba, WMCAF-Net improves the IoU, Dice score, and accuracy by 3.73, 3.28, and 0.48 percentage points, respectively, further validating the superiority of the proposed approach. Moreover, WMCAF-Net achieves the best IoU, Dice, and ACC while maintaining competitive SE and PC.

Table 2 reports the five-fold cross-validation results on Dataset B. WMCAF-Net achieves the highest IoU of 72.89% and SE of 85.42% among the compared methods, while obtaining competitive Dice, PC, and ACC scores of 82.27%, 83.79%, and 98.37%, respectively.

Compared with HCMNet, WMCAF-Net achieves higher IoU and SE while requiring substantially fewer floating-point operations, with 1.29 G FLOPs compared with 48.82 G FLOPs for HCMNet. These results demonstrate that WMCAF-Net maintains strong overall segmentation performance under the five-fold cross-validation protocol while requiring relatively low theoretical floating-point operations.

Table 1. Segmentation performance comparison on the BUSI dataset (mean ± standard deviation).

Model	FLOPs (G)	Params (M)	IoU (%)	Dice (%)	SE (%)	PC (%)	ACC (%)
UNet [9]	10.13	4.32	69.89 ± 1.53	79.69 ± 1.15	83.07 ± 2.18	81.52 ± 1.96	96.50 ± 0.09
UNeXt [24]	0.57	1.47	64.46 ± 1.03	77.48 ± 0.85	77.50 ± 1.48	78.52 ± 1.15	96.14 ± 0.09
DCSAU-Net [27]	6.92	2.60	62.68 ± 0.51	70.75 ± 0.32	76.30 ± 1.61	72.06 ± 0.80	94.54 ± 0.23
Rolling-Unet [25]	2.102	1.782	67.33 ± 0.45	79.62 ± 0.48	78.51 ± 0.43	82.26 ± 0.85	96.54 ± 0.04
CMUNeXt [26]	5.679	3.149	71.27 ± 0.40	80.17 ± 0.25	82.21 ± 1.02	83.70 ± 0.46	96.63 ± 0.09
TransFuse [16]	8.65	26.17	72.11 ± 0.63	81.50 ± 0.49	**86.60 ± 1.02**	80.38 ± 1.43	96.40 ± 0.32
UWT-Net [43]	8.37	7.61	66.34 ± 1.06	79.56 ± 0.73	76.23 ± 1.04	83.60 ± 0.36	94.63 ± 0.16
VMC-UNet [33]	6.55	23.65	69.43 ± 0.36	79.34 ± 0.70	83.27 ± 2.15	80.61 ± 2.10	96.48 ± 0.08
FMamba [34]	1.93	5.91	71.74 ± 1.10	80.19 ± 0.82	83.50 ± 2.14	81.96 ± 2.17	96.32 ± 0.38
DD-Mamba [35]	3.81	16.26	71.71 ± 0.08	79.80 ± 0.88	83.18 ± 0.76	81.69 ± 0.09	96.04 ± 0.28
HCMNet [41]	48.82	26.97	73.39 ± 0.70	80.91 ± 0.93	84.60 ± 0.55	80.33 ± 1.71	96.10 ± 0.25
Ours	1.29	10.77	**75.47 ± 0.26**	**83.47 ± 0.12**	85.73 ± 0.27	**85.30 ± 0.60**	**96.80 ± 0.18**

Note: Bold values indicate the best performance for each evaluation metric among all compared methods.

Table 2. Segmentation performance comparison on Dataset B using five-fold cross-validation (mean ± standard deviation across the five folds).

Model	FLOPs (G)	Params (M)	IoU (%)	Dice (%)	SE (%)	PC (%)	ACC (%)
UNet [9]	10.13	4.32	68.93 ± 3.05	77.59 ± 3.01	79.78 ± 4.86	80.09 ± 1.69	98.35 ± 0.43
UNeXt [24]	0.57	1.47	62.57 ± 2.60	71.72 ± 3.13	71.87 ± 4.64	82.85 ± 1.33	98.21 ± 0.34
DCSAU-Net [27]	6.92	2.60	68.03 ± 8.55	77.70 ± 7.44	79.44 ± 6.53	80.70 ± 6.60	98.18 ± 0.82
Rolling-Unet [25]	2.102	1.782	65.25 ± 5.00	74.54 ± 4.26	81.76 ± 2.58	74.88 ± 4.41	98.21 ± 0.41
CMUNeXt [26]	5.679	3.149	69.27 ± 2.31	78.08 ± 2.29	78.78 ± 4.11	83.08 ± 2.97	98.35 ± 0.31
TransFuse [16]	8.65	26.17	71.36 ± 0.88	79.67 ± 0.81	85.19 ± 3.39	78.26 ± 1.75	**98.57 ± 0.36**
UWT-Net [43]	8.37	7.61	54.98 ± 3.57	66.42 ± 3.65	75.19 ± 5.26	66.74 ± 4.75	97.73 ± 0.48
VMC-UNet [33]	6.55	23.65	67.41 ± 3.84	78.22 ± 3.87	77.44 ± 4.02	83.33 ± 3.96	98.38 ± 0.45
FMamba [34]	1.93	5.91	56.81 ± 1.40	67.83 ± 1.05	77.73 ± 3.33	69.64 ± 2.88	97.65 ± 0.26
DD-Mamba [35]	3.81	16.26	63.47 ± 2.87	73.23 ± 2.34	78.98 ± 2.86	76.89 ± 4.20	98.12 ± 0.44
HCMNet [41]	48.82	26.97	70.96 ± 0.58	79.70 ± 0.30	83.35 ± 2.25	81.16 ± 2.41	98.48 ± 0.26
Ours	1.29	10.77	**72.89 ± 2.17**	**82.27 ± 2.04**	**85.42 ± 3.14**	**83.79 ± 3.88**	98.37 ± 0.36

Note: Bold values indicate the best performance for each evaluation metric among all compared methods.

Furthermore, to evaluate the practical inference efficiency of WMCAF-Net, we measured its inference latency, throughput, and peak GPU memory usage on an NVIDIA GeForce RTX 4090D GPU. The results show that WMCAF-Net achieves an inference latency of 6.51 ms, a throughput of 174.49 images/s, and a peak GPU memory usage of 210.19 MiB, indicating its potential for efficient breast ultrasound image segmentation.

4.4.2. Qualitative Evaluation

Figure 2 provides visualizations of the segmentation results on the BUSI. Each sample is arranged in multiple columns: the first column corresponds to the original ultrasound image, the second column shows the ground truth annotation, the third column presents the segmentation result produced by WMCAF-Net, and the remaining columns display the predictions generated by representative comparison methods.

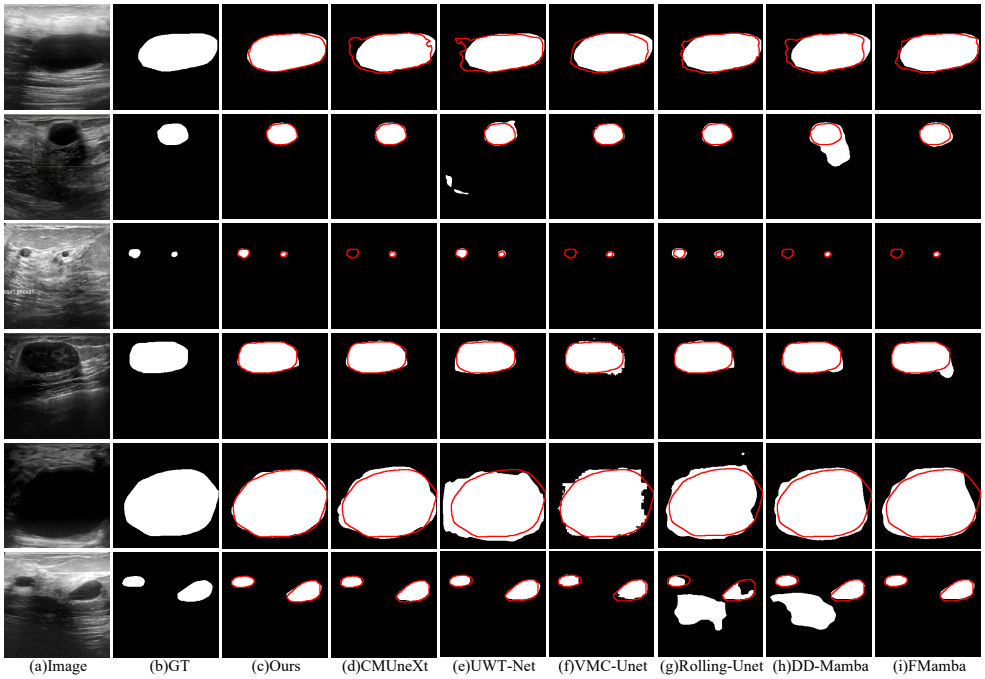


Figure 2. Visualization of segmentation results on BUSI. (**a**) Original ultrasound images; (**b**) ground truth; (**c**) results of our proposed WMCAF-Net; (**d**–**i**) results predicted by CMUNeXt, UWT-Net, VMC-UNet, Rolling-UNet, DD-Mamba, and FMamba, respectively. White regions represent the predicted segmentation masks, while red contours indicate the lesion boundaries in the ground truth.

From the visual comparisons, it can be observed that WMCAF-Net is able to more accurately delineate the overall morphology of lesion regions in most samples, while maintaining better regional integrity and consistency under complex background conditions. In contrast, some competing methods tend to suffer from missed detections, false positives, or boundary deviations, particularly in cases involving small lesions, adjacent lesion separation, or ambiguous boundaries. These observations indicate that the proposed method achieves more accurate and spatially consistent lesion segmentation in the evaluated BUSI cases.

Figure 3 presents visualizations of the segmentation results on Dataset B. Each row presents a set of representative test samples, including the original ultrasound image, the ground truth, as well as the predictions produced by the proposed method and several competing approaches.

From the visualization results, it can be observed that WMCAF-Net is able to accurately localize small-scale lesions in most cases while maintaining the integrity and stability of the predicted regions. In contrast, some competing methods tend to produce incomplete lesion regions, local missed detections, or false positives in low-contrast or weak-boundary scenarios. Notably, even for lesions with small sizes or irregular shapes, the proposed method still generates segmentation results that are more consistent with the ground truth, indicating stronger adaptability to complex backgrounds and fine-grained structures. These results further demonstrate that the proposed method can effectively handle challenging segmentation cases in Dataset B.

Although WMCAF-Net does not achieve the highest value for every individual metric, it achieves strong IoU and Dice performance on both datasets, indicating favorable overall region-overlap quality and segmentation consistency. Different evaluation metrics reflect different prediction tendencies of segmentation models. A higher SE usually indicates stronger lesion recall capability, but it may also be accompanied by more false-positive predictions, leading to a certain degree of over-segmentation. TransFuse shows prominent

performance in terms of SE; however, its relatively lower PC, IoU, and Dice scores suggest that this method may tend to enlarge the predicted regions. In contrast, a higher PC indicates stronger false-positive suppression, but it may also result in missed lesion regions or incomplete segmentation. Combined with the visualization results in Figure 3, it can be observed that although VMC-UNet achieves a relatively high PC, it still suffers from insufficient lesion-region coverage in some samples. Therefore, the superiority of a segmentation model should not be judged solely by a single metric. Compared with methods that achieve the highest SE or PC alone, WMCAF-Net provides a more balanced trade-off between lesion-region coverage and false-positive suppression, as further demonstrated by its consistently higher IoU and Dice scores. These results indicate that, under complex breast ultrasound imaging conditions, the proposed method can achieve more reliable, stable, and spatially consistent lesion segmentation.

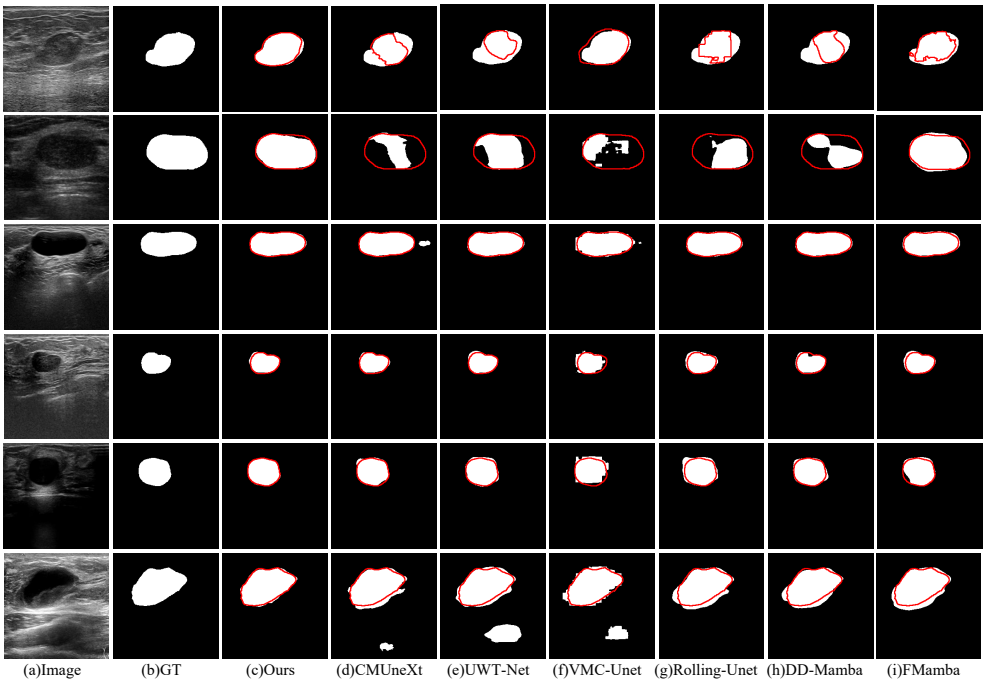


Figure 3. Visualization of segmentation results on Dataset B. (**a**) Original ultrasound images; (**b**) ground truth; (**c**) results of the proposed WMCAF-Net; (**d**–**i**) results predicted by CMUNeXt, UWT-Net, VMC-UNet, Rolling-UNet, DD-Mamba, and FMamba, respectively. Predicted lesion regions are shown in white, while red contours delineate the ground truth boundaries.

4.4.3. Statistical Significance Analysis

To further assess the statistical reliability of the performance differences between WMCAF-Net and competing methods, image-level statistical significance analyses were conducted using the Wilcoxon signed-rank test. IoU and Dice were emphasized as the primary segmentation metrics, as they directly quantify the overlap between predicted lesion regions and ground-truth annotations. Sensitivity (SE), precision (PC), and accuracy (ACC) were considered supplementary evaluation metrics, with their statistical comparisons interpreted as exploratory analyses to provide additional insights into segmentation performance. To account for multiple pairwise comparisons, the Holm correction was applied separately to each metric within each dataset. Both raw p-values and Holm-adjusted p-values are reported in Tables 3 and 4. Statistical significance was determined at a Holm-adjusted p-value threshold of 0.05.

Table 3. Statistical significance analysis of WMCAF-Net against competing methods on the BUSI dataset across five evaluation metrics (IoU, Dice, SE, PC, and ACC). Raw p-values and Holm-adjusted p-values are reported.

Compared Model	Metric	Raw p-Value	Holm-Adjusted p-Value	Sig.
UNet [9]	IoU	2.14×10^{-4}	1.92×10^{-3}	**
	Dice	2.37×10^{-4}	2.03×10^{-3}	**
	SE	1.54×10^{-6}	1.54×10^{-5}	***
	PC	0.05	0.15	ns
	ACC	8.20×10^{-4}	6.56×10^{-3}	**
UNeXt [24]	IoU	7.12×10^{-8}	7.83×10^{-7}	***
	Dice	7.12×10^{-8}	7.83×10^{-7}	***
	SE	4.31×10^{-7}	4.74×10^{-6}	***
	PC	6.49×10^{-3}	0.03	*
	ACC	9.51×10^{-8}	1.05×10^{-6}	***
DCSAU-Net [27]	IoU	2.20×10^{-3}	0.01	*
	Dice	1.77×10^{-3}	0.01	*
	SE	0.44	0.44	ns
	PC	1.33×10^{-5}	1.46×10^{-4}	***
	ACC	3.33×10^{-5}	3.00×10^{-4}	***
Rolling-Unet [25]	IoU	9.75×10^{-3}	0.03	*
	Dice	9.75×10^{-3}	0.03	*
	SE	1.41×10^{-3}	5.65×10^{-3}	**
	PC	3.11×10^{-3}	0.02	*
	ACC	0.02	0.06	ns
CMUNeXt [26]	IoU	1.89×10^{-5}	1.89×10^{-4}	***
	Dice	1.68×10^{-5}	1.68×10^{-4}	***
	SE	8.97×10^{-6}	7.18×10^{-5}	***
	PC	2.43×10^{-4}	2.43×10^{-3}	**
	ACC	1.90×10^{-6}	1.90×10^{-5}	***
TransFuse [16]	IoU	0.02	0.03	*
	Dice	0.02	0.03	*
	SE	8.70×10^{-3}	0.03	*
	PC	3.13×10^{-4}	2.82×10^{-3}	**
	ACC	7.16×10^{-3}	0.04	*
UWT-Net [43]	IoU	1.65×10^{-3}	0.01	*
	Dice	1.65×10^{-3}	0.01	*
	SE	6.26×10^{-5}	3.76×10^{-4}	***
	PC	3.99×10^{-3}	0.02	*
	ACC	0.02	0.06	ns
VMC-UNet [33]	IoU	2.20×10^{-3}	0.01	*
	Dice	2.25×10^{-3}	0.01	*
	SE	1.05×10^{-4}	5.26×10^{-4}	***
	PC	0.12	0.15	ns
	ACC	8.37×10^{-3}	0.04	*
FMamba [34]	IoU	4.33×10^{-4}	3.46×10^{-3}	**
	Dice	2.25×10^{-4}	2.03×10^{-3}	**
	SE	2.84×10^{-6}	2.55×10^{-5}	***
	PC	0.07	0.15	ns
	ACC	0.05	0.07	ns
DD-Mamba [35]	IoU	2.56×10^{-3}	0.01	*
	Dice	2.35×10^{-3}	0.01	*
	SE	3.78×10^{-5}	2.65×10^{-4}	***
	PC	0.02	0.08	ns
	ACC	6.12×10^{-3}	0.04	*
HCMNet [41]	IoU	0.03	0.03	*
	Dice	0.02	0.03	*
	SE	0.13	0.25	ns
	PC	1.02×10^{-3}	8.19×10^{-3}	**
	ACC	0.03	0.07	ns

Note: Pairwise comparisons against WMCAF-Net ($n = 65$) were Holm-adjusted separately for each metric (11 comparisons). Significance based on adjusted p-values: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, and ns $p \geq 0.05$.

Table 4. Statistical significance analysis of WMCAF-Net against competing methods on Dataset B across five evaluation metrics (IoU, Dice, SE, PC, and ACC). Raw p-values and Holm-adjusted p-values are reported.

Compared Model	Metric	Raw p-Value	Holm-Adjusted p-Value	Sig.
UNet [9]	IoU	4.81×10^{-8}	3.85×10^{-7}	***
	Dice	1.21×10^{-7}	9.66×10^{-7}	***
	SE	9.52×10^{-3}	0.07	ns
	PC	1.48×10^{-3}	8.90×10^{-3}	**
	ACC	1.28×10^{-8}	1.41×10^{-7}	***
UNeXt [24]	IoU	1.76×10^{-7}	1.06×10^{-6}	***
	Dice	1.76×10^{-7}	1.23×10^{-6}	***
	SE	5.17×10^{-10}	5.17×10^{-9}	***
	PC	0.06	0.24	ns
	ACC	0.06	0.12	ns
DCSAU-Net [27]	IoU	2.49×10^{-4}	7.46×10^{-4}	***
	Dice	2.09×10^{-4}	6.26×10^{-4}	***
	SE	1.08×10^{-5}	9.70×10^{-5}	***
	PC	0.27	0.82	ns
	ACC	0.02	0.06	ns
Rolling-Unet [25]	IoU	2.92×10^{-7}	1.46×10^{-6}	***
	Dice	1.76×10^{-7}	1.23×10^{-6}	***
	SE	0.50	1.00	ns
	PC	1.16×10^{-7}	1.16×10^{-6}	***
	ACC	7.54×10^{-3}	0.04	*
CMUNeXt [26]	IoU	6.66×10^{-7}	2.67×10^{-6}	***
	Dice	4.96×10^{-6}	1.98×10^{-5}	***
	SE	0.89	1.00	ns
	PC	2.14×10^{-7}	1.93×10^{-6}	***
	ACC	5.54×10^{-8}	5.54×10^{-7}	***
TransFuse [16]	IoU	4.02×10^{-3}	8.04×10^{-3}	**
	Dice	4.95×10^{-3}	9.90×10^{-3}	**
	SE	8.46×10^{-4}	6.77×10^{-3}	**
	PC	0.29	0.82	ns
	ACC	5.32×10^{-3}	0.03	*
UWT-Net [43]	IoU	1.90×10^{-10}	2.09×10^{-9}	***
	Dice	5.81×10^{-10}	6.39×10^{-9}	***
	SE	0.02	0.10	ns
	PC	8.46×10^{-8}	9.31×10^{-7}	***
	ACC	1.68×10^{-5}	1.52×10^{-4}	***
VMC-UNet [33]	IoU	5.14×10^{-8}	3.85×10^{-7}	***
	Dice	2.05×10^{-7}	1.23×10^{-6}	***
	SE	2.25×10^{-14}	2.48×10^{-13}	***
	PC	0.74	0.82	ns
	ACC	0.01	0.06	ns
FMamba [34]	IoU	1.15×10^{-8}	1.04×10^{-7}	***
	Dice	1.23×10^{-8}	1.11×10^{-7}	***
	SE	0.84	1.00	ns
	PC	5.58×10^{-5}	4.47×10^{-4}	***
	ACC	1.87×10^{-4}	1.50×10^{-3}	**
DD-Mamba [35]	IoU	8.03×10^{-9}	8.03×10^{-8}	***
	Dice	8.27×10^{-9}	8.27×10^{-8}	***
	SE	0.02	0.11	ns
	PC	1.05×10^{-3}	7.37×10^{-3}	**
	ACC	2.13×10^{-3}	0.01	*
HCMNet [41]	IoU	0.28	0.28	ns
	Dice	0.40	0.40	ns
	SE	0.26	1.00	ns
	PC	0.01	0.05	ns
	ACC	0.44	0.44	ns

Note: Pairwise comparisons against WMCAF-Net ($n = 163$) were Holm-adjusted separately for each metric (11 comparisons). Significance based on adjusted p-values: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, and ns $p \geq 0.05$.

For the BUSI dataset, the statistical analysis was performed using the image-level predictions of 65 test images obtained from models trained with random seed 41. The corresponding results are presented in Table 3.

For Dataset B, the statistical analysis was performed using 163 image-level predictions aggregated from the held-out test sets of the five-fold cross-validation. The corresponding results are presented in Table 4.

Overall, the results demonstrate that WMCAF-Net achieves statistically significant improvements in the primary segmentation metrics (IoU and Dice) compared with most competing methods. Specifically, statistically significant improvements in both IoU and Dice are observed against all eleven competing methods on BUSI and ten of the eleven competing methods on Dataset B after Holm correction (adjusted $p < 0.05$). For the supplementary metrics (SE, PC, and ACC), the exploratory analyses reveal statistically significant differences in several comparisons, although the significance varies across metrics and competing methods. These findings provide statistical evidence supporting the effectiveness of WMCAF-Net in improving segmentation overlap quality, while the supplementary analyses offer additional insights into its segmentation performance.

4.4.4. Cross-Dataset Evaluation

To further evaluate the transferability of WMCAF-Net across different ultrasound datasets, we conducted bidirectional cross-dataset evaluation experiments. Specifically, for each evaluation setting, the source dataset was first divided into training and validation sets with a ratio of 8:2. Each model was independently trained using three random seeds, namely 41, 42, and 43. For each run, the model was optimized using only the training set, and the checkpoint with the lowest validation loss was selected. The selected model was then directly evaluated on the target dataset, which was not involved in training, validation, or model selection. The results are reported as the mean and standard deviation across the three independent training runs, with the ± values indicating the standard deviation across runs.

The quantitative results are presented in Table 5. When trained on Dataset B and externally tested on the BUSI dataset, WMCAF-Net achieved a mean IoU of 47.19% and a mean Dice score of 57.33%. In the reverse setting, where the model was trained on BUSI and externally tested on Dataset B, WMCAF-Net obtained a mean IoU of 53.64% and a mean Dice score of 62.76%.

Compared with the within-dataset evaluation results, the overlap-related metrics (IoU and Dice) decreased in cross-dataset testing, indicating that distribution differences between ultrasound datasets may affect segmentation performance.

Table 5. Cross-dataset evaluation results of different segmentation methods (mean ± standard deviation).

Model	Dataset B → BUSI					BUSI → Dataset B				
	IoU (%)	Dice (%)	SE (%)	PC (%)	ACC (%)	IoU (%)	Dice (%)	SE (%)	PC (%)	ACC (%)
UNet [9]	44.26 ± 2.65	53.51 ± 2.97	51.23 ± 3.59	69.22 ± 1.73	93.41 ± 0.25	59.56 ± 3.44	67.11 ± 3.94	72.30 ± 4.19	69.17 ± 3.43	**97.86 ± 0.13**
UNext [24]	33.58 ± 0.77	42.49 ± 0.98	40.91 ± 1.90	63.23 ± 1.63	92.74 ± 0.10	53.82 ± 1.49	63.67 ± 1.36	82.77 ± 0.95	60.75 ± 2.87	95.93 ± 0.59
DCSAU-Net [27]	28.77 ± 1.83	36.01 ± 2.08	33.18 ± 1.96	57.82 ± 0.91	92.48 ± 0.15	56.25 ± 0.78	65.62 ± 0.97	84.73 ± 0.63	60.37 ± 1.04	95.01 ± 0.43
Rolling-Unet [25]	39.05 ± 1.10	46.84 ± 0.85	45.54 ± 0.81	63.06 ± 0.77	93.05 ± 0.07	62.44 ± 0.28	71.07 ± 0.23	84.67 ± 1.78	67.21 ± 0.84	96.98 ± 0.24
CMUNext [26]	34.18 ± 3.08	42.58 ± 4.32	41.69 ± 5.22	64.68 ± 2.43	92.73 ± 0.31	60.39 ± 1.61	68.99 ± 1.31	82.91 ± 2.97	65.74 ± 1.46	96.82 ± 0.63
TransFuse [16]	**59.22 ± 4.97**	**68.82 ± 4.93**	**74.51 ± 17.11**	73.65 ± 5.25	**95.57 ± 1.36**	65.64 ± 1.04	74.95 ± 1.23	**92.59 ± 1.03**	68.88 ± 1.13	97.35 ± 0.24
UWT-Net [43]	32.12 ± 0.07	40.75 ± 0.14	39.25 ± 0.18	62.14 ± 0.16	92.72 ± 0.01	57.56 ± 2.75	66.56 ± 3.05	75.13 ± 4.69	68.00 ± 4.64	97.65 ± 0.41
VMC-UNet [33]	49.46 ± 4.43	59.15 ± 3.83	53.76 ± 4.86	78.09 ± 1.12	94.11 ± 0.40	**65.80 ± 0.63**	**75.50 ± 0.32**	87.35 ± 0.84	**71.31 ± 0.36**	97.63 ± 0.32
FMamba [34]	34.82 ± 0.44	42.79 ± 0.45	40.76 ± 0.48	61.04 ± 0.18	92.84 ± 0.03	55.91 ± 1.10	64.67 ± 1.40	85.89 ± 0.80	59.87 ± 1.48	95.20 ± 0.51
DD-Mamba [35]	34.68 ± 0.82	41.55 ± 0.93	39.00 ± 1.08	57.80 ± 1.45	92.71 ± 0.10	53.65 ± 3.52	63.09 ± 3.18	86.41 ± 3.51	58.02 ± 4.57	94.70 ± 1.16
HCMNet [41]	50.35 ± 6.64	59.42 ± 6.99	55.09 ± 8.05	**76.24 ± 0.63**	93.59 ± 0.87	64.17 ± 0.23	72.81 ± 0.22	89.06 ± 0.80	67.96 ± 0.77	96.23 ± 0.17
Ours	47.19 ± 1.03	57.33 ± 1.18	52.85 ± 1.08	74.66 ± 1.39	93.69 ± 0.14	53.64 ± 3.69	62.76 ± 3.76	89.05 ± 2.75	56.89 ± 3.80	94.33 ± 1.12

Note: Bold values indicate the best performance for each evaluation metric among all compared methods.

4.5. Ablation Study

4.5.1. Block-Wise Ablation Analysis

Wave-Mamba and CAFM are two key components of the proposed framework. To evaluate their individual and joint contributions, we conduct block-wise ablation experiments on the BUSI dataset. The baseline model adopts the U-shaped architecture of FMamba, while using the same channel configuration as WMCAF-Net, i.e., 32, 64, 128, 256, and 512 channels across the five stages. This configuration is used consistently for all variants in the ablation study to ensure a controlled comparison of the introduced modules. In the Wave-Mamba configuration, the original FreqEM module in the baseline is replaced by the proposed Wave-Mamba module rather than being directly added on top of the baseline architecture.

We gradually introduce the proposed modules and analyze their effects on BUS image segmentation performance. The quantitative results are summarized in Table 6. Compared with the baseline, incorporating CAFM improves IoU and Dice by 4.66 and 5.06 percentage points, respectively, while SE, PC, and ACC increase by 6.10, 3.48, and 0.07 percentage points, respectively. These results indicate that CAFM improves encoder–decoder feature fusion and contributes to more accurate lesion segmentation.

Table 6. Block-wise ablation study results on the BUSI dataset.

Method	FLOPs (G)	Params (M)	IoU (%)	Dice (%)	SE (%)	PC (%)	ACC (%)
Baseline	1.14	2.64	66.54 ± 0.81	75.12 ± 0.58	78.97 ± 0.69	77.13 ± 0.83	96.25 ± 0.01
Baseline + CAFM	1.22	6.86	71.20 ± 1.01	80.18 ± 0.85	85.07 ± 0.32	80.61 ± 1.06	96.32 ± 0.34
Baseline + Wave-Mamba	1.21	6.55	73.68 ± 0.31	81.74 ± 0.40	**86.42 ± 1.81**	81.64 ± 0.64	96.46 ± 0.23
Baseline + Wave-Mamba + CAFM	1.29	10.77	**75.47 ± 0.26**	**83.47 ± 0.12**	85.73 ± 0.27	**85.30 ± 0.60**	**96.80 ± 0.18**

Note: Bold values indicate the best performance for each evaluation metric among all compared methods.

Replacing the original FreqEM module with Wave-Mamba leads to improvements of 7.14 and 6.62 percentage points in IoU and Dice, respectively, compared with the baseline. SE, PC, and ACC are also improved by 7.45, 4.51, and 0.21 percentage points, respectively, indicating that wavelet decomposition and directional high-frequency modeling contribute to more effective feature representation.

When Wave-Mamba and CAFM are jointly employed, the model achieves the highest IoU, Dice, PC, and ACC scores of 75.47%, 83.47%, 85.30%, and 96.80%, respectively. Compared with the baseline, these correspond to improvements of 8.93, 8.35, 8.17, and 0.55 percentage points in IoU, Dice, PC, and ACC, respectively.

To assess the statistical significance of the ablation results, we performed paired comparisons using predictions from the same 65 BUSI test images, generated by models trained with random seed 41 and selected based on the lowest validation loss. A two-sided Wilcoxon signed-rank test was applied, followed by Holm correction across the five pairwise comparisons separately for each metric. IoU and Dice were treated as primary metrics, while SE, PC, and ACC were analyzed as supplementary metrics. As shown in Table 7, all five comparisons showed statistically significant differences in IoU and Dice after Holm correction ($p_{\text{Holm}} < 0.05$). The supplementary metrics showed significance in selected comparisons. These results provide statistical support for the contributions of Wave-Mamba and CAFM to improving segmentation overlap performance.

Table 7. Statistical significance analysis of ablation experiments based on Wilcoxon signed-rank test with Holm correction.

Comparison	Metric	Raw *p*-Value	Holm-Adjusted *p*-Value	Sig.
Baseline vs. Baseline + CAFM	IoU	1.29×10^{-3}	2.58×10^{-3}	**
	Dice	9.78×10^{-4}	1.96×10^{-3}	**
	SE	4.74×10^{-4}	2.37×10^{-3}	**
	PC	0.815	1.00	ns
	ACC	0.019	0.075	ns
Baseline vs. Baseline + Wave-Mamba	IoU	4.52×10^{-3}	4.52×10^{-3}	**
	Dice	5.21×10^{-3}	5.21×10^{-3}	**
	SE	0.017	0.051	ns
	PC	0.760	1.00	ns
	ACC	0.161	0.161	ns
Baseline vs. Full	IoU	1.81×10^{-6}	9.04×10^{-6}	***
	Dice	1.49×10^{-6}	7.43×10^{-6}	***
	SE	7.37×10^{-4}	2.95×10^{-3}	**
	PC	0.031	0.125	ns
	ACC	1.90×10^{-4}	9.52×10^{-4}	***
Baseline + CAFM vs. Full	IoU	6.99×10^{-5}	2.80×10^{-4}	***
	Dice	9.44×10^{-5}	3.77×10^{-4}	***
	SE	0.115	0.115	ns
	PC	3.45×10^{-3}	0.017	*
	ACC	0.053	0.106	ns
Baseline + Wave-Mamba vs. Full	IoU	1.88×10^{-4}	5.64×10^{-4}	***
	Dice	1.83×10^{-4}	5.49×10^{-4}	***
	SE	0.021	0.051	ns
	PC	0.067	0.200	ns
	ACC	0.034	0.101	ns

Note: All pairwise comparisons were performed on the same BUSI test set ($n = 65$ paired observations). Holm's method was applied separately to each metric to adjust for multiple comparisons. Statistical significance was determined using the Holm-adjusted *p*-values: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, and ns $p \geq 0.05$.

Bidirectional cross-dataset ablation experiments were conducted between BUSI and Dataset B to investigate the contributions of Wave-Mamba and CAFM under distribution shifts. Specifically, the models were trained on either Dataset B or BUSI as the source dataset and directly evaluated on the other dataset to assess their segmentation performance under different data distributions. Four configurations were considered: Baseline, Baseline + CAFM, Baseline + Wave-Mamba, and the complete WMCAF-Net. By comparing these configurations in both transfer directions, we examine the individual contributions and combined effects of the proposed modules under cross-dataset conditions.

As shown in Table 8, in the Dataset B → BUSI transfer direction, the Baseline achieved an IoU of 41.42% and a Dice score of 49.55%. With CAFM, these metrics increased to 45.32% and 54.50%, respectively, while incorporating Wave-Mamba yielded 45.48% and 54.59%. When both modules were incorporated, the complete model further improved the IoU and Dice to 47.19% and 57.33%, representing gains of 5.77 and 7.78 percentage points over the Baseline, respectively. Moreover, the complete model achieved the best SE, PC, and ACC among the four configurations, indicating that both modules contribute to improved segmentation performance in this transfer direction.

As shown in Table 9, in the BUSI → Dataset B transfer direction, the Baseline achieved an IoU of 47.75% and a Dice score of 57.77%. Adding CAFM increased these metrics to 50.02% and 60.60%, while adding Wave-Mamba yielded 51.13% and 60.50%, respectively. The complete model further achieved an IoU of 53.64% and a Dice score of 62.76%, corresponding to improvements of 5.89 and 4.99 percentage points over the Baseline. Meanwhile,

the SE of the complete model reached 89.05%, although its ACC was slightly lower than that of the Baseline.

Table 8. Cross-dataset block-wise ablation results for Dataset B → BUSI (mean ± standard deviation).

Method	IoU	Dice	SE	PC	ACC
Baseline	0.4142 ± 0.0034	0.4955 ± 0.0025	0.4505 ± 0.0082	0.6921 ± 0.0249	0.9345 ± 0.0006
Baseline + CAFM	0.4532 ± 0.0027	0.5450 ± 0.0032	0.4977 ± 0.0072	0.7392 ± 0.0149	0.9358 ± 0.0005
Baseline + Wave-Mamba	0.4548 ± 0.0033	0.5459 ± 0.0036	0.4996 ± 0.0100	0.7360 ± 0.0163	0.9368 ± 0.0004
Baseline + Wave-Mamba + CAFM	**0.4719 ± 0.0103**	**0.5733 ± 0.0118**	**0.5285 ± 0.0108**	**0.7466 ± 0.0139**	**0.9369 ± 0.0014**

Note: Bold values indicate the best performance for each evaluation metric among all compared methods.

Table 9. Cross-dataset block-wise ablation results for BUSI → Dataset B (mean ± standard deviation).

Method	IoU	Dice	SE	PC	ACC
Baseline	0.4775 ± 0.0049	0.5777 ± 0.0052	0.7248 ± 0.0103	0.5472 ± 0.0077	**0.9635 ± 0.0015**
Baseline + CAFM	0.5002 ± 0.0067	0.6060 ± 0.0066	0.8444 ± 0.0216	0.5392 ± 0.0099	0.9531 ± 0.0014
Baseline + Wave-Mamba	0.5113 ± 0.0145	0.6050 ± 0.0171	0.8828 ± 0.0355	0.5441 ± 0.0157	0.9351 ± 0.0061
Baseline + Wave-Mamba + CAFM	**0.5364 ± 0.0369**	**0.6276 ± 0.0376**	**0.8905 ± 0.0275**	**0.5689 ± 0.0380**	0.9433 ± 0.0112

Note: Bold values indicate the best performance for each evaluation metric among all compared methods.

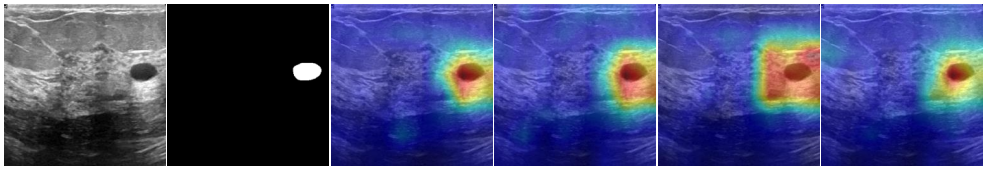
Overall, as shown in Tables 8 and 9, the results indicate that both Wave-Mamba and CAFM can improve cross-dataset segmentation performance to some extent, and their combination yields better overall performance. This suggests that differentiated frequency feature modeling based on wavelet decomposition and adaptive cross-scale feature fusion offer complementary benefits and help improve segmentation performance under different data distributions.

4.5.2. Ablation Study on Cross and Diagonal Convolutions in Wave-Mamba

Horizontal–vertical cross convolution (HVC) and diagonal cross convolution (DCC) are two key components of the high-frequency branch in the proposed Wave-Mamba module, designed to enhance complementary directional features for improved lesion boundary representation. To evaluate their effectiveness, we conduct ablation experiments on the BUSI dataset by replacing HVC and/or DCC with standard 3 × 3 convolutions while keeping the remaining network architecture and training settings unchanged.

Figure 4 presents heatmap visualizations under different convolution configurations. Retaining only HVC or DCC produces localized or directionally biased activation responses, whereas their combination yields more concentrated and complete responses around lesions while reducing irrelevant background activations. These observations suggest that joint directional modeling provides a more comprehensive representation of lesion-related high-frequency structures.

The quantitative results are presented in Table 10. The full configuration achieves the highest IoU and Dice scores of 75.47% and 83.47%, respectively, compared with 73.87% and 82.16% when both modules are replaced with standard 3 × 3 convolutions. Neither single-module configuration outperforms the two-standard-convolution configuration, indicating the importance of complementary directional modeling. Furthermore, the full configuration requires only 10.77 M parameters and 1.29 G FLOPs, compared with 18.15 M parameters and 1.48 G FLOPs for the two-standard-convolution configuration. This corresponds to reductions of 40.66% in parameters and 12.84% in FLOPs, alongside improvements of 1.60 and 1.31 percentage points in IoU and Dice, respectively. These results indicate that combining HVC and DCC improves segmentation accuracy while reducing model complexity and theoretical computational cost.



(a) (b) (c) (d) (e) (f)

Figure 4. Visualization of activation heatmaps under different convolutional configurations in the Wave-Mamba module. From left to right: (**a**) original image, (**b**) ground truth, (**c**) 3 × 3 standard convolutions replacing both HVC and DCC, (**d**) 3 × 3 standard convolution with HVC, (**e**) 3 × 3 standard convolution with DCC, and (**f**) the combination of HVC and DCC. Warmer colors denote stronger activation responses.

Table 10. Ablation study on HVC and DCC in the Wave-Mamba module on the BUSI dataset.

HVC	DCC	FLOPs (G)	Params (M)	IoU (%)	Dice (%)	SE (%)	PC (%)	ACC (%)
×	✓	1.38	14.46	73.33 ± 0.63	81.50 ± 0.63	85.13 ± 0.70	83.61 ± 1.17	96.47 ± 0.32
✓	×	1.38	14.46	73.36 ± 0.71	81.42 ± 0.61	84.40 ± 2.04	83.79 ± 0.78	96.46 ± 0.18
×	×	1.48	18.15	73.87 ± 0.10	82.16 ± 0.06	84.76 ± 0.22	84.02 ± 0.22	96.53 ± 0.21
✓	✓	1.29	10.77	**75.47 ± 0.26**	**83.47 ± 0.12**	**85.73 ± 0.27**	**85.30 ± 0.60**	**96.80 ± 0.18**

Note: ✓ and × indicate whether a component is enabled or disabled, respectively. Bold values indicate the best results for each metric.

4.5.3. Ablation Study on Channel and Spatial Attention in CAFM

To analyze the contributions of channel attention and spatial attention in the CAFM module to segmentation performance, we conduct ablation experiments on the BUSI dataset. With the remaining network architecture and training settings kept unchanged, five comparison configurations are constructed in addition to the complete CAFM: additive fusion, concatenation fusion, channel attention only, spatial attention only, and CA-guided spatial attention. In the additive fusion and concatenation fusion configurations, both channel attention and spatial attention are disabled, and the two configurations differ only in the feature fusion operation, using element-wise addition and channel-wise concatenation, respectively. The experimental results are summarized in Table 11.

Table 11. Ablation study on channel and spatial attention components in the CAFM module on the BUSI dataset.

Fusion Strategy	Channel	Spatial	IoU (%)	Dice (%)	SE (%)	PC (%)	ACC (%)
Additive fusion	×	×	72.80 ± 0.74	81.11 ± 0.37	83.36 ± 1.50	83.76 ± 1.18	96.46 ± 0.34
Concatenation fusion	×	×	73.10 ± 0.34	81.10 ± 0.33	83.72 ± 0.80	82.69 ± 0.16	96.76 ± 0.34
Channel attention only	✓	×	73.46 ± 0.70	82.10 ± 0.82	85.23 ± 1.50	83.99 ± 1.32	96.66 ± 0.21
Spatial attention only	×	✓	71.43 ± 0.62	79.80 ± 0.47	81.51 ± 2.13	84.23 ± 0.87	96.39 ± 0.06
Channel attention-guided spatial attention	✓ †	✓	73.16 ± 0.51	81.64 ± 0.28	84.76 ± 0.18	83.71 ± 0.14	96.66 ± 0.02
CAFM	✓	✓	**75.47 ± 0.26**	**83.47 ± 0.12**	**85.73 ± 0.27**	**85.30 ± 0.60**	**96.80 ± 0.18**

Note: ✓ and × indicate whether a component is enabled or disabled, respectively. Bold values indicate the best performance for each evaluation metric. † Channel recalibration is used only to guide the generation of spatial attention weights, while the original encoder and decoder features are retained for the final spatially weighted fusion.

Compared with the complete CAFM, using only channel attention decreases IoU and Dice by 2.01 and 1.37 percentage points, respectively, whereas using only spatial attention leads to larger decreases of 4.04 and 3.67 percentage points. When CAFM is replaced with additive fusion, IoU and Dice decrease by 2.67 and 2.36 percentage points, respectively, while concatenation fusion results in decreases of 2.37 percentage points for

both metrics. The complete CAFM achieves the best overall segmentation performance among all configurations.

Notably, the spatial-attention-only configuration achieves an IoU of 71.43%, lower than additive fusion (72.80%) and concatenation fusion (73.10%). This may be attributed to the lack of channel-wise recalibration, as redundant background responses in shallow features may interfere with spatial attention weight generation. To further investigate this explanation, we introduce a channel attention-guided spatial attention configuration, where channel-recalibrated features are used to generate spatial attention weights, while the original encoder and decoder features are retained for final fusion. Under this setting, IoU and Dice improve from 71.43% and 79.80% to 73.16% and 81.64%, respectively. These results suggest that channel recalibration facilitates more effective spatial attention weight generation. Furthermore, the complete CAFM achieves higher IoU and Dice scores of 75.47% and 83.47%, respectively, suggesting that channel recalibration not only facilitates more effective spatial attention but also improves the feature representations directly involved in cross-scale fusion.

The activation heatmaps in Figure 5 provide further qualitative observations consistent with the quantitative results. Additive fusion and concatenation fusion produce relatively dispersed responses and retain noticeable background activations, suggesting that direct feature fusion may introduce redundant information from shallow encoder features. Channel attention strengthens the responses around the lesion region, but the activations are mainly concentrated on locally salient areas and may not sufficiently cover the complete lesion structure. Spatial attention adjusts the spatial response distribution, while some irrelevant background activations remain. In contrast, the complete CAFM produces more concentrated and complete activation responses around the lesion region while more effectively suppressing irrelevant background responses. Overall, the targeted CA-guided spatial attention experiment, together with the quantitative results and heatmap visualizations, suggests that channel recalibration and spatial selection provide complementary contributions to encoder–decoder feature fusion, and their joint integration leads to improved segmentation performance.

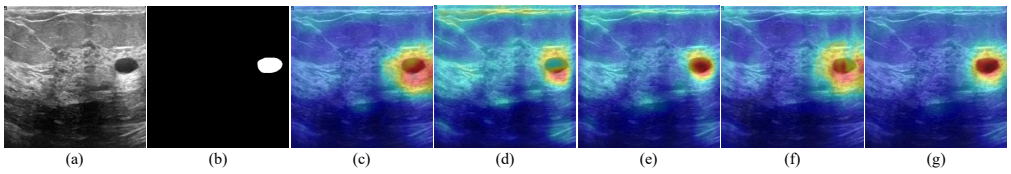


Figure 5. Visualization of activation heatmaps under different fusion strategies in the CAFM ablation study. From left to right: (**a**) original image, (**b**) ground truth, (**c**) additive fusion, (**d**) concatenation fusion, (**e**) channel attention only, (**f**) spatial attention only, and (**g**) the complete CAFM. Warmer colors denote stronger activation responses.

4.5.4. Ablation Study on Wavelet Bases in Wave-Mamba

To investigate the impact of different wavelet bases on Wave-Mamba, we evaluate four representative bases, Haar, db2, sym2, and bior2.2, while keeping the network architecture, channel configuration, input resolution, and training settings consistent. The results on the BUSI dataset are presented in Table 12.

As shown in Table 12, Haar achieves the best segmentation performance, with an IoU of 75.47% and a Dice score of 83.47%. Compared with Haar, bior2.2 achieves 1.43 and 1.14 percentage points lower IoU and Dice, respectively. All four configurations have the same parameter count (10.7736 M), while their measured FLOPs range from 1.2916 G to 1.3010 G.

Table 12. Comparison of different wavelet bases on the BUSI dataset.

Wavelet Basis	Params (M)	FLOPs (G)	IoU (%)	Dice (%)	SE (%)	PC (%)	ACC (%)
db2	10.7736	1.2951	72.48 ± 1.64	81.07 ± 2.02	85.24 ± 2.50	82.24 ± 1.80	96.19 ± 0.64
sym2	10.7736	1.2951	73.31 ± 0.30	82.03 ± 0.69	85.60 ± 1.92	83.52 ± 1.11	96.43 ± 0.45
bior2.2	10.7736	1.3010	74.04 ± 0.43	82.33 ± 0.31	85.70 ± 1.06	83.46 ± 0.48	96.75 ± 0.20
Haar (Ours)	10.7736	1.2916	**75.47 ± 0.26**	**83.47 ± 0.12**	**85.73 ± 0.27**	**85.30 ± 0.60**	**96.80 ± 0.18**

Note: Bold values indicate the best performance for each evaluation metric among all compared methods.

The Haar wavelet is preferred because its short filter support enables efficient computation and localized detection of abrupt intensity changes. Its high-pass filter behaves similarly to a local first-order difference operator, potentially helping capture fine-grained structural variations near lesion boundaries in breast ultrasound images. In contrast, longer wavelet filters may distribute local responses over broader spatial neighborhoods. Considering its superior segmentation performance and lowest measured FLOPs, Haar is adopted as the default wavelet basis in Wave-Mamba.

4.5.5. Ablation Study on the Number of Channels

We conduct channel ablation experiments to investigate the effect of network width on segmentation performance. As reported in Table 13, three representative channel configurations are evaluated, $\{16, 32, 64, 128, 256\}$, $\{32, 64, 128, 256, 512\}$, and $\{48, 96, 192, 384, 768\}$, corresponding to narrower, standard, and wider network variants, respectively. These configurations maintain the same architectural structure and channel scaling pattern, allowing us to examine the sensitivity of segmentation performance to network width under consistent experimental settings.

Table 13. Ablation study of different channel configurations in the proposed network.

Number of Channels	IoU (%)	Dice (%)	SE (%)	PC (%)	ACC (%)
{16, 32, 64, 128, 256}	72.19 ± 0.83	80.85 ± 0.91	85.50 ± 1.03	82.80 ± 0.30	96.29 ± 0.25
{32, 64, 128, 256, 512}	**75.47 ± 0.26**	**83.47 ± 0.12**	**85.73 ± 0.27**	**85.30 ± 0.60**	**96.80 ± 0.18**
{48, 96, 192, 384, 768}	71.40 ± 0.72	80.21 ± 0.85	84.57 ± 2.48	82.16 ± 0.97	96.38 ± 0.38

Note: Bold values indicate the best performance for each evaluation metric among all compared methods.

The results show that the default configuration achieves the best performance across all five evaluation metrics. Compared with the narrower configuration, it improves IoU and Dice by 3.28 and 2.62 percentage points, respectively, suggesting that insufficient channel capacity may limit feature representation. Further increasing the channel width does not improve performance; instead, IoU and Dice decrease by 4.07 and 3.26 percentage points, respectively. These findings indicate that increasing network width does not necessarily lead to better segmentation performance under the current experimental settings. Therefore, we adopt $\{32, 64, 128, 256, 512\}$ as the default channel configuration, as it achieves the best overall performance among the evaluated configurations.

4.5.6. Ablation Study of Loss-Weight Configurations

To investigate the effect of loss-weight selection on segmentation performance, we conducted an ablation study using four combinations of BCE and Dice loss weights on the BUSI dataset. Specifically, the evaluated configurations were $(\lambda_{\text{BCE}}, \lambda_{\text{Dice}}) \in \{(0.5, 0.5), (0.5, 1.0), (1.0, 0.5), (1.0, 1.0)\}$. All other experimental settings were kept unchanged to ensure a consistent comparison. The results are summarized in Table 14.

Table 14. Ablation study of different BCE and Dice loss weights on the BUSI dataset (mean ± standard deviation).

BCE Weight	Dice Weight	IoU (%)	Dice (%)	SE (%)	PC (%)	ACC (%)
0.5	0.5	72.03 ± 0.68	80.55 ± 0.60	86.83 ± 1.52	79.46 ± 0.94	95.67 ± 0.38
0.5	1.0	**75.47 ± 0.26**	**83.47 ± 0.12**	85.73 ± 0.27	**85.30 ± 0.60**	**96.80 ± 0.18**
1.0	0.5	72.21 ± 0.30	80.44 ± 0.56	83.74 ± 2.86	81.23 ± 1.46	96.05 ± 0.31
1.0	1.0	73.31 ± 0.57	81.65 ± 0.75	**87.12 ± 1.40**	81.18 ± 1.13	96.03 ± 0.43

Note: Bold values indicate the best performance for each evaluation metric among all compared methods.

As shown in Table 14, the configuration with $\lambda_{BCE} = 0.5$ and $\lambda_{Dice} = 1.0$ achieves the highest mean IoU (75.47%), Dice (83.47%), precision (85.30%), and accuracy (96.80%) among the four configurations. In contrast, the configuration with equal weights of 1.0 achieves the highest sensitivity (87.12%), indicating a stronger tendency to identify lesion pixels. However, its lower IoU, Dice, and precision scores suggest that the improvement in lesion recall does not translate into better overall region-overlap performance.

These results indicate that the relative contributions of BCE and Dice loss affect the balance between lesion detection and segmentation quality. Among the evaluated configurations, assigning a lower weight to BCE loss while retaining the full weight of Dice loss provides the best overall performance in terms of region overlap and precision, while maintaining competitive sensitivity. Therefore, we adopt $\lambda_{BCE} = 0.5$ and $\lambda_{Dice} = 1.0$ in the final WMCAF-Net configuration.

5. Discussion

Due to the diverse morphology, ambiguous boundaries, and susceptibility to noise and artifacts of breast lesions, breast ultrasound image segmentation remains a challenging task. In this paper, we propose WMCAF-Net, which integrates wavelet-based frequency-domain modeling with efficient global dependency modeling and cross-scale adaptive feature fusion. Experimental results on both BUSI and Dataset B demonstrate that WMCAF-Net achieves good overall segmentation performance compared with representative existing methods, highlighting the effectiveness of the proposed framework for breast ultrasound image segmentation.

The visualization results further show that WMCAF-Net may still exhibit local under-segmentation, boundary deviations, and background false positives in challenging cases, particularly for small lesions or lesions with weak and ambiguous boundaries. Representative failure cases are presented in Figure 6. These observations suggest that structural and frequency-domain information extracted from single-modal ultrasound images may not always be sufficient to distinguish lesion regions from complex backgrounds.

To address these limitations, future research will focus on improving feature modeling and fusion strategies. Although wavelet decomposition enables frequency-specific feature modeling, high-frequency components may still contain speckle noise that interferes with weak boundary representation. Therefore, adaptive high-frequency enhancement and noise suppression mechanisms will be explored to improve lesion boundary delineation. Furthermore, CAFM may still struggle to distinguish lesions from surrounding tissues with similar grayscale and texture characteristics. Future work will investigate boundary-aware and uncertainty-guided adaptive fusion to better emphasize lesion-relevant features and suppress background interference.

The cross-dataset evaluation results indicate that WMCAF-Net remains sensitive to distribution differences between ultrasound datasets. Variations in imaging devices, acquisition conditions, lesion characteristics, and image distributions may affect the learned feature representations and reduce segmentation performance on unseen datasets. Although the

proposed frequency-aware feature modeling and adaptive fusion strategies achieve strong performance under within-dataset evaluation settings, their effectiveness under domain shifts remains limited. Improving domain robustness through more diverse training data, domain-adaptive learning, and distribution-aware feature learning will be an important direction for future work.

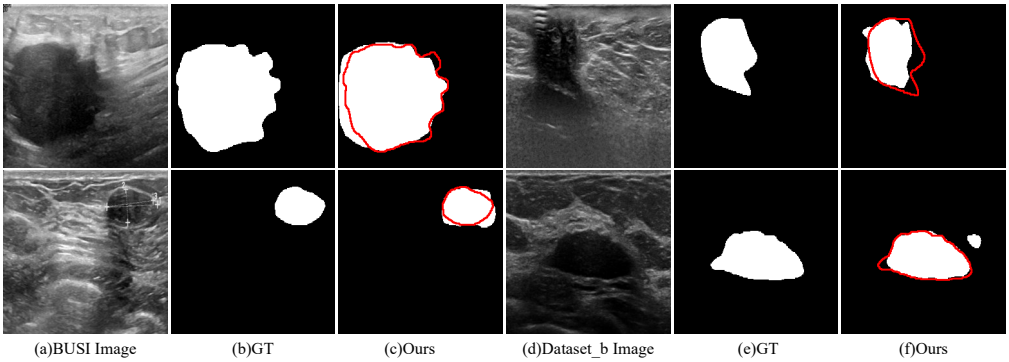


Figure 6. Representative failure cases of WMCAF-Net on the BUSI and Dataset B datasets. Columns (**a**–**c**) show the original BUSI images, ground-truth masks, and predictions of WMCAF-Net, respectively, while columns (**d**–**f**) show the corresponding results on Dataset B. White regions denote the predicted lesion masks, and red contours indicate the ground-truth boundaries.

Another limitation of this study is that the publicly available BUSI dataset does not provide complete patient-level identification information, making strict patient-level data partitioning infeasible based on the available metadata. Future studies will further evaluate WMCAF-Net using larger datasets with reliable patient identifiers to assess its performance under patient-independent evaluation protocols.

Overall, WMCAF-Net demonstrates strong segmentation performance on both BUSI and Dataset B, highlighting the effectiveness of integrating wavelet-based frequency-domain modeling, efficient global dependency modeling, and cross-scale adaptive fusion. The experimental results demonstrate the benefits of combining global structural representation with local boundary enhancement and adaptive feature fusion for breast ultrasound image segmentation. Future work will further explore advanced feature learning, boundary-aware modeling, and domain-robust learning strategies to enhance segmentation accuracy and generalization across diverse ultrasound imaging conditions.

6. Conclusions

Motivated by the challenges posed by diverse lesion appearances, strong noise interference, and indistinct structures in breast ultrasound image segmentation, this work introduces a segmentation framework termed WMCAF-Net, which combines frequency-domain modeling with global contextual learning. Within a unified framework, WMCAF-Net jointly incorporates wavelet-based frequency decomposition, visual state space modeling, and cross-scale feature correction mechanisms, enabling effective joint modeling of global structural information and local fine-grained lesion details. The proposed Wave-Mamba module decomposes features into different frequency bands using a wavelet transform and applies global modeling and directional enhancement strategies to low- and high-frequency components, respectively, thereby improving the representation of lesion contours and internal structures under complex ultrasound backgrounds. In addition, the proposed CAFM module performs adaptive correction of encoder and decoder features during skip connections, alleviating cross-layer feature distribution discrepancies and further enhancing detail recovery.

Experimental results on BUSI and Dataset B demonstrate that WMCAF-Net achieves strong segmentation performance under within-dataset evaluation settings. However, cross-dataset evaluation indicates that the model remains sensitive to distribution differences between ultrasound datasets, highlighting the need for further investigation into domain generalization. Ablation studies further indicate that the Wave-Mamba and CAFM modules improve feature representation and fusion from different perspectives and jointly contribute to the overall performance improvement. Overall, WMCAF-Net demonstrates strong segmentation performance on breast ultrasound images, providing a promising solution for lesion segmentation under complex ultrasound imaging conditions.

Author Contributions: Conceptualization, X.F.; methodology, X.F.; software, X.F.; validation, Y.L. and L.X.; formal analysis, J.L.; investigation, Y.L.; resources, H.D.; data curation, X.F.; writing—original draft preparation, X.F.; writing—review and editing, H.D.; visualization, X.F.; supervision, H.D. and Q.W.; project administration, H.D.; funding acquisition, H.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Hunan Provincial Natural Science Foundation of China (2023JJ50096), Hunan Provincial University Key Laboratory of Artificial Intelligence Application and System Security (Xiangjiaotong [2025] 233), the 14th Five-Year Plan Key Disciplines and Application-oriented Special Disciplines of Hunan Province (Xiangjiaotong [2022] 351), and the Science and Technology Innovation Program of Hunan Province (2016TP1020).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used in this study are publicly available. The Breast Ultrasound Images (BUSI) dataset is available at https://www.kaggle.com/datasets/sabahesaraki/breast-ultrasound-images-dataset (accessed on 4 November 2026). Dataset B is available at https://github.com/openmedlab/Awesome-Medical-Dataset/blob/main/resources/Breast_Ultrasound_DatasetB.md (accessed on 4 November 2026). The source code will be made publicly available at https://github.com/seaking007/WMCAF-Net upon publication (accessed on 4 November 2026).

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

References

1. Bray, F.; Laversanne, M.; Sung, H.; Ferlay, J.; Siegel, R.L.; Soerjomataram, I.; Jemal, A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA A Cancer J. Clin.* **2024**, *74*, 229–263. [CrossRef] [PubMed]
2. Ginsburg, O.; Yip, C.H.; Brooks, A.; Cabanes, A.; Caleffi, M.; Yataco, J.A.D.; Gyawali, B.; McCormack, V.; de Anderson, M.M.; Mehrotra, R.; et al. Breast cancer early detection: A phased approach to implementation. *Cancer* **2020**, *126*, 2379–2393. [CrossRef] [PubMed]
3. Dan, Q.; Zheng, T.; Liu, L.; Sun, D.; Chen, Y. Ultrasound for breast cancer screening in resource-limited settings: Current practice and future directions. *Cancers* **2023**, *15*, 2112. [CrossRef] [PubMed]
4. Huang, K.; Zhang, Y.; Cheng, H.D.; Xing, P. Trustworthy breast ultrasound image semantic segmentation based on fuzzy uncertainty reduction. *Healthcare* **2022**, *10*, 2480. [CrossRef] [PubMed]
5. Khaledyan, D.; Marini, T.J.; M. Baran, T.; O'Connell, A.; Parker, K. Enhancing breast ultrasound segmentation through fine-tuning and optimization techniques: Sharp attention UNet. *PLoS ONE* **2023**, *18*, e0289195. [CrossRef] [PubMed]
6. Liu, Y.; Cheng, H.D.; Huang, J.; Zhang, Y.; Tang, X. An effective approach of lesion segmentation within the breast ultrasound image based on the cellular automata principle. *J. Digit. Imaging* **2012**, *25*, 580–590. [CrossRef] [PubMed]
7. Chen, G.; Li, L.; Zhang, J.; Dai, Y. Rethinking the unpretentious U-Net for medical ultrasound image segmentation. *Pattern Recognit.* **2023**, *142*, 109728. [CrossRef]
8. Mall, P.K.; Singh, P.K.; Srivastav, S.; Narayan, V.; Paprzycki, M.; Jaworska, T.; Ganzha, M. A comprehensive review of deep neural networks for medical image processing: Recent developments and future opportunities. *Healthc. Anal.* **2023**, *4*, 100216. [CrossRef]
9. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241. [CrossRef]

10. Zhou, Z.; Rahman Siddiquee, M.M.; Tajbakhsh, N.; Liang, J. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 3–11. [CrossRef] [PubMed]
11. Diakogiannis, F.I.; Waldner, F.; Caccetta, P.; Wu, C. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS J. Photogramm. Remote Sens.* **2020**, *162*, 94–114. [CrossRef]
12. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017.
13. Shamshad, F.; Khan, S.; Zamir, S.W.; Khan, M.H.; Hayat, M.; Khan, F.S.; Fu, H. Transformers in medical imaging: A survey. *Med. Image Anal.* **2023**, *88*, 102802. [CrossRef] [PubMed]
14. Pu, Q.; Xi, Z.; Yin, S.; Zhao, Z.; Zhao, L. Advantages of transformer and its application for medical image segmentation: A survey. *Biomed. Eng. Online* **2024**, *23*, 14. [CrossRef] [PubMed]
15. 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* **2021**, arXiv:2102.04306.
16. Zhang, Y.; Liu, H.; Hu, Q. Transfuse: Fusing transformers and cnns for medical image segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 14–24. [CrossRef]
17. Gu, A.; Dao, T. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv* **2023**, arXiv:2312.00752.
18. Ruan, J.; Li, J.; Xiang, S. VM-UNet: Vision Mamba UNet for medical image segmentation. *ACM Trans. Multimed. Comput. Commun. Appl.* **2025**. [CrossRef]
19. Wu, R.; Liu, Y.; Liang, P.; Chang, Q. H-vmunet: High-order vision mamba unet for medical image segmentation. *Neurocomputing* **2025**, *624*, 129447. [CrossRef]
20. Wu, R.; Liu, Y.; Ning, G.; Liang, P.; Chang, Q. UltraLight VM-UNet: Parallel Vision Mamba significantly reduces parameters for skin lesion segmentation. *Patterns* **2025**, *6*, 101298. [CrossRef] [PubMed]
21. Lan, Y.; Zhang, X. Real-time ultrasound image despeckling using mixed-attention mechanism based residual UNet. *IEEE Access* **2020**, *8*, 195327–195340. [CrossRef]
22. Satapathy, L.; Das, P.; Shatapathy, A.; Patel, A. Bio-medical image denoising using wavelet transform. *Int. J. Recent Technol. Eng.* **2019**, *8*, 2874–2879.
23. Drozdzal, M.; Vorontsov, E.; Chartrand, G.; Kadoury, S.; Pal, C. The importance of skip connections in biomedical image segmentation. In *Deep Learning and Data Labeling for Medical Applications*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 179–187. [CrossRef]
24. Valanarasu, J.M.J.; Patel, V.M. Unext: Mlp-based rapid medical image segmentation network. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2022*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 23–33. [CrossRef]
25. Liu, Y.; Zhu, H.; Liu, M.; Yu, H.; Chen, Z.; Gao, J. Rolling-unet: Revitalizing mlp's ability to efficiently extract long-distance dependencies for medical image segmentation. *Proc. Aaai Conf. Artif. Intell.* **2024**, *38*, 3819–3827. [CrossRef]
26. Tang, F.; Ding, J.; Quan, Q.; Wang, L.; Ning, C.; Zhou, S.K. Cmunext: An efficient medical image segmentation network based on large kernel and skip fusion. In *Proceedings of the 2024 IEEE International Symposium on Biomedical Imaging (ISBI)*; IEEE: New York, NY, USA, 2024; pp. 1–5. [CrossRef]
27. Xu, Q.; Ma, Z.; Duan, W. DCSAU-Net: A deeper and more compact split-attention U-Net for medical image segmentation. *Comput. Biol. Med.* **2023**, *154*, 106626. [CrossRef] [PubMed]
28. Dosovitskiy, A. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
29. Tay, Y.; Dehghani, M.; Bahri, D.; Metzler, D. Efficient transformers: A survey. *ACM Comput. Surv.* **2022**, *55*, 109. [CrossRef]
30. Liu, X.; Zhang, C.; Huang, F.; Xia, S.; Wang, G.; Zhang, L. Vision Mamba: A Comprehensive Survey and Taxonomy. *IEEE Trans. Neural Netw. Learn. Syst.* **2026**, *37*, 505–525. [CrossRef] [PubMed]
31. Liu, Y.; Tian, Y.; Zhao, Y.; Yu, H.; Xie, L.; Wang, Y.; Ye, Q.; Jiao, J.; Liu, Y. Vmamba: Visual state space model. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 103031–103063. [CrossRef]
32. Liu, J.; Yang, H.; Zhou, H.Y.; Xi, Y.; Yu, L.; Li, C.; Liang, Y.; Shi, G.; Yu, Y.; Zhang, S.; et al. Swin-umamba: Mamba-based unet with imagenet-based pretraining. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2024*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 615–625. [CrossRef]
33. Wang, D.; Zhao, W.; Cui, K.; Zhu, Y. VMC-UNet: A Vision Mamba-CNN U-Net for Tumor Segmentation in Breast Ultrasound Image. *Int. J. Imaging Syst. Technol.* **2024**, *34*, e23222. [CrossRef]
34. Liu, S.; Lin, Y.; Liu, D.; Wang, P.; Zhou, B.; Si, F. Frequency-Enhanced Lightweight Vision Mamba Network for Medical Image Segmentation. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 5006412. [CrossRef]
35. Liu, S.; Zhao, M.; Lin, Y. Dual-domain guided vision Mamba network for medical image segmentation. *Expert Syst. Appl.* **2025**, *291*, 128512. [CrossRef]

36. Zhang, D. Wavelet Transform. In *Fundamentals of Image Data Mining: Analysis, Features, Classification and Retrieval*; Texts in Computer Science; Springer International Publishing: Cham, Switzerland, 2019; pp. 35–44. [CrossRef]
37. Azad, R.; Kazerouni, A.; Sulaiman, A.; Bozorgpour, A.; Aghdam, E.K.; Jose, A.; Merhof, D. Unlocking fine-grained details with wavelet-based high-frequency enhancement in transformers. In *Machine Learning in Medical Imaging*; Springer: Berlin/Heidelberg, Germany, 2023; pp. 207–216. [CrossRef]
38. Zhou, Y.; Huang, J.; Wang, C.; Song, L.; Yang, G. Xnet: Wavelet-based low and high frequency fusion networks for fully-and semi-supervised semantic segmentation of biomedical images. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 21085–21096. [CrossRef]
39. Meng, W.; Zhu, X.; Li, Y. YNet: Medical Image Segmentation Model Based on Wavelet Transform Boundary Enhancement. In *Computational Visual Media*; Springer: Berlin/Heidelberg, Germany, 2025; pp. 91–108. [CrossRef]
40. Sun, J.; Chen, K.; Wu, X.; Xu, Z.; Wang, S.; Zhang, Y. MSM-UNet: A medical image segmentation method based on wavelet transform and multi-scale Mamba-UNet. *Expert Syst. Appl.* **2025**, *288*, 128241. [CrossRef]
41. Xiong, Y.; Shu, X.; Liu, Q.; Yuan, D. HCMNet: A Hybrid CNN-Mamba Network for Breast Ultrasound Segmentation for Consumer Assisted Diagnosis. *IEEE Trans. Consum. Electron.* **2025**, *71*, 8045–8054. [CrossRef]
42. Liu, Y.; Yang, Y.; Jiang, Y.; Zhao, X.; Xie, Z. FABRF-Net: A frequency-aware boundary and region fusion network for breast ultrasound image segmentation. *Inf. Fusion* **2025**, *123*, 103299. [CrossRef]
43. Zhang, P.; Ouyang, X.; Peng, R. UWT-Net: Mining Low-Frequency Feature Information for Medical Image Segmentation. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2025*; Springer: Cham, Switzerland, 2025; pp. 615–624. [CrossRef]
44. Vilimek, D.; Kubicek, J.; Golian, M.; Jaros, R.; Kahankova, R.; Hanzlikova, P.; Barvik, D.; Krestanova, A.; Penhaker, M.; Cerny, M.; et al. Comparative analysis of wavelet transform filtering systems for noise reduction in ultrasound images. *PLoS ONE* **2022**, *17*, e0270745. [CrossRef] [PubMed]
45. Namsopa, P.; Koonkarnkhai, S.; Kovintavewat, P.; Dinsakul, H.; Suwansawang, S. A Study and Exploration of Discrete Wavelet Transform for Speckle Noise Reduction in Ultrasound Images. In Proceedings of the 2023 International Technical Conference on Circuits/Systems, Computers, and Communications (ITC-CSCC), Jeju, Republic of Korea, 25–28 June 2023; pp. 1–5. [CrossRef]
46. Al-Dhabyani, W.; Gomaa, M.; Khaled, H.; Fahmy, A. Dataset of breast ultrasound images. *Data Brief* **2020**, *28*, 104863. [CrossRef] [PubMed]
47. Yap, M.H.; Goyal, M.; Osman, F.; Martí, R.; Denton, E.; Juette, A.; Zwiggelaar, R. Breast ultrasound region of interest detection and lesion localisation. *Artif. Intell. Med.* **2020**, *107*, 101880. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.