ESKNet: An enhanced adaptive selection kernel convolution for breast tumors segmentation

Gongping Chen, Jianxun Zhang, Yuming Liu, Jingjing Yin, Xiaotao Yin, Liang Cui, Yu Dai

Abstract—Breast cancer is one of the common cancers that endanger the health of women globally. Accurate target lesion segmentation is essential for early clinical intervention and postoperative follow-up. Recently, many convolutional neural networks (CNNs) have been proposed to segment breast tumors from ultrasound images. However, the complex ultrasound pattern and the variable tumor shape and size bring challenges to the accurate segmentation of the breast lesion. Motivated by the selective kernel convolution, we introduce an enhanced selective kernel convolution for breast tumor segmentation, which integrates multiple feature map region representations and adaptively recalibrates the weights of these feature map regions from the channel and spatial dimensions. This region recalibration strategy enables the network to focus more on high-contributing region features and mitigate the perturbation of less useful regions. Finally, the enhanced selective kernel convolution is integrated into U-net with deep supervision constraints to adaptively capture the robust representation of breast tumors. Extensive experiments with twelve state-of-the-art deep learning segmentation methods on three public breast ultrasound datasets demonstrate that our method has a more competitive segmentation performance in breast ultrasound images. The source code is publicly available at: https://github.com/CGPxy/ESKNet.

Index Terms—Breast tumors, Ultrasound images, Selection kernel convolution, Attention module, Deep supervision.

I. INTRODUCTION

Breast cancer is one of the most common malignant tumors in women, which seriously threatens women’s health and even their lives [1]–[3]. Regular early screening is critical to make treatment plans and improving survival rate due to the strong concealment and many inducements of breast cancer [4]. At present, ultrasound imaging is widely used in the clinical diagnosis of breast diseases due to its advantages of being non-invasive, cheap, and fast [5]. Unfortunately, because of the complexity of ultrasound images, even experienced radiologists are difficult to mark the lesion region accurately and quickly, as shown in Fig. 1. To overcome this problem, various computer-aided diagnosis systems (CAD) have been developed to assist doctors in interpreting breast ultrasound images [6]. It is well known that medical image segmentation can help to locate and evaluate pathological regions [8]–[10]. Therefore, it is one of the key steps of CAD to segment the breast lesions automatically and accurately from ultrasound images.

Accurate segmentation of lesion regions from breast ultrasound images has always been a subject of extensive research [11]–[13]. Previously, manual prior methods were often used to fit the contours of breast tumors [14]. Xue et al. pointed out that the limited representation ability of artificial prior can easy to cause the misrecognition of breast lesions in complex ultrasound images [14]. In addition, the method based on manual prior requires a lot of time and manpower. Recently, many representative CNNs have been successfully and widely used in medical image segmentation [15]–[17]. Among them, fully convolutional network (FCN) [18] and U-net [19] are the two most widely used network architectures in medical image segmentation. Almajalid et al. used U-net to segment breast lesions from ultrasound images for the first time [20]. In the same year, Yap et al. comprehensively evaluated the segmentation performance of U-net, FCN-AlexNet, and patch-based LeNet on breast ultrasound images [21]. Similarly, Mishra et al. used FCN to design a deep supervision network for ultrasound image segmentation [22]. However, pattern complexity and intensity similarity between the surrounding
tissues (i.e., background) and lesion regions (i.e., foreground) increase the difficulty in breast lesion segmentation [4]. Therefore, it is difficult to acquire satisfactory segmentation results by simply applying existing networks (such as U-net and FCN) on breast ultrasound images, as shown in Fig. 1.

For breast ultrasound images, accurate breast lesions segmentation is still a challenging task due to the following interference factors: (i) similar intensity distribution and blurred boundary, especially in malignant lesions; (ii) significant morphology and position variation of breast tumors [23]. To deal with the challenges bring by complex breast ultrasound images, the segmentation network design should not only be able to adapt to breast tumors with different scales, but also need to improve the focus on the lesion regions. The benefits of attention mechanisms and multi-scale convolution have been demonstrated in many low-level tasks [24], [25]. The attention mechanism biases the allocation of the most informative feature expressions and simultaneously suppresses the less useful ones [26]. Multi-scale convolution improves the representation ability of objects by using different convolution kernels to extract interesting features under different receptive fields [27], [28]. The CNN architecture based on attention mechanism and multi-scale convolution has been widely used in breast ultrasound image segmentation [14]. For example, Yan et al. proposed the attention-enhanced U-net (AE U-net) with hybrid dilated convolution based on attention U-net to segment the breast tumor from ultrasound images [30]. Zhuang et al. introduced dilated convolution and residual learning in the Att U-net to capture features under different receptive fields [31]. However, the use of dilated convolutions on deeper convolutional layers cannot capture sufficient contextual information [14]. To better capture the multi-scale information of the breast tumors, Punn et al. replaced the convolution blocks of Att U-net with the residual blocks constructed by the inception convolution layer [32]. Abraham et al. constructed a new U-shaped network (MADU-net) to segment breast ultrasound images by introducing a multi-scale image input pyramid and deep supervision mechanism into the Att U-net [33]. Multi-image inputs can provide more fine-grained feature maps, but introducing too many low-level feature maps will affect the characterize high-level semantic features and reduce the performance of the network. Although these methods improve the segmentation accuracy of breast tumors to varying degrees, they still have two obvious limitations: (i) The multi-scale information is more dependent on the artificially set convolution kernel size and cannot adaptively capture the multi-scale information of breast tumors [34]. (ii) They tend to use a single attention mechanism to calibrate objective features.

Recently, Li et al. designed a selective kernel convolution (SK) to adaptively select useful feature information under different receptive fields, as shown in Fig. 2(a) [35]. To adaptively capture the feature information of breast tumors under different receptive fields, Byra et al. developed a selective kernel U-net (SKNet) to segment breast tumors using the selective kernel convolution module [35]. Although the introduction of the selective kernel convolution module improves the segmentation accuracy of breast lesions, there are still two obvious limitations: (i) ignoring the calibration of spatial dimension features; (ii) reducing the correlation of features in the module. To overcome this limitation, we first introduce spatial attention into the selective kernel convolution module to calibrate spatial dimension features, as shown in Fig. 2(b). Then, residual learning is added to the selective kernel convolution module to strengthen the relevance of long-distance features, as shown in Fig. 2(b). Finally, we use the enhanced selective kernel convolution module (ESK) to construct a novel deep supervised U-net (Named as ESKNet) to segment breast lesions adaptively from ultrasound images. In general, the method proposed in this paper has the following contributions:

- First, an enhanced selective kernel convolution is designed, which not only adaptively selects features under different scale receptive fields from the channel and spatial dimensions but also further strengthens the relevance of remote feature information.
- Second, a novel deep supervision U-net integrating the enhanced selective kernel convolution module is developed to segment breast lesions from ultrasound images. The network can improve breast lesions segmentation accuracy by learning generic representations from breast ultrasound images.
- Moreover, extensive experiments on three public breast ultrasound datasets demonstrate that our approach consistently improves the segmentation accuracy of breast lesions, outperforming the strong baseline and state-of-the-art medical image segmentation methods.

II. RELATED WORKS

A. CNNs for Breast Ultrasound Segmentation

Deep CNNs have been successfully applied to lesion segmentation in breast ultrasound images [36], [37]. To segment the small tumor accurately from breast ultrasound images, Shareef et al. designed a small tumor-aware network. This method uses different multi-scale convolution blocks to integrate the context information and high-resolution features of breast tumors, which improves the accuracy of small breast tumor segmentation [38]. Lei et al. proposed a boundary-regularized deep convolutional encoder-decoder network to alleviate the challenge of segmenting whole breast ultrasound images [39]. Xue et al. developed a deep CNN equipped with a global guidance block and breast lesion boundary detection modules for boosting breast lesion segmentation [40]. Huang et al. proposed a boundary-rendering network for breast lesions segmentation by a differentiable boundary selection module and a GCN-based boundary rendering module [40]. However, obtaining accurate boundaries from heavily cascaded or shadow-occluded regions remains challenging. Tong et al. utilized residual convolution blocks to replace convolution blocks of Att U-net to segment breast tumors [41]. Subsequently, Zhuang et al. introduced dilated convolution layers based on the work of Tong et al. [41] to capture features under different receptive fields [31]. However, the objective feature captured on deeper convolutional layers lacks contextual information [14]. Moon et al. proposed an ensemble...
CNN architecture for the CAD system to diagnose breast ultrasound images [42]. However, this method is limited by the segmentation performance of existing networks on breast ultrasound images. Wang et al. used deep supervision strategy constraints on the feature maps captured at each stage of U-net to segment breast lesions [43]. However, excessively introducing deep supervision constraints will not improve the performance of the network and will increase the network parameters. In this study, we develop a novel U-net with deep supervision to segment breast lesions, in which the deep supervision mechanism was only added to the decoding stage.

B. Attention Mechanisms

CNN has become the preferred tool for medical image segmentation tasks [44]. However, these architectures learn filters capturing local spatial patterns along all the input channels and generate feature maps jointly encoding the spatial and channel information [45]. Inspired by human visual attention, many attention algorithms have been developed to strengthen the representation ability of CNNs. Oktay et al. developed a spatial attention module to weigh low-level semantic features and high-level instance features, which has been successfully applied [46]. An architectural component called squeeze & exception (SE) block, which is used to recalibrate the feature map to emphasize useful channels, was developed by Hu et al. [47]. Inspired by the SE block, Roy et al. proposed to have concurrent spatial and channel SE blocks (scSE) that recognize the feature maps separately along channel and space, and then combine the output [45]. Similarly, Zhong et al. developed a new squeeze-and-attention network (SANet) for semantic segmentation tasks based on SE block [48]. While much effort is put into improving this joint encoding of spatial and channel information, these methods cannot adaptively select useful features under different receptive fields. In order to alleviate this challenge, Li et al. designed a selective kernel convolution (SK) to adaptively select useful feature information at different scales from the channel dimension [34]. In our work, we introduce an enhanced selected kernel convolution (ESK), which can simultaneously calibrate the features under different receptive fields from the dimensions of space and channels to improve the representation ability.

III. METHOD

Fig. 3 is an illustration of our developed deep supervision U-net with enhanced selective kernel convolution (ESKNet) for breast lesions segmentation. Our ESKNet has the same core architecture as U-net [19] mainly including four downsampling, four upsampling, and four skip-connections. The difference is that an enhanced selective kernel convolution module (ESK) is proposed to replace the original convolution layer to better adapt to the segmentation of breast lesions. Convolutional layers with different kernel sizes in the ESK can provide receptive fields with different scales, which can improve the adaptability of the network to different input images. As shown in Fig. 2(b), the ESK can learn more robust representations from breast ultrasound images through channel dimension and spatial dimension constraints. In addition, the residual connection in each ESK module can enhance the relevance of long-distance features and improve the robustness of the network. To further refine the segmentation results, we use ground-truth masks to constrain each decoding stage.

A. Enhanced Selective Kernel Convolution (ESK)

As shown in Fig. 2(b). The original selection kernel convolution module can be roughly regarded as a multi-scale convolution block with different kernel sizes and a channel
attention block. The developed enhanced selection kernel convolution module mainly consists of three parts: a multi-scale convolution block with different kernel sizes, a channel attention module, and a spatial attention module. Specifically, the given intermediate feature maps $F \in \mathbb{R}^{c \times h \times w}$ first undergo two parallel convolution operations, namely $5 \times 5$ convolution and $3 \times 3$ dilated convolution. The dilated rate of the dilated convolution is 3. The feature map captured by two convolution operations can be denoted as:

$$F_1 = W_{5 \times 5} \times F$$

$$F_2 = W_{3 \times 3} \times F$$

where $F \in \mathbb{R}^{c \times h \times w}$ denotes the given intermediate feature maps, $W_{5 \times 5}$ and $W_{3 \times 3}$ denotes the matrix of $5 \times 5$ convolution and $3 \times 3$ dilated convolution, respectively. $F_1 \in \mathbb{R}^{c \times h \times w}$ and $F_2 \in \mathbb{R}^{c \times h \times w}$ denote the feature map captured by $5 \times 5$ convolution and $3 \times 3$ dilated convolution, respectively. Subsequently, $F_1 \in \mathbb{R}^{c \times h \times w}$ and $F_2 \in \mathbb{R}^{c \times h \times w}$ are integrated along the dimensions of the channel and input to the channel attention module (see Section 3.2) and the spatial attention module (see Section 3.3). The merged feature maps $F_M \in \mathbb{R}^{c \times h \times w}$ can be represented as:

$$F_M = F_1 \oplus F_2$$

where $\oplus$ denotes the element-wise addition. Finally, a new set of feature maps

$$F_{CS} \in \mathbb{R}^{c \times h \times w}$$

is obtained by combining the feature maps obtained from the channel attention module and the spatial attention module with the input feature maps $F \in \mathbb{R}^{c \times h \times w}$,

$$F_{CS} = F \oplus (SAM(F_M) \oplus CAM(F_M))$$

where $\oplus$ denotes the element-wise addition. $SAM(\cdot)$ and $CAM(\cdot)$ denote the spatial attention module and channel attention module, respectively.

### B. Channel Attention Module

As shown in Fig. 2(a) and Fig. 2(b), the channel attention used in this paper is the same as the channel attention in the SK module. The channel attention main purposes is to select more useful lesion features through the calibration of channel dimensions. Specifically, the channel-wise statistics are first obtained through a global average pooling (GAP) operation. The obtained feature can be expressed as:

$$S_C = GAP(F_M) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_M(i, j)$$

Then, the feature map $S_C$ is input to a fully connected layer followed by a batch-normalization layer and a ReLU layer to produce a set new feature map:

$$Z_C = \delta_r(B(W_{fc} \cdot S_C))$$

where $W_{fc} \in \mathbb{R}^{32 \times 1}$ represents the matrix of fully connected (FC) layers, the size of the matrix dimension is $32$. $B(\cdot)$ is the Batch-Normalization operation. $\delta_r(\cdot)$ represents the ReLU activation operation. The feature map $Z_C$ again undergoes a fully connected operation to obtain a new feature map,

$$Z = W_{fc} \cdot Z_C$$

where $W_{fc} \in \mathbb{R}^{C \times 1}$ represents the matrix of fully connected (FC) layers, the size of the matrix dimension is $C$. Finally, the feature map $Z$ is executed through sigmoid activation to obtain the channel attention map:

$$\beta = \sigma_s(Z)$$

In this paper, $\beta \in [0, 1]^{c \times 1 \times 1}$ is taken as the channel attention map of $F_2$, $1 - \beta \in [0, 1]^{c \times 1 \times 1}$ is regarded as the channel attention map $F_1$. Each value of the channel attention maps indicates the importance of channel information at the corresponding voxel in $F_1 / F_2$. The feature maps obtained after $F_1$ and $F_2$ are calibrated by the channel activation maps can be expressed as:

$$F_{C1} = (1 - \beta) \cdot F_1$$

$$F_{C2} = \beta \cdot F_2$$
where $F_{C1} \in \mathbb{R}^{c \times h \times w}$ and $F_{C2} \in \mathbb{R}^{c \times h \times w}$ are the final outputs of the channel attention module. During the calibration of channel features, the channel activation map is reshaped to the same dimension as the calibration feature map.

### C. Spatial Attention Module

The calibration on the channel clarifies the importance of each channel, but does not emphasize the location of the objective. Woo et al. pointed out that it is more beneficial to extract useful features to perform calibration operations simultaneously on the dimensions of channels and spaces [51]. In this paper, we introduce a spatial attention mechanism into the original SK module to calibrate spatial features, as shown in Fig. 2(b). Similar to the channel attention in the SK module, the spatial attention we designed also includes two branches. Especially, the feature maps $F_M$ undergo a ReLU activation, a $1 \times 1$ convolution operation, and a sigmoid activation, to obtain the spatial attention map:

$$\alpha = \sigma_s(W \cdot \delta_r(F_M))$$  \hspace{1cm} (11)

where $\delta_r(\cdot)$ and $\sigma_s(\cdot)$ represent the ReLU activation and sigmoid activation, respectively. $W \in \mathbb{R}^{1 \times h \times w}$ represents the matrix of $1 \times 1$ convolution. In this paper, $\alpha \in [0, 1]^{1 \times h \times w}$ is taken as the spatial attention map of $F_2$, $1 - \alpha \in [0, 1]^{1 \times h \times w}$ is regarded as the spatial attention map of $F_1$. Each value of the spatial attention map indicates the importance of spatial information at the corresponding voxel in $F_1/F_2$. The feature maps obtained after $F_1$ and $F_2$ are calibrated by the spatial activation maps can be expressed as:

$$F_{S1} = (1 - \alpha) \cdot F_1$$  \hspace{1cm} (12)

$$F_{S2} = \alpha \cdot F_2$$  \hspace{1cm} (13)

where $F_{S1} \in \mathbb{R}^{c \times h \times w}$ and $F_{S2} \in \mathbb{R}^{c \times h \times w}$ are the final outputs of the spatial attention module. During the calibration of spatial features, the spatial activation map is reshaped to the same dimension as the calibration feature map. Finally, the calibrated feature maps and the input feature maps are merged and input to the next stage,

$$F_{CS} = F \oplus F_{C1} \oplus F_{C2} \oplus F_{S1} \oplus F_{S2}$$  \hspace{1cm} (14)

where $\oplus$ denotes the element-wise addition. $F_{CS} \in \mathbb{R}^{c \times h \times w}$ denotes the feature map obtained by the enhanced selection kernel convolution module.

### D. Deep Supervision

To make the segmentation results closer to the ground-truth masks, we use the deep supervision strategy to constrain the decoding features, as shown in Fig. 3. By introducing deep supervision constraints, the decoder can learn to predict more accurate segmentation results scale-by-scale. Specifically, the feature maps captured at each decoding stage are first processed by a $1 \times 1$ convolution. Then, a sigmoid activation layer is used to predict the segmentation results. Finally, the predicted segmentation masks are up-sampled for comparison with the ground-truth masks. From Fig. 3, we can see that our method can predict five segmentation masks, which can be expressed as:

$$S_i = U^{16/i} (\sigma_s(W \cdot F_{Di}))$$  \hspace{1cm} (15)

where $S_i$ represents the predicted mask of the $i$th decoding stage, $W \in \mathbb{R}^{1 \times h \times w}$ represents the matrix of $1 \times 1$ convolution. $U^{16/i}(\cdot)$ and $\sigma_s(\cdot)$ represent the up-sampled operation and sigmoid activation operation, respectively. Although every prediction mask is up-sampled to the same size as the input image, the last prediction mask $S_5$ has the highest accuracy and hence is taken as the final output of the method.

### E. Loss Function

In this paper, binary cross-entropy (BCE) is used as the loss function. The loss of the method can be denoted as:

$$\mathcal{L} = \sum_{i=1}^{5} \ell_{BCE}^i$$  \hspace{1cm} (16)

where $\ell_{BCE}^i$ denote the segmentation loss of the $i$th decoding stage.

### IV. MATERIALS AND EXPERIMENTS

#### A. Datasets

In this paper, three widely used public breast ultrasound datasets are used to evaluate the segmentation network performance. The first breast ultrasound dataset (denotes BUSI) is constructed by Al-Dhabyani et al. [49]. The second breast ultrasound dataset used in this paper named Dataset B is collected by Yap et al. [50]. The third public breast ultrasound dataset is the STU provided by Zhuang et al. [51]. A more detailed description of the three public datasets is given in Table 1.

#### B. Experimental Settings

In this paper, we use k-fold cross-validation to carry out the experimental analysis. In the ablation and comparison experiments, we implement four-fold cross-validation on BUSI and Dataset B, respectively. In the robustness analysis of the network, we execute four-fold and three-fold cross-validation on benign and malignant lesions, respectively. Finally, we use Dataset B and STU to conduct external validation experiments.
Table 2
The quantitative evaluation results of different network components on BUSI and Dataset B.

| Architectures                  | BUSI                          | Dataset B                      |
|--------------------------------|-------------------------------|--------------------------------|
|                                | Jaccard | Precision | Recall | Specificity | Dice  | Jaccard | Precision | Recall | Specificity | Dice  |
| Baseline U-net                 | 60.70±2.36 | 71.88±2.41 | 76.30±2.48 | 96.18±0.55 | 70.10±2.20 | 79.80±2.05 | 81.47±2.97 | 97.30±0.42 | 73.99±4.24 |
| U-net with SK module           | 68.10±1.63 | 78.62±1.66 | 79.53±1.93 | 97.33±0.45 | 76.92±1.57 |
| U-net with ESK module          | 69.67±2.10 | 79.26±1.99 | 81.47±2.97 | 97.40±0.42 | 78.68±2.05 |
| Deep supervision U-net with ESK module (Ours) | **70.20±2.28** | **79.57±1.65** | **82.41±2.84** | **97.47±0.35** | **78.71±2.37** |

| Baseline U-net                 | 58.44±4.26 | 70.27±6.11 | 75.32±2.85 | 94.44±0.40 | 68.20±4.23 |
| U-net with SK module           | 64.25±4.01 | 75.27±6.70 | 79.36±2.50 | 98.68±0.39 | 73.53±4.05 |
| U-net with ESK module          | 67.53±2.38 | 77.25±2.89 | 79.93±3.86 | 97.87±0.25 | 76.34±1.97 |
| Deep supervision U-net with ESK module (Ours) | **71.65±2.39** | **81.01±3.91** | **82.66±1.40** | **99.01±0.35** | **79.92±2.21** |

Table 3
The segmentation results (mean ± std) of different competing methods on BUSI and Dataset B. We perform four-fold cross-validation on BUSI and Dataset B, respectively. Asterisks indicate that the difference between our method and the competing method is significant using a paired student’s t-test. (*: p < 0.05).

| Methods                  | BUSI                          | Dataset B                      |
|--------------------------|-------------------------------|--------------------------------|
|                          | Jaccard | Precision | Recall | Specificity | Dice  | Jaccard | Precision | Recall | Specificity | Dice  |
| U-net [19]               | 60.70±2.36 | 71.88±2.41 | 76.30±2.48 | 96.18±0.55 | 70.10±2.20 | 79.80±2.05 | 81.47±2.97 | 97.30±0.42 | 73.99±4.24 |
| STAN [38]                | 64.10±3.05 | 73.96±3.30 | 78.39±2.16 | 96.64±0.67 | 73.04±2.95 |
| Att U-net [40]           | 57.09±1.22 | 78.78±4.67 | 66.97±4.08 | 98.67±0.83 | 67.99±1.18 |
| RDAU-net [41]            | 63.75±3.36 | 71.25±4.11 | 78.90±1.35 | 96.63±0.76 | 71.94±3.46 |
| U-net++ [52]             | 61.38±1.73 | 79.68±3.07 | 71.44±2.77 | 97.04±0.54 | 71.58±2.09 |
| MADU-net [33]            | 61.62±2.69 | 73.77±2.90 | 76.87±2.58 | 94.60±0.62 | 71.35±2.67 |
| U-net3+ [37]             | 63.03±2.79 | 71.89±3.28 | 80.58±2.48 | 96.19±0.68 | 71.85±2.73 |
| SegNet [54]              | 67.31±1.87 | 76.09±2.00 | 80.85±1.03 | 96.99±0.53 | 75.64±1.80 |
| AE U-net [30]            | 64.57±2.91 | 74.44±3.74 | 79.00±2.11 | 98.60±0.54 | 73.47±3.03 |
| SANet [48]               | 65.96±2.78 | 74.84±4.18 | 80.76±2.30 | 96.75±0.70 | 74.46±2.76 |
| scSEU-net [45]           | 67.68±2.28 | 78.95±2.73 | 79.58±1.14 | 97.26±0.48 | 76.67±2.20 |
| SKNet [45]               | 68.10±1.63 | 78.62±1.66 | 79.53±1.93 | 97.33±0.45 | 76.92±1.57 |
| Ours                     | 70.20±2.28 | 79.57±1.65 | 82.41±2.84 | 97.47±0.35 | 78.71±2.37 |
|                          | 71.65±2.39 | 81.01±3.91 | 82.66±1.40 | 99.01±0.35 | 79.92±2.21 |

on the already trained segmentation networks. Since the STU dataset contains too few images, it is only used in external validation experiments.

In the experiment, the optimizer we used is Adam with default hyperparameters. In the training process, the initial learning rate is 0.001, and every 10 epochs are reduced by 50%. When the learning rate is less than 0.0001, the learning rate is set to 0.0001. The epoch size and batch size are set to 50 and 12, respectively. The open-source TensorFlow-2.0.6 is used to structure our segmentation network. Our development language and accelerator are Python 3.6 and NVIDIA RTX 3090, respectively.

C. Evaluation Metrics

In this paper, five widely-used segmentation metrics are used to evaluate the segmentation performance of different methods on breast lesions. They are Jaccard, Precision, Recall, Specificity, and Dice [55].

V. RESULTS

A. Ablation Study

To fully verify the effectiveness of different network components (such as the SK module, ESK module, and deep supervision), we conduct the four-fold cross-validation ablation experiment on BUSI and Dataset B, respectively. In the experiment, the original U-net is used as the benchmark network.

The quantitative evaluation results of different components on BUSI and Dataset B are shown in Table 2. According to the results in Table 2, several conclusions can be drawn. First, the introduction of these components significantly improves the segmentation performance of the original U-net on breast lesions. Second, according to the comparison between the SK and ESK module, it can be concluded that the calibration of spatial dimension features through spatial attention is conducive to more robust feature selection. Third, the addition of the deep supervision mechanism can further guide the network to predict more precise segmentation masks. To sum up, these components designed in this paper play an essential role in improving the segmentation accuracy of breast lesions.

B. Comparison Experiments

In this study, twelve methods for ultrasound image or medical image segmentation are used to compare with our methods. These state-of-the-art deep learning segmentation methods are U-net [19], Att U-net [40], U-net++ [52], U-net3+ [37], scSEU-net [45], SANet [48], STAN [38], SKNet [45], AE U-Net [30], RDAU-Net [31], MADU-net [33]. As the extension of U-Net, U-Net++ and U-net3+ connect the encoder and decoder sub-networks through a series of nested dense skip pathways to further improve the target characterization capability. SegNet is also one of the methods widely used in medical image segmentation by recording the location information of the objective. The evaluation results of
various methods on BUSI and Dataset B are shown in Table 3. Compared with the twelve state-of-the-art segmentation methods, the method proposed in this paper achieves the best segmentation results. In the segmentation of BUSI, the values of the five evaluation indicators are 70.20%, 79.57%, 82.41%, 97.47%, and 78.71%, respectively. The values of the five evaluation indicators on Dataset B are 71.65%, 81.01%, 82.66%, 99.01%, and 79.92%, respectively. To highlight the improvement of our method on the segmentation results, we perform paired student’s t-test with the second results, and the p-value ($p < 0.05$) indicates our method has significant improvement in five indicators.

In addition, we illustrate the P-R curves and the ROC curves of different segmentation methods on BUSI and Dataset B in Fig. 4. The P-R curve represents the confidence level that the true positive and false positive classes are predicted correctly. The ROC curve represents the confidence level that a method predicts correctly. Compared to other methods, our method achieves the highest AUC values on both BUSI and Dataset B. The AUC values of our method on BUSI and Dataset B are 0.9149 and 0.9395, respectively. According to the comparison of P-R and ROC curves, it can be concluded that our method achieves the highest confidence level in the segmentation of BUSI and Dataset B.

The comparison with the state-of-the-art segmentation methods in five quantitative evaluation indicators fully demonstrates the competitiveness of our method. In addition, the analysis based on P-R and ROC curves also shows the feasibility of our method in breast lesion segmentation. In order to compare the segmentation results of each method more intuitively, the segmentation results of breast tumors are visualized. The visual segmentation masks of various methods are shown in Fig. 5. In general, according to Table 3 and Fig. 5, we can draw three key points: (i) Compared with other methods, the method proposed in this paper can better deal with different input images. As shown in Fig. 5, our method is less disturbed by the disease category, lesion morphology, and surrounding tissues. This shows that the method proposed in this paper has good adaptability and can extract robust breast tumor representation. (ii) Although various optimized U-nets improve the segmentation accuracy of breast lesions to varying degrees, especially SKNet, scSEU-net, etc., these variants still have room for further improvement. Although different segmentation methods have achieved relatively competitive results in individual evaluation indicators, their overall performance in BUSI and Dataset B is not ideal, as shown in Table 3. (iii)
Fig. 5. The segmentation results of different methods on BUSI and Dataset B. Some methods fail to segment breast tumors on individual ultrasound images.

Due to the complexity of ultrasound modalities, these variant networks still have seriously missed detections and false detections on individual images, and even the segmentation fails, as shown in the second and fifth rows. (iv) Our method also has some missing and false detection in the segmentation of breast lesions. Compared with other methods, the segmentation result of this method has fewer missing false detection regions and is closer to the ground-true mask.

VI. DISCUSSION

In this study, we propose deep supervision U-net with enhanced selective kernel convolution (ESKNet) for breast
The segmentation results (mean ± std) of benign and malignant lesions in BUSI by different methods. We perform four-fold cross-validation on the benign lesions and three-fold cross-validation on the malignant lesions. Asterisks indicate that the difference between our method and the competing method is significant using a paired student’s t-test. (*p < 0.05).

| Methods       | Jaccard | Precision | Recall | Specificity | Dice |
|---------------|---------|-----------|--------|-------------|------|
| U-net [19]    | 61.53±3.98 | 74.97±2.80 | 73.97±5.81 | 97.72±0.59 | 70.49±3.23 |
| STAN [30]     | 64.20±2.73  | 73.37±3.70  | 79.91±1.27  | 97.65±0.52  | 71.98±2.98  |
| Att U-net [46] | 65.03±2.05  | 75.24±1.68  | 79.44±2.84  | 97.68±0.62  | 73.30±2.00  |
| RDAU-net [31] | 64.70±2.17  | 72.54±1.57  | 79.36±0.98  | 97.79±0.28  | 72.70±1.62  |
| U-net++ [52]  | 68.25±2.75  | 75.93±3.66  | 81.58±1.09  | 97.74±0.62  | 75.56±2.79  |
| MADU-net [43] | 66.74±2.10  | 76.74±2.94  | 79.97±1.64  | 97.75±0.60  | 74.82±2.26  |
| U-net3+ [53]  | 67.63±1.86  | 75.58±2.88  | 81.07±1.27  | 97.72±0.58  | 75.07±2.10  |
| SegNet [54]   | 67.89±3.31  | 76.96±3.11  | 79.57±2.21  | 97.98±0.46  | 75.47±2.91  |
| AE U-net [30] | 67.89±1.96  | 77.17±3.63  | 80.54±1.25  | 97.95±0.60  | 75.77±1.82  |
| SANet [45]    | 65.96±2.78  | 74.84±4.18  | 80.76±2.30  | 96.75±0.70  | 74.46±2.76  |
| scSEU-net [35] | 71.33±2.45  | 80.76±2.69  | 81.58±2.45  | 98.20±1.47  | 78.97±2.44  |
| SKNet [6]     | 69.91±2.11  | 79.15±2.05  | 81.54±2.17  | 98.06±0.52  | 77.84±2.98  |
| Ours          | 72.73±2.12  | 81.50±2.62  | 82.69±0.40  | 98.29±0.40  | 80.17±1.79  |

| Methods       | Jaccard | Precision | Recall | Specificity | Dice |
|---------------|---------|-----------|--------|-------------|------|
| U-net [19]    | 51.11±2.62  | 64.96±2.55  | 68.86±4.27  | 93.63±1.28  | 63.47±2.38  |
| STAN [30]     | 50.1±2.38   | 62.74±3.72  | 70.96±6.75  | 93.69±1.56  | 62.50±1.97  |
| Att U-net [46] | 51.12±2.35  | 61.62±0.97  | 72.57±2.17  | 93.12±1.00  | 62.95±2.14  |
| RDAU-net [31] | 51.63±1.62  | 60.85±5.01  | 71.89±2.55  | 93.87±1.45  | 64.22±4.24  |
| U-net++ [52]  | 54.03±3.03  | 65.50±2.94  | 73.43±2.10  | 93.73±1.31  | 65.52±2.75  |
| MADU-net [43] | 54.12±2.96  | 67.46±3.40  | 72.36±5.05  | 93.94±1.25  | 65.77±2.58  |
| U-net3+ [53]  | 54.77±3.55  | 65.78±2.66  | 74.38±3.21  | 93.82±1.06  | 66.19±3.37  |
| SegNet [54]   | 54.89±1.78  | 63.79±2.65  | 77.25±4.02  | 94.00±1.14  | 65.90±1.97  |
| AE U-net [30] | 55.38±1.77  | 67.87±3.81  | 72.88±3.47  | 94.43±1.33  | 66.50±1.52  |
| SANet [45]    | 57.55±1.28* | 67.49±0.80  | 75.91±3.91  | 94.21±1.12  | 68.45±1.05* |
| scSEU-net [35] | 56.21±2.16  | 67.75±4.77  | 73.99±3.22  | 94.29±1.13  | 67.39±1.96  |
| SKNet [6]     | 57.06±2.42  | 69.59±4.20* | 73.58±6.75  | 94.65±1.49* | 68.19±2.28  |
| Ours          | 59.63±1.57  | 71.52±3.40  | 74.71±3.21  | 95.24±1.31  | 70.43±1.32  |

A. Robustness Analysis

Robustness on Benign and Malignant Lesions: There are great differences between benign tumors and malignant tumors in morphology and intensity distribution. Generally speaking, benign lesions have regular morphology, but the tumor size varies greatly. The shape of malignant lesions is irregular, and the internal energy distribution is uneven and the boundary is more blurred. To further evaluate the robustness of the method in breast lesions segmentation, we conducted comparative experiments on benign and malignant breast tumors. Specially, we perform four-fold and three-fold cross-validation on benign and malignant lesions, respectively. According to the results in Table 4, the sensitivity of the same segmentation methods to different lesions can be obtained.

B. Limitations and Future Works

Although the proposed method has achieved promising performance on three different datasets, some limitations and future work should be taken into consideration: (i) Although the design of ESKNet has further improved the segmentation accuracy of different lesions to a certain extent and obtained the most competitive segmentation results. In addition, the p-value based on the T-test further demonstrates the robustness and superiority of our method.

External validation: In addition to the influence of the type of breast lesions, there are also great differences between ultrasound images collected by different devices. These differences will also seriously affect the network segmentation performance (tend to fail to achieve ideal segmentation results on external test data). In this paper, STU provided by Zhuang et al. [31] is used as external test data to further evaluate the generalization capability of each segmentation network. Table 5 shows the evaluation results of external tests after different segmentation methods are trained on BUSI and Dataset B, respectively. The method proposed in this paper has achieved the best results in two external validation experiments. In the external experiment of BUSI, the values of the five evaluation indicators are 77.65%, 90.23%, 86.49%, 98.48%, and 86.79%, respectively. The values of the five evaluation indicators on Dataset B are 72.63%, 79.23%, 92.01%, 96.57%, and 82.72%, respectively. Although some methods achieved competitive results in K-fold cross-validation experiments, they did not perform well in external validation data (such as scSEU-net). This shows that the generalization ability and robustness of these methods are limited. Fig. 6 and Fig. 7 show the P-R curve and ROC curve of STU on BUSI and Dataset B, respectively. It can be seen from Fig. 6 and Fig. 7 that our method achieves the highest confidence level in the external validation experiment of each fold. According to Table 5, Fig. 6, and Fig. 7, our method has better generalization performance in breast lesion segmentation.
Table 5
The segmentation results (mean ± std) of different competing methods on the external test data STU. Asterisks indicate that the difference between our method and the competing method is significant using a paired student’s t-test. (*p < 0.05).

| Methods        | Jaccard | Precision | Recall | Specificity | Dice  | Jaccard | Precision | Recall | Specificity | Dice  |
|----------------|---------|-----------|--------|-------------|-------|---------|-----------|--------|-------------|-------|
| STU on BUSI    |         |           |        |             |       | STU on Dataset B |         |        |             |       |
| U-net [19]     | 62.40±5.06 | 73.11±4.40 | 82.47±2.34 | 95.53±0.85 | 74.09±4.62 | 58.90±3.75 | 66.27±5.46 | 86.88±1.60 | 94.54±0.74 | 71.41±3.67 |
| STAN [38]      | 70.92±2.21 | 82.47±1.65 | 84.20±1.86 | 97.09±0.32 | 81.33±1.90 | 59.5±3.31 | 64.38±3.73 | 90.72±0.41 | 94.59±0.57 | 70.76±3.16 |
| Att U-net [46] | 56.44±2.97 | 65.20±4.12 | 84.21±1.10 | 94.35±0.54 | 69.81±2.72 | 52.65±2.29 | 59.26±2.86 | 86.35±1.29 | 93.41±0.41 | 65.19±2.73 |
| RDAU-net [31]  | 69.36±4.57 | 79.82±6.29 | 86.07±1.61 | 96.78±1.02 | 79.89±4.01 | 60.11±3.02 | 63.49±2.86 | 91.37±1.07 | 94.70±0.45 | 72.40±2.74 |
| U-net++ [52]   | 61.69±3.60 | 70.70±5.02 | 86.01±1.01 | 95.27±0.58 | 73.73±3.59 | 59.18±4.21 | 64.86±5.36 | 89.67±1.59 | 94.33±0.83 | 70.70±4.19 |
| MADU-net [33]  | 59.36±1.92 | 69.24±2.44 | 82.66±0.85 | 94.91±0.42 | 72.27±1.65 | 58.11±3.24 | 66.05±3.69 | 86.17±0.99 | 94.35±0.47 | 70.32±0.01 |
| U-net+++ [53]  | 61.19±2.58 | 68.31±3.14 | 86.26±0.76 | 94.81±0.35 | 72.76±1.86 | 61.51±1.78 | 67.12±2.22 | 89.96±1.01 | 94.49±0.31 | 72.96±1.69 |
| SegNet [54]    | 75.5±1.33  | 86.96±0.21 | 86.32±1.40 | 97.93±1.11 | 85.21±0.99 | 62.70±3.09 | 66.57±3.05 | 91.36±0.38 | 95.04±0.49 | 73.50±3.62 |
| AE U-net [30]  | 71.92±0.65 | 83.10±0.45 | 84.86±0.84 | 97.35±0.16 | 82.12±0.60 | 62.46±2.44 | 66.79±2.95 | 91.24±0.42 | 95.00±0.39 | 74.41±2.30 |
| SANet [48]     | 69.89±2.12 | 79.58±2.90 | 86.28±1.52 | 96.78±0.80 | 79.62±2.13 | 63.99±5.42 | 69.91±7.58 | 91.03±2.23 | 95.37±0.80 | 74.99±5.16 |
| scSEU-net [45] | 71.97±0.73 | 86.15±0.61 | 83.01±1.18 | 97.71±0.12 | 82.84±0.66 | 55.12±5.83 | 60.87±6.76 | 89.73±1.63 | 93.85±0.60 | 67.30±5.93 |
| SKNet [35]     | 76.23±0.78 | 92.61±0.64 | 84.60±1.27 | 98.31±0.31 | 85.56±0.70 | 66.94±3.15 | 71.99±4.08 | 91.44±1.08 | 95.40±0.51 | 78.29±3.05 |
| Ours           | 77.65±0.70 | 90.23±0.93 | 86.49±0.79 | 98.48±0.26 | 86.79±0.56 | 72.63±2.03 | 79.23±3.39 | 92.01±0.81 | 96.57±0.53 | 82.72±1.70 |

Fig. 6. P-R and ROC curves for external verification of STU after training on BUSI.

accuracy of breast tumors, it still does not completely overcome the limitations of the complexity of ultrasound pattern on clinical practice. As shown in Fig. 8, Our method also has some missing and false detection in the segmentation of breast lesions. Consequently, in the future, we will explore some efficient mechanisms (such as boundary attention mechanism or boundary generation algorithm) to deal with this issue. (ii) The spatial and channel attention mechanism of the enhanced selective kernel convolution (ESK) module improves the representation ability of breast tumors, but inevitably increases the complexity of the network compared with the original selective kernel convolution (SK) module. In the future, it is necessary to further reduce the complexity of modules while ensuring the robustness of the network.

VII. CONCLUSION

This paper develops a novel enhanced selective kernel convolution (ESK) based on selective kernel convolution (SK), which can adaptively select useful objective features from the dimensions of space and channel. Then, we use the ESK module to construct a deep supervision U-net to segment breast lesions. Extensive experiments with several state-of-the-art deep learning segmentation methods on three public breast ultrasound datasets demonstrate that our method has a more competitive segmentation performance in breast ultrasound images. The source code is publicly available at: [https://github.com/CGPxy/ESKNet](https://github.com/CGPxy/ESKNet).

REFERENCES

[1] B. Lei, S. Huang, H. Li, R. Li, C. Bian, Y. H. Chou, J. Qin, P. Zhou, X. Gong, and J. Z. Cheng, “Self-co-attention neural network for anatomy segmentation in whole breast ultrasound,” Medical Image Analysis, vol. 64, p. 101753, 2020.

[2] Y. Liu, Shengfen and Gardezi2019 Wang, X. Yang, B. Lei, L. Liu, S. X. Li, D. Ni, and T. Wang, “Deep Learning in Medical Ultrasound Analysis: A Review,” Engineering, vol. 5, no. 2, pp. 261–275, 2019.
Fig. 7. P-R and ROC curves for external verification of STU after training on Dataset B.

Fig. 8. Some worst segmentation cases of the proposed method. In addition, the segmentation results of the three compared methods are also shown. The red lines represent the ground-truth masks.

[3] S. J. S. Gardezi, A. Elazab, B. Lei, and T. Wang, “Breast cancer detection and diagnosis using mammographic data: Systematic review,” Journal of medical Internet research, vol. 21, no. 7, p. e14464, 2019.

[4] 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.

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

[6] P. Tang, X. Yang, Y. Nan, S. Xiang, and Q. Liang, “Feature pyramid nonlocal network with transform modal ensemble learning for breast tumor segmentation in ultrasound images,” IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 68, no. 12, pp. 3549–3559, 2021.

[7] H. Li, D. Chen, W. H. Nailon, M. E. Davies, and D. I. Laurenson, “Dual convolutional neural networks for breast mass segmentation and diagnosis in mammography,” IEEE Transactions on Medical Imaging, vol. 41, no. 1, pp. 3–13, 2021.

[8] H. Du, J. Wang, M. Liu, Y. Wang, and E. Meijering, “SwinPA-Net: Swin Transformer-Based Multiscale Feature Pyramid Aggregation Network for Medical Image Segmentation,” IEEE Transactions on Neural Networks and Learning Systems, 2022.

[9] S. Wang, K. Sun, L. Wang, L. Qu, F. Yan, Q. Wang, and D. Shen, “Breast Tumor Segmentation in DCE-MRI With Tumor Sensitive Synthesis,” IEEE Transactions on Neural Networks and Learning Systems, 2021.

[10] D. Konar, S. Bhattacharyya, B. K. Panigrahi, and E. C. Behrman, “Ouitri-inspried fully self-supervised shallow quantum learning network for brain tumor segmentation,” IEEE Transactions on Neural Networks and Learning Systems, 2021.

[11] Z. Wang, “Deep learning in medical ultrasound image segmentation: A review,” arXiv preprint arXiv:2002.07703, 2020.

[12] L. Abdelrahman, M. Al Ghamdi, F. Collado-Mesa, and M. Abdel-Mottaleb, “Convolutional neural networks for breast cancer detection in mammography: A survey,” Computers in biology and medicine, vol. 131, p. 104248, 2021.

[13] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, “Image segmentation using deep learning: A survey,” IEEE transactions on pattern analysis and machine intelligence, 2021.

[14] C. Xue, L. Zhu, H. Fu, X. Hu, X. Li, H. Zhang, and P. Heng, “Global guidance network for breast lesion segmentation in ultrasound images,” Medical image analysis, vol. 70, p. 101989, 2021.

[15] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P. F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, and S. Wirkert, “mu-net: Self-adapting framework for u-net-based medical image segmentation,” arXiv preprint arXiv:1809.10486, 2018.

[16] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, “U-net and its variants for medical image segmentation: A review of theory and applications,” IEEE Access, 2021.

[17] H. Fu, G. Wang, W. Lei, W. Xu, Q. Zhao, S. Zhang, K. Li, and S. Zhang, “HMRNet: High-and-Multi-Resolution Network With Bidirectional Feature Calibration for Brain Structure Segmentation in Radiotherapy,” IEEE Journal of Biomedical and Health Informatics, 2022.

[18] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.

[19] 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.

[20] R. Almajilad, J. Shan, Y. Du, and M. Zhang, “Development of a deep-learning-based method for breast ultrasound image segmentation,” in
2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE, 2018, pp. 1103–1108.

[21] 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.

[22] 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.

[23] G. Chen, Y. Dai, and J. Zhang, “C-Net: Cascaded Convolutional Neural Network with Global Guidance and Refinement Residuals for Breast Ultrasound Image Segmentation,” Computer Methods and Programs in Biomedicine, p. 107086, 2022.

[24] K. Tomar, D. Jha, M. A. Riegler, H. D. Johansen, D. Johansen, J. Rittscher, P. Halvorsen, and S. Ali, “Fanet: A feedback attention network for improved biomedical image segmentation,” IEEE Transactions on Neural Networks and Learning Systems, 2022.

[25] A. Elmoufidi, “Deep Multiple Instance Learning for Automatic Breast Cancer Assessment using Digital Mammography,” IEEE Transactions on Instrumentation and Measurement, 2022.

[26] W. Qian, W. Bu, W. Zhu, W. Li, and Q. B. T. Hu, “Eca-net: Efficient channel attention for deep convolutional neural networks,” in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020. Computer Vision and Pattern Recognition (CVPR), 2020, pp. 11,531–11,539.

[27] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs,” IEEE Trans Pattern Anal Mach Intell, vol. 40, no. 4, pp. 834–848, 2018.

[28] A. O. Joshua, F. V. Nelwamondo, and G. Mabuza-Hocquet, “Blood Vessel Segmentation from Fundus Images Using Modified U-net Convolutional Neural Network,” Journal of Image and Graphics, vol. 8, no. 1, pp. 21–25, 2020.

[29] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature pyramid networks for object detection,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2117–2125.

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

[31] Z. Zhang, 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.

[32] N. S. Punn and S. Agarwal, “RCA-Utnet: A residual cross-spatial attention-guided inception U-Net model for tumor segmentation in breast ultrasound imaging,” Machine Vision and Applications, vol. 33, no. 2, pp. 1–10, 2022.

[33] 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 15th International Symposium on Biomedical Imaging (ISBI). IEEE, 2019, pp. 683–687.

[34] 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.

[35] 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.

[36] E. H. Houssein, M. M. Emam, A. A. Ali, and P. N. Suganthan, “Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review,” Expert Systems with Applications, vol. 167, p. 114161, 2021.

[37] M. Xian, Y. Zhang, H.-D. Cheng, F. Xu, B. Zhang, and J. Ding, “A deep automatic breast ultrasound image segmentation: A survey,” Pattern Recognition, vol. 79, pp. 340–355, 2018.

[38] B. Shareef, M. Xian, and A. Vakanski, “Stan: Small tumor-aware network for breast ultrasound image segmentation,” in 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE, 2020, pp. 75–78.

[39] B. Lei, S. Huang, R. Li, C. Bhan, H. Li, Y.-H. Chou, and J.-Z. Cheng, “Segmentation of breast anatomy for automated whole breast ultrasound images with boundary regularized convolutional encoder-decoder network,” Neurocomputing, vol. 321, pp. 178–186, 2018.