Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Automatic and precise segmentation of medical ultrasound images based on concurrent multi-scale adaptive attention

Gongping Chen

cgp110@mail.nankai.edu.cn

Tianjin Medical University

Qiaolu Wang

Sichuan Technology and Business University

Mengran Fan

Tianjin Medical University

Hongxin Zhang

Tianjin People's Hospital, The First Affiliated Hospital of Nankai University

Ningning Liu

Tianjin Medical University Second Hospital

Yang Zhou

Sichuan Technology and Business University

Research Article

Keywords:

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11752016/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Automatic and precise segmentation of medical ultrasound images based on concurrent multi-scale adaptive attention

Gongping Chen[a,b,1], Qiaolu Wang[c,1], Mengran Fan[a], Hongxin Zhang[d], Ningning Liu[e,*], Yang Zhou[c,*]

[a]*The School of Biomedical Engineering and Technology, Tianjin Medical University, Tianjin, China*
[b]*The department of Medical Imaging, Tianjin Medical University General Hospital, Tianjin, China*
[c]*The College of Artificial Intelligence and Electrical Engineering, Sichuan Technology and Business University, Chengdu, China*
[d]*The department of Ultrasonography, The First Affiliated Hospital of Nankai University, Nankai University, Tianjin, China*
[e]*The department of Ultrasonography, Tianjin Medical University Second Hospital, Tianjin, China*

Abstract

Many convolutional neural networks have been proposed for medical ultrasound image segmentation. Unfortunately, the convolution operation with fixed receptive fields cannot capture robust target characterizations from complex ultrasound images. Moreover, severe heterogeneous structures, similar intensity distributions, ambiguous boundaries, and variable morphologies also bring great challenges to the precise segmentation of medical ultrasound images. In this paper, a unique deep U-shaped framework with the concurrent adaptive attention module is constructed for the robust segmentation of medical ultrasound image. The concurrent adaptive attention module designed in this paper includes two branches of adaptive attention: channel feature adaptive and spatial feature adaptative. During the calibration of the two attention branches, they can adaptively select robust characterizations captured from different convolutional kernels and reduce the introduction of irrelevant features. In addition, the design of multiple convolutional kernels in this attention module not only attenuates the influence of target morphology on the network but also strengthens the representational power of objective regions. The proposed method achieved the best results on four medical ultrasound datasets, with Dice scores of 0.7944, 0.8053, 0.9383, and 0.9082, respectively. Furthermore, the proposed method produced probability-values less than 0.05 based on the Student's t-test for all evaluation metrics. The robustness of our approach on medical ultrasound image segmentation task is perfectly demonstrated through the extensive and reliable experimental results with more than ten state-of-the-art deep learning segmentation approaches on four datasets. Moreover, the comparison with several classical and advanced attention modules further evidences the effectiveness of the concurrent adaptive attention module.

Keywords: Medical ultrasound image, Breast tumor, Kidney cyst, Concurrent adaptive attention, Deep Learning.

1. Introduction

Ultrasound Ultrasound imaging is a common form of medical imaging in clinical applications with the benefits of being highly real-time and non-invasive [1]. In clinical diagnosis, medical ultrasound image segmentation can help the doctor to localize and evaluate tissues or lesions. However, the perturbation of various factors (e.g., heterogeneous structures, similar surrounding tissues, weak boundaries, irregular morphology, artifacts, and object discrepancy) severely reduces the accuracy and speed of manual marking by radiologists, as shown in Fig. 1 [2]. Therefore, the development of automatic and accurate medical ultrasound image segmentation algorithms is very valuable to improve the efficiency and accuracy of lesion diagnosis in clinical practice [3, 4].

To segment target tissues or lesion regions automatically in the medical ultrasound image, many semi-automatic segmentation algorithms have been proposed [5, 6]. For drawing the outline of tumor regions on the breast ultrasound image, Huang et al. employed the typical level-set algorithm [7]. Similarly, Wang et al. employed the level-set module

*Corresponding authors (cgp110@mail.nankai.edu.cn).
[1]These authors contributed equally to this work.

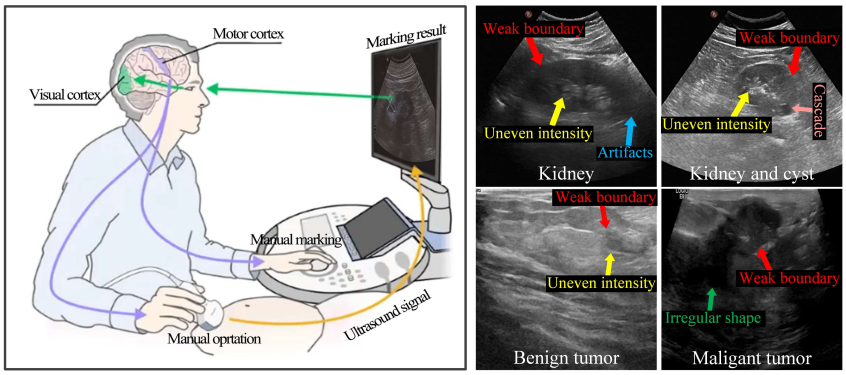

Fig. 1. An illustration of the clinical diagnosis process by the radiologist. The radiologist manually controls the ultrasound probe while marking the objective regions. In addition, various interference factors seriously affect the segmentation accuracy and efficiency.

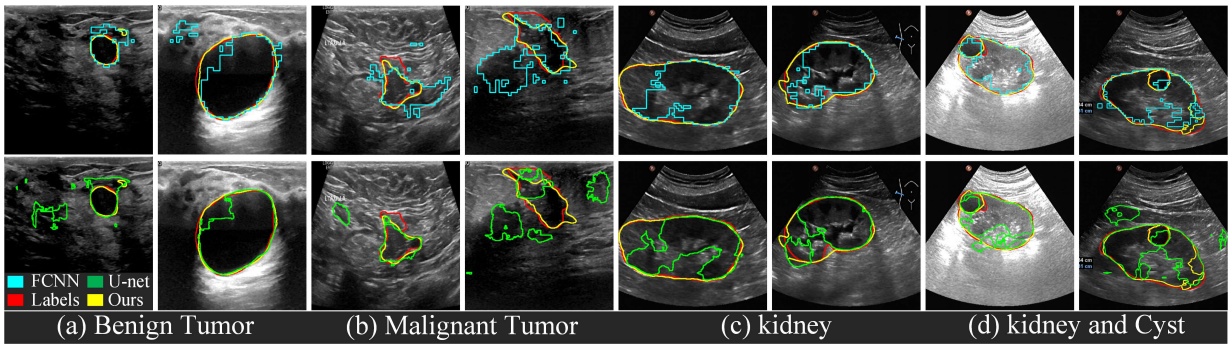

Fig. 2. The diversity of lesions, the complexity of the background, and the limitations of image quality severely restrict the precise and faster segmentation of ultrasound images. Under the perturbation of various factors, simple automatic segmentation algorithms cannot achieve satisfactory results.

to evolve the renal parenchymal area in renal ultrasound images [8]. For optimizing the segmentation accuracy of the level set methodology on ultrasound images, shape prior is added to the evolution action of level-set algorithms. Yang et al. proposed a distance-regularized level-set segmentation framework with shape optimization to segment kidney ultrasound images [9]. However, the level-set algorithm that uses the intensity region information of the image to drive the contour evolution is more sensitive to ultrasound images with uneven intensity. Compared with manual marking,the emergence of semi-automatic segmentation methods has optimized the segmentation procedure and and efficiency of ultrasound images, but still requires manual initialization procedures.

In recent years, CNNs have been extensively investigated and achieved satisfactory performance on the automatic segmentation task of medical images [10]. Zhang et al. designed a deep learning framework for lymph node ultrasound using two FCN networks [11]. Wu et al. suggested a cascaded FCN framework to achieve the automatic measurement of fetal head circumference in prenatal ultrasound images [12]. However, the FCN-based automatic segmentation method loses the location information of the target during the pooling procedure. To capture positional information, Ronneberger proposed the U-net with symmetric architecture [13]. Behboodi et al. evaluated the contribution of U-net directly on simulated and real medical ultrasound image segmentation tasks [14]. Subsequently, a more systematic analysis of U-net and FCN architectures performance for medical ultrasound image segmentation was presented by Yap et al. [15].The emergence of FCN and U-net significantly improves the segmentation efficiency and accuracy compared to existing semi-automatic segmentation algorithms. However, the complexity of ultrasound patterns still severely affects the results of these simple networks for lesion region segmentation, as shown in Fig. 2. In order to further improve the segmentation performance of medical ultrasound images, many optimization strategies have been added to FCN and U-net. Mishra et al. presented a depth-supervised FCN focusing on the objective boundary to refine the segmentation process of the network [16]. Kim et al. constructed a deep learning network with the multi-scale input block and hybrid loss function to segment coronary arteries from ultrasound images [17]. Although existing

deep learning algorithms have shown excellent segmentation performance on ultrasound images, these architectures are often performed convolution actions with fixed receptive fields. Li et al. pointed out that convolution operations on a single receptive field cannot fully capture object features and proposed the SKNet network to select target features along the channel dimension in multi-scale receptive fields [18]. However, ultrasound images of different tissues or lesions are significantly variable, which also requires multi-scale receptive field features in spatial dimensions to cope with this challenge. Subsequently, Chen et al. were inspired by SKNet and attempted to design multi-scale receptive field adaptive selective attention AAU-net [19] and ESKNet [20] along the channel and spatial dimensions. However, the proposed AAU-net cascade channel and spatial features, ignoring adaptive selection in multi-receptive fields along both channel and spatial dimensions simultaneously. Although ESKNet employs a parallel architecture, it uses only two scales of receptive fields, which have limited feature capture capabilities. Furthermore, ESKNet introduces additional deep supervisory constraints and does not optimize the network's backbone architecture, resulting in greater computational overhead.

In this work, a unique concurrent adaptive attention module (CAAM) with multi-kernel convolution is first developed to enhance the generalization ability of segmentation methods to multi-variable input ultrasound images, as shown in Fig. 3. Subsequently, we constructed a novel U-shaped network using the developed CAAM module. Compared with the existing typical attention methods SKNet, ESKNet and AAU-net, this method has further improved its performance in ultrasound segmentation tasks. Compared to AAU-net and ESKNet, the computational overhead has been further reduced. The most important innovations of this study include four aspects:

- First, a concurrent adaptive attention module (CAAM) is proposed, which not only captures object characterizations under multiple receptive fields but also can adaptively calibrate the captured features from the channel and spatial dimensions.
- Second, a novel U-shaped network with the CAAM block is developed to segment diverse medical ultrasound images.
- Third, extensive experiments with state-of-the-art CNN modules and attention models on the available dataset fully illustrated the robustness of our approach in ultrasound images segmentation.
- Moreover, we constructed a publicly available kidney and cyst ultrasound dataset (KCUS). The dataset contains clinical images of 300 patients. These images were manually annotated by experienced urologists.

2. Related Works

2.1. Breast ultrasound segmentation

In terms of women, the incidence of breast cancer ranks first among cancersand its impact is enormous [21]. To improve the characterization of tumor regions, many optimization strategies have been applied to tumor lesion segmentation. For example, a novel attention-enhanced U-net (AE U-net) was presented by Yan et al. to achieve the automatic segmentation of breast lesion regions [22]. For capturing tumor features under the larger receptive field, Zhuang et al. constructed an RDAU-Net [23]. In this network, dilated convolution and residual connection were used jointly to construct the framework. Lyu et al. proposed AMS-PAN by combining multi-scale modules and attention mechanisms to realize breast ultrasound image segmentation automatically [24]. Shareef et al. constructed an aware network using dense multi-scale convolution to overcome the challenge of segmenting small breast tumors [25]. Inspired by M-net, Abraham et al. by aggregating input pyramid blocks and deep supervision blocks designed a novel MADU-net to automatically segment tumor areas [26]. However, the fixed convolutional kernel size in existing methods can only capture the objective features under the single receptive field [27]. Subsequently, a new selective kernel U-shaped network (SKNet) was constructed by Byra et al. for segmenting breast ultrasound images by integrating a selective kernel network [27]. Although the propose of SKNet has refined the segmentation performance of the tumor region, the effect of spatial dimensional calibration on segmentation accuracy is ignored. In this study, our method can adaptively select robust lesion representations under different receptive fields to enhance the characterization of the tumor region.

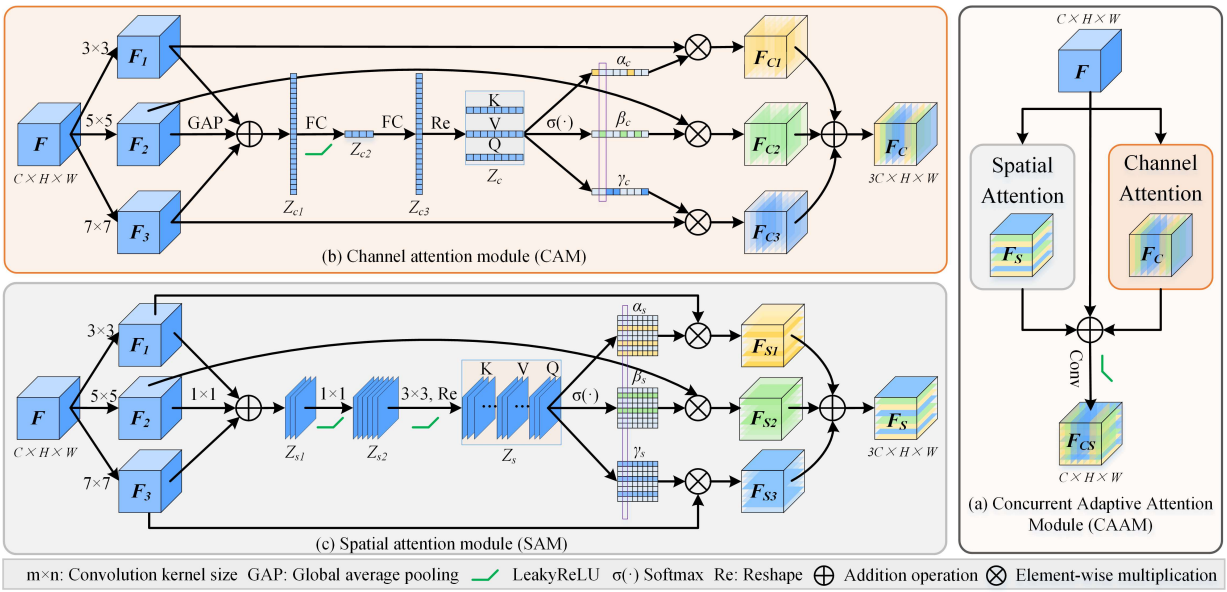

Fig. 3. The illustration of the concurrent adaptive attention module (CAAM), which is composed of three parts: spatial attention block, channel attention block, and residual connections. Three branches of adaptive calibration are included in channel attention and spatial attention, respectively (see Section 3).

2.2. Kidney ultrasound segmentation

Automated and accurate kidney ultrasound image segmentation is very helpful in the clinical assessment of kidney function and detection of kidney disease [28]. Achieving accurate kidney ultrasound image segmentation remains a challenge due to the perturbation of the various kidney shapes and the uneven intensity distribution [29, 30]. To alleviate the disturbance of kidney morphology, Yin et al. achieved the automatic segmentation of renal ultrasound images by regression of renal boundaries [28]. Chen et al. constructed a structural feature and detail feature fusion network (SDFNet) to automatically segment the kidneys [30]. In this work, they designed a kidney boundary detection module to constrain the extraction of kidney structural features. Alex et al. presented a YSegNet combined with a boundary extraction framework [31]. As shown in Fig. 2, fitting kidney boundaries from blurred, cascaded, and crippled ultrasound images is an extremely challenging job. To further improve the segmentation accuracy of kidneys, Chen et al. developed a multi-branch convolutional neural network (MBANet) using a three-branch encoder with shared weights to extract robust kidney characterizations [32]. In the same year, a network integrating multi-scale inputs, pyramidal modules, and deep supervised constraints (MSDSNet) was constructed to segment kidneys automatically [33]. Although directly aggregating the features from multiple convolutional kernels can improve the the representation capability of the network, it inevitably introduces a large number of features that are not relevant to the object [27]. In this study, our method is able to choose more robust target features adaptively in terms of channel and spatial dimensions and suppress the introduction of irrelevant features.

3. Method

3.1. Overview

To accurately segment the objective tissue or the lesion region from the complex medical ultrasound image, we developed a concurrent adaptive attention model (CAAM) that adaptively selects robust object characterizations from channel dimension and spatial dimension, as shown in Fig. 3. Subsequently, we constructed a unique U-shaped architecture using the CAAM module to implement the automatic segmentation of medical ultrasound images. As shown in Fig. 4, the U-shaped network contains fifteen CAAM units, seven down-sampling procedures, seven skip-connection operations, and seven up-sampling programs. In the encoding stage, the filter sizes of CAAM units are 32, 64, 64, 128, 128, 256, 256, 512. Unlike the existing U-shaped network, each encoding and decoding stage of our

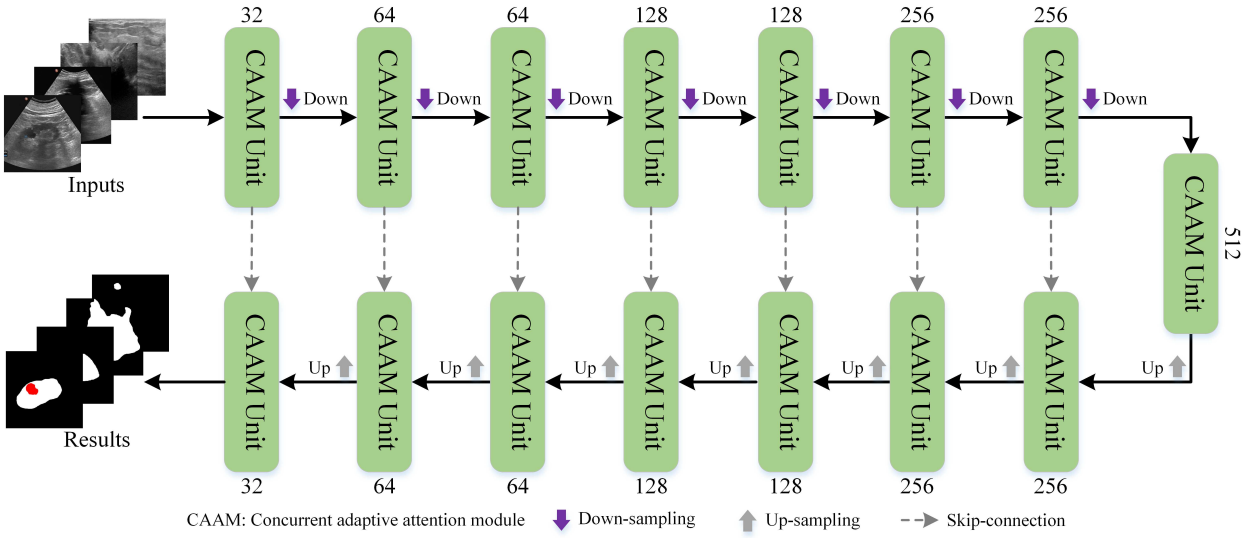

Fig. 4. The U-shaped network constructed using the concurrent adaptive attention module (CAAM).

network only consists of one CAAM module. As shown in Fig. 3(a), the CAAM is composed of channel attention, spatial attention, and residual connection. Specifically, the feature map F input to the CAAM is subjected to channel-wise calibration and spatial-wise calibration to obtain the feature maps F_s and F_c, respectively. subsequently, the input feature F is concatenated with the calibrated feature maps F_s and F_c. Finally, the merged feature map is executed with a convolution operation, a batch normalization, and an activation operation to obtain the final output of CAAM.

$$F_c = CAM(F), \tag{1}$$

$$F_s = SAM(F), \tag{2}$$

$$F_{cs} = \delta_r(B(W \cdot (F \oplus F_c \oplus F_s))), \tag{3}$$

where $CAM(\cdot)$ and $SAM(\cdot)$ are defined as the channel attention module and the spatial attention module, respectively. $\oplus$ represents as the concatenate operation. W is the matrix of the convolution operation, $B(\cdot)$ indicates the batch normalization action, and $\delta_r(\cdot)$ denotes the LeakyReLU operation. The detailed description of $SAM(\cdot)$ and $CAM(\cdot)$ are demonstrated in Section 3.

3.2. The adaptive calibration of channel Dimension

The calibration procedure of the channel attention, as shown in Fig. 3(b). The calibration procedure of channel attention is described as follows. The input feature maps F are first subject to three kernel-size convolution operations to obtain the object feature maps F_1, F_2, and F_3 with different receptive fields. Subsequently, we use the global average pooling operation to compress the feature maps F_1, F_2, and F_3 from the three receptive fields, respectively. Finally, these compressed feature maps are subjected to a combined operation to obtain a new set of feature vectors:

$$Z_{c1} = GAP(F_1) \oplus GAP(F_2) \oplus GAP(F_3), \tag{4}$$

where $Z_{c1} \in \mathbb{R}^{3C\times 1}$ represents the feature vector obtained by fusing three branches. $\oplus$ denotes the concatenate action.

Then, the feature vector Z_{c1} is performed by the full convolution operation, the LeakyReLU activation, the full convolution operation, and the reshaped process to generate a new group of feature vectors:

$$Z_c = Re(W_{fc2} \cdot (\delta_r(W_{fc1} \cdot Z_{c1}))), \tag{5}$$

where $W_{fc2} \in \mathbb{R}^{3C\times 1}$ is the matrix of the second full convolution operation, $Re(\cdot)$ indicates the reshape action, $\delta_r(\cdot)$ denotes the LeakyReLU operation, and $W_{fc1} \in \mathbb{R}^{d\times 1}$ represents the matrix of the first full convolution operation. d is squeezed dimension, which can be denoted as:

$$d = \max(C/r, \mathrm{L}), \tag{6}$$

where r is a reduction rate and its value is 8. L represents the minimum value of the reduction rate, which was set to 32 in our study [18].

Finally, Z_c is performed the softmax activation to obtain the channel-wise activation maps of F_1, F_2, and F_3, respectively. They can be expressed as:

$$\alpha_c = \frac{e^{K_c}}{e^{K_c} + e^{V_c} + e^{Q_c}}, \tag{7}$$

$$\beta_c = \frac{e^{V_c}}{e^{K_c} + e^{V_c} + e^{Q_c}}, \tag{8}$$

$$\gamma_c = \frac{e^{Q_c}}{e^{K_c} + e^{V_c} + e^{Q_c}}, \tag{9}$$

$$\alpha_{ci} + \beta_{ci} + \gamma_{ci} = 1, \tag{10}$$

where K_c, V_c, and $Q_c \in \mathbb{R}^{C\times 1}$. α_c, β_c, and $\gamma_c \in \mathbb{R}^{C\times 1}$ indicate the channel activation maps of F_1, F_2, and F_3. Note that α_{ci} is the $i-th$ element of α_c. The significance of the channel-wise feature in $F_1/F_2/F_3$ can be represented by the value in α_c, β_c, and γ_c. Subsequently, we combined the calibrated feature maps as an output of channel attention,

$$F_c = \alpha_c \cdot F_1 \oplus \beta_c \cdot F_2 \oplus \gamma_c \cdot F_3, \tag{11}$$

where $F_c \in \mathbb{R}^{3C\times H\times W}$ is the output feature map of the channel module, and $\oplus$ denotes the concatenate action.

3.3. The adaptive calibration of spatial Dimension

As shown in Fig. 3(c), the spatial attention module also consists of three calibration branches. Similar to channel attention, the feature map F of the input spatial attention first undergoes three kernel-size convolution operations to obtain the objective feature maps F_1, F_2, and F_3 under different receptive fields. spatial attention calibration is performed in the spatial dimension. Therefore, we compress the feature maps F_1, F_2, and F_3 by utilizing three convolution operations with filter size 1 instead of global average pooling. Subsequently, we merge the three-compressed feature maps to obtain a set of new feature maps:

$$Z_{s1} = W \cdot F_1 \oplus W \cdot F_2 \oplus W \cdot F_3, \tag{12}$$

where $Z_{s1} \in \mathbb{R}^{3\times H\times W}$ indicates the feature map obtained by combining three branches, and $W \in \mathbb{R}^{1\times H\times W}$ denotes the matrix of 1×1 convolution action.

Then, the feature map Z_{s1} is performed by the 1×1 convolution operation, the LeakyReLU activation, the 3×3 convolution operation, the LeakyReLU activation, and the reshaped process to generate a new group of feature vectors:

$$Z_s = Re(\delta_r(W_{3\times 3} \cdot (\delta_r(W_{1\times 1} \cdot Z_{s1})))), \tag{13}$$

where $Re(\cdot)$ indicates the reshape operation, $\delta_r(\cdot)$ denotes the LeakyReLU activation, $W_{3\times 3} \in \mathbb{R}^{1\times H\times W}$ is the matrix of the 3×3 convolution operation, and $W_{1\times 1} \in \mathbb{R}^{1\times H\times W}$ represents the matrix of the 1×1 convolution action. The filter size of the 3×3 convolution is set to d, and its definition can be found in Section 3.

Finally, Z_s is performed the softmax activation to obtain the spatial-wise activation maps of F_1, F_2, and F_3. Similarly, we use α_s, β_s, and γ_s represent the spatial activation map of F_1, F_2, and F_3, respectively. They can be expressed as:

$$\alpha_s = \frac{e^{K_s}}{e^{K_s} + e^{V_s} + e^{Q_s}}, \tag{14}$$

$$\beta_s = \frac{e^{V_s}}{e^{K_s} + e^{V_s} + e^{Q_s}}, \tag{15}$$

$$\gamma_s = \frac{e^{Q_s}}{e^{K_s} + e^{V_s} + e^{Q_s}}, \tag{16}$$

$$\alpha_{si} + \beta_{si} + \gamma_{si} = 1, \tag{17}$$

where K_s, V_s, and $Q_s \in \mathbb{R}^{C\times H\times W}$. α_s, β_s, and $\gamma_s \in \mathbb{R}^{C\times H\times W}$ indicate the spatial activation maps of F_1, F_2, and F_3. Note that α_{si} is the $i-th$ element of α_s. The significance of the spatial-wise feature in $F_1/F_2/F_3$ can be represented by the value in α_s, β_s, and γ_s. Subsequently, we combined the calibrated feature maps as an output of spatial attention,

$$F_s = \alpha_s \cdot F_1 \oplus \beta_s \cdot F_2 \oplus \gamma_s \cdot F_3, \tag{18}$$

where $F_s \in \mathbb{R}^{3C\times H\times W}$ is the final output by fusing the calibrated spatial feature.

4. Datasets and Experiments

4.1. Breast ultrasound Images

We performed the segmentation of breast tumors on two publicly available breast ultrasound datasets. The first breast ultrasound dataset is named BUSI, which consists of 780 images acquired by Al-Dhabyani et al. using two ultrasound devices [34]. In BUSI, the numbers of benign and malignant masses are 437 and 210, respectively. The remaining are normal breast ultrasound images. The second breast ultrasound dataset is named Dataset B, which consists of 163 images [35]. The benign and malignant masses in Dataset B are 110 and 53. In this work, these breast ultrasound images were resize to 384×384. These two breast ultrasound datasets are publicly available.In this paper, the four-fold cross-validation experiments were implemented on BUSI and dataset B, respectively.

4.2. Kidney ultrasound Images

We constructed two kidney ultrasound datasets by retrospective study under the ethical regulation (Tianjin Medical University ethics board, No.KY2025K114, 2025/2/26). The first kidney ultrasound dataset (named KUS) contains 300 ultrasound images of 300 patients used only for kidney tissue region segmentation. These kidney ultrasound images are collected by Esaote MyLab and Philips EPIQ7 ultrasound devices. In KUS, the case images acquired by single ultrasound device are 150. The second kidney ultrasound dataset (named KCUS) is used for kidney and cyst segmentation, which includes 300 ultrasound images from 300 patients collected by Esaote MyLab, Hitachi, and Philips EPIQ7 ultrasound machines. In KCUS, the case images acquired by single ultrasound device are 100. In KUS and KCUS, the kidney and cyst masks were obtained by two clinicians using Labelme manual annotation. These two doctors have extensive clinical experience and specialize in the interpretation of ultrasound images. Similarly, these kidney ultrasound images were resized to 384×384. In the experiments, KUS and KCUS datasets are executed with three-fold cross-validation training, respectively. The KUS and KCUS datasets are available by email contact.

Table 1
Ablation study of different network components. The top 2 scores are marked with bold and green text, respectively. CAM: channel attention module, SAM: spatial attention module, CAAM: concurrent adaptive attention module.

	BUSI				KCUS				Complexity	
Methods	Jaccard	Precision	Recall	Dice	Jaccard	Precision	Recall	Dice	Params	GFLOPs
U-net (baseline)	60.70±2.36	71.88±2.41	76.30±2.48	70.10±2.20	71.34±0.88	82.58±1.00	84.58±0.96	82.03±0.70	7.85	62.82
U-net + CAM with two branches	68.10±1.63	78.62±1.66	79.53±1.93	76.92±1.57	80.51±1.33	88.41±0.34	90.56±1.49	88.41±0.92	37.23	295.74
U-net + CAM with three branches	68.75±1.99	77.99±2.49	81.89±0.94	77.25±2.09	81.79±1.62	89.47±1.27	90.91±1.21	89.15±1.11	80.2	637.61
U-net + SAM with two branches	67.26±3.11	76.59±3.41	78.74±2.33	75.28±3.04	80.84±1.40	87.52±1.76	91.28±0.35	88.00±1.37	39.01	338.5
U-net + SAM with three branches	67.67±2.20	76.18±1.77	81.19±2.86	75.55±2.33	81.89±1.28	88.29±1.82	91.37±0.94	88.68±1.09	82.88	701.74
U-net + CAAM (No simplified)	70.93±2.38	80.14±2.01	82.03±2.46	79.28±2.17	83.39±1.08	91.22±0.67	91.47±0.77	90.21±0.80	86.75	726.8
U-net + CAAM (Simplified)	**71.29±2.07**	**80.87±2.02**	**82.22±2.00**	**79.44±2.07**	**84.38±1.14**	**91.57±0.16**	**91.73±1.42**	**90.82±0.95**	85.75	343.65

4.3. Evaluation metrics

We use eight metrics to quantitatively evaluate experimental results. They are Dice, Specificity, Precision, Jaccard, Recall, ABD, ASSD, and HD [19]. In BUSI and Dataset B, it can occur that the tumor region is completely misidentified as background. Therefore, the three boundary-based evaluation indexes of ABD, ASSD and HD are only used in KUS and KCUS.

4.4. Experimental settings

During the training process, we conduct data augmentation on the training data for each fold. Our data augmentation procedures consist of blurring, adding Gaussian noise, flipping, random gamma transformation, elastic transformation, and random rotation. The detailed parameter settings of our data augmentation procedures are publicly available after the acceptance of our manuscript. In this work, the training data of each fold is expanded by 20 times (Assuming that the original training data of the first fold has 200 images and the training data after data augmentation contains 4000 images.). The data augmentation algorithm is available in our resource code (Theresource code are available by email contact.). All comparison methods are adequately retrained on the same training data. In all comparison experiments, the prediction masks are not subjected to any post-processing.

In this study, we use binary-cross entropy (BCE) to train our segmentation network. Adam is defined as the optimizer of the segmentation network (learning rate=0.001). 50 and 10 are chosen as the values for epochs and batch size, respectively. We use keras integrated in tensorflow to build the segmentation method in this paper. The version of tensorflow is 2.6.0. python 3.6 is our development language. We use NVIDIA RTX 3090 as our calculation resource.

5. Experimental Results

5.1. Architecture ablation

We performed ablation experiments on BUSI and KCUS to demonstrate the performance of different network components. The U-net network was used as the benchmark network. According to the quantitative evaluation indicators demonstrated in Table 1, several findings can be drawn: (1) The introduction of the CAM module and the SAM module significantly improves the characterization capability of the baseline. Comparing to the baseline, the addition of CAM with three branches improves the four metrics by 13.3%, 8.5%, 7.3%, and 10.2% on BUSI and by 14.6%, 8.3%, 7.5%, and 8.7% on KCUS. Compared to the baseline, the introduction of SAM with three branches improves the four metrics by 11.5%, 6.0%, 6.4%, and 7.8% on BUSI and by 14.8%, 6.9%, 8.0%, and 8.1% on KCUS. (2) The segmentation capability of CAM is slightly better than the SAM module. Compared to channel attention, spatial attention has more network parameters and computational costs. (3) In the CAM and SAM modules, the characterization capability of the three branches is better than the two branches. The design of the three-branch architecture significantly increases the network parameters and computational overhead compared to the two-branch framework. (4) Based on the quantitative segmentation results of "U-net + CAAM (No simplified)", it can be obtained that the concurrent introduction of spatial and channel attention blocks could help the network to achieve better representation capability.

The complexity of the CAAM module leads to the construction of U-shaped networks with more network parameters and larger computational costs, as shown in "U-net + CAAM (No simplified)". In this work, we conducted some simplified activities for the "U-net + CAAM (No simplified)". Specifically, we first reduced the number of CAAM units from two to one for each convolutional module in the U-shaped network (see Section 3.1). Subsequently, we increased the depth of the network (from 9 to 15) to balance the characterization capability of the U-shaped framework (see Section 3.1). Compared with the original "U-net + CAAM (No simplified)", the introduction of the simplified activity not only boosts the characterization ability of the developed network but also significantly discounts the computational overhead, as shown in "U-net + CAAM (Simplified)". Compared to the baseline, our simplified approach improves the four metrics by 17.5%, 12.5%, 7.6%, and 13.3% on BUSI and by 18.3%, 10.9%, 8.5%, and 10.7% on KCUS.

5.2. Comparison with state-of-the-art methods

Thirteen SOTA deep-learning methods are used to the comparison experiments in this paper, which include U-net [13], SegNet [36], Att U-net [37], U-net++ [38], UNETR [39], STAN [25], AE U-Net [22], RDAU-Net [23], MADU-net [26], SKU-net [27], SDFNet [30], MBANet [32], and MSDSNet [33]. Among them, U-net, SegNet, U-net++, and Att U-net are the typical network used to segment medical images. UNETR is an architecture based on transformer attention. STAN, AE U-Net, RDAU-Net, MADU-net, and SKU-net are five SOTA breast ultrasound segmentation methods. SDFNet, MBANet, and MSDSNet are three SOTA kidney ultrasound segmentation methods. Based on the open-source codes, we retrain these comparison approaches on the same datasets as our model. The prediction masks

Table 2
The quantitative evaluation results (mean ± std) on BUSI and Dataset B datasets. The top 2 scores are marked with bold and green text, respectively. "†" represents the medicine-specific method.

	BUSI					Dataset B				
Methods	Dice	Jaccard	Recall	Specificity	Precision	Dice	Jaccard	Recall	Specificity	Precision
U-net	70.10±2.2	60.70±2.36	76.30±2.48	96.18±0.55	71.88±2.41	68.20±4.23	58.44±4.26	75.32±2.85	98.44±0.40	70.27±6.11
Att U-net	67.99±1.18	57.09±1.22	66.97±4.08	96.87±0.83	78.78±4.67	69.30±4.07	59.93±4.53	76.15±4.21	98.43±0.33	70.40±6.05
UNETR	72.99±1.67	63.86±1.78	80.21±1.60	96.65±0.65	73.44±2.85	69.64±5.28	59.11±5.76	78.17±4.31	98.04±0.53	70.39±3.15
U-net++	71.58±2.09	61.38±1.73	71.44±2.77	97.04±0.54	79.68±3.07	69.77±5.30	61.19±5.86	79.64±3.84	98.44±0.41	68.32±5.73
SegNet	75.64±1.80	67.31±1.87	79.85±1.03	96.99±0.53	76.09±2.00	72.16±1.52	62.83±2.20	80.15±3.90	98.59±0.30	71.72±1.70
STAN†	73.04±2.95	64.10±3.05	78.39±2.16	96.64±0.67	73.96±3.30	66.06±4.24	57.09±3.92	69.95±6.17	98.58±0.47	67.71±3.11
RDAU-net†	71.94±3.46	63.75±3.36	78.90±1.35	96.63±0.76	71.25±4.11	68.22±4.94	58.17±4.91	73.55±5.28	98.37±0.39	70.49±4.26
MADU-net†	71.35±2.67	61.62±2.69	76.87±2.58	96.40±0.62	73.77±2.90	72.32±3.14	63.09±3.04	79.24±1.72	98.61±0.36	73.70±5.08
AE U-net†	73.47±3.03	64.57±2.91	79.00±2.11	96.80±0.54	74.44±3.74	72.23±2.14	62.37±2.16	78.97±2.29	98.67±0.28	72.27±1.91
SKU-net†	76.92±1.57	68.10±1.63	79.53±1.93	97.33±0.45	78.62±1.66	73.53±4.05	64.25±4.01	79.36±2.50	98.68±0.39	75.27±6.70
Ours	**79.44±2.07**	**71.29±2.07**	**82.22±2.00**	**97.60±0.52**	**80.87±2.02**	**80.53±1.16**	**71.89±2.30**	**83.58±2.16**	**99.04±0.31**	**81.82±1.00**
P-value	4.91E-09	2.20E-11	2.13E-03	4.53E-04	3.73E-02	1.58E-04	6.06E-05	3.16E-02	2.50E-02	4.59E-05

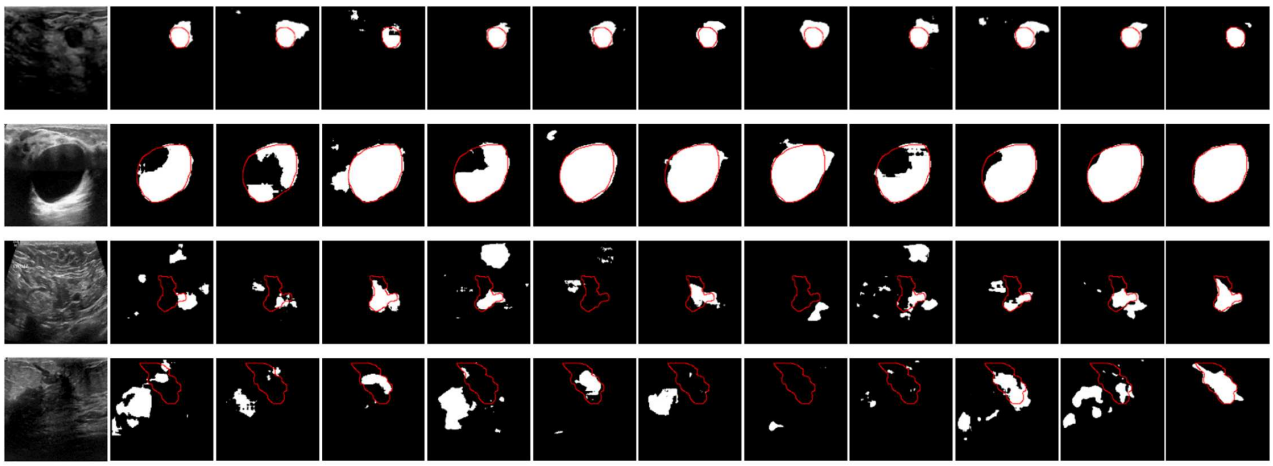

Fig. 5. The visualization of prediction mask of different methods on BUSI and Dataset B. These images suffer from interference of variable shapes, weak boundaries, cascades, homogeneous structures, and similar surrounding tissue.

from all methods are not exposed to any post-processing activities to ensure absolute objectivity of the experimental results.

The quantitative results of BUSI and Dataset B are illustrated in Table 2. The quantitative evaluation results indicated that the proposed approach obtained the most satisfactory segmentation results on breast ultrasound images. In addition, we conducted the statistical analysis of t-tests for the quantitative assessment results of the different methodologies. Specifically, we adjudicated the method excellence by the p-value obtained from the t-test with the second methods. The p-value ($p < 0.05$) based on t-tests indicate that the scores of our method on these indicators are significantly different from the comparison method. Although UNETR, U-net++, and SegNet achieved second results in individual metrics, they scored unsatisfactorily in other indexes, as shown in Table 2. It is noteworthy that SKU-net achieved the second result on multiple metrics and it is a competitive methodology. Except for Att U-net, MADU-net, and Ours, the other methods achieved better segmentation performance than Dataset B on BUSI. This demonstrates that these methods are more sensitive to the input images and less robust. This demonstrates that these methods are more responsive to the variable tumors and less robust. The visualized prediction masks of different comparison

Table 3
The quantitative evaluation results (mean ± std) on KUS and KCUS datasets. The top 2 scores are marked with bold and green text, respectively. "†" represents the medicine-specific method.

Methods	U-net	Att U-net	UNETR	U-net++	SegNet	AE U-net	SKU-net	SDFNet†	MSDSNet†	MBANet†	Ours	P-value
Dice	91.57±0.86	92.10±0.92	89.20±1.15	91.78±1.01	91.90±0.60	90.79±2.61	92.91±0.25	91.88±1.39	92.91±0.55	93.35±0.87	**93.83±0.31**	3.80E-02
Jaccard	85.00±1.39	85.99±1.50	81.42±1.67	85.59±1.53	85.78±0.98	83.90±4.03	87.16±0.42	85.81±1.94	87.28±0.80	88.03±1.28	**88.64±0.53**	4.88E-02
Recall	91.11±0.79	93.70±0.32	91.44±1.46	93.22±0.22	92.31±0.89	90.91±5.88	93.36±0.93	94.43±1.00	92.84±1.36	94.03±0.61	**94.48±0.73**	4.97E-02
Specificity	98.77±0.60	98.50±0.46	97.91±0.65	98.28±0.51	98.67±0.63	98.54±0.44	98.72±0.37	98.35±0.63	98.82±0.57	98.70±0.57	**98.98±0.26**	4.86E-02
Precision	93.03±1.86	91.56±1.99	88.22±2.31	90.33±1.69	92.66±1.47	91.98±1.22	93.23±1.36	90.74±3.12	93.90±1.79	93.30±1.65	**94.16±1.17**	4.82E-02
HD	57.38±17.83	52.14±8.22	82.73±5.34	57.42±4.92	42.43±13.41	53.32±33.08	31.57±9.00	31.26±8.50	23.79±6.57	24.27±11.03	**15.64±4.93**	9.57E-04
ASSD	2.09±0.99	1.03±0.04	1.95±0.77	0.79±0.21	0.71±0.04	2.17±2.42	0.95±0.27	0.87±0.17	1.07±0.69	0.72±0.09	**0.50±0.05**	2.46E-02
ABD	9.62±3.43	6.39±0.97	10.36±1.63	6.28±0.81	5.03±0.85	8.36±5.13	5.33±0.55	5.84±0.59	5.16±1.21	4.24±0.68	**3.60±0.38**	2.65E-02
Dice	82.03±0.70	82.52±1.44	66.49±4.12	84.01±2.35	87.32±1.56	87.39±1.34	88.41±0.92	82.76±1.34	89.31±0.58	88.12±1.51	**90.82±0.95**	6.44E-04
Jaccard	71.34±0.88	72.15±1.71	51.60±4.63	74.93±2.96	79.94±1.52	79.43±1.38	80.51±1.33	73.13±1.57	82.19±0.63	81.05±1.69	**84.38±1.14**	5.38E-05
Recall	84.58±0.96	88.58±0.81	72.43±7.01	91.45±0.67	90.89±0.34	91.08±0.73	90.56±1.49	89.74±0.80	91.32±1.01	91.56±0.42	**92.03±0.42**	3.84E-03
Specificity	97.75±0.05	97.48±0.20	95.74±0.71	97.63±0.33	98.40±0.11	98.25±0.14	98.45±0.09	97.54±0.14	98.65±0.07	98.42±0.13	**98.84±0.04**	1.93E-03
Precision	82.58±1.00	80.40±1.76	66.90±7.03	81.20±2.86	87.19±1.99	86.37±1.77	88.41±0.34	80.23±1.47	89.66±0.72	87.27±2.37	**91.57±0.16**	4.91E-04
HD	104.93±13.90	118.88±5.76	166.11±21.46	101.25±10.87	70.42±13.33	61.28±9.24	63.08±11.26	99.77±1.74	51.30±5.58	48.85±4.97	**29.85±4.14**	1.41E-06
ASSD	5.82±0.66	4.22±0.76	13.16±5.41	2.68±0.25	2.51±0.23	2.60±0.37	2.35±0.53	3.77±0.21	2.09±0.55	2.09±0.43	**1.92±0.37**	3.15E-02
ABD	19.01±1.57	18.58±1.68	32.43±4.72	13.59±0.91	9.18±0.72	10.04±0.74	10.00±0.50	15.34±0.49	8.74±0.43	7.93±0.34	**6.81±0.63**	6.56E-03

methods on Dataset B and BUSI are shown in Fig.5. Obviously, the diversity of tumor morphology has the least effect on our segmentation results and the prediction masks that are closest to ground-truth labels are obtained. In addition, our method can largely alleviate the perturbation of similar surrounding tissue, boundary cascade, and weak boundaries to obtain a more refined tumor contour.

Table 3 illustrates the evaluation result of different comparison methods on KUS and KCUS. The quantitative evaluation results indicated that the proposed approach obtained the most satisfactory prediction results on KUS and KCUS. Similarly, we performed paired t-tests with the second result to emphasize the improvements of our approach. The p-values ($p < 0.05$) in Table 3 demonstrate that our method has a significant improvement in these evaluation metrics compared to the comparison method. For the KUS dataset, SDFNet obtained the second result in the recall index, but the remaining index scores are not ideal. MSDSNet and MBANet achieved second results in several metrics, which indicates their potential for kidney ultrasound image segmentation tasks. It is worth noting that the segmentation results of UNETR on KUS and KCUS are very unsatisfactory, which represents that transform attention cannot learn adequate kidney and cyst characteristics from ultrasound images. The main reason for this phenomenon is the uneven intensity distribution within the kidney ultrasound images. The segmentation task is more complex for KCUS compared to the KUS dataset. Therefore, the various segmentation methods score significantly higher on KUS than on KCUS. According to the p-values obtained from t-tests performed on KUS and KCUS, it could be concluded that our approach has a more noticeable improvement in the evaluation indicator scores on KCUS. This also indicates that our approach can not only segment out the kidney well but also cope with the segmentation of cysts with better characterization ability. Fig.6 presents the visualized prediction masks of different methods on KUS and KCUS. As can be seen from Fig.6, our method can largely mitigate the perturbations of various factors and achieve more satisfying segmentation results. In addition, our method not only obtained the most optimal kidney mask but also the satisfactory cyst mask as illustrated in the third to sixth rows.

6. Discussion

From the quantitative evaluation results and the visualized prediction masks, it can be seen that the U-shaped network with the CAAM module achieves the most satisfactory performance on BUSI, Dataset B, KUS, and KCUS

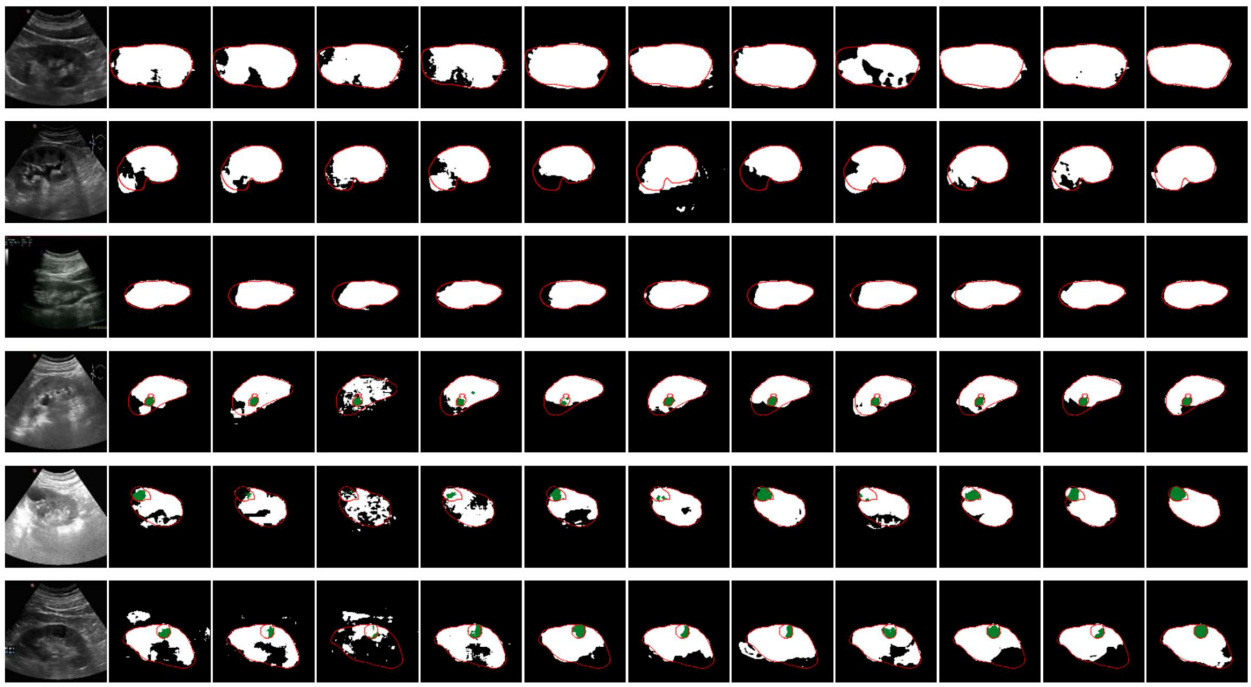

Fig. 6. The visualization of KUS and KCUS prediction masks by different approaches. These images suffer from interference of variable shapes, weak boundaries, cascades, homogeneous structures, and Uneven intensity distribution.

ultrasound datasets. This demonstrates that the CAAM module designed in this paper has good robustness and stability for ultrasound segmentation tasks. From the quantitative evaluation results presented in Table 2 and Table 3, we can conclude that SegNet achieves the most competitive segmentation competence among non-medicine-specific approaches. This indicates that recording the location information of the object during the encoding process is beneficial to improve the segmentation precision of tissue or lesion regions. According to the segmentation performance of UNETR on breast ultrasound and kidney ultrasound, it can be concluded that transformer attention is more sensitive to ultrasound images with uneven intensity distribution. In the segmentation of breast ultrasound images, the segmentation performance of two medicine-specific methods (STAN and RDAU-net) is not satisfactory. This indicates that they have limited generalization ability and robustness. In terms of the kidney ultrasound image segmentation task, the segmentation performance of the SDFNet is not satisfactory. This suggests that directly using structural features and boundary constraints to acquire refined prediction masks in ultrasound images with weak boundaries and cascading is a challenging task. In general, most medicine-specific methods achieved better segmentation performance than the non-medicine-specific approach.

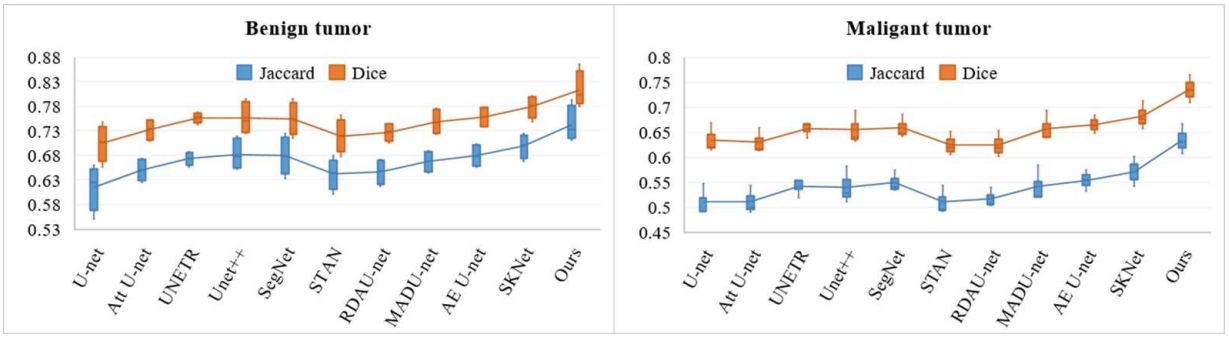

Fig. 7. The quantitative evaluation results of different segmentation methods on benign and malignant breast tumors.

6.1. Robustness analysis

Compared with benign breast tumors, malignant breast tumors have a more irregular morphology and blurred boundaries. We evaluated the representational ability of different methods to respond to malignant tumors and benign tumors.Three-fold cross-validation experiments and four-fold cross-validation experiments were performed on malignant lesions and benign lesions, respectively. Fig.7 the numbers of the different segmentation methods on Dice and Jaccard. The mean numbers of our approach on Dice and Jaccard for benign tumor segmentation are 81.58 and 74.68, respectively. The mean numbers of our approach on Dice and Jaccard for malignant tumor segmentation are 73.55 and 63.38, respectively. Our method achieved the highest score at Jaccard and Dice. Fig.8 demonstrates the ROC curves and P-R curves on benign and malignant tumor segmentation tasks. According to the P-R curve of benign tumors, we could drawn that Att U-net can cause seriously missed detection of breast tumors. It can be seen from the ROC curve that our method achieves the highest score on the AUC indicator. In general, the prediction results of our method on tumor lesion segmentation with high confidence.

6.2. Compared with existing attention modules

We conducted comparison experiments with several attention networks to further analyze the superiority of CAAM module. These attention networks include AAU-net [19], AGNet [40], SANet [41], SENet [42], scSENet [43], CBAM [44], and ECA-Net [45]. Our experimental setup is the same as AAU-net [19]. As shown in Table 4, our approach obtains satisfactory quantitative scores on BUSI and Dataset B. Moreover, the p-value also validates the excellence of our methodology. It is worth noting that AAU-net and ECA-Net also adopt the idea of adaptive feature selection and achieve more satisfactory segmentation results in the compared methods. This indicates that the adaptive selection of objective features in the convolution process is beneficial for the improvement of medical ultrasound image segmentation accuracy. The characterization ability of SENet, CBAM and scSENet on BUSI is similar. On Dataset B, the characterization ability of scSENet is better than SENet. This indicates that aggregating features in both channel and spatial dimensions is beneficial for network performance improvement but there is still room for improvement. In this study, our developed concurrent adaptive attention module not only aggregates objective features of multiple dimensions but also adaptively calibrates the features of each dimension. The above analysis further illustrates the superiority of concurrent adaptive attention module for medical ultrasound image segmentation.

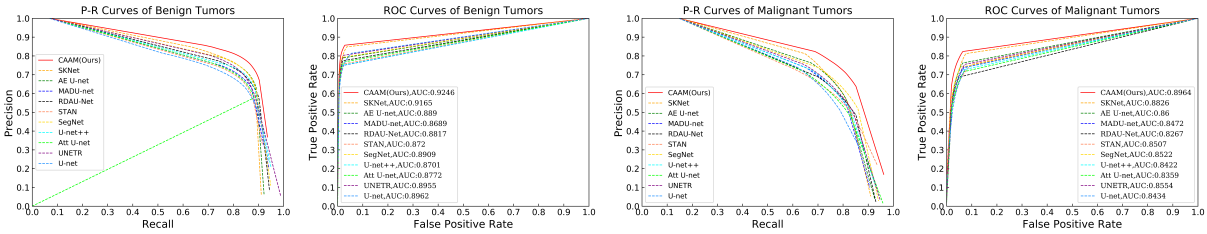

Fig. 8. The P-R and ROC curves of different segmentation methods on benign and malignant breast tumors.

6.3. Limitations and future works

There are still some limitations of our method that need further optimization. (1) Our approach has over-segmentation or under-segmentation problems on several ultrasound images, as shown in Fig.5 and Fig.6. In addition, the P-R curve and ROC curve also reflect that our method can cause false detections and missed detections. Especially for malignant breast tumor segmentation, the approach designed in this study still has much room for improvement. (2) Although the complexity of the network is considered when constructing the architecture, the introduction of the CAAM module still inevitably increases the computational cost, as shown in Fig.9. In the future, two directions can be further investigated to address the above limitations: (1) Introducing global prior and boundary constraints to guide the training of segmentation networks. First, the introduction of global prior knowledge can help the segmentation method to better distinguish object regions from the surrounding organization. Then, boundary constraints are used to refine the prediction results. (2) During the multiscale feature selection, we rank the target features. Subsequently, the bottom-ranked features are removed to reduce the computational cost.

Table 4
The quantitative evaluation results (mean ± std) of several attention mechanisms on BUSI and Dataset B. The top 2 scores are marked with bold and green text, respectively. "†" represents the medicine-specific method.

	BUSI					Dataset B				
Methods	Dice	Jaccard	Recall	Specificity	Precision	Dice	Jaccard	Recall	Specificity	Precision
AGNet	71.35±2.10	62.23±2.02	78.80±2.00	96.30±0.68	72.11±2.77	73.30±1.03	64.15±1.35	80.19±4.30	98.70±0.31	74.35±4.72
SANet	74.46±2.76	65.96±2.78	80.76±2.30	96.75±0.70	74.84±4.18	71.84±7.96	63.26±7.77	77.81±5.07	98.61±0.57	72.40±10.19
SENet	76.71±2.88	67.75±3.09	80.04±3.07	97.15±0.71	78.70±3.51	70.45±5.30	60.77±6.41	78.18±6.77	98.49±0.27	71.95±8.50
CBAM	76.85±2.01	67.51±2.01	80.43±1.23	97.21±0.52	78.82±2.32	71.65±3.52	63.49±6.41	80.76±5.68	98.57±0.23	72.96±3.87
scSENet	76.67±2.20	67.68±2.28	79.58±1.14	97.26±0.48	78.95±2.73	71.30±4.24	62.17±5.03	79.16±5.79	98.52±0.22	71.31±5.44
ECA-Net	77.10±2.17	68.17±2.21	80.73±1.91	97.20±0.53	78.59±2.77	72.09±4.74	63.23±5.26	78.53±4.99	98.62±0.26	72.65±4.19
AAU-net†	77.51±0.68	68.82±0.44	81.10±0.52	97.57±0.24	79.61±1.07	78.14±2.41	69.10±2.98	82.22±3.84	98.82±0.35	78.83±2.40
ESKNet†	78.71±2.37	70.20±2.28	**82.41±2.84**	97.47±0.35	79.57±1.65	79.92±2.21	71.65±2.39	82.66±1.40	99.01±0.35	81.01±3.91
Ours	**79.44±2.07**	**71.29±2.07**	82.22±2.00	**97.60±0.52**	**80.87±2.02**	**80.53±1.16**	**71.89±2.30**	**83.58±2.16**	**99.04±0.31**	**81.82±1.00**
P-value	6.28E-04	5.85E-04	-	1.07E-04	1.63E-04	4.34E-02	4.83E-02	3.63E-02	3.01E-02	7.15E-03

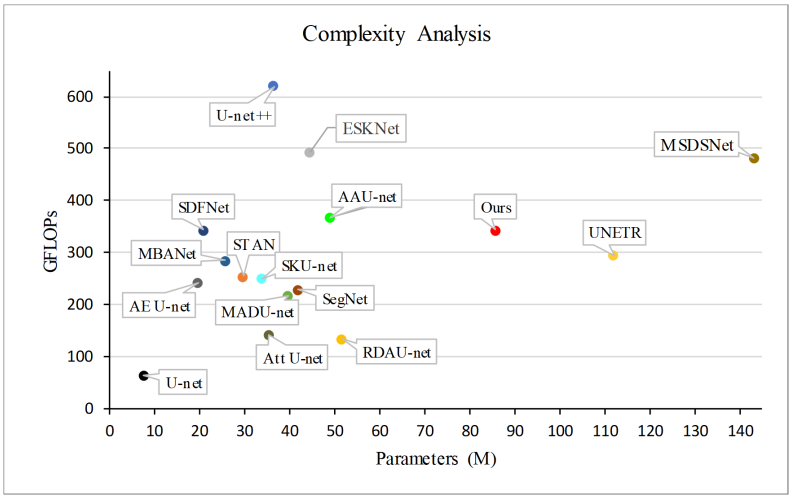

Fig. 9. Statistical analysis of the complexity.

7. Conclusion

To respond to the limitations of fixed receptive field convolution, variable morphology, and heterogeneous structures for medical ultrasound image segmentation, we designed a concurrent adaptive attention module (CAAM). The CAAM module can adaptively select robust object characterizations under multiple receptive fields from both spatial dimension and channel dimension. Subsequently, we constructed a novel U-shaped network using the CAAM module to achieve automatic segmentation of medical ultrasound images. The robustness and effectiveness of our approach on medical ultrasound image segmentation task is perfectly demonstrated through the extensive and reliable experimental results with more than ten SOTA segmentation approaches on four ultrasound datasets. Moreover, the comparison with several classical and advanced attention modules further evidences the superiority of the CAAM module. In general, the approach proposed in this study acquires the most satisfactory characterization ability on the medical ultrasound image segmentation task.

Acknowledgments

This work is supported by the Tianjin Science and Technology Major Project (25ZXRGGX00260) and the Tianjin Municipal Education Commission (2025KJ005). The deep-learning model training and testing in this work are

supported by the High performance Computing Platform of Tianjin Medical University.

Conflict of Interest Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author Contributions Statement

Gongping Chen: Conceptualization, Methodology, Software, Formal analysis, Investigation, Visualization, Writing – original draft. Qiaolu Wang: Data curation, Validation, Formal analysis, Writing – review and editing. Mengran Fan: Data curation, Investigation, Validation, Writing – review and editing. Hongxin Zhang: Data curation, Investigation, Validation, Writing – review and editing. Ningning Liu: Supervision, Project administration, Methodology, Writing – review and editing.Yang Zhou: Supervision, Project administration, Methodology, Writing – review and editing. All authors read and approved the final manuscript.

Data Availability Statement

The two breast ultrasound datasets used in this paper are already publicly available. The two kidney ultrasound datasets and the code used in this paper can be obtained by contacting the corresponding author via email.

References

[1] R. Huang, M. Lin, H. Dou, Z. Lin, Q. Ying, X. Jia, W. Xu, Z. Mei, X. Yang, and Y. Dong, "Boundary-rendering Network for Breast Lesion Segmentation in Ultrasound Images," *Medical Image Analysis*, p. 102478, 2022.

[2] Z. Ning, Y. Mao, X. Xu, Q. Feng, S. Zhong, and Y. Zhang, "Dc-net: Decomposing and coupling saliency map for lesion segmentation in ultrasound images," *Engineering Applications of Artificial Intelligence*, vol. 148, p. 110355, 2025.

[3] J. A. Noble and D. Boukerroui, "Ultrasound image segmentation: a survey," *IEEE Trans Med Imaging*, vol. 25, no. 8, pp. 987–1010, 2006.

[4] Q. Huang, L. Zhao, G. Ren, X. Wang, C. Liu, and W. Wang, "Nag-net: Nested attention-guided learning for segmentation of carotid lumen-intima interface and media-adventitia interface," *Computers in Biology and Medicine*, vol. 156, p. 106718, 2023.

[5] Q. Huang, D. Wang, Z. Lu, S. Zhou, J. Li, L. Liu, and C. Chang, "A novel image-to-knowledge inference approach for automatically diagnosing tumors," *Expert Systems with Applications*, vol. 229, p. 120450, 2023.

[6] Y. Luo, Q. Huang, and L. Liu, "Classification of tumor in one single ultrasound image via a novel multi-view learning strategy," *Pattern Recognition*, vol. 143, p. 109776, 2023.

[7] Y.-L. Huang, Y.-R. Jiang, D.-R. Chen, and W. K. Moon, "Level set contouring for breast tumor in sonography," *Journal of digital imaging*, vol. 20, pp. 238–247, 2007.

[8] H. Wang, J. E. Pulido, Y. Song, S. L. Furth, C. Tu, C. Zhang, C. Li, and G. E. Tasian, "Segmentation of renal parenchymal area from ultrasoundl images using level set evolution," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014, pp. 4703–4706.

[9] F. Yang, J. Gu, Y. Xie, T. Wen, and W. Qin, "A shape-optimized framework for kidney segmentation in ultrasound images using NLTV denoising and DRLSE," *BioMedical Engineering OnLine*, vol. 11, no. 1, p. 82, 2012.

[10] H. R. Torres, P. Morais, B. Oliveira, C. Birdir, M. Rüdiger, J. C. Fonseca, and J. L. Vilaça, "A review of image processing methods for fetal head and brain analysis in ultrasound images," *Computer Methods and Programs in Biomedicine*, p. 106629, 2022.

[11] Y. Zhang, M. Ying, Y. Lin, A. T. Ahuja, and D. Z. B. T. I. I. C. o. B. . B. Chen, "Coarse-to-Fine Stacked Fully Convolutional Nets for lymph node segmentation in ultrasound images," in *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2016, pp. 443–448.

[12] L. Wu, X. Yang, S. Li, T. Wang, and N. B. T. Dong, "Cascaded Fully Convolutional Networks for automatic prenatal ultrasound image segmentation," in *IEEE 14th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2017, pp. 663–666.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[14] B. Behboodi and H. Rivaz, "Ultrasound segmentation using U-Net: learning from simulated data and testing on real data," *IEEE*, 2019.

[15] M. H. Yap, G. Pons, J. Martí, S. Ganau, M. Sentís, R. Zwiggelaar, A. K. Davison, and R. Martí, "Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 4, pp. 1218–1226, 2018.

[16] Deepak, Mishra, Santanu, Chaudhury, Mukul, Sarkar, Arvinder, Singh, and Soin, "Ultrasound Image Segmentation: A Deeply Supervised Network with Attention to Boundaries." *IEEE Transactions on Bio Medical Engineering*, vol. 66, no. 6, pp. 1637–1648, 2018.

[17] S. Kim, Y. Jang, B. Jeon, Y. Hong, and H. Chang, *Fully Automatic Segmentation of Coronary Arteries Based on Deep Neural Network in Intravascular Ultrasound Images*. Intravascular Imaging and Computer Assisted Stenting and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, 2018.

[18] X. Li, W. Wang, X. Hu, and J. Yang, "Selective kernel networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 510–519.

[19] G. Chen, L. Li, Y. Dai, J. Zhang, and M. H. Yap, "AAU-Net: An Adaptive Attention U-Net for Breast Lesions Segmentation in Ultrasound Images," *IEEE Transactions on Medical Imaging*, vol. 42, no. 5, pp. 1289–1300, 2023.

[20] G. Chen, L. Zhou, J. Zhang, X. Yin, L. Cui, and Y. Dai, "Esknet: An enhanced adaptive selection kernel convolution for ultrasound breast tumors segmentation," *Expert Systems with Applications*, pp. 123 265–123 265, 2024.

[21] Z. Ning, S. Zhong, Q. Feng, W. Chen, and Y. Zhang, "SMU-Net: Saliency-guided Morphology-aware U-Net for Breast Lesion Segmentation in Ultrasound Image," *IEEE transactions on medical imaging*, 2021.

[22] Y. Yan, Y. Liu, Y. Wu, H. Zhang, Y. Zhang, and L. Meng, "Accurate segmentation of breast tumors using AE U-net with HDC model in ultrasound images," *Biomedical Signal Processing and Control*, vol. 72, no. PA, p. 103299, 2022.

[23] Z. Zhuang, N. Li, A. N. Joseph Raj, V. G. V. Mahesh, and S. Qiu, "An RDAU-NET model for lesion segmentation in breast ultrasound images," *PloS one*, vol. 14, no. 8, p. e0221535, 2019.

[24] Y. Lyu, Y. Xu, X. Jiang, J. Liu, X. Zhao, and X. Zhu, "AMS-PAN: Breast ultrasound image segmentation model combining attention mechanism and multi-scale features," *Biomedical Signal Processing and Control*, vol. 81, p. 104425, 2023.

[25] B. Shareef, M. Xian, and A. Vakanski, "Stan: Small tumor-aware network for breast ultrasound image segmentation," in *IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2020, pp. 1–5.

[26] N. Abraham and N. M. B. T. Khan, "A Novel Focal Tversky Loss Function With Improved Attention U-Net for Lesion Segmentation," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2019, pp. 683–687.

[27] M. Byra, P. Jarosik, A. Szubert, M. Galperin, H. Ojeda-Fournier, L. Olson, M. O'Boyle, C. Comstock, and M. Andre, "Breast mass segmentation in ultrasound with selective kernel U-Net convolutional neural network," *Biomedical Signal Processing and Control*, vol. 61, 2020.

[28] S. Yin, Q. Peng, H. Li, Z. Zhang, X. You, K. Fischer, S. L. Furth, G. E. Tasian, and Y. Fan, "Automatic kidney segmentation in ultrasound images using subsequent boundary distance regression and pixelwise classification networks," *Med Image Anal*, vol. 60, p. 101602, 2020.

[29] H. R. Torres, P. Morais, B. Oliveira, C. Birdir, M. Rüdiger, J. C. Fonseca, and J. L. Vilaça, "Kidney segmentation in ultrasound, magnetic resonance and computed tomography images: A systematic review," *Comput Methods Programs Biomed*, vol. 157, pp. 49–67, 2018.

[30] G. Chen, Y. Dai, J. Zhang, X. Yin, and L. Cui, "SDFNet: Automatic segmentation of kidney ultrasound images using multi-scale low-level structural feature," *Expert Systems with Applications*, vol. 185, p. 115619, 2021.

[31] D. M. Alex, D. Abraham Chandy, A. Hepzibah Christinal, A. Singh, and M. Pushkaran, "YSegNet: a novel deep learning network for kidney segmentation in 2D ultrasound images," *Neural Computing and Applications*, vol. 34, no. 24, pp. 22 405–22 416, 2022.

[32] G. Chen, Y. Dai, J. Zhang, X. Yin, and L. Cui, "MBANet: Multi-branch aware network for kidney ultrasound images segmentation," *Computers in biology and medicine*, vol. 141, p. 105140, 2022.

[33] G. Chen, J. Yin, Y. Dai, J. Zhang, X. Yin, and L. Cui, "A novel convolutional neural network for kidney ultrasound images segmentation." *Computer methods and programs in biomedicine*, vol. 218, p. 106712, 2022.

[34] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data in brief*, vol. 28, p. 104863, 2020.

[35] M. H. Yap, M. Goyal, F. Osman, R. Martí, E. Denton, A. Juette, and R. Zwiggelaar, "Breast ultrasound region of interest detection and lesion localisation," *Artificial Intelligence in Medicine*, vol. 107, p. 101880, 2020.

[36] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *IEEE Trans Pattern Anal Mach Intell*, vol. 39, no. 12, pp. 2481–2495, 2017.

[37] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning Where to Look for the Pancreas," in *Medical Imaging with Deep Learning*, 2018.

[38] Z. Zhou, M. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1856–1867, 2020.

[39] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. R. Roth, and D. Xu, "UNETR: Transformers for 3D Medical Image Segmentation," in *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2022, pp. 1748–1758.

[40] S. Zhang, H. Fu, Y. Yan, Y. Zhang, Q. Wu, M. Yang, M. Tan, and Y. Xu, "Attention guided network for retinal image segmentation," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2019, pp. 797–805.

[41] Z. Zhong, Z. Q. Lin, R. Bidart, X. Hu, I. B. Daya, Z. Li, W.-S. Zheng, J. Li, and A. Wong, "Squeeze-and-attention networks for semantic segmentation," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),*, 2020, pp. 13 062–13 071.

[42] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-Excitation Networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 8, 2020.

[43] A. G. Roy, N. Navab, and C. Wachinger, "Concurrent spatial and channel 'squeeze and excitation'in fully convolutional networks," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2018, pp. 421–429.

[44] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.

[45] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 11 531–11 539.