

Article

NSBR-Net: A Novel Noise Suppression and Boundary Refinement Network for Breast Tumor Segmentation in Ultrasound Images

Yue Sun , Zhaohong Huang, Guorong Cai, Jinhe Su * and Zheng Gong * 

College of Computer Engineering, Jimei University, Xiamen 361021, China; yuesun@jmu.edu.cn (Y.S.); jmuhez@126.com (Z.H.); guorongcai.jmu@gmail.com (G.C.)

* Correspondence: sujh@jmu.edu.cn (J.S.); zheng.gong@jmu.edu.cn (Z.G.)

Abstract: Breast tumor segmentation of ultrasound images provides valuable tumor information for early detection and diagnosis. However, speckle noise and blurred boundaries in breast ultrasound images present challenges for tumor segmentation, especially for malignant tumors with irregular shapes. Recent vision transformers have shown promising performance in handling the variation through global context modeling. Nevertheless, they are often dominated by features of large patterns and lack the ability to recognize negative information in ultrasound images, which leads to the loss of breast tumor details (e.g., boundaries and small objects). In this paper, we propose a novel noise suppression and boundary refinement network, NSBR-Net, to simultaneously alleviate speckle noise interference and blurred boundary problems of breast tumor segmentation. Specifically, we propose two innovative designs, namely, the Noise Suppression Module (NSM) and the Boundary Refinement Module (BRM). The NSM filters noise information from the coarse-grained feature maps, while the BRM progressively refines the boundaries of significant lesion objects. Our method demonstrates superior accuracy over state-of-the-art deep learning models, achieving significant improvements of 3.67% on Dataset B and 2.30% on the BUSI dataset in mDice for testing malignant tumors.

Keywords: breast tumor segmentation; ultrasound images; deep learning; transformer



Citation: Sun, Y.; Huang, Z.; Cai, G.; Su, J.; Gong, Z. NSBR-Net: A Novel Noise Suppression and Boundary Refinement Network for Breast Tumor Segmentation in Ultrasound Images. *Algorithms* **2024**, *17*, 257. <https://doi.org/10.3390/a17060257>

Academic Editor: Frank Werner

Received: 11 May 2024

Revised: 4 June 2024

Accepted: 7 June 2024

Published: 12 June 2024



Copyright: © 2024 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Breast tumors are a prevalent health concern for women that significantly impacts their well-being and lives. As a result, regular breast screening and diagnosis play a crucial role in formulating effective treatment plans and improving survival rates. Due to the flexibility and convenience of ultrasound imaging, it has become a conventional modality for breast tumor screening. In recent years, many deep learning methods based on ultrasound images have been proposed for breast tumor segmentation. However, complex ultrasound patterns continue to pose the following challenges: (1). blurred boundaries caused by low contrast between the foreground and background; (2). segmentation disruption due to speckle noise (as illustrated in Figure 1).

The impressive non-linear learning capability has led to significant successes in medical image segmentation with Full Convolution Network (FCN) and U-Net [1,2]. Motivated by this, many deep learning approaches have emerged for segmenting breast tumors from ultrasound images. In 2018, Almajalid et al. [3] were the first to systematically evaluate the impact of different FCN variants on breast tumor segmentation and achieve segmentation results that outperformed traditional methods. AAU-Net [4] integrates a hybrid adaptive attention module instead of the conventional convolution block, enhancing feature extraction across diverse receptive fields. NU-Net [5] utilizes sub-networks of varying depths with shared weights to attain robust representations of breast tumors.

Transformer has garnered scholarly attention for its attention mechanism and the complete elimination of convolution. Subsequent investigations [6–8] have explored the

integration of transformer structures in image recognition. Notably, ViT [9] stands out as the pioneering work to apply a pure transformer to image classification, substantiating the viability of transformer architectures for computer vision tasks. In the realm of medical image segmentation, the efficacy of vision transformers has been substantiated by PVT-CASCADE [10] and DuAT [11]. DuAT proposes a Dual-Aggregation Transformer Network to address the challenge of capturing both global and local spatial features, while PVT-CASCADE introduces an innovative attention-based decoder leveraging the multi-stage feature representation of the vision transformer. These advancements underscore the transformative impact of transformer architectures in the medical image field.

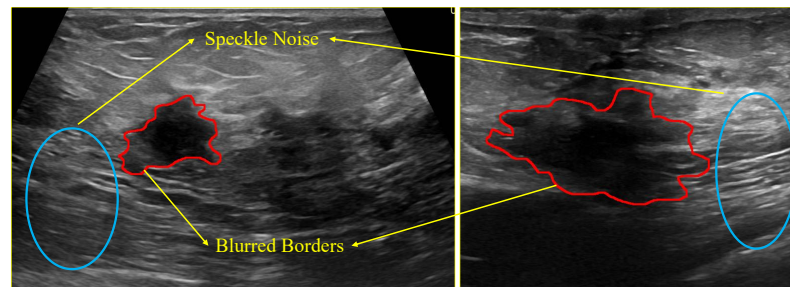


Figure 1. Challenges in a breast ultrasound image segmentation task. The red lines are the boundaries of the breast tumors. Inside and outside the boundaries are where the boundaries are blurred. The areas inside the blue circles are typical regions of speckle noise, which refers to irregular, distinct brightness and darkness distribution.

To further address the issue of blurred boundaries caused by breast tumors, the following optimization strategies have been considered: expanding the receptive field and the attention mechanisms has been widely used. The dilated convolution operation is a commonly used strategy to expand the receptive field. For example, Hu et al. [12] obtained the large receptive field of breast tumors by using dilated convolutions in deeper network layers. In terms of attention mechanisms, Lee et al. [13] proposed a channel attention module to further improve the performance of U-Net for breast tumor segmentation. Yan et al. [14] proposed an attention enhanced U-Net with hybrid dilated convolution, merging dilated convolutions with an attention mechanism. Although progress has been made by these methods, the optimization paradigm from fine to coarse granularity struggles to capture prominent object regions in deeper convolutional layers, where object regions and boundaries stand as two crucial distinguishing features between normal tissue and breast tumors. Thus, we propose an iteratively enhanced Boundary Refinement Module (BRM) based on a global map, emphasizing a pattern from coarse to fine granularity. Our motivation arises from clinical practice, where clinicians initially approximate the location of a breast tumor and then meticulously extract its silhouette mask based on local features. In the NSBR-Net model, we adopt a two-step approach: first predicting the coarse region and then implicitly modeling the boundaries using axial reverse attention. There are two advantages to this strategy, including better learning ability and improved generalization capability.

In ultrasound imaging, the inherent nature of speckle noise tends to degrade image quality and complicates the distinction between breast tissue and noise artifacts [15], making accurate tumor detection more challenging. Moreover, speckle noise significantly impacts segmentation accuracy by propagating across various convolutional layers at different scales. Current methods primarily leverage the concept of deep supervision to develop refined networks [16], exploring neighboring decisions to correct potential errors induced by speckle noise. However, we propose addressing noise influence from a more fundamental perspective by introducing “frequency”. In an intriguing experiment, we examined the network’s performance variation when eliminating high-frequency information (detail and noise) [17] in deeper layers. We used the mainstream method UNet [18] to evaluate the impact of high frequencies on breast tumor segmentation in the BUSI testset

(ultrasound images, including both benign and malignant breast tumors) [19]. Building upon [20], we employed multiple pooling operations on the last two stages of the UNet architecture to filter out the high-frequency information and keep only the low-frequency information. As shown in Figure 2, we observed a substantial improvement in model performance when the network solely contained low-frequency information, indicating that speckle noise within high-frequency information disrupts spatial consistency. To address this phenomenon, we introduced a Noise Suppression Module, decoupling high- and low-frequency information in feature maps and denoising the high-frequency components with a Gaussian filter. While following prior works' principles, NSBR-Net also incorporates a deep supervision mechanism.

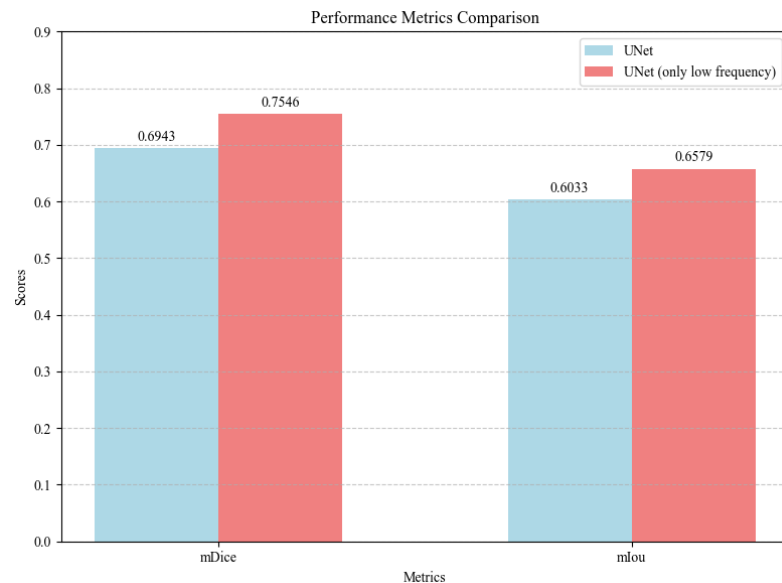


Figure 2. The network's performance variation when eliminating high-frequency information. The metrics, including the mean Dice coefficient (mDice) and mean intersection over union (mIoU), both assess the internal consistency of objects within the segmentation results.

Our method, built upon a transformer-based encoder, incorporates BRM and NSM for breast tumor segmentation in ultrasound images. Its efficacy was validated through extensive experiments on a breast ultrasound dataset and resulted in highlight significant improvement over existing methods. Our contributions include the following:

- We present a novel breast tumor segmentation framework, termed NSBR-Net. Unlike existing CNN-based methods, we adopt the pyramid vision transformer as an encoder to extract more robust features.
- To support our framework, we introduce two simple modules. Specifically, NSM is utilized to suppress speckle noise within high-frequency information, while BRM performs boundary refinement based on coarse regions.
- Comparative experiments juxtaposed with leading-edge medical image segmentation models demonstrate the superior efficacy of our method on a breast ultrasound dataset.

This paper is structured as follows: Section 2 outlines the dataset, method (including NSM and BRM), loss function, and experimental settings (including evaluation protocols and implementation details). Section 3 presents a qualitative and quantitative comparison of different methods and their corresponding analyses. Section 4 discusses the results, limitations, and future research directions, while Section 5 is dedicated to the conclusion.

2. Materials and Methods

2.1. Datasets

We conducted experiments on two widely used public breast ultrasound datasets, the BUSI dataset [19] and Dataset B [21]. The BUSI dataset contains 780 images acquired by two types of ultrasound equipment (LOGIQ E9 ultrasound and LOGIQ E9 Agile ultrasound system) in Baheya Hospital. The average image size of these images is 500×500 pixels. Radiologists manually segmented the boundaries of breast lesions in each ultrasound image using Matlab. Dataset B contains 163 breast US images from different women, which were scanned with a Siemens ACUSON Sequoia C512 system 17L5 HD linear array transducer (8.5 MHz) at the UDIAT Diagnostic Centre of the Parc Tau'II Corporation, Sabadell (Spain), in 2012. In the ground truth (GT) images, the boundaries of breast lesions were delineated by professional radiologists. The STU dataset comprises 42 BUS (breast ultrasound) images, each with an average size of 128×128 pixels. These images were obtained by the Imaging Department of the First Affiliated Hospital of Shantou University using a GE Voluson E10 ultrasonic diagnostic system. Due to the limited number of images in the STU dataset, it was solely employed as external validation data to assess the segmentation network's generalization performance.

2.2. Method

An overview of NSBR-Net is shown in Figure 3. Upon inputting an ultrasound image, we initially extracted four levels of feature maps of various scales sequentially utilizing the pyramid vision transformer (PVT) block [22]. We input the feature maps from the last three stages into NSM individually for the suppression of speckle noise, followed by utilizing a parallel partial decoder (PD) [23] to generate high-level semantic global maps. Lastly, a set of reverse axial attention mechanisms was employed to refine the tumor boundaries progressively. Detailed expositions of NSM and BRM are presented as follows. The overall pseudocode of our method is presented in Algorithm 1.

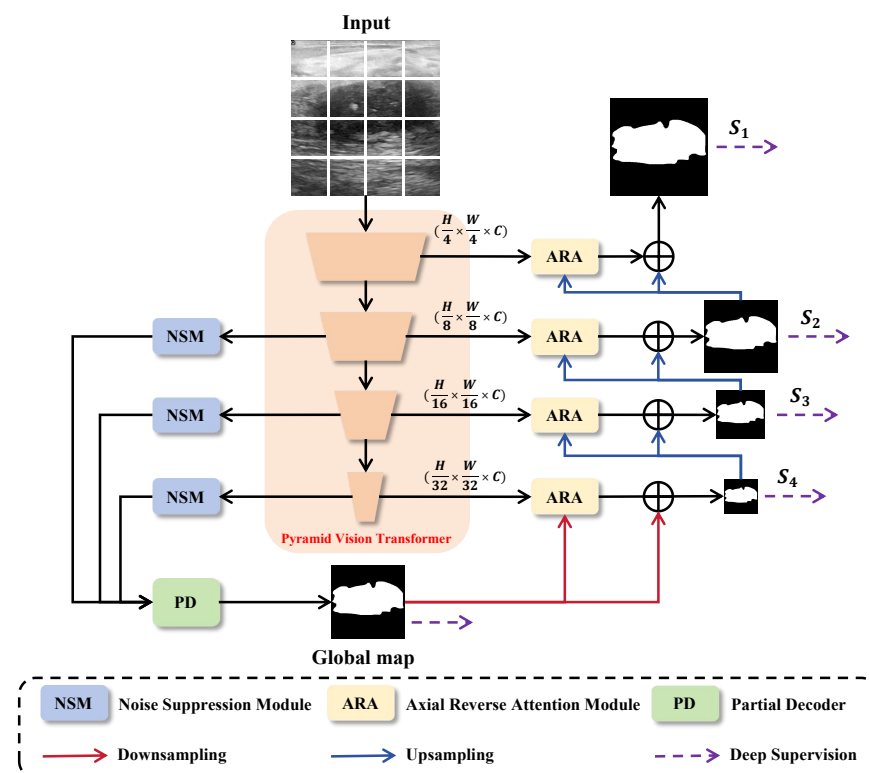


Figure 3. The framework of our proposed NSBR-Net primarily comprises the pyramid vision transformer, partial decoder (PD) [23], Noise Suppression Module, and Boundary Refinement Module.

Algorithm 1 Pseudocode of NSBR-Net in a Pytorch-like Style

```

#Operator: BI, Bilinear interpolation
# Pass input through the backbone network
pvt = backbone(x)
# Translayer: Apply convolutional, batch normalization, and ReLU
pvt_transformed = [Translayer (pvt[i]) for i in range (4)]
# NSM modules
nsm_outputs = [NSM (pvt_transformed[i]) for i in range (3)]
# partial decoder(PD)
pd_output = PD (*nsm_outputs)
# BRM modules
global_maps[4] = BI (pd_output)
stage_outputs[4] = BRM (pvt_transformed[4],global_maps[4]) + global_maps[4]
for i in range (3, 0, -1):
    global_maps[i] = BI (stage_outputs[i + 1])
    stage_outputs[i] = BRM (pvt_transformed[i],global_maps[i]) + global_maps[i]
predictions = [BI (stage_outputs[i]) for i in range (5)]
return predictions

```

2.2.1. Noise Suppression Module

Speckle noise is a complex physical characteristic observed in ultrasound images that frequently poses challenges in accurate object localization. The adoption of frequency representation presents a novel approach to discerning differences between categories, potentially revealing insights overlooked by human visual perception. To mitigate this, we propose a Noise Suppression Module (NSM) that considers speckle noise suppression from a frequency perspective, as illustrated in Figure 4.

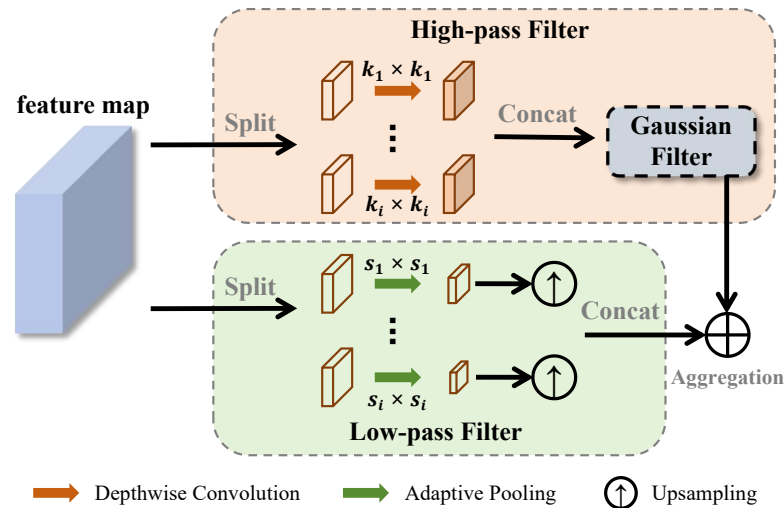


Figure 4. Overall architecture of NSM. It is composed of a low-pass filter (LPF) and a high-pass filter (HPF).

Low-pass filter (LPF). Low-frequency components occupy most of the energy in the absolute image and represent most of the semantic information. A low-pass filter allows signals below the cutoff frequency to pass, while signals above the cutoff frequency are obstructed. Thus, we employed typical average pooling as a low-pass filter. However, the cutoff frequencies of different images are different. To adapt to this, we utilized channel splitting [20], where the feature map is partitioned into multiple groups. This allowed us to apply different kernels and strides to each group, thereby generating low-pass filters. For the m th group, we have

$$LPF_m(v^m) = Up(\Gamma_{s \times s}(v^m)), \quad (1)$$

where $Up(\cdot)$ represents upsampling, $\Gamma_{s \times s}$ denotes the adaptive average pooling with the output size of $s \times s$, $s \in \{1, 2, 3, 6\}$, and v^m stands for the m th channel split from the input feature map, $m \in \{1, 2, 3, 4\}$.

High-pass filter (HPF). High-frequency information is crucial to preserve details in segmentation. As a typical high-pass operator, convolution can filter out irrelevant low-frequency redundant components to retain favorable high-frequency components. The high-frequency components determine the image quality, and the cutoff frequency of the high pass for each image is different. Similar to the LPF, we partitioned the feature map into n groups. For each group, we used a convolution layer with different kernels to simulate the cutoff frequencies in different high-pass filters. For the n th group, we have

$$HPF_n(v^n) = \Lambda_{k \times k}(v^n), \quad (2)$$

where $\Lambda_{k \times k}$ denotes the depthwise convolution layer with a kernel size of $k \times k$, $k \in \{1, 3, 5, 7\}$, and v^n stands for the n th channel split from the input feature map, $n \in \{1, 2, 3, 4\}$. The continuous accumulation of speckle noise within the internal high frequencies often yields adverse effects on the extracted high-frequency information. Therefore, we employed Gaussian filtering on the high-frequency features to effectively eliminate noise.

$$W(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}, \quad (3)$$

$$G(x, y) = \sum_{i=0}^k \sum_{j=0}^k W(i, j) \cdot I(x+i, y+j), \quad (4)$$

where $W(x, y)$ represents the Gaussian weight matrix and σ is the standard deviation of the Gaussian function. $G(x, y)$ represents the value of the Gaussian function at the spatial coordinates (x, y) , k stands for the window size of the Gaussian filter, and I represents the high-frequency feature by the pooling filter. The final output, F_{NSM} , is obtained by summing the denoised high-frequency information with the low-frequency information:

$$F_{NSM} = [HPF_n(v^n)] + G([LPF_n(v^n)]), \quad (5)$$

where $[\cdot]$ indicates the concatenate operation, which integrates feature maps of different cutoff frequencies.

2.2.2. Boundary Refinement Module

As discussed above, our global map S_g was derived from the deepest segment of the network, achieved via partial decoders (PD), which can only capture a relatively rough location of the breast tumor, without structural details. To address this issue, we propose a Boundary Refinement Module (BRM) to progressively mine discriminative breast tumors by erasing foreground objects, as illustrated in Figure 5. Our method can sequentially mine complementary regions and details by erasing the existing estimated tumor regions from high-level side-output features, where the existing estimation is upsampled from the deeper layer. Simultaneously, we introduce axial attention [24] for further saliency analysis of intermediate features, F_m^i , $i \in \{1, 2, 3, 4\}$, which can be represented by the following equation:

$$F_{Axial} = Attention_{axial}(F_m^i). \quad (6)$$

This consideration primarily addresses the complexity of ultrasound images, requiring increased focus on the object regions. The reverse attention weight $W_{Reverse}$ is de facto for salient object detection in the computer vision community [23,25], and can be formulated as

$$W_{Reverse} = \Theta(\sigma(Up(S_g))), \quad (7)$$

where $Up(\cdot)$ denotes an upsampling operation, $\sigma(\cdot)$ is the Sigmoid function, and $\Theta(\cdot)$ is a reverse operation subtracting the input from matrix $\mathbf{E}$, in which all the elements are 1. It is worth noting that the erasing strategy driven by reverse attention can eventually refine the imprecise and coarse estimation into an accurate and complete prediction map. Finally, we obtain the output boundary refinement features F_{BRM} by multiplying the axial attention output feature F_{Axial} by a reverse attention weight $W_{Reverse}$, as below:

$$F_{BRM} = W_{Reverse} \times F_{Axial}. \quad (8)$$

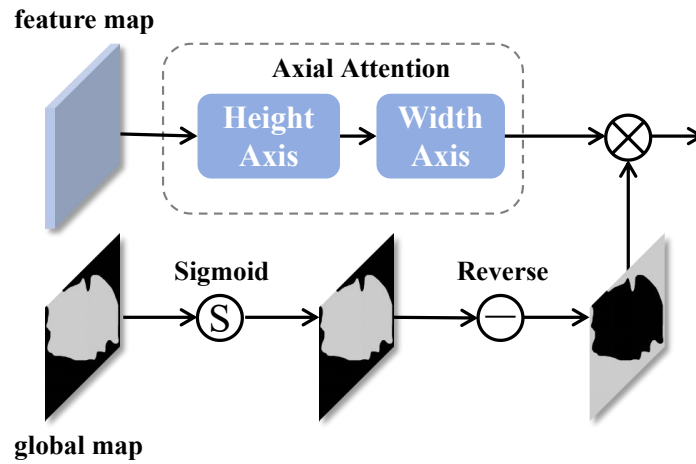


Figure 5. Overall architecture of BRM, which contains reverse attention and axial attention.

2.3. Loss Function

Our loss function comprises the summation of weighted intersection over union (IoU) losses and weighted binary cross entropy (BCE) losses across multiple output layers, represented by the following equation:

$$\mathcal{L}(S, G) = \mathcal{L}_{IoU}^w(S, G) + \mathcal{L}_{BCE}^w(S, G), \quad (9)$$

where S stands for the side-output; G represents the ground truth; and $\mathcal{L}_{IoU}^w(S, G)$ and $\mathcal{L}_{BCE}^w(S, G)$, respectively, denote the weighted IoU loss [26] and weighted binary cross entropy (BCE) loss [27]. Thus, the total loss for the proposed method can be formulated as

$$L_{total} = \mathcal{L}(S_g, G) + \sum_{i=1}^4 \mathcal{L}(S_i, G), \quad (10)$$

where S_g is the global map; S_i represents side-outputs of different scales.

2.4. Experimental Settings

2.4.1. Evaluation Protocols

For quantitative comparison, we report three widely used metrics, including the mean Dice coefficient (mDice), mean intersection over union (mIoU), and mean absolute error (MAE). mDice and mIoU focus on the internal consistency of objects, while MAE represents the average value of the absolute error between the prediction and ground truth.

2.4.2. Implementation Details

We utilized a pre-trained PVT [22] model on ImageNet [28] as the backbone and conducted end-to-end training employing the AdamW optimizer [29]. The initial learning rate was set to 1×10^{-4} and the weight decay was adjusted to 1×10^{-4} too. Further, we resized the input images to 352×352 with a mini-batch size of 8 for 100 epochs. Given the diverse scales of objects in medical imaging, multi-scale training was adopted following

previous work [30]. For AAU-Net, the result was adopted from [4], and we report results for the other models using our re-implementation. For other comparative models within this paper, we derived the results using open-source code, running and testing the models under the same experimental settings. All experiments were carried out utilizing PyTorch [31] on a singular NVIDIA GeForce RTX 3060 GPU boasting 12 GB of memory. For speckle noise in ultrasound images, we did not perform any preprocessing for denoising on the dataset images beforehand. Instead, we solely relied on our method for noise suppression.

3. Results

3.1. Comparison with State-of-the-Art Methods

The quantitative evaluation results on the BUSI dataset are presented in Table 1. For breast tumors with indistinct boundaries, NSBR-Net achieves consistent improvements against baseline models under comparable mDice, mIoU, and MAE values. In clinical diagnosis, the segmentation and localization of malignant tumors are paramount. For malignant tumors, these metrics are improved by 2.3%, 2.39%, and 0.23% compared to the second-best results. Compared to AAU-Net [4] in terms of mDice, our model achieves a 5.32% improvement in malignant tumors.

Table 1. Quantitative comparison of different methods on BUSI [19] to validate our model’s learning ability. ↑ denotes the higher the better, and ↓ denotes the lower the better. Red indicates the best results and blue represents the second-best results.

Method	All			Benign			Malignant		
	mDice ↑	mIoU ↑	MAE ↓	mDice ↑	mIoU ↑	MAE ↓	mDice ↑	mIoU ↑	MAE ↓
UNet [18]	0.6943	0.6033	0.0496	0.7219	0.6362	0.0380	0.6232	0.5183	0.0798
Attention U-Net [32]	0.6934	0.6016	0.0509	0.7247	0.6374	0.0378	0.6125	0.5092	0.0845
UNet++ [1]	0.7023	0.6070	0.0509	0.7212	0.6301	0.0398	0.6538	0.5476	0.0796
UNet3+ [2]	0.7055	0.6139	0.0493	0.7358	0.6433	0.0388	0.6487	0.5414	0.0765
PraNet [23]	0.7698	0.6847	0.0413	0.7841	0.7037	0.0320	0.7330	0.6272	0.0654
DoubleU-Net [33]	0.7735	0.6870	0.0461	0.8016	0.7179	0.0333	0.7010	0.5885	0.0790
UACANet [25]	0.7473	0.6650	0.0442	0.7593	0.6773	0.0353	0.7163	0.6089	0.0672
SANet [34]	0.7708	0.6842	0.0458	0.7929	0.7074	0.0351	0.7136	0.6065	0.0732
UNext [35]	0.7171	0.6258	0.0436	0.7366	0.6509	0.0332	0.6668	0.5613	0.0702
CaraNet [36]	0.7769	0.6968	0.0383	0.7947	0.7199	0.0287	0.7289	0.6267	0.0633
AAU-Net [4]	0.7751	0.6882	–	0.8088	0.7333	–	0.7154	0.6060	–
DuAT [11]	0.8017	0.7163	0.0406	0.7864	0.7037	0.0314	0.7185	0.6104	0.0767
PVT-CASCADE [10]	0.8118	0.7270	0.0380	0.8374	0.7582	0.0245	0.7456	0.6465	0.0619
NSBR-Net (Ours)	0.8183	0.7350	0.0355	0.8375	0.7601	0.0262	0.7686	0.6704	0.0596

We demonstrate the qualitative performance of our model on Dataset B in Table 2. Our model achieves good performance across the entire dataset, with an additional improvement of 3.67% in mDice specifically in the testing of malignant tumors compared to the transformer-based method, DuAT [11].

The visualization of our method and comparative methods on the BUSI dataset and Dataset B is shown in Figure 6. Our method has the best performance in segmenting tumors. For instance, in the case of blurred tumor boundaries (first row and third row), other methods exhibit significant instances of false negatives, while our method addresses this issue well. Specifically, the reason lies in the BRM’s ability to extract complementary regions and details of tumors and enhance the saliency analysis of intermediate features through the attention mechanism, enabling increased focus on object regions. Similarly, under speckle noise (second row and fourth row) conditions, NSBR-Net exhibits no issues with false positives.

Table 2. Quantitative comparison of different methods on Dataset B [21] to validate our model’s learning ability. $\uparrow$ denotes the higher the better, and $\downarrow$ denotes the lower the better. Red indicates the best results and blue represents the second-best results.

Method	All			Benign			Malignant		
	mDice $\uparrow$	mIoU $\uparrow$	MAE $\downarrow$	mDice $\uparrow$	mIoU $\uparrow$	MAE $\downarrow$	mDice $\uparrow$	mIoU $\uparrow$	MAE $\downarrow$
UNet [18]	0.7339	0.6462	0.0251	0.7496	0.6581	0.0174	0.7034	0.6233	0.0401
Attention U-Net [32]	0.7407	0.6590	0.0231	0.7594	0.6786	0.0164	0.7045	0.6211	0.0362
UNet++ [1]	0.7345	0.6493	0.0210	0.7596	0.6776	0.0141	0.6858	0.5945	0.0345
UNet3+ [2]	0.6769	0.5944	0.0247	0.6634	0.5798	0.0213	0.7031	0.6227	0.0313
PraNet [23]	0.7681	0.6714	0.0178	0.8008	0.7010	0.0100	0.7047	0.6140	0.0331
DoubleU-Net [33]	0.7672	0.6765	0.0192	0.7933	0.7016	0.0118	0.7164	0.6277	0.0333
UACANet [25]	0.7600	0.6711	0.0185	0.7917	0.7044	0.0109	0.6984	0.6065	0.0331
SANet [34]	0.7535	0.6722	0.0195	0.7726	0.6899	0.0118	0.7165	0.6378	0.0344
UNext [35]	0.6948	0.5984	0.0208	0.6966	0.5986	0.0139	0.6912	0.5982	0.0343
CaraNet [36]	0.7742	0.6971	0.0171	0.8051	0.7300	0.0096	0.7144	0.6331	0.0315
AAU-Net [4]	0.7814	0.6910	–	–	–	–	–	–	–
DuAT [11]	0.8046	0.7219	0.0161	0.8464	0.7616	0.0085	0.7233	0.6449	0.0308
PVT-CASCADE [10]	0.8119	0.7321	0.0180	0.8586	0.7819	0.0099	0.7213	0.6355	0.0338
NSBR-Net (Ours)	0.8148	0.7308	0.0174	0.8431	0.7624	0.0100	0.7600	0.6694	0.0317

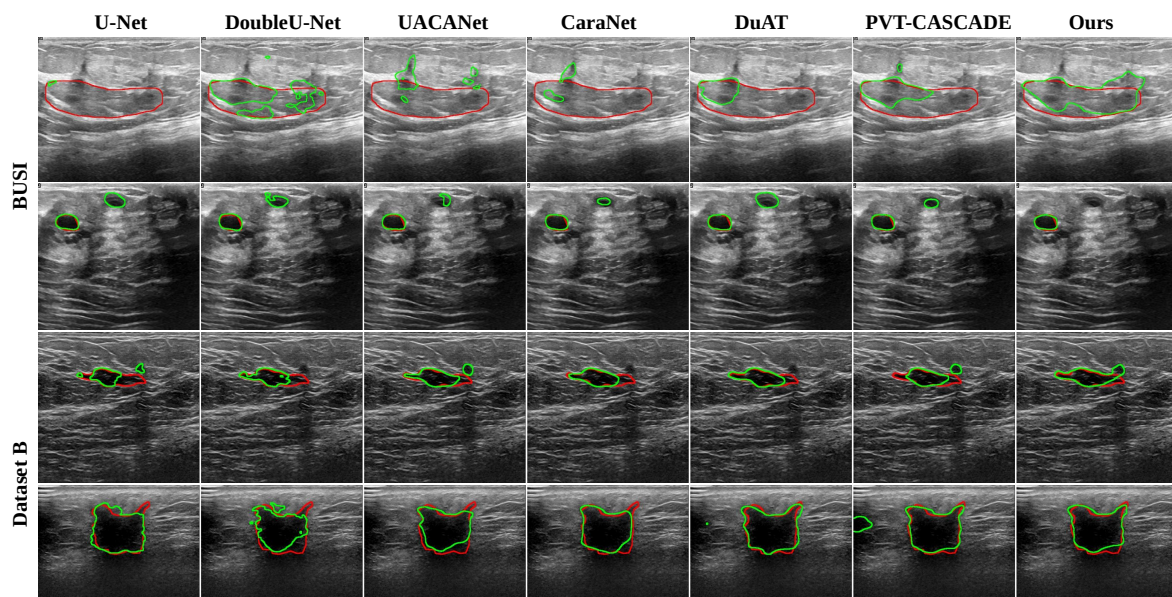


Figure 6. Qualitative comparison of different methods on BUSI [19] (first row and second row) and Dataset B [21] (third row and fourth row). The red curve is the ground truth boundary. The green curve is the segmentation results of these methods.

3.2. Robustness Analysis

External Validation. Due to variations between different datasets, a model may perform well on the training dataset but not generalize effectively to external data. To further assess the robustness of the proposed method in this paper, we utilized STU [37] as an external dataset to evaluate models trained on Dataset B [21]. As demonstrated in Table 3, our method outperforms others on three evaluation metrics in the external validation dataset. Compared to the second-best results, these metrics show improvements of 1.38%, 2.06%, and 0.39%, respectively. These findings suggest that our method exhibits insensitivity to input data variations and has good generalization capabilities.

Table 3. Performance of models trained on Dataset B [21] and evaluated on STU [37]. $\uparrow$ denotes the higher the better and $\downarrow$ denotes the lower the better. Red indicates the best results and blue represents the second-best results.

Method	mDice $\uparrow$	mIoU $\uparrow$	MAE $\downarrow$
UNet [18]	0.7838	0.6834	0.0492
Attention U-Net [32]	0.7476	0.6380	0.0561
CaraNet [36]	0.7708	0.6635	0.0434
DuAT [11]	0.8739	0.7842	0.0275
NSBR-Net (Ours)	0.8877	0.8048	0.0236

3.3. Ablation Study

Effectiveness of Different Network Components. In Table 4, we employ a transformer-based encoder combined with a partial decoder (PD) as our baseline. Note that our partial decoder is only deployed on the high-level features, which achieve an mDice score of 78.79%. This not only showcases the effectiveness of the transformer encoder but also highlights a natural advantage over conventional CNN-based methods. We further investigate the contribution of the Noise Suppression Module. We observe that adding NSM improves the baseline performance, increasing the mDice score from 78.79% to 80.31%. These improvements suggest that introducing an NSM component can enable our model to enhance the quality of global maps. We verify the performance enhancement after integrating the Boundary Refinement Module. A noticeable improvement of 1.93% in the mDice score compared to the baseline is observed. This substantiates that BRM enables our model to accurately distinguish breast tumors. Finally, by simultaneously integrating the two primary components, we achieve a performance boost of 3.04%. This indicates that the fusion of high-quality coarse-grained information with refined boundary recovery is crucial and indispensable for localizing breast tumors.

Table 4. Ablation study on the effectiveness of different components on the BUSI dataset. Red values enclosed in parentheses refer to improvement compared with partial decoder.

PD	NSM	BRM	mDice (%)	mIoU (%)
✓			78.79	70.16
✓	✓		80.31 (+1.52)	71.49 (+1.33)
✓		✓	80.72 (+1.93)	72.10 (+1.94)
✓	✓	✓	81.83 (+3.04)	73.50 (+3.34)

To further validate the boundary refinement capability of BRM, we present the visualization of ablative experiment results in Figure 7. We selected tumors of varying sizes for analysis, with the first and third rows representing small tumors and the second and fourth rows representing large tumors. It can be observed that regardless of the size of the tumor, the segmentation results of the model incorporating BRM (second and fourth columns) are better fitted compared to those without BRM (third column). This further confirms that BRM indeed refines the boundary segmentation results, enabling more accurate localization of lesion boundaries, which is crucial for clinical applications.

Quantitative Comparison of Variants of NSM. As reported in Table 5, when showcasing variations in the NSM module, we aimed to demonstrate the impact of frequency information on speckle noise suppression. It is noteworthy that retaining only the low-pass filter within the NSM resulted in exceptional performance, surpassing an mDice score of 81%. However, introducing high-frequency information led to performance degradation, indicating that the cumulative noise error within the high-frequency data impaired the model's performance. Additionally, leveraging Gaussian filtering on top of the high-frequency filter effectively mitigated noise errors, preserving valuable high-frequency information and contributing to a 0.76% performance boost.

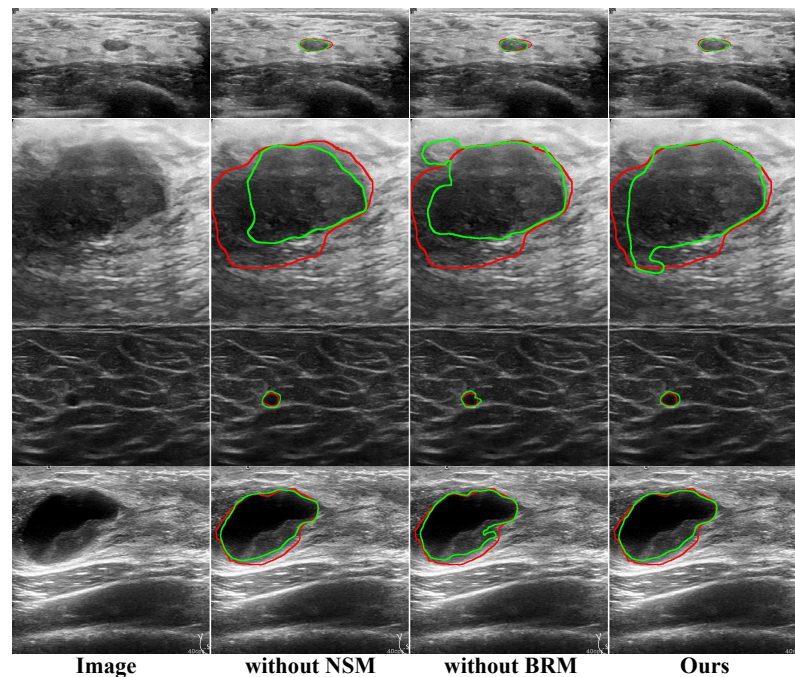


Figure 7. Visualization results of the ablative experiment on BUSI [19] (first row and second row) and Dataset B [21] (third row and fourth row). The red curve is the ground truth boundary. The green curve is the segmentation results of different components.

Table 5. Quantitative comparison of variants of NSM on BUSI dataset. Red values enclosed in parentheses refer to improvement compared with our model with only low-pass filter within NSM, while the blue ones indicates reductions.

LPF	HPF (Without Denoising)	HPF	mDice (%)	mIoU (%)
✓			81.07	72.34
✓	✓		80.31 (−0.76)	71.49 (−0.85)
✓		✓	81.83 (+0.76)	73.50 (+1.16)

Quantitative Comparison of Backbone Architecture. As reported in Table 6, we present the performance of our proposed NSBR-Net method with different backbone architectures on the BUSI dataset. Specifically, we compare three different backbone architectures: UNet [18], Res2Net [38], and PVT (ours) [22], using three evaluation metrics: mDice, mIoU, and MAE. The results demonstrate the effectiveness of our method, showing adaptability and robustness across all backbones. Furthermore, the comparison indicates that the transformer-based encoder is capable of extracting more robust features, leading us to ultimately select PVT as the backbone.

Table 6. Quantitative comparison of backbone architecture on BUSI dataset. ↑ denotes the higher the better and ↓ denotes the lower the better. Red indicates the best results and blue represents the second-best results.

Backbone	mDice ↑	mIoU ↑	MAE ↓
UNet [18]	0.7359	0.6395	0.8758
Res2Net [38]	0.7641	0.6777	0.0413
PVT (Ours) [22]	0.8183	0.7350	0.0355

4. Discussion

Breast cancer is a prevalent gynecological disease which poses a significant threat to women's health. With the development of deep learning, intelligent analysis based on

ultrasound imaging is increasingly widely used in clinical pre-screening and is becoming the mainstream trend. However, the segmentation of breast tumors suffers from the inherent limitations of ultrasound images, including the presence of speckle noise and the issue of indistinct boundaries in malignant tumors.

Speckle noise is inevitable in ultrasound images, causing strong interference during neural network training, thereby reducing the model's generalization ability. Simultaneously, the blurred boundaries of malignant tumors also lead to decreased model accuracy, which is detrimental to intelligent diagnosis. To address these issues, we discussed how to effectively suppress speckle noise from a frequency perspective and designed a coarse-to-fine paradigm, namely NSBR-Net, which shows outstanding segmentation performance and brings new insights into ultrasound image analysis.

The proposed breast tumor segmentation model exhibited superior accuracy over competing algorithms, achieving mDice, mIoU, and MAE scores of 81.83%, 73.50%, and 3.55% on the BUSI dataset and 81.48%, 73.08%, and 1.74% on Dataset B, respectively. Compared to the state-of-the-art transformer-based method, there was a significant improvement of 3.67% on Dataset B and 2.30% on BUSI in mDice, in the testing of malignant tumors, which holds great significance for the clinical diagnosis of cancer. Moreover, as can be seen in Figure 6, NSBR-Net showed no issues of false positives, unlike other segmentation models, which are obviously affected by speckle noise. We attribute this capability to our NSM module, which filters out noise information from coarse-grained feature mappings while preserving detailed boundary information.

Considering the significant individual differences in breast tumors (shown in Figure 6), one of our future research directions will focus on developing appropriate data augmentation algorithms to expand the sample space and enhance the generalization capabilities of our model further. Since our model is designed for two-dimensional ultrasound images, another future direction is to extend the methodology to three-dimensional images. Furthermore, considering the requirement for real-time performance in clinical diagnostic assistance, enhancing the computational efficiency of the model is also a crucial direction for future research. Therefore, we plan to conduct in-depth evaluations and optimizations of the model's computational performance in our future studies.

Compared to ultrasound, mammography is irreplaceable due to its ability to clearly detect tiny calcifications within breast tissue. Thus, mammography is also a commonly employed method for breast cancer screening, with a wealth of AI-related research conducted in this area, including machine learning methods [39] and deep learning methods [40]. Digital Breast Tomosynthesis (DBT), which significantly mitigates the issue of missed detections caused by overlapping breast fibroglandular tissue in mammography, has also emerged as a widely adopted new technology. By combining deep learning methods with other breast cancer screening techniques, our future endeavors aim to broaden the scope of our research, facilitating more comprehensive and accurate diagnostic tools for breast pathology assessment.

5. Conclusions

This paper proposed a novel noise suppression and boundary refinement network (NSBR-Net) for breast tumor segmentation, which utilizes a pyramid vision transformer backbone as the encoder to explicitly extract more powerful and robust features. The core idea is to suppress the cumulative high-frequency internal errors caused by speckle noise and optimize the boundary refinement process from a practical perspective. We validated NSBR-Net's effectiveness through quantitative and qualitative comparisons with current cutting-edge models, demonstrating its superior accuracy overall. We anticipate that this research will inspire further innovative approaches to address the segmentation of breast tumors in ultrasound images.

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

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 42371457, Grant 41971424, Grant 42301468, and Grant 61902330; in part by the Key Project of the Natural Science Foundation of Fujian Province, China, under Grant 2022J02045; in part by the Natural Science Foundation of Fujian Province, China, under Grant 2022J01337, Grant 2022J01819, Grant 2023J01801, Grant 2023J01799, Grant 2022J05157, and Grant 2022J011394; in part by the Natural Science Foundation of Xiamen, China, under Grant 3502Z20227048 and Grant 3502Z20227049; and in part by the Startup Fund of Jimei University under Grant ZQ2022031.

Institutional Review Board Statement: Ethical review and approval were waived for this study because all data used in this study are from public datasets.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

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

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
NSBR-Net	Novel Noise Suppression and Boundary Refinement Network
NSM	Noise Suppression Module
BRM	Boundary Refinement Module
FCN	Fully Convolutional Network
CNN	Convolutional Neural Network
BUS	breast ultrasound
PD	partial decoder
HPF	high-pass filter
LPF	low-pass filter
IoU	intersection over union
BCE	binary cross entropy

References

1. Zhou, Z.; Rahman Siddiquee, M.M.; Tajbakhsh, N.; Liang, J. Unet++: A nested u-net architecture for medical image segmentation. In Proceedings of the Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Granada, Spain, 20 September 2018; pp. 3–11.
2. Huang, H.; Lin, L.; Tong, R.; Hu, H.; Zhang, Q.; Iwamoto, Y.; Han, X.; Chen, Y.W.; Wu, J. Unet 3+: A full-scale connected unet for medical image segmentation. In Proceedings of the ICASSP 2020—2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 1055–1059.
3. Almajalid, R.; Shan, J.; Du, Y.; Zhang, M. Development of a deep-learning-based method for breast ultrasound image segmentation. In Proceedings of the 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 17–20 December 2018; pp. 1103–1108.
4. Chen, G.; Li, L.; Dai, Y.; Zhang, J.; Yap, M.H. AAU-Net: An Adaptive Attention U-Net for Breast Lesions Segmentation in Ultrasound Images. *IEEE Trans. Med. Imaging* **2023**, *42*, 1289–1300. [[CrossRef](#)] [[PubMed](#)]
5. Chen, G.; Li, L.; Zhang, J.; Dai, Y. Rethinking the unpretentious U-net for medical ultrasound image segmentation. *Pattern Recognit.* **2023**, *142*, 109728. [[CrossRef](#)]
6. Hu, H.; Zhang, Z.; Xie, Z.; Lin, S. Local relation networks for image recognition. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 3464–3473.
7. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-local neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7794–7803.
8. Zhao, H.; Jia, J.; Koltun, V. Exploring self-attention for image recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10076–10085.

9. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
10. Rahman, M.M.; Marculescu, R. Medical image segmentation via cascaded attention decoding. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 2–7 January 2023; pp. 6222–6231.
11. Tang, F.; Xu, Z.; Huang, Q.; Wang, J.; Hou, X.; Su, J.; Liu, J. DuAT: Dual-aggregation transformer network for medical image segmentation. In Proceedings of the Chinese Conference on Pattern Recognition and Computer Vision (PRCV), Shenzhen, China, 14–17 October 2023; pp. 343–356.
12. Hu, Y.; Guo, Y.; Wang, Y.; Yu, J.; Li, J.; Zhou, S.; Chang, C. Automatic tumor segmentation in breast ultrasound images using a dilated fully convolutional network combined with an active contour model. *Med. Phys.* **2019**, *46*, 215–228. [[CrossRef](#)] [[PubMed](#)]
13. Lee, H.; Park, J.; Hwang, J.Y. Channel attention module with multiscale grid average pooling for breast cancer segmentation in an ultrasound image. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **2020**, *67*, 1344–1353.
14. Yan, Y.; Liu, Y.; Wu, Y.; Zhang, H.; Zhang, Y.; Meng, L. Accurate segmentation of breast tumors using AE U-net with HDC model in ultrasound images. *Biomed. Signal Process. Control* **2022**, *72*, 103299. [[CrossRef](#)]
15. Maity, A.; Pattanaik, A.; Sagnika, S.; Pani, S. A comparative study on approaches to speckle noise reduction in images. In Proceedings of the 2015 International Conference on Computational Intelligence and Networks, Odisha, India, 12–13 January 2015; pp. 148–155.
16. Qi, W.; Wu, H.; Chan, S. Mdf-net: A multi-scale dynamic fusion network for breast tumor segmentation of ultrasound images. *IEEE Trans. Image Process.* **2023**, *32*, 4842–4855. [[CrossRef](#)]
17. Fan, L.; Zhang, F.; Fan, H.; Zhang, C. Brief review of image denoising techniques. *Vis. Comput. Ind. Biomed. Art* **2019**, *2*, 7. [[CrossRef](#)]
18. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; pp. 234–241.
19. Al-Dhabyani, W.; Gomaa, M.; Khaled, H.; Fahmy, A. Dataset of breast ultrasound images. *Data Brief* **2020**, *28*, 104863. [[CrossRef](#)] [[PubMed](#)]
20. Dong, B.; Wang, P.; Wang, F. Head-free lightweight semantic segmentation with linear transformer. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; Volume 37, pp. 516–524.
21. Yap, M.H.; Pons, G.; Marti, J.; Ganau, S.; Sentis, M.; Zwigglelaar, R.; Davison, A.K.; Marti, R. Automated breast ultrasound lesions detection using convolutional neural networks. *IEEE J. Biomed. Health Inform.* **2017**, *22*, 1218–1226. [[CrossRef](#)] [[PubMed](#)]
22. Wang, W.; Xie, E.; Li, X.; Fan, D.P.; Song, K.; Liang, D.; Lu, T.; Luo, P.; Shao, L. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 568–578.
23. Fan, D.P.; Ji, G.P.; Zhou, T.; Chen, G.; Fu, H.; Shen, J.; Shao, L. Pranut: Parallel reverse attention network for polyp segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Lima, Peru, 4–8 October 2020; pp. 263–273.
24. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 6000–6010.
25. Kim, T.; Lee, H.; Kim, D. Ucanet: Uncertainty augmented context attention for polyp segmentation. In Proceedings of the 29th ACM International Conference on Multimedia, Virtual Event, 20–24 October 2021; pp. 2167–2175.
26. Mátyus, G.; Luo, W.; Urtasun, R. Deeproadmapper: Extracting road topology from aerial images. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 3438–3446.
27. De Boer, P.T.; Kroese, D.P.; Mannor, S.; Rubinstein, R.Y. A tutorial on the cross-entropy method. *Ann. Oper. Res.* **2005**, *134*, 19–67. [[CrossRef](#)]
28. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
29. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. *arXiv* **2017**, arXiv:1711.05101.
30. Dong, B.; Wang, W.; Fan, D.P.; Li, J.; Fu, H.; Shao, L. Polyp-pvt: Polyp segmentation with pyramid vision transformers. *arXiv* **2021**, arXiv:2108.06932.
31. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. Pytorch: An imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 8026–8037.
32. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention u-net: Learning where to look for the pancreas. *arXiv* **2018**, arXiv:1804.03999.
33. Jha, D.; Riegler, M.A.; Johansen, D.; Halvorsen, P.; Johansen, H.D. Doubleu-net: A deep convolutional neural network for medical image segmentation. In Proceedings of the 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS), Rochester, MN, USA, 28–30 July 2020; pp. 558–564.
34. Wei, J.; Hu, Y.; Zhang, R.; Li, Z.; Zhou, S.K.; Cui, S. Shallow attention network for polyp segmentation. In Proceedings of the Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, 27 September–1 October 2021; Springer: Berlin/Heidelberg, Germany, 2021; pp. 699–708.

35. Valanarasu, J.M.J.; Patel, V.M. Unext: Mlp-based rapid medical image segmentation network. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Singapore, 18–22 September 2022; pp. 23–33.
36. Lou, A.; Guan, S.; Ko, H.; Loew, M.H. CaraNet: Context axial reverse attention network for segmentation of small medical objects. In Proceedings of the Medical Imaging 2022: Image Processing, San Diego, CA, USA, 20–24 February 2022; Volume 12032, pp. 81–92.
37. Zhuang, Z.; Li, N.; Joseph Raj, A.N.; Mahesh, V.G.; Qiu, S. An RDAU-NET model for lesion segmentation in breast ultrasound images. *PLoS ONE* **2019**, *14*, e0221535. [[CrossRef](#)] [[PubMed](#)]
38. Gao, S.H.; Cheng, M.M.; Zhao, K.; Zhang, X.Y.; Yang, M.H.; Torr, P. Res2net: A new multi-scale backbone architecture. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 652–662. [[CrossRef](#)] [[PubMed](#)]
39. Cai, S.; Liu, P.Z.; Luo, Y.M.; Du, Y.Z.; Tang, J.N. Breast microcalcification detection algorithm based on contourlet and asvm. *Algorithms* **2019**, *12*, 135. [[CrossRef](#)]
40. Sakaida, M.; Yoshimura, T.; Tang, M.; Ichikawa, S.; Sugimori, H. Development of a Mammography Calcification Detection Algorithm Using Deep Learning with Resolution-Preserved Image Patch Division. *Algorithms* **2023**, *16*, 483. [[CrossRef](#)]

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.