



# Residual Feedback Network for Breast Lesion Segmentation in Ultrasound Image

Ke Wang<sup>1,2</sup>, Shujun Liang<sup>1,2</sup>, and Yu Zhang<sup>1,2</sup>(✉)

<sup>1</sup> School of Biomedical Engineering,  
Southern Medical University,  
Guangzhou 510515, Guangdong, China  
yuzhang@smu.edu.cn

<sup>2</sup> Guangdong Provincial Key Laboratory of Medical Image Processing, Southern  
Medical University, Guangzhou 510515, Guangdong, China

**Abstract.** Accurate lesion segmentation in breast ultrasound (BUS) images is of great significance for the clinical diagnosis and treatment of breast cancer. However, precise segmentation on missing/ambiguous boundaries or confusing regions remains challenging. In this paper, we proposed a novel residual feedback network, which enhances the confidence of the inconclusive pixels to boost breast lesion segmentation performance. In the proposed network, a residual representation module is introduced to learn the residual representation of missing/ambiguous boundaries and confusing regions, which promotes the network to make more efforts on those hardly-predicted pixels. Moreover, a residual feedback transmission strategy is designed to update the input of the encoder blocks by combining the residual representation with original features. This strategy could enhance the regions including hardly-predicted pixels, which makes the network can further correct the errors in initial segmentation results. Experimental results on three datasets (3813 images in total) demonstrate that our proposed network outperforms the state-of-the-art segmentation methods. Our code is available at <https://github.com/mniwk/RF-Net>.

**Keywords:** Ultrasound image · Breast lesion segmentation · Residual feedback network

## 1 Introduction

Breast cancer is the leading cancer in women, which disrupts the lives of millions of women in the world [1]. Early screening of breast cancer can reduce mortality by up to 20% [2]. Ultrasound imaging is one of the popularly screening methods

**Electronic supplementary material** The online version of this chapter ([https://doi.org/10.1007/978-3-030-87193-2\\_45](https://doi.org/10.1007/978-3-030-87193-2_45)) contains supplementary material, which is available to authorized users.

to detect breast lesions due to its advantages of non-radiation, real-time visualization, cost-effectiveness and non-invasive diagnosis [3,4]. Segmenting breast lesion is a crucial step in computer-aided diagnosis system, which can provide interpretive information for different functional tissues discrimination and breast cancer diagnosis [4,5]. However, automatic precise lesion segmentation in BUS images is a challenging task due to (1) missing/ambiguous boundaries caused by speckle noise and the similar visual appearance between lesions and non-lesion background; (2) large variety of breast lesions, which come in various shapes, sizes and locations; (3) significant individual differences in breast structures between patients [5–7].

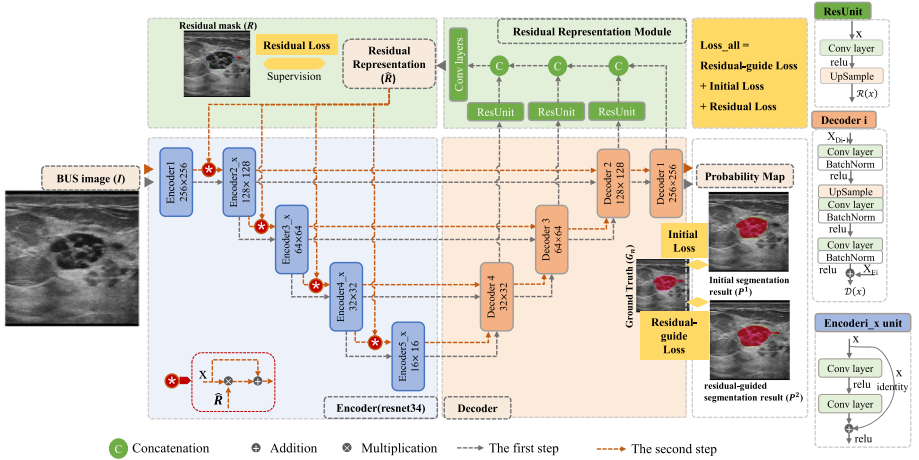
Automated segmentation of breast lesions in BUS images has been studied for many years. The existing methods can be roughly divided into two categories: traditional methods and deep learning methods [8,9]. Manual cropping tumor-centered regions of interest (ROIs) and complex preprocessing and/or postprocess steps are commonly required in traditional methods [10,11]. Luo Y et al. [10] performed the interaction to firstly determine the ROI and combined region- and edge-based information to overcome the problem of over-segmentation and under-segmentation in BUS images. Huang Q et al. [11] also cropped the ROIs, utilized simple linear iterative clustering and k-nearest neighbor algorithms to obtain the final segmentation. Deep-learning-based methods have triumphed in segmentation tasks [12]. For BUS images, Han L et al. [13] proposed a semi-supervised segmentation method using a dual-attention-fusion block to extract features of the lesion regions and non-lesion regions separately, which enhanced the discriminative ability to discriminate lesions. Vakanski A et al. [14] introduced attention blocks into a U-Net architecture and learned feature representations that prioritize spatial regions with high saliency levels. Zhu L et al. [15] proposed a novel second-order sub-region pooling network (S<sup>2</sup>P-Net) for boosting the breast lesion segmentation in ultrasound images, which aggregated the multi-context information from the whole image and multiple sub-regions. Despite promising segmentation results of existing methods, precise segmentation in the confusing regions and unclear boundaries of the breast BUS images segmentation is still challenging.

In this paper, we propose a novel residual feedback network based on encoder-decoder architecture to boost breast lesion segmentation by learning residual representation of hardly-predicted pixels, and feeding it into encoder blocks, to enhance the confidence of the hardly-predicted pixels. The contributions of our work can be summarized as follows: (1) A supervised residual representation module is designed to learn the residual representations of missing/ambiguous boundaries and confusing regions. (2) A residual feedback transmission strategy is designed to update the input of encoder blocks by feeding residual representations back and combining it with original features. This strategy could enhance the weight of regions including hardly-predicted pixels, which makes the network can further correct the errors in initial segmentation results. (3) Experimental results on three public datasets (3813 images in total) demonstrate that our proposed network outperforms the state-of-the-art segmentation methods.

## 2 Method

### 2.1 Overview

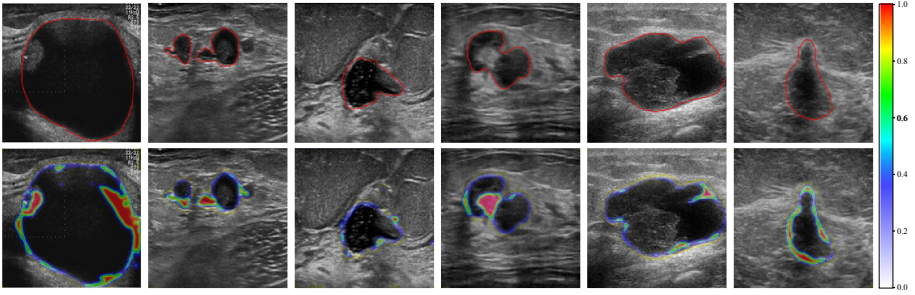
Figure 1 illustrates the architecture of the proposed residual feedback network for lesion segmentation in BUS images. In the proposed network, an encoder-decoder architecture is used as the baseline. The proposed network includes two steps: (1) in the first step, we perform an encoder-decoder architecture to generate initial segmentation results under the input of BUS images. And then, a residual representation module is designed to take the features of decoder blocks as inputs and learn the residual information about low-confident and error-predicted pixels, which is supervised by the residual masks (the difference between the ground truths and the initial segmentation results). (2) In the second step, a residual feedback transmission strategy provides an opportunity to correct the errors by utilizing the residual representation generated during the first step, and then the preferable residual-guided segmentation results can be obtained by reusing the encoder-decoder baseline. It worth to note that the proposed network is trained in an end-to-end manner without introducing any extra-parameter.



**Fig. 1.** The overview of our proposed network for lesion segmentation in BUS images. ResNet-34 is adopted as the Encoder path with pre-trained parameters. Our network includes two steps. The first step (gray dotted arrows) is used to generate the initial segmentation results and learn the residual representation of missing/ambiguous boundaries and confusing regions. The second step is used to (red dotted arrows) feed the residual representation into encoder path and generate more precise segmentation results. (Color figure online)

## 2.2 Encoder-Decoder Baseline

In this study, we modify the ResNet-34 as the encoder path by retaining the first five blocks, to extract detailed context features of BUS images. We also apply the skip connection in U-Net architecture to transport the detailed information from encoder to decoder, aiming to remedy the information loss owing to consecutive pooling and striding convolutional operations [18]. As shown in Fig. 1, each encoder block contains  $x^i$  Encoder <sub>$i$</sub>  units (where  $x^i \in \{3, 4, 6, 3\}$  and  $i \in \{2, 3, 4, 5\}$ ). Each Encoder <sub>$i$</sub>  unit consists of two  $3 \times 3$  convolutional layers, except Encoder<sub>1</sub> unit which contains one  $7 \times 7$  convolutional layer. The decoder path comprises four convolutional blocks. As illustrated in Fig. 1, each decoder block contains an element-wise addition operation, a  $1 \times 1$  convolution layer, an upsampling layer, a  $3 \times 3$  convolution layer, and a  $1 \times 1$  convolution layer, consecutively. And the upsampling layer is used to ensure the same scale with corresponding encoder blocks by linear interpolation.



**Fig. 2.** Examples of the residual representations. The red contours in the first row correspond the ground truths. The residual representation maps in the second row are marked from blue to red, which indicate the low to high probabilities. (Color figure online)

## 2.3 Residual Representation Module

In general, the predicted probabilities of missing/ambiguous boundaries or confusing regions are low in segmented map, but salient in the residual image (the difference between the ground truths and the segmented results, see Fig. 2). The goal of the proposed residual representation module is to learn the residual representations of missing/ambiguous boundaries and confusing regions, and encourage the network to explore more information about hardly-predicted pixels.

In the first step, we obtain the initial segmentation results  $P^1$  with the input of BUS images  $I$ . Then, we introduce the residual representation module, which integrates the features of multiple decoder blocks to learn the residual representations ( $\hat{R}$ ) under the supervision of the residual masks ( $R$ , the difference

between the ground truths  $G$  and the segmented results  $P^1$ ), and generate confident representation maps for the inconclusive pixels.

The architecture of the proposed residual representation module is shown in Fig. 1, and the residual units are used to extract the context information from decoder blocks. Each residual unit (ResUnit) contains a  $3 \times 3$  convolutional layer and an upsampling layer that makes the feature maps with the same size of  $256 \times 256$ . Then the residual representations can be generated by two consecutive convolutional layers with kernel sizes of  $3 \times 3$  and  $1 \times 1$ . It can be observed in Fig. 2 that the missing/ambiguity boundaries and the confusing regions can be easily localized under the guidance of the residual representations.

## 2.4 Residual Feedback Transmission and Loss Function

Existing segmentation methods principally design a module to predict the boundaries or a loss function for boundaries restriction [19–22], which is a forward processing and only focuses on the boundary pixels. In this study, the missing/ambiguity boundaries and the confusing regions can be easily located with the help of the residual representation module. Under the inspiration of [20], we adopt a residual feedback transmission strategy to make the proposed network further correct the errors in initial segmentation results.

After the residual representations  $\hat{R}$  and the initial segmentation results  $P^1$  being obtained in the first step, the residual representations are fed into each encoder block to weight the low-confidence and wrongly predicted pixels in  $P^1$  by combining with the original image features with element-wise addition and channel-wise multiplication (see the second step (red dotted arrows) in Fig. 1). Then the second step is still performed on the same network architecture as the first step, and outputs the presents residual-guided segmentation results  $P^2$ .

For training our network in an end-to-end manner, the final loss  $L$  consists of three sub-losses for the initial segmentation results ( $L_{wbl}\{G, P^1\}$ ), the residual-guided segmentation results ( $L_{wbl}\{G, P^2\}$ ), and the residual representation  $L_{wbce}\{R, \hat{R}\}$  respectively:

$$L = L_{wbl}\{G, P^1\} + L_{wbl}\{G, P^2\} + L_{wbce}\{R, \hat{R}\} \quad (1)$$

where  $L_{wbl}$  and  $L_{wbce}$  denote the weighted-balanced loss function [27] and the weighted binary cross-entropy loss function, respectively. And  $L_{wbl}$  and  $L_{wbce}$  are expressed as follows:

$$L_{wbl} = 1 - w \frac{\sum_{n=1}^{N_1} p_n y_n}{\sum_{n=1}^{N_1} (p_n + y_n)} - (1 - w) \frac{\sum_{n=1}^{N_0} (1 - p_n)(1 - y_n)}{\sum_{n=1}^{N_0} (2 - p_n - y_n)} \quad (2)$$

$$L_{wbce} = -w \sum_{n=1}^{N_1} y_n \log(p_n) - (1 - w) \sum_{n=1}^{N_0} (1 - y_n) \log(1 - p_n) \quad (3)$$

where  $w = N_0/(N_1 + N_0)$  is the weighted-balanced parameter proposed in our previous work [27]. Specifically,  $y_n$  and  $p_n$  denote the ground truth and predicted probability of pixels, respectively.  $N_0$  and  $N_1$  denote the number of the pixels

where  $y_n = 0$  and  $y_n = 1$ . For each BUS image in the test phase, the segmentation results can be generated recurrently for more precise performance owing to the proposed feedback transmission strategy.

**Implementation Details.** We implement our network on Pytorch and train it on one GeForce Titan GPU with the mini-batch size of 16. We adopt an adaptive gradient algorithm (Adagrad) optimizer with momentum of 0.9 to optimize and update the network parameters with 70 epochs. The learning rate is  $2e^{-4} \times 0.95^{epoch_i}$ , where  $epoch_i$  is the  $i$ -th epoch. The network parameters are initialized by “he\_normal” [23].

### 3 Experiments

**Datasets.** Three BUS image datasets are used in this study to evaluate the effectiveness of the proposed network. The first one (datasetA) contains two public datasets: (a) the dataset is available at <https://www.ultrasoundcases.info/> derived from SonoSkills and Hitachi Medical Systems Europe, which has a total of 2499 tumor images from 554 patients, including benign lesions (1061 BUS images) and malignant lesions (1438 BUS images) (b) BUSI [24], from the Baheya Hospital for Early Detection and Treatment of Women’s Cancer (Cairo, Egypt), which contains 210 images with benign lesions and 437 images with malignant lesions and 133 normal images without lesions. The second one, UDIAT (datasetB), containing 110 images with benign lesions and 53 images with malignant lesions, is a public dataset provided by [5], which is collected from the UDIAT Diagnostic Center of the Parc Tauli Corporation, Sabadell (Spain). The third one, datasetC, is available at <https://radiopaedia.org/> and contains 504 tumor images, including benign lesion (242 BUS images) and malignant lesions (262 BUS images). We perform five-fold cross-validation on datasetA and report the validated results, then use datasetB and datasetC as two test datasets to illustrate the effectiveness of the proposed model based on five models trained on datasetA.

**Evaluation Metrics.** For quantitatively comparing different methods, we employ widely-used segmentation metrics, including dice similarity coefficient (DSC), jaccard index (JI), recall, precision, Hausdorff distance (HD, in pixel) and average surface distance (ASD, in pixel) [7, 9, 25]. For DSC, JI, Recall, and Precision, larger values indicate better performance, while HD and ASD vice versa. We evaluate the overlap areas between the predicted lesions and the delineated lesions by radiologists with DSC, JI, recall, and precision, and measure the Euclidean distance of boundaries with HD and ASD [7].

#### 3.1 Segmentation Performance

We compare our proposed method with four state-of-the-art methods, including U-Net [17], DeepLabV3 [26], CE-Net [18], and S<sup>2</sup>P-Net [15]. For a fair

comparison, we adopt the same encoder (ResNet-34) for DeepLabV3 and CE-Net, and validate their performance with/without pretrain parameters [28] (DeepLabV3<sup>w/wo</sup> and CE-Net<sup>w/wo</sup> in Table 1, 2, 3). Our proposed method and the comparison methods are trained and validated on datasetA (reporting mean results of the five-fold cross-validation), and tested on datasetB and datasetC. All the networks are fine-tuned for best training parameters.

**Table 1.** Comparison with different methods on datasetA (mean  $\pm$  variance).

Methods	DSC	JI	Recall	Precision	HD	ASD
U-Net	84.33 $\pm$ 0.99	76.42 $\pm$ 1.01	86.01 $\pm$ 0.81	87.10 $\pm$ 1.54	30.12 $\pm$ 1.55	11.1 $\pm$ 1.50
DeepLabV3 <sup>wo</sup>	84.16 $\pm$ 0.48	75.69 $\pm$ 0.68	86.29 $\pm$ 0.49	85.70 $\pm$ 0.58	28.08 $\pm$ 0.87	9.03 $\pm$ 0.39
DeepLabV3 <sup>w</sup>	85.97 $\pm$ 0.71	77.85 $\pm$ 0.95	87.49 $\pm$ 0.64	87.40 $\pm$ 1.61	25.58 $\pm$ 0.99	8.36 $\pm$ 0.30
CE-Net <sup>wo</sup>	84.59 $\pm$ 0.97	76.38 $\pm$ 1.14	86.46 $\pm$ 1.30	86.27 $\pm$ 2.11	28.34 $\pm$ 1.58	8.62 $\pm$ 0.24
CE-Net <sup>w</sup>	86.32 $\pm$ 0.75	78.39 $\pm$ 0.86	88.02 $\pm$ 0.72	87.23 $\pm$ 1.27	26.30 $\pm$ 1.48	7.58 $\pm$ 0.55
S <sup>2</sup> P-Net	84.65 $\pm$ 0.72	76.63 $\pm$ 0.64	85.94 $\pm$ 0.70	87.41 $\pm$ 0.99	29.23 $\pm$ 1.69	10.7 $\pm$ 1.33
Our method	<b>86.91 <math>\pm</math> 0.55</b>	<b>79.24 <math>\pm</math> 0.61</b>	<b>88.22 <math>\pm</math> 0.77</b>	<b>88.40 <math>\pm</math> 1.55</b>	<b>24.44 <math>\pm</math> 0.94</b>	<b>7.68 <math>\pm</math> 0.59</b>

**Table 2.** Comparison with different methods on datasetB (mean  $\pm$  variance).

Methods	DSC	JI	Recall	Precision	HD	ASD
U-Net	74.20 $\pm$ 1.02	65.99 $\pm$ 1.13	83.54 $\pm$ 1.33	78.01 $\pm$ 2.81	46.83 $\pm$ 4.07	33.62 $\pm$ 5.62
DeepLabV3 <sup>wo</sup>	74.87 $\pm$ 2.28	65.58 $\pm$ 2.48	85.09 $\pm$ 1.51	71.97 $\pm$ 2.65	25.26 $\pm$ 3.17	9.24 $\pm$ 1.27
DeepLabV3 <sup>w</sup>	78.94 $\pm$ 1.49	69.70 $\pm$ 1.66	87.02 $\pm$ 0.65	77.13 $\pm$ 2.11	23.39 $\pm$ 1.62	10.89 $\pm$ 1.92
CE-Net <sup>wo</sup>	74.13 $\pm$ 1.94	65.65 $\pm$ 1.95	86.24 $\pm$ 1.95	71.25 $\pm$ 2.29	28.84 $\pm$ 3.25	11.23 $\pm$ 1.57
CE-Net <sup>w</sup>	79.61 $\pm$ 1.76	71.00 $\pm$ 1.87	88.62 $\pm$ 1.34	77.07 $\pm$ 2.50	24.74 $\pm$ 3.53	9.65 $\pm$ 1.98
S <sup>2</sup> P-Net	77.42 $\pm$ 1.13	69.04 $\pm$ 1.10	85.42 $\pm$ 2.15	77.30 $\pm$ 1.95	29.57 $\pm$ 2.18	16.36 $\pm$ 2.41
Our method	<b>81.79 <math>\pm</math> 0.76</b>	<b>73.09 <math>\pm</math> 0.64</b>	<b>90.07 <math>\pm</math> 1.00</b>	<b>78.61 <math>\pm</math> 0.97</b>	<b>18.06 <math>\pm</math> 0.87</b>	<b>6.54 <math>\pm</math> 0.29</b>

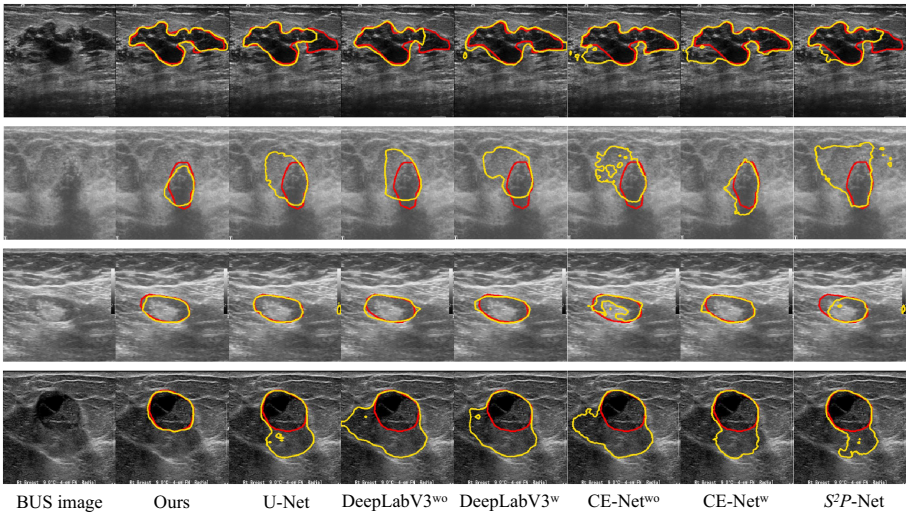
From the results in Table 1, 2, 3, our proposed method achieves highest performance in terms of DSC, JI, Recall and Precision, and outperforms all the competitors. In addition, our method also achieves the competitive values of HD and ASD with a high level of robustness, which demonstrates that our model can achieve more accurate boundary localization. The improvement of these metrics on all three datasets indicates that our model also has better generalization ability than all the competitors.

Figure 3 shows the segmentation results produced by different methods. Apparently, our method can detect missing/ambiguity boundaries accurately and better segment the whole breast lesion regions. The superior performance of our method, with significantly improved results ( $p < 0.05$  for paired t-test in Tables 1, 2 and 3), indicates that our residual representation module and the residual feedback transmission strategy can help distinguish some pseudo-lesions from positive lesions, and generate more precise boundaries. Moreover, it is noteworthy that our model achieves superior performance on malignant lesion with posterior acoustic shadowing in the second row of Fig. 3.



**Table 3.** Comparison with different methods on datasetC (mean  $\pm$  variance).

Methods	DSC	JI	Recall	Precision	HD	ASD
U-Net	$83.42 \pm 0.65$	$75.53 \pm 0.66$	$85.82 \pm 1.01$	$87.08 \pm 0.76$	$34.11 \pm 2.13$	$14.44 \pm 1.98$
DeepLabV3 <sup>wo</sup>	$83.80 \pm 0.91$	$75.53 \pm 1.16$	$86.75 \pm 1.06$	$84.70 \pm 0.93$	$27.96 \pm 1.16$	$8.50 \pm 0.89$
DeepLabV3 <sup>w</sup>	$85.20 \pm 0.53$	$77.33 \pm 0.74$	$87.79 \pm 0.51$	$86.02 \pm 1.12$	$25.80 \pm 1.09$	$8.13 \pm 0.96$
CE-Net <sup>wo</sup>	$84.17 \pm 0.55$	$75.98 \pm 0.63$	$87.16 \pm 1.25$	$85.01 \pm 1.20$	$28.34 \pm 0.91$	$8.01 \pm 0.40$
CE-Net <sup>wo</sup>	$86.13 \pm 0.43$	$78.46 \pm 0.47$	$88.94 \pm 0.68$	$86.41 \pm 1.16$	$26.30 \pm 1.48$	$7.58 \pm 0.55$
S <sup>2</sup> P-Net	$84.01 \pm 0.35$	$76.10 \pm 0.41$	$85.93 \pm 0.45$	$87.36 \pm 0.44$	$32.92 \pm 0.87$	$13.63 \pm 0.95$
Our method	<b><math>87.00 \pm 0.23</math></b>	<b><math>79.54 \pm 0.23</math></b>	<b><math>88.99 \pm 0.79</math></b>	<b><math>87.87 \pm 0.79</math></b>	<b><math>23.66 \pm 0.51</math></b>	<b><math>6.77 \pm 0.51</math></b>



**Fig. 3.** The segmentation results (yellow) produced by different methods and the ground truths (red) of random test subjects. (Color figure online)

### 3.2 Ablation Study

Three ablation studies experiments are conducted to verify the setups in our network design. Using the encoder-decoder architecture (in Fig. 1) as baseline, and we add three setups which boost the segmentation performance in BUS images to the baseline in a step-by-step manner: 1) the encoder based on ResNet34 with pretrain parameters [28] is denoted as ‘Pretrain’; 2) the residual representation module (denoted as ‘RSM’); 3) the residual feedback transmission strategy (denoted as ‘RFT’) which is the guidance for more accurate segmentation results.

From the segmentation results in Table 4, it can be seen that all three setups help improve the segmentation performance on three datasets ( $p < 0.05$ , paired t-test). Compared with the baseline, the model with pretrained parameters shows better performance, especially on the testing datasets (improvement approximately 5.77% and 1.98% for DSC on datasetB and datasetC). This reflects



**Table 4.** Segmentation results on three datasets to validate the effectiveness of three setups in our method (mean $\pm$ variance).

Pretrain		×	✓	✓	✓
RSM		×	×	✓	✓
RFT		×	×	×	✓
DatasetA	DSC	84.60 $\pm$ 0.58	85.76 $\pm$ 0.49	86.46 $\pm$ 0.48	<b>86.91 <math>\pm</math> 0.55</b>
	Recall	86.63 $\pm$ 0.96	86.95 $\pm$ 0.43	87.65 $\pm$ 0.37	<b>88.22 <math>\pm</math> 0.77</b>
	Precision	86.12 $\pm$ 0.91	87.89 $\pm$ 1.24	88.25 $\pm$ 0.84	<b>88.34 <math>\pm</math> 1.55</b>
	HD	28.09 $\pm$ 1.13	26.55 $\pm$ 0.68	25.54 $\pm$ 0.56	<b>24.44 <math>\pm</math> 0.94</b>
DatasetB	DSC	74.06 $\pm$ 0.83	79.83 $\pm$ 0.89	80.75 $\pm$ 0.44	<b>81.79 <math>\pm</math> 0.76</b>
	Recall	84.41 $\pm$ 1.54	87.13 $\pm$ 0.35	89.12 $\pm$ 0.87	<b>90.07 <math>\pm</math> 1.00</b>
	Precision	72.20 $\pm$ 1.54	78.54 $\pm$ 1.93	78.36 $\pm$ 0.92	<b>78.61 <math>\pm</math> 0.97</b>
	HD	30.63 $\pm$ 3.25	22.64 $\pm$ 3.17	20.49 $\pm$ 2.14	<b>18.06 <math>\pm</math> 0.87</b>
DatasetC	DSC	84.04 $\pm$ 0.41	86.02 $\pm$ 0.28	86.44 $\pm$ 0.22	<b>87.00 <math>\pm</math> 0.23</b>
	Recall	86.33 $\pm$ 0.54	87.71 $\pm$ 0.56	88.26 $\pm$ 0.64	<b>88.99 <math>\pm</math> 0.79</b>
	Precision	85.73 $\pm$ 0.75	87.71 $\pm$ 0.78	87.74 $\pm$ 0.60	<b>87.87 <math>\pm</math> 0.79</b>
	HD	28.38 $\pm$ 0.50	26.15 $\pm$ 0.98	25.12 $\pm$ 1.85	<b>23.66 <math>\pm</math> 0.51</b>

that the models pretrained on ImageNet dataset have sensitive and transferable parameters, which are more robust on different testing datasets. In addition, we observe that employing the RSM yields significant improvements in term of recall (0.70%, 1.99%, 0.55% on datasetA, datasetB, datasetC). This achievement can be explained by learning the representations of missing/ambiguous boundaries and confusing regions in residual representation module, and encouraging the network to explore more information about hardly-predicted pixels. Further, observed from Table 4, our model adopting RFT achieve more accurate segmentation results in terms of all metrics, especially on the boundary-related metrics HD. It proves that the RFT, which feedbacks the learned representation into the encoder, can enhance the weight of the hardly-predicted pixels, and make the network iteratively correct the errors in segmentation process.

## 4 Discussion and Conclusion

This paper proposes a novel residual feedback network for breast lesion segmentation on ultrasound image and aims to focus on hardly-predicted pixels. The key idea is to design a supervised residual representation module, which extracts multi-scale features from the decoder blocks and generates residual representations representing the semantic context of inconclusive pixels. And a residual feedback transmission strategy is designed to embed the residual representations into the encoder blocks to increase the saliency of hardly-predicted pixels, which makes our proposed network can further correct the errors in initial segmentation results. Experimental results have evaluated on three datasets, and sig-

nificantly outperforms the state-of-the-art methods. Specifically, the DSC score is 86.91% for datasetA, 81.79% for datasetB, and 87.00% for datasetC, which proves our model has better generalization ability on testing datasets (datasetB and datasetC). Detailed discussion of typical failed cases of our model will be included in Supplementary Material.

**Acknowledgements.** This work was supported in part by the National Natural Science Foundation of China under Grant 61971213 and Grant 61671230, in part by the Basic and Applied Basic Research Foundation of Guangdong Province under Grant 2019A1515010417, and in part by the Guangdong Provincial Key Laboratory of Medical Image Processing under Grant No.2020B1212060039. The authors have no relevant conflicts of interest to disclose.

## References

1. Ahmad, A. (ed.): Breast Cancer Metastasis and Drug Resistance. AEMB, vol. 1152. Springer, Cham (2019). <https://doi.org/10.1007/978-3-030-20301-6>
2. Berg, W.A., et al.: Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer. *JAMA* **299**(18), 2151–2163 (2008)
3. Sahiner, B., et al.: Malignant and benign breast masses on 3D US volumetric images: effect of computer-aided diagnosis on radiologist accuracy. *Radiology* **242**(3), 716–724 (2007)
4. Xu, Y., et al.: Medical breast ultrasound image segmentation by machine learning. *Ultrasonics* **91**, 1–9 (2019)
5. Yap, M.H., et al.: Automated breast ultrasound lesions detection using convolutional neural networks. *IEEE J. Biomed. Health Inform.* **22**(4), 1218–1226 (2018)
6. Lei, B., et al.: Segmentation of breast anatomy for automated whole breast ultrasound images with boundary regularized convolutional encoder-decoder network. *Neurocomputing* **321**, 178–186 (2018)
7. Xing, J., et al.: Lesion segmentation in ultrasound using semi-pixel-wise cycle generative adversarial nets. *IEEE ACM Trans. Comput. Biol. Bioinform.* (2020). <https://ieeexplore.ieee.org/abstract/document/9025227>
8. Saeed, J.N.: A Survey of Ultrasonography Breast Cancer Image Segmentation Techniques. *Infinite Study* (2020)
9. Xian, M., et al.: Automatic breast ultrasound image segmentation: a survey. *Pattern Recognit.* **79**, 340–355 (2018)
10. Luo, Y., et al.: A novel segmentation approach combining region- and edge-based information for ultrasound images. *BioMed Res. Int.* **2017** (2017). <https://www.hindawi.com/journals/bmri/2017/9157341/>
11. Huang, Q., et al.: Segmentation of breast ultrasound image with semantic classification of superpixels. *Med. Image Anal.* **61**, 101657 (2020)
12. Havaei, M., Davy, A., Warde-Farley, D., et al.: Brain tumor segmentation with deep neural networks. *Med. Image Anal.* **35**, 18–31 (2017)
13. Han, L., et al.: Semi-supervised segmentation of lesion from breast ultrasound images with attentional generative adversarial network. *Comput. Methods Programs Biomed.* **189**, 105275 (2020)
14. Vakanski, A., et al.: Attention-enriched deep learning model for breast tumor segmentation in ultrasound images. *Ultrasound Med. Biol.* **46**(10), 2819–2833 (2020)

15. Zhu, L., et al.: A second-order subregion pooling network for breast lesion segmentation in ultrasound. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12266, pp. 160–170. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-59725-2\\_16](https://doi.org/10.1007/978-3-030-59725-2_16)
16. He, K., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
17. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
18. Gu, Z., et al.: Ce-net: context encoder network for 2d medical image segmentation. *IEEE Trans. Med. Imaging* **38**(10), 2281–2292 (2019)
19. Wang, S., et al.: Boundary coding representation for organ segmentation in prostate cancer radiotherapy. *IEEE Trans. Med. Imaging* **40**(1), 310–320 (2020)
20. Feng, M., et al.: Attentive feedback network for boundary-aware salient object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1623–1632 (2019)
21. Wei, J., et al.: F<sup>3</sup>Net: Fusion, Feedback and Focus for Salient Object Detection. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 07, pp. 12321–12328 (2020)
22. Zhang, R., Li, G., Li, Z., Cui, S., Qian, D., Yu, Y.: Adaptive context selection for polyp segmentation. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12266, pp. 253–262. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-59725-2\\_25](https://doi.org/10.1007/978-3-030-59725-2_25)
23. He, K.M., et al.: Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: International Conference on Computer Vision, pp. 1026–1034 (2015)
24. Al-Dhabyani, W., et al.: Dataset of breast ultrasound images. *Data Brief* **28**, 104863 (2020)
25. Guo, Y., et al.: A novel breast ultrasound image segmentation algorithm based on neutrosophic similarity score and level set. *Comput. Methods Programs Biomed.* **123**, 43–53 (2016)
26. Chen, L.C., et al.: Rethinking atrous convolution for semantic image segmentation. preprint [arXiv:1706.05587](https://arxiv.org/abs/1706.05587) (2017)
27. Wang, K., Liang, S.J., et al.: Breast ultrasound image segmentation: a coarse-to-fine fusion convolutional neural network. *Med. Phys.* (2021). <https://doi.org/10.1002/mp.15006>
28. Xie, S.N., et al.: Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1492–1500 (2017)