



## DSEU-net: A novel deep supervision SEU-net for medical ultrasound image segmentation

Gongping Chen<sup>a</sup>, Yuming Liu<sup>a</sup>, Jiang Qian<sup>b</sup>, Jianxun Zhang<sup>a</sup>, Xiaotao Yin<sup>c</sup>, Liang Cui<sup>d</sup>, Yu Dai<sup>a,\*</sup>

<sup>a</sup> The College of Artificial Intelligence, Nankai University, Tianjin, China

<sup>b</sup> The Department of Information Management Center, Anhui Business College, Wuhu, China

<sup>c</sup> The Department of Urology, Fourth Medical Center of PLA General Hospital, Beijing, China

<sup>d</sup> The Department of Urology, Civil Aviation General Hospital, Beijing, China



### ARTICLE INFO

#### Keywords:

Medical ultrasound  
Breast tumor  
Kidney  
Attention mechanism  
Deep supervision

### ABSTRACT

The automatic and accurate medical ultrasound image segmentation has been a challenging task due to the coupled interference of various internal and external factors. In recent years, CNN techniques have been widely and successfully used in medical image segmentation. Motivated by this, this paper proposes a novel squeeze-and-excitation attention U-net with deep supervision (DSEU-net) for medical ultrasound image segmentation. Specifically, a deeper U-net is first used as a benchmark network to capture sufficient target feature information from complex ultrasound images. Then, the squeeze-and-excitation (SE) block is regarded as the bond between encoder and decoder to enhance the attention to useful object regions. Moreover, the introduction of SE block not only strengthens the association of useful information at a distance, but also suppresses the introduction of irrelevant information. Finally, the deep supervised constraints are added to the decoding stage of the network to refine the prediction masks of ultrasound images. Extensive experimental results on three clinical ultrasound datasets show that DSEU-net has better robustness and superiority in ultrasound image segmentation. In the segmentation of the first breast ultrasound dataset (BUSI), the values of Jaccard, Precision, Recall, Specificity and Dice are 70.36%, 79.73%, 82.70%, 97.42% and 78.51%, respectively. The values of Jaccard, Precision, Recall, Specificity and Dice for our method on the second breast ultrasound dataset (Dataset B) are 73.17%, 82.58%, 84.02%, 99.05% and 81.50%, respectively. For the segmentation of kidney ultrasound dataset (KUS), the values of Jaccard, Precision, Recall, Specificity, Dice, HD, ASSD and ABD are 89.47, 94.77, 94.36, 99.10, 94.32, 12.42, 0.48 and 3.44, respectively. Comparing with the original U-net, DSEU-net improved on average 8.28% and 12.55% on five metrics for two breast ultrasound data. DSEU-net improved on average 54.81% on eight metrics for the kidney ultrasound dataset.

### 1. Introduction

Ultrasound imaging is one of the most prevalent imaging techniques and has become an important part of clinical medical imaging (Xian et al., 2018). Compared with other imaging techniques, ultrasonography not only has the advantages of being non-invasive, sensitive, real-time and inexpensive, but also can rapidly determine the cause and location of morbidity to improve clinical diagnostic efficiency (Singh et al., 2020). Medical image segmentation is one of the important steps in computer-aided diagnosis, which involves extracting useful objective tissues or regions from the whole image to help physicians in further

diagnosis and treatment (Houssein et al., 2021). However, the internal and external perturbations in the ultrasound imaging process increase the difficulty of accurate segmentation of ultrasound images, and it is difficult for even experienced radiologists to accurately and quickly landmark lesion regions, as shown in Fig. 1.

The medical ultrasound image segmentation task has been extensively investigated (Siddique et al., 2021). Early on, grabcut method, level-set method and shape prior model are the more representative ultrasound image segmentation methods. Grabcut method is often implemented based on the texture features of the objective tissue (Mahmud et al., 2015). Grabcut method has the advantage of being

\* Corresponding author.

E-mail addresses: [cgp110@stu.xhu.edu.cn](mailto:cgp110@stu.xhu.edu.cn) (G. Chen), [Qianjiang@abc.edu.cn](mailto:Qianjiang@abc.edu.cn) (J. Qian), [zhangjx@nankai.edu.cn](mailto:zhangjx@nankai.edu.cn) (J. Zhang), [daiyu@nankai.edu.cn](mailto:daiyu@nankai.edu.cn) (Y. Dai).

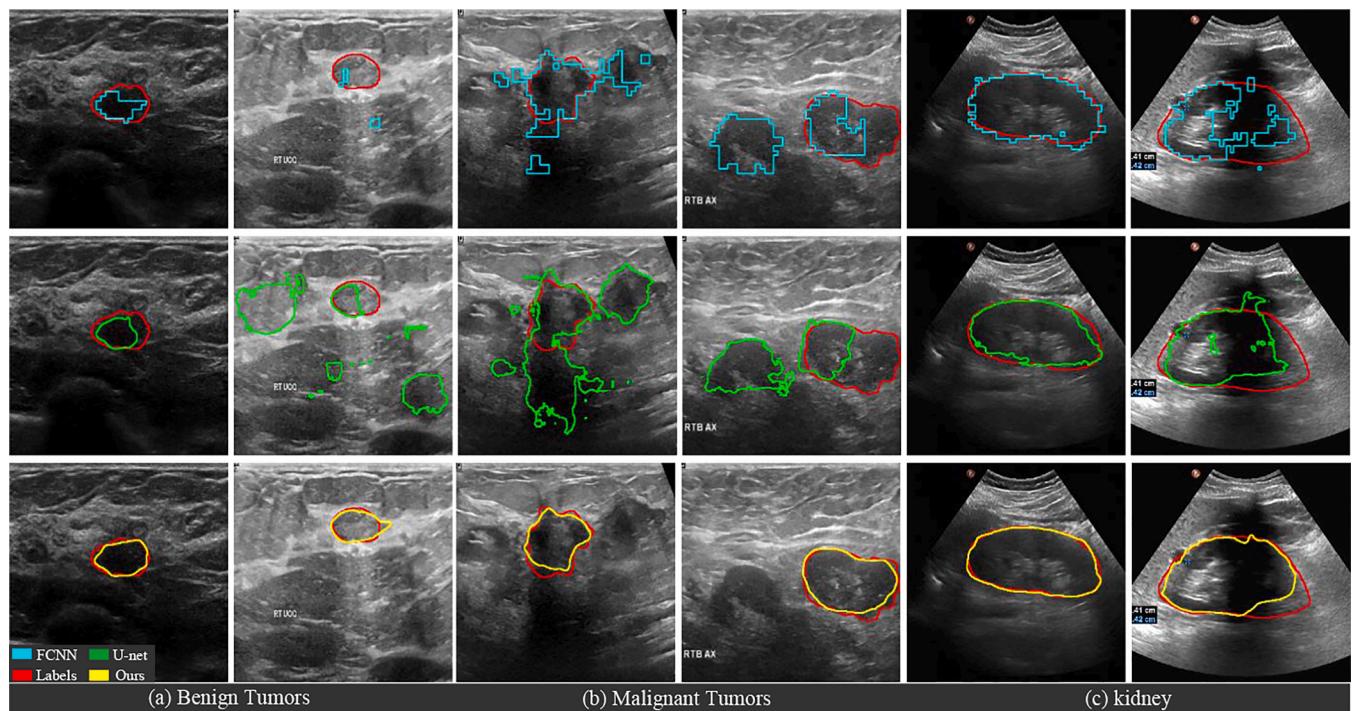
simple and easy to understand, but the segmentation results are often perturbed by various degradation factors (Noble and Boukerroui, 2006). In addition, heterogeneous structures and similar surrounding tissues can seriously affect the segmentation accuracy of grabcut methods. The level-set method is one of the classical methods of the deformable model, which achieves segmentation by evolving the objective tissue contour (Wang et al., 2014). However, the level set method requires the initial position of the objective tissue, and the evolution process is also more time consuming (Chen et al., 2022c). In addition, the blurred boundary of the objective tissue in the ultrasound image increases the difficulty of evolution. In order to speed up the evolution process and improve the segmentation accuracy, the shape prior of the objective tissue is often introduced to guide the segmentation process (Mendoza et al., 2013). However, the most of shape priors often rely on manual obtaining (Yin et al., 2020). Therefore, how to achieve automatic segmentation of ultrasound images needs further research.

In recent years, the development of deep convolutional neural networks (CNN) has greatly enhanced the nonlinear learning capability of targets and has been successfully used in several vision tasks. Motivated by this, CNN techniques have been applied to the automatic segmentation of medical ultrasound images (Liu et al., 2019). In 2016, Zhang et al. achieved the automatic segmentation of lymph node ultrasound images by two FCN models (Zhang et al., 2016). Similarly, Wu et al. used cascaded FCNs to segment prenatal ultrasound images (Wu et al., 2017). In 2018, Mishra et al. refined the segmentation results of ultrasound images by performing ground-true boundary and ground-true mask supervision on the encoding process of FCNs (Deepak et al., 2018). Compared with traditional methods, the use of FCN speeds up the segmentation of ultrasound images and improves the segmentation accuracy. However, FCN itself has two obvious limitations: (1) the segmentation results are not fine enough and the boundaries are irregular. (2) The logical relationship between pixels, i.e., spatial consistency, is not considered.

Besides FCN, U-net based on the code-decode architecture is also one of the widely used network frameworks. Abundant research works have demonstrated that U-net with encode-decode architecture has better

performance in medical image segmentation. Motivated by this, Kim et al. implemented automatic segmentation of coronary artery ultrasound images by introducing multiscale input modules and hybrid loss functions in U-net (Kim et al., 2018). In 2019, Zhuang et al. designed RDAU-Net to segment breast tumor ultrasound images by introducing residual connections in U-Net (Zhuang et al., 2019). In 2020, Lee et al. achieved automatic segmentation of breast tumors in ultrasound images by introducing a channel attention module with multi-scale averaging pooling operation in U-Net (Lee et al., 2020). Compared with the FCN framework, U-net can further improve the segmentation accuracy of ultrasound images by integrating low-level and high-level semantic information. Most importantly, the various optimization strategies introduced in the U-net framework can also refine the segmentation results. However, most approaches tend to employ 9 or shallower layers of U-net to segment ultrasound images. Due to the complexity of ultrasound patterns, it is necessary to utilize a deeper network framework to capture sufficient target features (Chen et al., 2023b). The most important point is that many methods are often designed with only one type of ultrasound image in mind, and their robustness and generalization are not strong.

To alleviate the above challenges and reduce the perturbation of ultrasound image segmentation by ultrasound pattern complexity and target morphological diversity, a novel novel squeeze-and-excitation attention U-net with deep supervision (DSEU-net) is proposed. In this study, we first construct a deeper squeeze-and-excitation attention U-net (SEU-net) to improve the learning ability of objective tissues with the depth of 15. In the SEU-net, the squeeze-and-excitation (SE) block is used as the bond between encoder and decoder not only to strengthen the correlation of distant information, but also to reduce the perturbation of useless regions on the segmentation results. To further refine the segmentation results of the network, this study performs the supervised constraint on the decoding phase of SEU-net. The supervision of the SEU-net can guide the network to obtain more refined segmentation results. In summary, the main innovations of this work are as follows:



**Fig. 1.** Complex ultrasound patterns, similar intensity distributions, variable tissue morphology, and blurred boundaries bring great challenges for ultrasound image segmentation.

- In this work, a novel SE attention U-net with deep supervision is developed for medical ultrasound image automatic segmentation.
- The introduction of SE attention can help the network to extract more useful objective features and suppress useless objective features. In addition, the introduction of deep supervision mechanism can help the network learn to predict more accurate objective masks.
- The robustness and superiority of the DSEU-net developed in this paper are well demonstrated by the extensive comparison experiments conducted on three public clinical ultrasound datasets.

## 2. Method

### 2.1. Deeper benchmark U-net

Currently, U-net and its variants have been widely used for medical image segmentation (Chen et al., 2022b). However, these methods tend to use U-net with a depth of 4 or shallower as a benchmark framework. The effect of network depth on segmentation accuracy was analyzed on six medical images in the U-net++ framework proposed by Zhou et al., (Zhou et al., 2020). The six medical images were electron microscopy (EM), cell, nuclei, brain tumor, liver and lung nodule images. Compared to these medical images, ultrasound images have more complex imaging patterns, uneven intensity distribution and severe heterogeneous structure (Chen et al., 2022a). Although U-net++ can achieve similar or better segmentation accuracy on these six medical images using fewer convolution layers, the segmentation results on medical ultrasound images are unsatisfactory (Chen et al., 2023a). To better handle the ultrasound image segmentation task, this study utilized a U-net with the depth of 15 as the benchmark network to capture more objective features from the complex ultrasound images. This benchmark U-net contains seven down-sampling and seven up-sampling operations, and each convolution module consists of two  $3 \times 3$  convolution layers, two batch

normalization layers and two LeakyReLU activation layers. The filter sizes of the deeper benchmark U-net are 64, 128, 128, 256, 256, 512, 512, 1024, 512, 512, 256, 256, 128, 128 and 64, respectively.

### 2.2. Squeeze-and-excitation attention (SE) attention module

To improve the segmentation accuracy of U-net for objective tissues/regions, attentional mechanisms are widely used in U-net (Chen et al., 2022d; Hu et al., 2020; Roy et al., 2018; Wang et al., 2020; Zhang et al., 2019; Zhong et al., 2020). The existing attention mechanisms can be roughly classified into: spatial attention, channel attention and hybrid attention. The principle of spatial attention is to help the network extract important features from the spatial dimension, and the Att U-net is a typical example (Oktay et al., 2018). Similarly, the principle of channel attention mechanism is to select more useful target features and suppress irrelevant features from the channel's dimension, such as SE block (Hu et al., 2020). Hybrid attention mechanisms are often a combination of multiple attention modules to improve the segmentation performance of the network (Roy et al., 2018). Due to the uneven intensity distribution and similar surrounding tissues in ultrasound images, the objective tissue cannot be accurately constrained using spatial attention mechanisms. In this study, the SE block is used as the bond between the encoder and decoder to calibrate the target features from the channels' dimensions. Fig. 2 illustrates the SE framework. Specifically, the two sets of feature maps input to the SE module are first combined and subjected to a  $3 \times 3$  convolution operation to obtain a new feature map,

$$F_i = C_{3 \times 3}(F_i^E \oplus U(F_{i-1}^D)) \quad (1)$$

where  $C_{3 \times 3}$  and  $U(\cdot)$  represent  $3 \times 3$  convolution operation and up-sampling operation, respectively.  $F_i^E \in \mathbb{R}^{c \times h \times w}$  denotes the feature map captured at the  $i$ th encoding stage. Similarly,  $F_{i-1}^D \in \mathbb{R}^{c \times h \times w}$  represents

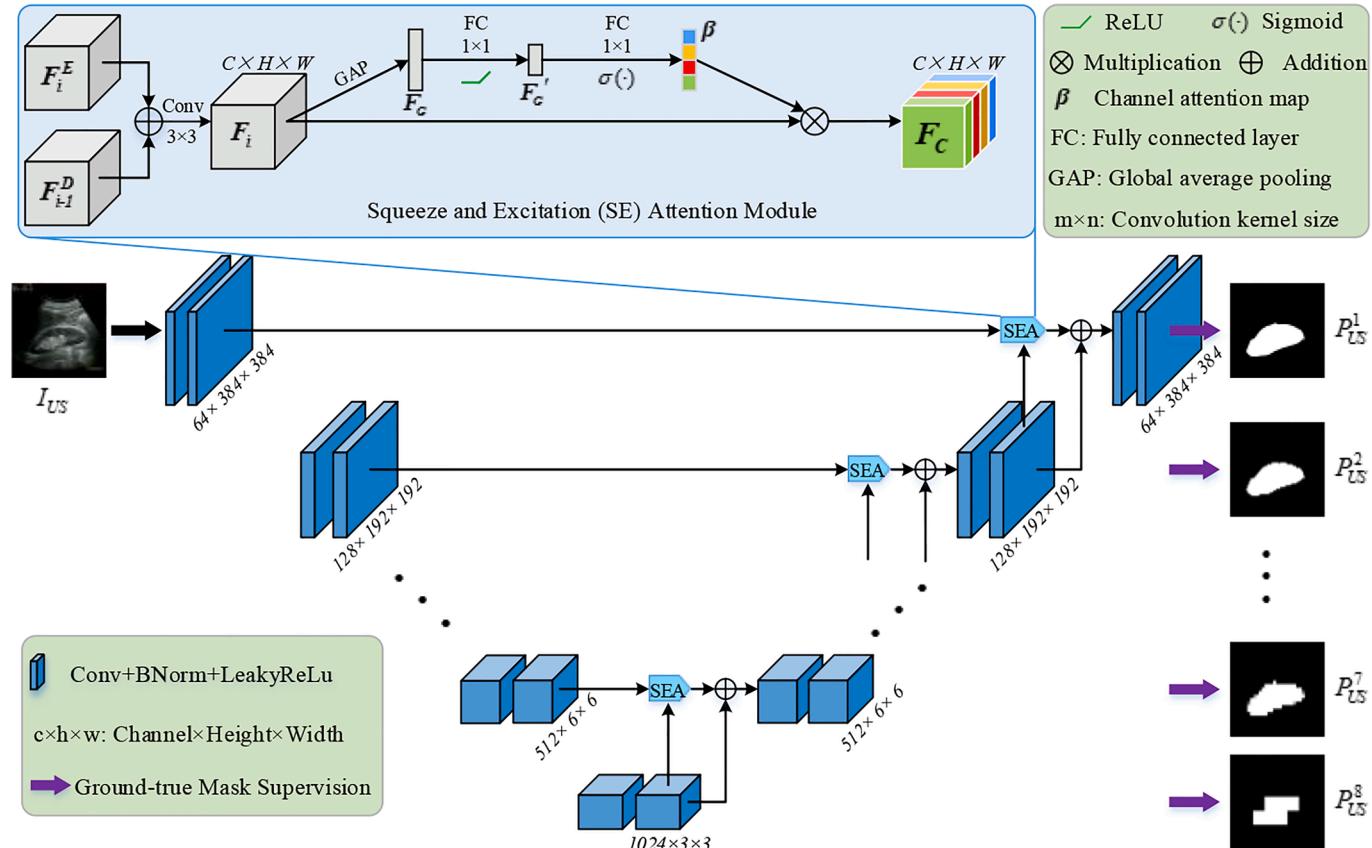


Fig. 2. The overall framework of DSEU-net, which mainly consists of deeper U-net with SE attention and deep supervision of ground-true masks.

the feature map captured at the  $i - 1$ th decoding stage. Subsequently,  $F_i \in \mathbb{R}^{c \times h \times w}$  undergoes a global pooling operation to generate a set of feature maps  $F_G \in \mathbb{R}^{c \times 1 \times 1}$  with size  $1 \times 1 \times 1$ . After  $F_G \in \mathbb{R}^{c \times 1 \times 1}$  undergoes a fully connected operation and a ReLU activation  $\sigma_r(\cdot)$  operation to produce a new set of feature maps,

$$F'_G = \sigma_r(W_G \cdot F_G) \quad (2)$$

where  $W_G \in \mathbb{R}^{\frac{c}{r} \times c}$  denotes the matrix of the fully connected layers.  $F'_G \in \mathbb{R}^{\frac{c}{r} \times 1 \times 1}$  is again subjected to a fully connected operation and a sigmoid activation  $\sigma(\cdot)$  operation to produce the channel attention map,

$$\beta = \sigma(W'_G \cdot F'_G) \quad (3)$$

where  $W'_G \in \mathbb{R}^{c \times \frac{c}{r}}$  denotes the matrix of the fully connected layers.  $\beta \in [0, 1]^{c \times 1 \times 1}$  is the channel attention map for  $F_i$ . Each value of  $\beta$  indicates the importance of channel information at the corresponding voxel in  $F_i$ . Finally, the channel attention map  $\beta$  is used to calibrate  $F_i$ . The feature map calibrated by  $\beta$  can be expressed as:

$$F_C = \beta \otimes F_i \quad (4)$$

where  $F_C \in \mathbb{R}^{c \times h \times w}$  denotes the output of the SE module.

### 2.3. Deep supervision guidance

It is well known that encoding-decoding network is the process of feature extraction followed by reconstruction. To enable the network to obtain more accurate ultrasound prediction results, the supervision of ground-true masks is added to each decoding stage of the method. Through the introduction of a deep supervision strategy, the method can predict the segmentation masks step-by-step. Specifically, the feature maps captured in each decoding stage are first fed into a convolution layer. Subsequently, a sigmoid activation operation is performed on the features captured in the convolution layer to obtain the prediction masks. Finally, the prediction masks of each decoding stage are performed an up-sampling operation to obtain the size of the input image. As shown in Fig. 2, inputting an ultrasound image into DSEU-net can generate eight prediction masks,

$$P_{US}^i = U(\sigma_s(W_{1 \times 1} \cdot F_D^i)), i = 1, 2, 3, 4, 5, 6, 7, 8 \quad (5)$$

where  $P_{US}^i$  denotes the prediction result of the  $i$ th decoding stage.  $U(\cdot)$  and  $\sigma_s(\cdot)$  represent up-sampling operation and sigmoid activation operation, respectively.  $W_{1 \times 1}$  indicates the matrix of the convolution layer with kernel size  $1 \times 1$ .  $F_D^i \in \mathbb{R}^{c \times h \times w}$  means the feature map captured at the  $i$ -th decoding stage. Since the last predicted mask has higher segmentation accuracy and more complete boundary, it is regarded as the final prediction result of DSEU-net.

### 2.4. Loss function

The binary cross entropy (BCE) is used as the loss function during the training process. As can be observed in Fig. 2, our proposed method uses eight ground-true masks to constrain the training of the network. During the training process, the predicted results of each decoding stage are performed an up-sampling operation to calculate the loss of the ground-true mask and the predicted mask. Therefore, the loss function of DSEU-net can be indicated as:

$$L = \sum_{i=1}^8 \ell_{seg}^i \quad (6)$$

where  $\ell_{seg}^i$  represents the loss between the predicted result of the  $i$ -th decoding stage and the ground-true mask.

## 3. Materials and experiments

### 3.1. Datasets

In this study, two public breast ultrasound datasets and one public kidney ultrasound dataset are used for comparative experimental analysis. The first breast ultrasound dataset (BUSI) was constructed by Al-Dhabayani et al., (Al-Dhabayani et al., 2020). A second public breast ultrasound dataset (Dataset B) was collected by Yap et al., (Yap et al., 2020). Kidney ultrasound data (KUS) was constructed by Chen et al. (2022a). The more detailed description of these ultrasound datasets is presented in reference (Chen et al., 2022b).

### 3.2. Experimental settings

To ensure the fairness of the experiment, this work performed the comparison experiment by k-fold cross validation. According to the sample size, BUSI and Dataset B are executed four-fold cross validation and KUS is implemented three-fold cross validation, respectively. This work chose Adam (initial learning rate is 1e-3) as the optimizer of the network. In the cross-validation experiment, 20% of the training data in each fold is considered as validation data to determine the conditions for network termination. Finally, the epoch size and batch size are set to 50 and 12, respectively. The development environment of our network is Ubuntu 20.04, python 3.6, TensorFlow 2.6.0, NVIDIA RTX 3090 GPUs.

### 3.3. Evaluation metrics

In this work, six region-based segmentation metrics, three boundary-based segmentation metrics and Kappa coefficient were used to evaluate the performance of different segmentation methods. The six region-based evaluation metrics are Jaccard, Precision, Recall, Specificity, Dice and MCC. The three boundary-based evaluation metrics are hausdorff distance (HD), average symmetric surface distance (ASDS) and average boundary distance (ABD). These metrics are described in detail in (Chen et al., 2021; Rezaei, 2021). Higher values of Jaccard, Precision, Recall, Specificity, Dice, MCC and Kappa indicate better prediction results. On the contrary, smaller values of HD, ASDS, and ABD represent more refined prediction masks.

## 4. Experimental results

### 4.1. Ablation experiments

To validate the effectiveness of the different network components, we performed ablation experiments on two ultrasound datasets (Dataset B and KUS). Specifically, we performed the four-fold cross-validation on Dataset B and the three-fold cross-validation on KUS. Table 1 demonstrates the quantitative evaluation results on both Dataset B and KUS datasets after adding different network components.

Based on the quantitative evaluation results in Table 1, we can find that the segmentation accuracy of ultrasound images can be gradually improved by increasing the depth, adding SE blocks, and introducing deep supervision constraints. This indicates that the introduction of these components is beneficial to the improvement of network segmentation performance. Compared with the demonstrates U-net, the average values of the five evaluation-metrics on Dataset B and KUS are improved by 7.63% and 2.87% by increasing the network depth. The mean values of the five evaluation parameters on Dataset B and KUS are improved by 3.50% and 0.55% after the SE block is added. The mean values of the five evaluation-indexes on Dataset B and KUS are further improved by 0.86% and 0.15% by adding deep supervision constraints to the decoding stage. The visualization results of the different network frameworks on the two datasets are presented in Fig. 3. As shown in Fig. 3, the introduction of different components can bring the predicted results closer to the real masks. In addition, the omission and false

**Table 1**

The segmentation results ( $\text{mean} \pm \text{std}$ ) of different framework components on Dataset B and KUS. ‘Deep U-net’ denotes increasing the depth of the baseline U-net. ‘SEU-net’ represents the introduction of SE attention module in ‘Deep U-net’. ‘DSEU-net’ represents introducing deep supervision strategy in ‘SEU-net’.

Components	Dataset B					KUS				
	Jaccard	Precision	Recall	Specificity	Dice	Jaccard	Precision	Recall	Specificity	Dice
U-net	58.44 ± 4.26	70.27 ± 6.11	75.32 ± 2.85	98.44 ± 0.40	68.20 ± 4.23	83.19 ± 2.69	90.81 ± 2.64	91.20 ± 0.66	98.33 ± 0.56	90.42 ± 1.79
Deep U-net	67.86 ± 1.87	76.59 ± 2.76	80.95 ± 2.13	98.65 ± 0.55	76.48 ± 1.77	88.22 ± 0.77	93.81 ± 1.69	93.97 ± 1.02	98.83 ± 0.45	93.55 ± 0.50
SEU-net	72.16 ± 1.62	81.14 ± 2.24	83.76 ± 3.68	98.98 ± 0.32	80.23 ± 0.93	89.18 ± 0.70	94.49 ± 1.14	94.31 ± 0.40	99.05 ± 0.27	94.17 ± 0.41
DSEU-net (Ours)	<b>73.17 ± 2.73</b>	<b>82.58 ± 2.28</b>	<b>84.02 ± 3.57</b>	<b>99.05 ± 0.19</b>	<b>81.50 ± 2.29</b>	<b>89.47 ± 0.73</b>	<b>94.77 ± 0.63</b>	<b>94.36 ± 0.44</b>	<b>99.10 ± 0.44</b>	<b>94.32 ± 0.20</b>

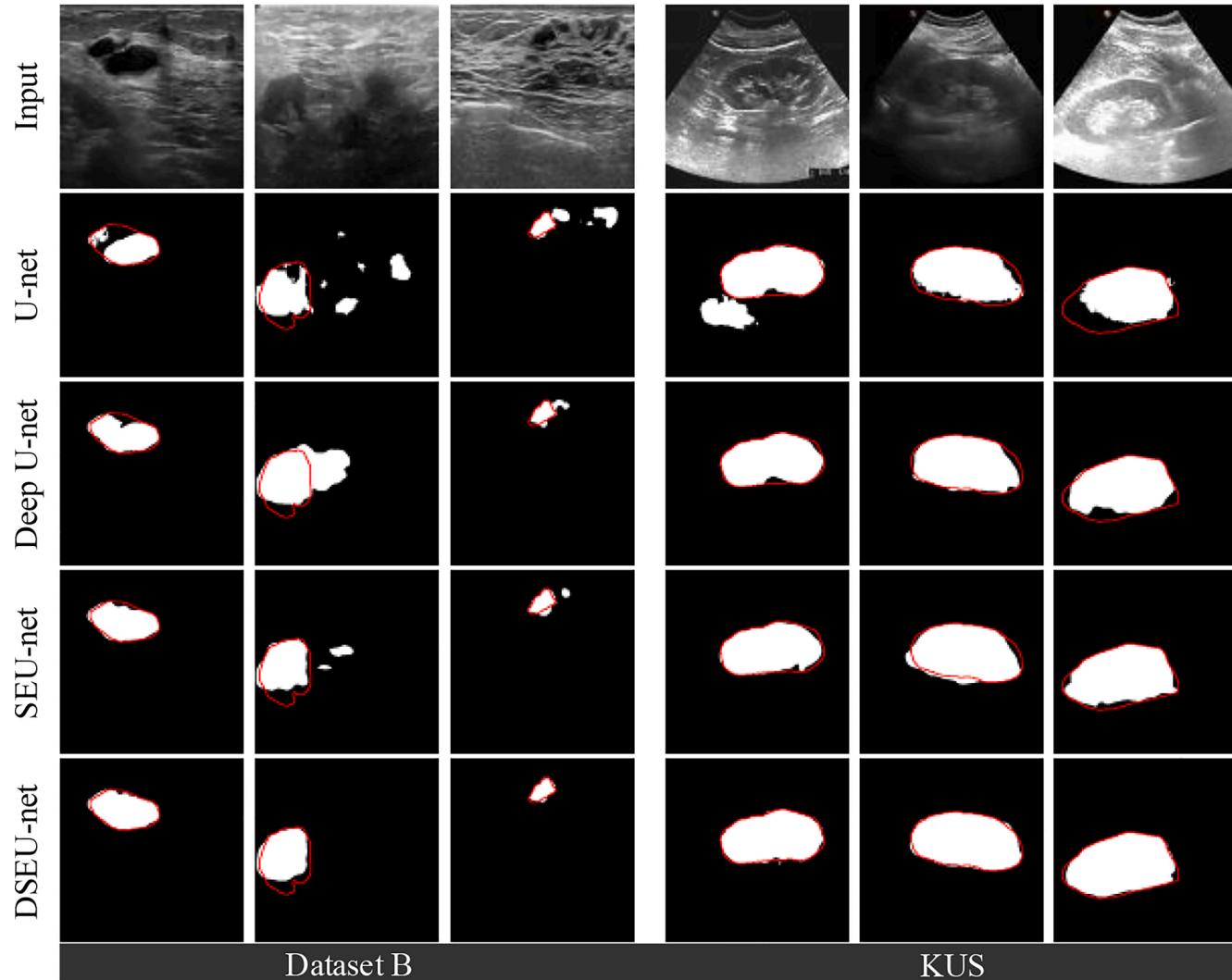


Fig. 3. The visualization of prediction results for different network components on two ultrasound datasets.

detection in the prediction results are further alleviated.

#### 4.2. Comparison with state-of-the-art segmentation methods

In this work, five the state-of-the-art medical image segmentation methods and five the state-of-the-art ultrasound image segmentation methods are used for the comparative experimental analysis. These methods used for the comparative analysis are U-net (Ronneberger et al., 2015), Att U-net (Oktay et al., 2018), U-net++ (Zhou et al., 2020), U-net3+ (Huang et al., 2020), SegNet (Badrinarayanan et al., 2017),

MADU-net (Abraham and Khan, 2019), RDAU-Net (Zhuang et al., 2019), AE U-Net (Yan et al., 2022) and SKU-net (Byra et al., 2020). In the comparative experiments, BUSI and Dataset B are performed for four-fold cross-validation and KUS is executed for three-fold cross-validation.

The quantitative evaluation results of BUSI and Dataset B are presented in Table 2. The quantitative segmentation results of KUS are illustrated in Table 3. It is worth noting that there are cases of segmentation failure in breast ultrasound images, and the boundary-based evaluation metrics cannot be applied. Therefore, the boundary-based evaluation metrics are only used for the quantitative evaluation of the

**Table 2**

The segmentation results (mean  $\pm$  std) of different competing methods on BUSI and Dataset B. We perform four-fold cross-validation on BUSI and Dataset B, respectively. The p-value of the paired t-test is less than 0.05 (\*:  $p < 0.05$ ).

Methods	BUSI					Dataset B				
	Jaccard	Precision	Recall	Specificity	Dice	Jaccard	Precision	Recall	Specificity	Dice
U-net	60.70 $\pm$ 2.36	71.88 $\pm$ 2.41	76.30 $\pm$ 2.48	96.18 $\pm$ 0.55	70.10 $\pm$ 2.20	58.44 $\pm$ 4.26	70.27 $\pm$ 6.11	75.32 $\pm$ 2.85	98.44 $\pm$ 0.40	68.20 $\pm$ 4.23
STAN	64.10 $\pm$ 3.05	73.96 $\pm$ 3.30	78.39 $\pm$ 2.16	96.64 $\pm$ 0.67	73.04 $\pm$ 2.95	57.09 $\pm$ 3.92	67.71 $\pm$ 3.11	69.95 $\pm$ 6.17	98.58 $\pm$ 0.47	66.06 $\pm$ 4.24
Att U-net	57.09 $\pm$ 1.22	78.78 $\pm$ 4.67	66.97 $\pm$ 4.08	96.87 $\pm$ 0.83	67.99 $\pm$ 1.18	59.93 $\pm$ 4.53	70.40 $\pm$ 6.05	76.15 $\pm$ 4.21	98.43 $\pm$ 0.33	69.30 $\pm$ 4.07
RDAU-net	63.75 $\pm$ 3.36	71.25 $\pm$ 4.11	78.90 $\pm$ 1.35	96.63 $\pm$ 0.76	71.94 $\pm$ 3.46	58.17 $\pm$ 4.91	70.49 $\pm$ 4.26	73.55 $\pm$ 5.28	98.37 $\pm$ 0.39	68.22 $\pm$ 4.94
U-net++	61.38 $\pm$ 1.73	79.68 $\pm$ 3.07*	71.44 $\pm$ 2.77	97.04 $\pm$ 0.54	71.58 $\pm$ 2.09	61.19 $\pm$ 5.86	68.32 $\pm$ 5.73	79.64 $\pm$ 3.84	98.44 $\pm$ 0.41	69.77 $\pm$ 5.30
MADU-net	61.62 $\pm$ 2.69	73.77 $\pm$ 2.90	76.87 $\pm$ 2.58	96.40 $\pm$ 0.62	71.35 $\pm$ 2.67	63.09 $\pm$ 3.04	73.70 $\pm$ 5.08	79.24 $\pm$ 1.72	98.61 $\pm$ 0.36	72.32 $\pm$ 3.14
U-net3+	63.03 $\pm$ 2.79	71.89 $\pm$ 3.28	79.58 $\pm$ 2.48	96.19 $\pm$ 0.68	71.85 $\pm$ 2.73	65.63 $\pm$ 5.26*	73.50 $\pm$ 6.21	80.29 $\pm$ 3.93*	98.60 $\pm$ 0.36	73.98 $\pm$ 4.72*
SegNet	67.31 $\pm$ 1.87	76.09 $\pm$ 2.00	79.85 $\pm$ 1.03*	96.99 $\pm$ 0.53	75.64 $\pm$ 1.80	62.83 $\pm$ 2.20	71.72 $\pm$ 1.70	80.15 $\pm$ 3.90	98.59 $\pm$ 0.30	72.16 $\pm$ 1.52
AE U-net	64.57 $\pm$ 2.91	74.44 $\pm$ 3.74	79.00 $\pm$ 2.11	96.80 $\pm$ 0.54	73.47 $\pm$ 3.03	62.37 $\pm$ 2.16	72.27 $\pm$ 1.91	78.97 $\pm$ 2.29	98.67 $\pm$ 0.28	72.23 $\pm$ 2.14
SKU-net	68.10 $\pm$ 1.63*	78.62 $\pm$ 1.66	79.53 $\pm$ 1.93	97.33 $\pm$ 0.45*	76.92 $\pm$ 1.57*	64.25 $\pm$ 4.01	75.27 $\pm$ 6.70*	79.36 $\pm$ 2.50	98.68 $\pm$ 0.39*	73.53 $\pm$ 4.05
Ours	70.36 $\pm$ 1.70	79.73 $\pm$ 1.71	82.70 $\pm$ 2.13	97.42 $\pm$ 0.60	78.51 $\pm$ 1.87	73.17 $\pm$ 2.73	82.58 $\pm$ 2.28	84.02 $\pm$ 3.57	99.05 $\pm$ 0.19	81.50 $\pm$ 2.29

**Table 3**

The segmentation results (mean  $\pm$  std) of different competing methods on the KUS. We perform three-fold cross-validation on the KUS. The p-value of the paired t-test is less than 0.05 (\*:  $p < 0.05$ ).

Methods	Jaccard	Precision	Recall	Specificity	Dice	HD	ASSD	ABD
U-net	83.19 $\pm$ 2.69	90.81 $\pm$ 2.64	91.20 $\pm$ 0.66	98.33 $\pm$ 0.56	90.42 $\pm$ 1.79	57.38 $\pm$ 17.83	2.09 $\pm$ 0.99	9.62 $\pm$ 3.43
STAN	86.01 $\pm$ 0.48	92.64 $\pm$ 1.19	92.51 $\pm$ 0.65	98.79 $\pm$ 0.42	92.12 $\pm$ 0.32	31.28 $\pm$ 8.22*	0.95 $\pm$ 0.37	5.57 $\pm$ 1.04
Att U-net	85.99 $\pm$ 1.50	91.56 $\pm$ 1.99	93.70 $\pm$ 0.32*	98.50 $\pm$ 0.46	92.10 $\pm$ 0.92	52.14 $\pm$ 8.22	1.03 $\pm$ 0.04	6.39 $\pm$ 0.97
RDAU-net	83.13 $\pm$ 1.02	93.53 $\pm$ 1.32*	85.80 $\pm$ 2.60	98.97 $\pm$ 0.50*	90.47 $\pm$ 0.56	36.37 $\pm$ 7.55	1.57 $\pm$ 0.47	6.98 $\pm$ 0.58
U-net++	85.59 $\pm$ 1.53	90.33 $\pm$ 1.69	93.22 $\pm$ 0.22	98.28 $\pm$ 0.51	91.78 $\pm$ 1.01	57.42 $\pm$ 4.92	0.79 $\pm$ 0.21	6.28 $\pm$ 0.81
MADU-net	86.16 $\pm$ 1.66	93.24 $\pm$ 1.46	91.54 $\pm$ 2.54	98.95 $\pm$ 0.49	92.27 $\pm$ 1.09	53.40 $\pm$ 13.11	0.96 $\pm$ 0.12	6.54 $\pm$ 1.39
U-net3+	86.35 $\pm$ 1.41	91.73 $\pm$ 2.07	93.33 $\pm$ 0.51	98.52 $\pm$ 0.49	92.32 $\pm$ 0.80	35.65 $\pm$ 8.87	0.93 $\pm$ 0.21	5.81 $\pm$ 1.27
SegNet	85.78 $\pm$ 0.98	92.66 $\pm$ 1.47	92.31 $\pm$ 0.89	98.67 $\pm$ 0.63	91.90 $\pm$ 0.60	42.43 $\pm$ 13.41	0.71 $\pm$ 0.04*	5.03 $\pm$ 0.85*
AEU-net	83.90 $\pm$ 4.03	91.98 $\pm$ 1.22	90.91 $\pm$ 5.88	98.54 $\pm$ 0.44	90.79 $\pm$ 2.61	53.32 $\pm$ 33.08	2.17 $\pm$ 2.42	8.36 $\pm$ 5.13
SKU-net	87.16 $\pm$ 0.42*	93.23 $\pm$ 1.36	93.36 $\pm$ 0.93	98.72 $\pm$ 0.37	92.91 $\pm$ 0.25*	31.57 $\pm$ 9.00	0.95 $\pm$ 0.27	5.33 $\pm$ 0.55
Ours	89.47 $\pm$ 0.73	94.77 $\pm$ 0.63	94.36 $\pm$ 0.44	99.10 $\pm$ 0.20	94.32 $\pm$ 0.43	12.42 $\pm$ 2.84	0.48 $\pm$ 0.08	3.44 $\pm$ 0.37

segmentation results of the kidney ultrasound images. As can be seen in [Table 2](#) and [Table 3](#), the DSEU-net designed in this paper achieved the best segmentation results on the three ultrasound datasets. the five quantitative evaluation results of DSEU-net on BUSI are 70.36%, 79.73%, 82.70%, 97.42% and 78.51%, respectively. The five quantitative evaluation results of DSEU-net on Dataset B are 73.17%, 82.58%, 84.02%, 99.05% and 81.50%, respectively. The eight quantitative evaluations of DSEU-net on KUS are 89.47%, 94.77%, 94.36%, 99.10%, 94.32%, 12.42, 0.48 and 3.44. Compared to the second result, the average values of the five metrics (Jaccard, Precision, Recall, Specificity and Dice) on BUSI, Dataset B and KUS are improved by 1.82%, 7.28% and 1.27%, respectively. Compared to the second result, the mean values of the three metrics (HD, ASSD, ABD) on KUS are reduced by 81.97%. To further estimate the improvement of these evaluation indexes by the method proposed in this paper, a paired t-test is performed. As shown in [Table 2](#) and [Table 3](#), the p-values of the t-tests demonstrated that DSEU-net had significant improvement on these quantitative evaluation metrics.

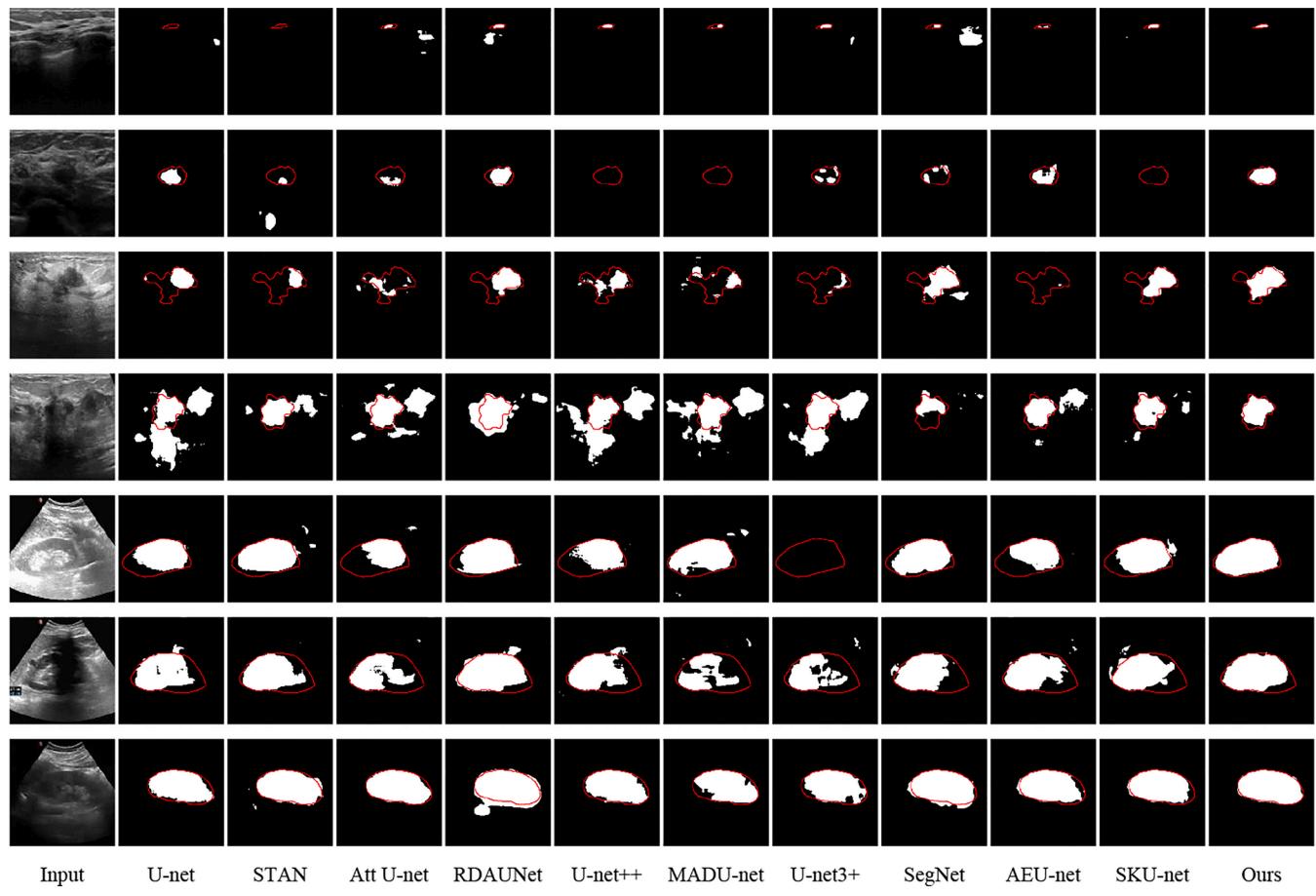
The visual segmentation results of different segmentation methods on BUSI, Dataset B and KUS datasets are displayed in [Fig. 4](#). It is obvious that the DSEU-net developed in this paper achieves the most satisfactory segmentation results on various ultrasound images with more complete objective masks. Due to the complexity of ultrasound images and the perturbation of similar surrounding tissues can cause serious missed and false detections in segmentation results, and even the target tissue may

be misidentified as the background leading to segmentation failure. Compared with other methods, this study greatly improves the ability of the network to cope with complex images by increasing the depth of the baseline network, introducing the SE attention module and deep supervision constraints. From the fourth and sixth row images, it can be seen that DSEU-net can largely mitigate the interference of fan-shaped shadow on the segmentation results. Based on the segmentation results in the second and third rows, it can be concluded that the similar surrounding tissue has the least effect on DSEU-net. Although the method proposed in this paper has some missed and false detections, it achieves the most competitive performance overall.

## 5. Discussion

### 5.1. Robustness analysis

To improve the segmentation accuracy of medical ultrasound images, a novel deep supervised attention U-net is proposed. To verify the effectiveness of each component in the DSEU-net, we conducted ablation experiments on two ultrasound datasets. According to [Table 1](#) and [Fig. 3](#), it is obvious that each key component design can boost the segmentation performance of the network. The superiority of DSEU-net for the ultrasound image segmentation is further validated by the extensive experiments conducted on the three ultrasound datasets with the state-of-the-art segmentation methods as shown in [Table 2](#), [Table 3](#) and [Fig. 4](#). Based



**Fig. 4.** The visualized prediction results of different segmentation methods on the three ultrasound datasets.

on the ablation experiments and comparison experiments, we can also draw two conclusions: (1) According to Table 1, Table 2 and Table 3, it is observed that simply increasing the depth of the network can achieve better segmentation performance than most methods on ultrasound images, which further demonstrates the necessary use of a deeper network framework to segment ultrasound images. (2) Compared with the benchmark U-net, various U-net variants have improved the segmentation accuracy of ultrasound images to different degrees. However, most variants of the network have small improvement on the segmentation accuracy. Therefore, it is necessary to design specific components according to ultrasound image characteristics.

## 5.2. Complexity analysis

Except for the network performance, the network parameter size and computational costs are two key points to be considered. To further compare the advantages of different segmentation frameworks, we conducted a statistical analysis of the network parameters and computational costs. Table 4 shows the statistical results of the complexity of different segmentation frameworks. It is clear that U-net has fewer network parameters and lower computational cost. On the contrary, DSEU-net has the most network parameters due to the addition of different network components (such as attention module, deep

supervision, etc.). It is worth noting that DSEU-net does not significantly increase the computational cost to achieve the fourth result. Although STAN, U-net++, MADU-net, U-net3+, SegNet, AE U-net and SKU-net have fewer network parameters, they require more computational costs. Comparing with these seven methods, the network parameters of our method are increased by 33.06, 26.23, 23.01, 24.02, 20.84, 43.13 and 28.84, respectively. On the contrary, the computational costs are decreased by 41.41, 407.67, 5.55, 668.52, 15.46, 28.12 and 38.05, respectively. Comparing to U-net, the network parameters are increased 8 times but the computational costs are increased only 3.4 times. The network parameters are increased 1.8 times compared to Att U-net but the computational costs are only increased 1.5 times. Through the above analysis, we can conclude that DSEU-net utilizes more network parameters to characterize the objective tissue, but does not significantly increase the computational cost. In general, the analysis of the complexity of different segmentation frameworks further demonstrates the superiority of the DSEU-net proposed in this work.

## 5.3. Comparison with attention modules

To fully evaluate the superiority of the SE attention module, we performed a comparative analysis with six attention modules used for image segmentation. These attention modules are spatial attention

**Table 4**

Complexity analysis of different segmentation methods in terms of network parameters and computational costs.

	U-net	STAN	Att U-net	RDAU-Net	U-net++	MADU-net	U-net3+	SegNet	AE U-net	SKU-net	Ours
Params (M)	7.85	29.80	35.56	51.76	36.63	39.85	38.84	42.02	19.73	34.02	62.86
GFLOPs	62.82	254.18	141.45	133.13	620.44	218.32	881.29	228.23	240.95	250.82	212.77

(SPA) (Oktay et al., 2018), attention guided block (AG) (Zhang et al., 2019), squeeze and attention block (SQA) (Zhong et al., 2020), efficient channel attention (ECA) (Wang et al., 2020), concurrent spatial and channel SE block (scSE) (Roy et al., 2018), hybrid adaptive attention module (HAAM) (Chen et al., 2022b).

In the experiments, the deeper U-net is considered as the benchmark framework. The segmentation results of these attention modules on the three datasets are demonstrated in Table 5 and Table 6, respectively. Based on the results in Table 5 and Table 6, it can be found that the introduction of the spatial attention mechanism improves the baseline network segmentation performance on BUSI, but reduces the baseline network segmentation performance on Dataset B and KUS. Similarly, the comparison between scSE and SE shows that the introduction of spatial attention does not improve the segmentation performance of the network for ultrasound images. Similar to the SPA module, the introduction of the AG and SQA modules improves the segmentation accuracy of BUSI, but does not perform well on Dataset B and KUS. This indicates that the AG and SQA modules have poor adaptability to different ultrasound images. Although the AG module achieved the first result in the Specificity and HD indexes, it performed generally in the remaining metrics. Comparing with SPA, AG and SQA, the addition of HAAM further improves the segmentation performance of the benchmark network on breast ultrasound images, but the segmentation performance on kidney ultrasound images is poor. The introduction of the ECA module improves the segmentation accuracy of the benchmark network for the three ultrasound images. It is worth noting that the ECA does not improve the segmentation accuracy of Dataset B as much as BUSI and KUS. According to the results of scSE and SE, it can be noticed that they can significantly improve the segmentation performance of the benchmark network on the three ultrasound datasets. Although the performance of the SE module is not satisfactory in the three metrics of Precision, Specificity and HD, it has the best segmentation performance overall. Moreover, the p-values of the t-tests demonstrated that SE module had significant improvement on these quantitative evaluation metrics, as shown in Table 5 and Table 6. Based on the comparison of different attention modules, we can draw several conclusions: (1) The SPA, AG and SQA models are more sensitive to different input images and the improvement of ultrasound image segmentation accuracy is limited. In general, they are unable to extract robust objective characterization from complex ultrasound images and are not suitable for medical ultrasound image segmentation tasks. (2) HAAM is able to cope with the task of breast ultrasound image segmentation, but is sensitive to kidney ultrasound images. (3) ECA achieved the third result among the compared attention methods, which shows that it has some potential for medical ultrasound image segmentation tasks, but there is still room for further improvement. (4) The scSE and SE modules have better robustness and generalization ability, which can significant improvement on ultrasound image segmentation accuracy. In summary, the introduction of the SE module can help the network to obtain the

optimal segmentation performance on ultrasound images.

#### 5.4. Limitations and future works

Extensive experiments on three public datasets have adequately demonstrated the robustness and superiority of the proposed DSEU-net for ultrasound image segmentation in this paper. Compared with existing methods, our method is able to mitigate the perturbations of various factors to reduce the occurrence of missed and false detections. However, the segmentation of ultrasound images by DSEU-net can still be further refined to obtain more complete prediction masks as shown in Fig. 3 and Fig. 4. In future research, we will overcome these challenges by the introduction of boundary constraints. The constraint on the prediction mask boundary can help us to obtain more refined and complete segmentation results. In addition, the introduction of different components will inevitably increase the network parameters. Therefore, a more lightweight and efficient framework for segmentation networks is also a need for further consideration in future work.

## 6. Conclusion

In this work, we constructed a DSEU-net to alleviate the challenges of ultrasound image segmentation by increasing the network depth, SE block, and deep supervision constraints. The ablation experiments of the network fully demonstrate the effectiveness and necessity of these network components to cope with ultrasound image segmentation. Comparative experimental analysis with the state-of-the-art segmentation methods on three ultrasound data also further validates the robustness of the DSEU-net for ultrasound image segmentation. In addition, the comparison of several attention mechanism modules illustrates the superiority of SE block. In summary, DSEU-net achieves the most competitive segmentation performance on ultrasound image segmentation.

#### CRediT authorship contribution statement

**Gongping Chen:** Methodology, Software, Writing – original draft, Formal analysis, Writing – review & editing. **Yuming Liu:** Conceptualization, Data curation, Investigation. **Jiang Qian:** Conceptualization, Data curation, Investigation. **Jianxun Zhang:** Conceptualization, Investigation, Visualization. **Xiaotao Yin:** Supervision, Validation, Resources, Data curation. **Liang Cui:** Validation, Resources, Data curation. **Yu Dai:** Conceptualization, Methodology, Investigation, Writing – review & editing, Funding acquisition.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

**Table 5**

The segmentation results (mean  $\pm$  std) of U-net with different attention modules on BUSI and Dataset B. The p-value of the paired t-test is less than 0.05 (\*:  $p < 0.05$ ).

		Deep U-net	SPA	AG	SQA	HAAM	ECA	scSE	SE
Dataset B	Jaccard	60.70 $\pm$ 2.36	67.62 $\pm$ 2.89	68.57 $\pm$ 1.00	68.02 $\pm$ 0.87	69.00 $\pm$ 2.33	69.24 $\pm$ 2.51*	67.68 $\pm$ 2.28	<b>69.74 <math>\pm</math> 2.34</b>
	Precision	71.88 $\pm$ 2.41	77.95 $\pm$ 3.49	78.82 $\pm$ 1.79	77.66 $\pm$ 1.89	78.43 $\pm$ 3.37	78.15 $\pm$ 2.18	<b>78.95 <math>\pm</math> 2.73</b>	78.41 $\pm$ 2.53
	Recall	76.30 $\pm$ 2.48	80.24 $\pm$ 1.17	81.14 $\pm$ 1.54*	80.03 $\pm$ 2.05	80.93 $\pm$ 2.09	81.71 $\pm$ 2.64	79.58 $\pm$ 1.14	<b>82.22 <math>\pm</math> 1.79</b>
	Specificity	96.18 $\pm$ 0.55	97.24 $\pm$ 0.54	<b>97.36 <math>\pm</math> 0.54</b>	97.30 $\pm$ 0.48	97.14 $\pm$ 0.60	97.25 $\pm$ 0.51	97.26 $\pm$ 0.48	97.19 $\pm$ 0.50
	Dice	70.10 $\pm$ 2.20	76.50 $\pm$ 3.03	77.07 $\pm$ 0.86	76.27 $\pm$ 0.87	77.42 $\pm$ 2.19*	77.41 $\pm$ 2.52	76.67 $\pm$ 2.20	<b>78.07 <math>\pm</math> 2.35</b>
	MCC	0.767 $\pm$ 0.02	0.756 $\pm$ 0.03	0.764 $\pm$ 0.01	0.756 $\pm$ 0.01	0.765 $\pm$ 0.02	0.766 $\pm$ 0.03	0.773 $\pm$ 0.03*	<b>0.778 <math>\pm</math> 0.02</b>
	Kappa	0.755 $\pm$ 0.02	0.745 $\pm$ 0.03	0.752 $\pm$ 0.01	0.745 $\pm$ 0.01	0.755 $\pm$ 0.02	0.755 $\pm$ 0.03	0.763 $\pm$ 0.03*	<b>0.768 <math>\pm</math> 0.02</b>
	Jaccard	67.86 $\pm$ 1.87	65.49 $\pm$ 3.90	65.34 $\pm$ 7.34	67.70 $\pm$ 2.24	68.40 $\pm$ 4.65	68.65 $\pm$ 3.25	70.24 $\pm$ 1.57*	<b>72.16 <math>\pm</math> 1.62</b>
	Precision	76.59 $\pm$ 2.76	75.44 $\pm$ 4.51	73.82 $\pm$ 6.31	77.19 $\pm$ 1.16	78.15 $\pm$ 3.01	77.96 $\pm$ 3.13	79.97 $\pm$ 4.14*	<b>81.14 <math>\pm</math> 2.24</b>
	Recall	80.95 $\pm$ 2.13	79.79 $\pm$ 2.00	80.35 $\pm$ 7.23	80.63 $\pm$ 2.32	81.41 $\pm$ 3.78	80.78 $\pm$ 4.18	82.81 $\pm$ 3.66*	<b>83.76 <math>\pm</math> 3.68</b>
	Specificity	98.65 $\pm$ 0.55	98.76 $\pm$ 0.27	98.61 $\pm$ 0.28	98.76 $\pm$ 0.27	98.81 $\pm$ 0.22	98.92 $\pm$ 0.29*	98.86 $\pm$ 0.23	<b>98.98 <math>\pm</math> 0.32</b>
	Dice	76.48 $\pm$ 1.77	74.51 $\pm$ 3.45	73.41 $\pm$ 7.07	76.03 $\pm$ 2.15	77.22 $\pm$ 4.05	76.79 $\pm$ 2.73	78.79 $\pm$ 2.28*	<b>80.23 <math>\pm</math> 0.93</b>
	MCC	0.776 $\pm$ 0.03	0.750 $\pm$ 0.03	0.743 $\pm$ 0.07	0.766 $\pm$ 0.02	0.777 $\pm$ 0.04	0.773 $\pm$ 0.03	0.794 $\pm$ 0.02*	<b>0.817 <math>\pm</math> 0.02</b>
	Kappa	0.766 $\pm$ 0.03	0.738 $\pm$ 0.04	0.727 $\pm$ 0.07	0.753 $\pm$ 0.02	0.765 $\pm$ 0.04	0.762 $\pm$ 0.03	0.781 $\pm$ 0.02*	<b>0.809 <math>\pm</math> 0.02</b>

**Table 6**

The segmentation results (mean  $\pm$  std) of U-net with different attention modules on the KUS. The p-value of the paired t-test is less than 0.05 (\*:  $p < 0.05$ ).

Methods	Jaccard	Precision	Recall	Specificity	Dice	MCC	Kappa	HD	ASSD	ABD
Deep U-net	88.22 $\pm$ 0.77	93.81 $\pm$ 1.69	93.97 $\pm$ 1.02	98.83 $\pm$ 0.45	93.55 $\pm$ 0.50	0.928 $\pm$ 0.01	0.925 $\pm$ 0.01	18.84 $\pm$ 3.07	0.58 $\pm$ 0.03	4.12 $\pm$ 0.43
SPA	87.47 $\pm$ 0.65	93.39 $\pm$ 1.61	93.59 $\pm$ 0.92	98.80 $\pm$ 0.39	93.15 $\pm$ 0.39	0.923 $\pm$ 0.01	0.921 $\pm$ 0.01	74.11 $\pm$ 2.85	0.67 $\pm$ 0.05	5.00 $\pm$ 0.57
AG	86.35 $\pm$ 1.59	92.47 $\pm$ 2.36	93.26 $\pm$ 0.72	98.63 $\pm$ 0.48	92.37 $\pm$ 1.02	0.915 $\pm$ 0.01	0.912 $\pm$ 0.01	13.02 $\pm$ 2.24	0.79 $\pm$ 0.16	4.74 $\pm$ 0.71
SQA	87.49 $\pm$ 0.74	92.81 $\pm$ 1.61	94.22 $\pm$ 1.04	98.65 $\pm$ 0.52	93.04 $\pm$ 0.52	0.923 $\pm$ 0.01	0.920 $\pm$ 0.01	22.55 $\pm$ 6.31	0.59 $\pm$ 0.06	4.47 $\pm$ 0.45
HAAM	88.02 $\pm$ 1.08	93.67 $\pm$ 1.89	93.92 $\pm$ 1.21	98.79 $\pm$ 0.49	93.42 $\pm$ 0.69	0.926 $\pm$ 0.01	0.924 $\pm$ 0.01	33.78 $\pm$ 17.82	0.56 $\pm$ 0.08	4.29 $\pm$ 0.64
ECA	88.33 $\pm$ 1.07	93.64 $\pm$ 1.50	94.28 $\pm$ 0.76	98.87 $\pm$ 0.38	93.61 $\pm$ 0.70	0.929 $\pm$ 0.01	0.926 $\pm$ 0.01	18.81 $\pm$ 5.57	0.56 $\pm$ 0.09	3.85 $\pm$ 0.51
scSE	88.75 $\pm$ 0.66*	94.04 $\pm$ 1.05*	94.30 $\pm$ 0.71*	98.97 $\pm$ 0.25*	93.90 $\pm$ 0.39*	0.932 $\pm$ 0.01*	0.930 $\pm$ 0.01*	16.08 $\pm$ 5.14	0.53 $\pm$ 0.06*	3.80 $\pm$ 0.48*
SE	89.18 $\pm$ 0.70	94.49 $\pm$ 1.14	94.31 $\pm$ 0.40	99.05 $\pm$ 0.27	94.17 $\pm$ 0.41	0.936 $\pm$ 0.01	0.935 $\pm$ 0.01	14.53 $\pm$ 0.94	0.50 $\pm$ 0.10	3.64 $\pm$ 0.38

the work reported in this paper.

## Data availability

Data will be made available on request.

## Acknowledgments

This work is supported by the National Natural Science Foundation of China (U1913207, 51875394) and Anhui Department of Education (KJ2021A1482).

## References

- Abraham, N., Khan, N.M.B.T., 2019. 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, pp. 683–687.
- Al-Dhabayani, W., Gomaa, M., Khaled, H., Fahmy, A., 2020. Dataset of breast ultrasound images. Data Br. 28, 104863.
- Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39, 2481–2495. <https://doi.org/10.1109/TPAMI.2016.2644615>
- Byra, M., Jarosik, P., Szubert, A., Galperin, M., Ojeda-Fournier, H., Olson, L., ... Andre, M. (2020). Breast mass segmentation in ultrasound with selective kernel U-Net convolutional neural network. *Biomedical Signal Processing and Control*, 61. <https://doi.org/10.1016/j.bspc.2020.102027>
- Chen, G., Dai, Y., & Zhang, J. (2023). RRCNet: Refinement residual convolutional network for breast ultrasound images segmentation. *Engineering Applications of Artificial Intelligence*, 117, Article 105601.
- Chen, G., Dai, Y., Zhang, J., 2022a. C-Net: Cascaded Convolutional Neural Network with Global Guidance and Refinement Residuals for Breast Ultrasound Images Segmentation. *Comput. Methods Programs Biomed.* 107086.
- Chen, G., Dai, Y., Zhang, J., & Yap, M. H. (2022). AAU-net: An Adaptive Attention U-net for Breast Lesions Segmentation in Ultrasound Images. *IEEE Transactions on Medical Imaging*.
- Chen, G., Dai, Y., Zhang, J., Yin, X., & Cui, L. (2022). MBANet: Multi-branch aware network for kidney ultrasound images segmentation. *Computers in Biology and Medicine*, 141, Article 105140.
- Chen, G., Dai, Y., Zhang, J., Yin, X., & Cui, L. (2021). SDFNet: Automatic segmentation of kidney ultrasound images using multi-scale low-level structural feature. *Expert Systems with Applications*, 185, Article 115619.
- Chen, G., Liu, Y., Dai, Y., Zhang, J., Cui, L., & Yin, X. (2022). In BAGNet: Bidirectional Aware Guidance Network for Malignant Breast lesions Segmentation (pp. 112–116). <https://doi.org/10.1109/ACIRSS55390.2022.9845607>
- Chen, G., Zhao, Y., Dai, Y., Zhang, J.-X., Yin, X.-T., Cui, L., & Qian, J. (2023). Asymmetric U-shaped network with hybrid attention mechanism for kidney ultrasound images segmentation. *Expert Systems with Applications*, 212, Article 118847.
- Deepak, Mishra, Santanu, Chaudhury, Mukul, Sarkar, Arvinder, Singh, Soin, 2018. Ultrasound Image Segmentation: A Deeply Supervised Network with Attention to Boundaries. *IEEE Transactions on Biomedical Engineering* 66, 1637–1648. 10.1109/TBME.2018.2877577.
- Houssein, E. H., Emam, M. M., Ali, A. A., & Suganthan, P. N. (2021). Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review. *Expert Systems with Applications*, 167, Article 114161. <https://doi.org/10.1016/j.eswa.2020.114161>
- Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E., 2020. Squeeze-and-Excitation Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42.
- Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., ... Wu, J. (2020). UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation. In: *ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1055–1059). <https://doi.org/10.1109/ICASSP40776.2020.9053405>
- Kim, S., Jang, Y., Jeon, B., Hong, Y., & Chang, H. (2018). 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*.
- Lee, H., Park, J., & Hwang, J. Y. (2020). Channel attention module with multiscale grid average pooling for breast cancer segmentation in an ultrasound image. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 67, 1344–1353.
- Liu, S., Wang, Y., Yang, X., Lei, B., Liu, L., Li, S. X., ... Wang, T. (2019). Deep Learning in Medical Ultrasound Analysis: A Review. *Engineering*, 5, 261–275. <https://doi.org/10.1016/j.eng.2018.11.020>
- Mahmud, W. M. H. W., Izaham, R. M. A. R., & Supriyanto, E. (2015). Boundary detection of kidney ultrasound image based on vector graphic approach. *ARPN Journal of Engineering and Applied Sciences*, 10, 8822–8829.
- Mendoza, C. S., Kang, X., Saifdar, N., Myers, E., Martin, A. D., Grisan, E., ... Linguraru, M. G. (2013). Automatic analysis of pediatric renal ultrasound using shape, anatomical and image acquisition priors. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer (pp. 259–266).
- Noble, J. A., & Boukerroui, D. (2006). Ultrasound image segmentation: A survey. *IEEE Transactions on Medical Imaging*, 25, 987–1010. <https://doi.org/10.1109/tmi.2006.877092>
- Oktay, O., Schlemper, J., Folgoc, L., Le, Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., Glocker, B., Rueckert, D., 2018. Attention U-Net: Learning Where to Look for the Pancreas, in: Medical Imaging with Deep Learning.
- Rezaei, Z. (2021). A review on image-based approaches for breast cancer detection, segmentation, and classification. *Expert Systems with Applications*, 182, Article 115204. <https://doi.org/10.1016/j.eswa.2021.115204>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer (pp. 234–241).
- Roy, A. G., Navab, N., & Wachinger, C. (2018). Concurrent spatial and channel ‘squeeze & excitation’ in fully convolutional networks. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer (pp. 421–429).
- Siddique, N., Paheding, S., Elkin, C. P., & Devabhaktuni, V. (2021). U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access*.
- Singh, V. K., Abdel-Nasser, M., Akram, F., Rashwan, H. A., Sarker, M. M. K., Pandey, N., ... Puig, D. (2020). Breast tumor segmentation in ultrasound images using contextual-information-aware deep adversarial learning framework. *Expert Systems with Applications*, 162, Article 113870. <https://doi.org/10.1016/j.eswa.2020.113870>
- Wang, H., Pulido, J. E., Song, Y., Furth, S. L., Tu, C., Zhang, C., ... Tasian, G. E. (2014). Segmentation of renal parenchymal area from ultrasound images using level set evolution. In: *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 4703–4706). <https://doi.org/10.1109/EMBC.2014.6944674>
- Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., & Hu, Q. (2020). ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 11531–11539). <https://doi.org/10.1109/CVPR42600.2020.01155>
- Wu, L., Yang, X., Li, S., Wang, T., & Dong, N. B. T. (2017). Cascaded Fully Convolutional Networks for automatic prenatal ultrasound image segmentation. In: *IEEE 14th International Symposium on Biomedical Imaging (ISBI)*. IEEE (pp. 663–666). <https://doi.org/10.1109/ISBI.2017.7950607>
- Xian, M., Zhang, Y., Cheng, H.-D., Xu, F., Zhang, B., & Ding, J. (2018). Automatic breast ultrasound image segmentation: A survey. *Pattern Recognition*, 79, 340–355.
- Yan, Y., Liu, Y., Wu, Y., Zhang, H., Zhang, Y., & Meng, L. (2022). Accurate segmentation of breast tumors using AE U-net with HDC model in ultrasound images. *Biomedical Signal Processing and Control*, 72, Article 103299. <https://doi.org/10.1016/j.bspc.2021.103299>

- Yap, M. H., Goyal, M., Osman, F., Martí, R., Denton, E., Juette, A., & Zwigelaar, R. (2020). Breast ultrasound region of interest detection and lesion localisation. *Artificial Intelligence in Medicine*, 107, Article 101880. <https://doi.org/10.1016/j.artmed.2020.101880>
- Yin, S., Peng, Q., Li, H., Zhang, Z., You, X., Fischer, K., ... Fan, Y. (2020). Automatic kidney segmentation in ultrasound images using subsequent boundary distance regression and pixelwise classification networks. *Medical Image Analysis*, 60, Article 101602. <https://doi.org/10.1016/j.media.2019.101602>
- Zhang, S., Fu, H., Yan, Y., Zhang, Y., Wu, Q., Yang, M., ... Xu, Y. (2019). Attention guided network for retinal image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer (pp. 797–805).
- Zhang, Y., Ying, M., Lin, Y., Ahuja, A.T., Chen, D.Z.B.T.-I.I.C. on B.& B., 2016. Coarse-to-Fine Stacked Fully Convolutional Nets for lymph node segmentation in ultrasound images, in: 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 443–448. 10.1109/BIBM.2016.7822557.
- Zhong, Z., Lin, Z. Q., Bidart, R., Hu, X., Daya, I. B., Li, Z., ... Wong, A. (2020). Squeeze-and-attention networks for semantic segmentation. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 13062–13071). <https://doi.org/10.1109/CVPR42600.2020.01308>
- Zhou, Z., Siddiquee, M., Tajbakhsh, N., & Liang, J. (2020). UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation. *IEEE Transactions on Medical Imaging*, 39, 1856–1867.
- Zhuang, Z., Li, N., Joseph Raj, A. N., Mahesh, V. G. V., & Qiu, S. (2019). An RDAU-NET model for lesion segmentation in breast ultrasound images. *PLoS One*, 14, e0221535.