Neurocomputing 493 (2022) 204–216

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

ASS-GAN: Asymmetric semi-supervised GAN for breast ultrasound image segmentation

Donghai Zhai [a], Bijie Hu [a,*], Xun Gong [a,b,*], Haipeng Zou [a], Jun Luo [c]

[a] School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu, Sichuan 610031, China
[b] Chains Collaboration and Information Support Technology Key Laboratory of Sichuan Province, Chengdu, Sichuan 610031, China
[c] School of Medicine UESTC, Sichuan Academy of Medical Sciences, Sichuan Provincial People's Hospital, Chengdu, Sichuan 610031, China

ARTICLE INFO

Article history:
Received 27 August 2021
Revised 9 January 2022
Accepted 3 April 2022
Available online 5 April 2022
Communicated by Zidong Wang

Keywords:
Breast ultrasound image
Lesion segmentation
Generative adversarial networks
Semi-supervised semantic segmentation

ABSTRACT

Ultrasound imaging is considered to be one of the important methods for diagnosing breast cancers, and lesion segmentation is an essential step in automatic computer-aided ultrasonic diagnosis. However, the high cost of ultrasound image labeling and the small amount of data in a single dataset hinder the progress of breast ultrasound (BUS) image segmentation algorithms. In this paper, we propose a novel asymmetric semi-supervised GAN (ASSGAN), which employs two generators and a discriminator for adversarial learning. The two generators can supervise each other, *i.e.*, they can generate reliable segmentation predicted masks as guidance for each other without labels. Therefore, the unlabeled cases can be used to effectively promote model training. To verify the proposed method, we compared it with fully supervised and semi-supervised methods on three public BUS datasets (DBUI, OASBUI, SPDBUI) and one dataset (SDBUI) that we collected. DBUI, OASBUI, SPDBUI and SDBUI contain 647, 200, 320 and 1805 cases respectively. The experimental results show that the proposed method has excellent performance under the condition of having a small number of labeled images. Compared with fully supervised methods, our method is higher by 4.16% ∼ 13.94% in *IoU*.

1. Introduction

Breast cancer is one of the leading causes of death among women worldwide [1]. It accounts for nearly a quarter of all cancer cases in women worldwide. Two-dimensional ultrasound imaging has the advantages of low cost, fast acquisition and noninvasive intervention. Therefore, it is often used as one of the first choice methods of modern medical screening for breast cancer. Clinically, doctors must judge the benign and malignant tumors by the aspect ratio, shape, boundary, and local gland structure of nodules that appear in ultrasound images. However, subjective errors from doctors and the limitations of their experiences often result in a high rate of erroneous diagnosis and misdiagnoses. In recent years, computer-aided detection (CAD) technology dominated by deep learning has developed rapidly. CAD technology not only improves the work efficiency greatly but also reduces the rate of missed diagnosis and misdiagnosis. However, the number of segmentation labeled by professional doctors is relatively small.

Although there are some segmentation networks [2–4] that perform well in natural images, their performance on BUS images still require improvement. There are three main challenges in breast ultrasound (BUS) image segmentation: (1) The edges of the lesion area in BUS images are generally blurred (as shown in Fig. 1), which makes it difficult for convolutional neural networks (CNNs) to learn the edge information. (2) The data annotations of BUS lesion areas are different from those of natural image segmentation, and doctors with professional knowledge are required for accurate annotation. Therefore, it is very costly to label BUS lesions. (3) The BUS images contain privacy with respect to patients. There are relatively few public BUS image datasets, and generally, the amount of data is small, and the corresponding labeled datasets are much smaller than those of natural image segmentation datasets. As far as we know, the dataset that Dhabyani et al. [5] released, DBUI (Dataset of Breast Ultrasound Images), has the largest number of cases, which includes only 647 cases.

Unlabeled data is easy to obtain and has great potential utilization value. Therefore, we considered using unlabeled data to solve the above challenges. However, unlabeled data lacks supervision signals in the process of deep learning training. So our motivation is how to construct the supervision signals for unlabeled data. We propose a semi-supervised adversarial segmentation network ASS-

* Corresponding authors.
E-mail addresses: JUG2020@my.swjtu.edu.cn (B. Hu), xgong@swjtu.edu.cn (X. Gong).

https://doi.org/10.1016/j.neucom.2022.04.021
0925-2312/

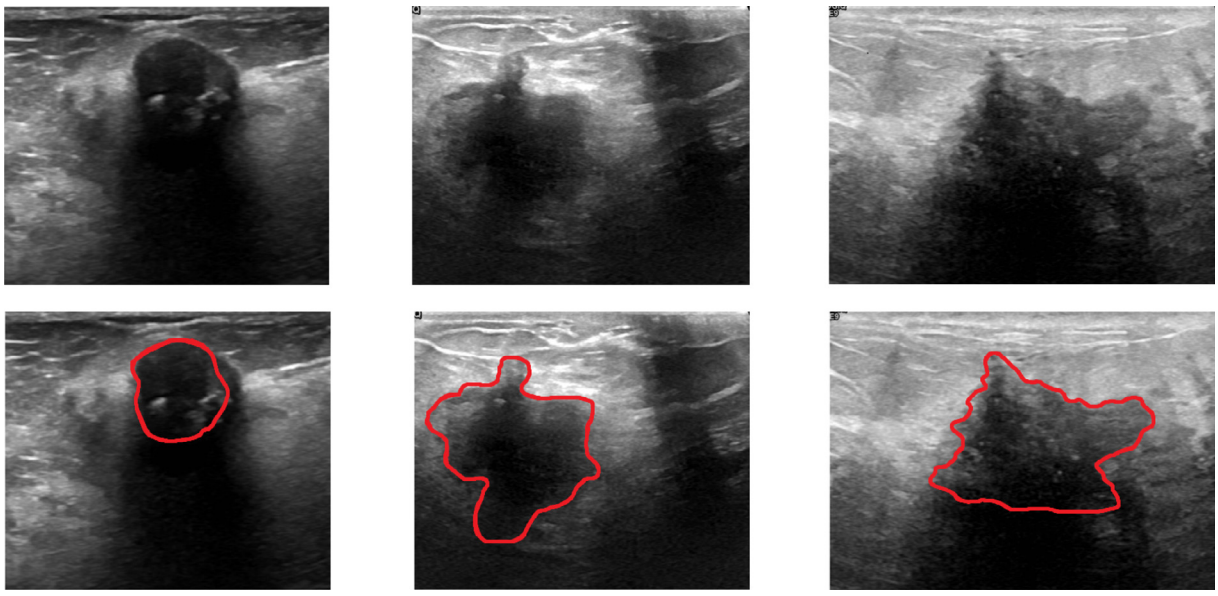

Fig. 1. Three ultrasound images of breast tumor lesions with blurred boundary. The first row is the ultrasound images. The second row is the outlines of the lesion areas, which are labeled by red lines.

GAN for BUS image segmentation. ASSGAN is an asymmetric generative adversarial network that is different from currently existing GANs. In ASSGAN we define two generators and a discriminator, shown in Fig. 2, instead of a pair composed of a generator and a discriminator, as commonly exist in GANs. The input of the discriminator is a segmentation mask, the discriminator responsible for judging whether the segmentation masks are real or not. After filtering the segmentation results of two generators by the discriminator, the masks classified as real masks are used as supervision signals for the other generator to achieve the purpose of using unlabeled images. Inspired by Hung's work [6], the proposed ASSGAN also uses a segmentation network as the generator. This approach is different from a typical generator that generates an image from a noise vector. Our generators learn the mapping from BUS images to semantic segmentation masks. In our method, the training stage consists of two steps. At the first step, we use labeled images to make the generator and discriminator have basic segmentation and discrimination ability. In this way, the discriminator can avoid misrecognition of the segmentation masks. Misrecognition refers to the use of a bad segmentation mask as a supervision signal, which might affect the generators' improvements. At the second step, we use the unlabeled images. The discriminator scores the generated masks from unlabeled images, and these masks with high scores are considered to be the real masks. Then the real masks are treated as the ground truth for two generators. In summary, our contributions are as follows:

(a) An asymmetric adversarial network structure in which two generators correspond to one discriminator is proposed. It make full use of unlabeled BUS images to improve segmentation performance. This method reduces the labor and financial costs of labeling, and training requires only limited labeled data.
(b) A mutual supervision module composed of two generators is proposed to provide supervision signals for each other to effectively use unlabeled data. Adversarial training is used

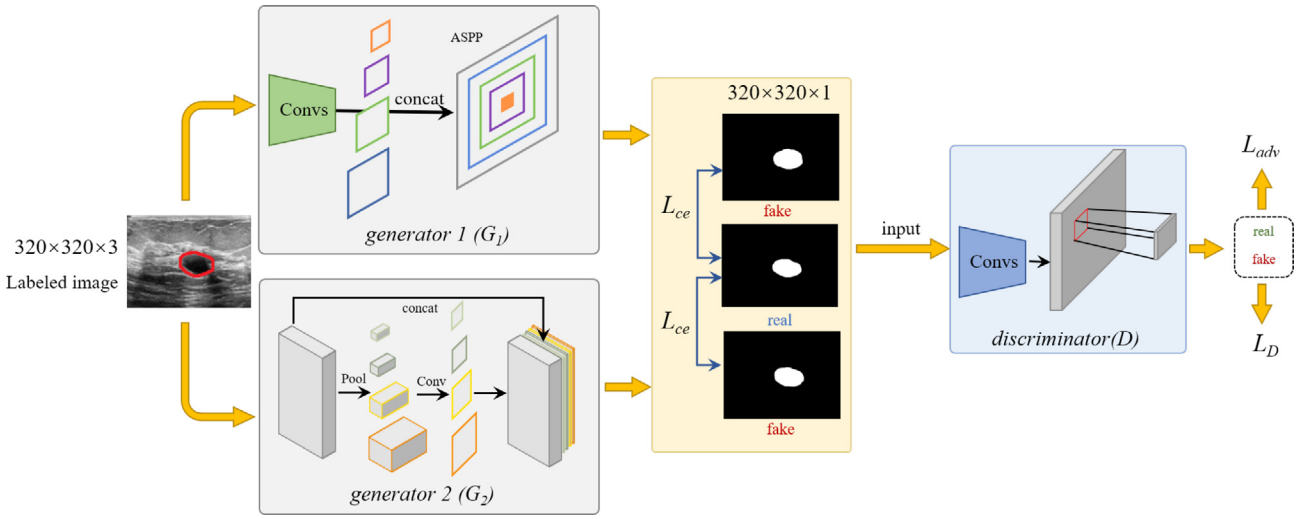

Fig. 2. The training process for the labeled images. G_1, G_2 generate the predicted masks to fool the discriminator and use L_{adv} as adversarial loss to constrain the generators to generate images that are closer to the ground truth. At the same time, the masks generated by G_1 and G_2 are also guided by the cross-entropy loss L_{ce} with the ground truth.

to further constrain the supervision signal to be closer to the ground truth, thereby promoting the training process of the mutual supervision module.

(c) Experimental results on four datasets demonstrate the effectiveness of the proposed adversarial framework using unlabeled data to improve segmentation performance.

The remainder of this paper is organized as follows. Section 2 includes the related work on BUS image segmentation, semi-supervised semantic segmentation, and generative adversarial networks. Section 3 elaborates the proposed method, including the network architecture, training details and so on. Section 4 describes the experimental analysis. Finally, we summarize our work.

2. Related work

2.1. BUS Image Segmentation

In recent years, CNNs have made considerable progress in the field of medical imaging. Yap et al. [7] studied three different CNNs for BUS tumor segmentation: a patch-based U-Net [8], LeNet [9], and a transfer learning approach with a pretrained FCN-AlexNet [10]. They also compared CNNs with traditional methods and found that the CNNs had better overall performance than the other four lesion detection algorithms (*i.e.*, radial gradient index, multifractal filtering, rule-based region ranking, and deformable part models). In addition, Huang et al. [11] trained a full convolutional network (FCN) by using information extended images. The BUS image was semantically divided into three categories: breast layer, tumor, and background. Finally, the structure information of the mammary layer was applied to the conditional random field to segment the breast tumor lesion, which was tested on 325 BUS images. Zhuang et al. [12] proposed a model based on a traditional U-Net, using residual units [18] instead of ordinary neural units to overcome performance degradation. Additionally, Zhuang et al. [12] combined the expansion convolution and the attention gate module to improve the learning ability of the model by increasing the receptive field while suppressing the background information. Shareef et al. [13] proposed a small tumor sensing network that integrates rich context information and high-resolution image features to solve the problem of difficult segmentation of small breast tumors. Xue et al. [14] developed a deep convolutional neural network equipped with a global guidance block (GGB) and breast lesion boundary detection (BD) modules for boosting the breast ultrasound lesion segmentation. Zhu et al. [15] propose a novel second-order subregion pooling network (S2P-Net) for boosting the breast lesion segmentation in ultrasound images. The network aggregated global features from the whole image and local information of subregions. Huang et al. [16] proposed a novel method to segment the breast tumor via semantic classification and merging patches. First, they extract the region of interest in the images, then use methods such as histogram equalization to enhance the images. Then, the images are divided into multiple superpixels. Finally, the result is refined by using k-nearest neighbor (KNN). Huang et al. [17] proposed a new segmentation method by dense prediction and local fusion of superpixels for breast anatomy. They transformed the small sample problem of segmentation into the large sample of classification by using superpixels.

Most of the above methods are models that are trained in a fully supervised way; however, the majority of their performances are affected by the amount of accurately labeled data. It is difficult to obtain BUS images with perfectly accurate labeling in practice. This paper proposes a novel semi-supervised method by using the idea of the generative adversarial network.

2.2. Semi-supervised learning

Semi-supervised and weakly supervised image segmentation methods have been proposed to solve the challenge of data scarcity. Li et al. [19] used a weighted combination of ordinary supervised loss with only labeled input and regularization loss of both labeled and unlabeled data. To make use of unlabeled data, this method made consistent predictions for the same input when training under different regularization conditions. Experiments showed that this method achieved good segmentation performance on three medical image segmentation tasks. To solve the problem of having only a limited number of manual annotations, Dai et al. [20] proposed a novel multitask learning framework for 3D brain image segmentation. This framework utilized a large number of automatically generated partial annotations together with a small set of manual annotations for the network training. In target detection area, Chen et al. [21] present a multi-task mean teacher model for semisupervised shadow detection by leveraging complementary information shadow regions, shadow edges, and shadow. These additional unlabeled data effectively boosting the shadow detection perfomance. But it might not work well for images with multiple and complex shadows. In image synthesis, Liu et al. [22] proposed a novel image dehazing framework collaborating with unlabeled real data. They encourage the coarse predictions and refinements of each disentangled component to be consistent between the student and teacher networks by using a consistency loss on unlabeled real data.

This study is partially inspired by Wang et al. [23]. They proposed a bottom-up and top-down iterative framework, which uses a classification network to initialize the seed points in such a way that the segmentation network focuses on the relevant areas around the seed points, and then uses the predicted mask to update the seed points. In the bottom-up step, they extracted common object features from the initial location and used the mined features to extend the area of the object. In the top-down step, the refined object region is used as supervision to train the segmentation network and predict the segmentation mask. In addition, the segmentation mask is used as the initial location and common object features were extracted from it. These two steps are repeated to gradually generate a more accurate segmentation mask. This study follows the iterative training method of a classification network and the segmentation network in Wang's method. There are some differences between our approaches: on the one hand, we use two segmentation networks to monitor and optimize each other. On the other hand, we employ the idea of a GAN to filter the segmentation masks by a discriminator. The masks that pass the discriminator are further used as supervision signals.

2.3. Generative Adversarial Network

Semi-supervised learning methods using GANs have been recently proposed in the literature. For example, Hung et al. [6] designed a fully convolutional discriminator to discover the confidence of the predicted results on unlabeled images. This step provided additional supervision signals. Finally, the accuracy of the semantic segmentation is improved by fusing the adversarial loss and the standard cross entropy loss. Compared with the existing methods that use weakly labeled images, this method shows better performance on the datasets of PASCAL VOC 2012 [24] and Cityscapes [25]. Our current method follows the adversarial training ideas of Hung et al. [6]. It also uses the segmentation network as the generator, and optimizes the generator through the back propagation of discriminator ratings.

In the field of medical imaging, the problem of data scarcity is especially serious. Therefore, many solutions using GAN have been proposed to solve this problem. Li et al. [26] proposed an adversar-

ial training method to segment brain tumors. This method used a discriminator to distinguish the generated mask from the real segmentation mask. The adversarial loss provided by the discriminator network is fed back to the generating network. This method not only reduced the differences between the synthetic labels and ground-truth labels, but also reinforced the spatial contiguity with high-order loss terms. In another example, Lahiri et al. [27] added a new unsupervised adversarial loss and a structured prediction based architecture. This method performed better than fully supervised neural networks for retinal vascular segmentation and had an extremely low annotation ratio (0.8% - 1.6% of the contemporary annotation size). Moeskops et al. [28] used a method of adversarial training to improve CNN-based brain MR image segmentation. They designed a loss function that drives the generator to produce a segmentation that was difficult to distinguish from manual segmentation. This loss function was optimized along with the traditional cross entropy loss. Moeskops et al. [28] evaluated two different sets of images and two different network architectures. The results showed that their method had better segmentation performance both visually and quantitatively under the *Dice* metric. Our method also utilizes the idea of the generative adversarial approach to solve the problem of data insufficiency. The proposed method consists of two generators and a discriminator, and feeds back the adversarial loss of the discriminator to the generator. In addition, the discriminator filters the segmentation masks of the two generators. The masks that pass the screening of the discriminator are called trusted masks. The two generators are optimized by calculating the cross entropy loss of the trusted mask for the supervision.

3. Methodology

3.1. ASSGAN for semi-supervised segmentation

Asymmetric semi-supervised GAN (ASSGAN) consists of three main parts, two generators (G_1, G_2) and a discriminator (D), as shown in Fig. 2. G_1 and G_2 are two segmentation networks that are responsible for generating masks. The input size of G_1, G_2 is $320 \times 320 \times 3$. The output segmentation masks have a size of $320 \times 320 \times C$, where C is the number of semantic categories, *e.g.*, in this study, $C = 1$ means category 1, denoting the lesion region. D is used to distinguish whether a mask is real or fake. For each iteration, in a batch, D takes the predicted masks, generated from G_1, G_2, and the ground-truth masks as inputs. The output of D is γ, which represents the probability that the mask is real. We divide the training process into two steps, a fully supervised step and a semi-supervised step.

Algorithm 1: Supervised Process

Input:
Labeled data X^L and ground truth Y;
Confidence coefficient threshold γ;
Output:
Predicted masks Z;
Confidence coefficient R;
Initialization:
Generator G_1, Generator G_2 and Discriminator D;
$N = 200$ 1: **for** epoch = 1 to N **do**
2: shuffle (labeled data)
3: divide labeled data into small batches with size M
4: **for** i = 1 to M **do**
5: Estimate $z_i^1 \Leftarrow G_1(x_i^l)$ via (1) with y_i
6: Estimate $z_i^2 \Leftarrow G_2(x_i^l)$ via (1) with y_i
7 : Optimize G_1, G_2 via Adam
8: Estimate $r_i^1, r_i^2 \Leftarrow D(z_i^1, z_i^2)$ via (3), (4)
9: Estimate $r_{y_i} \Leftarrow D(y_i)$ via (3)
10: update G_1, G_2, D using Adam
11: **end for**
12: **end for**

In the fully supervised step, only labeled images are used. Fig. 2 shows the process of the supervised part of ASSGAN, the pseudocode of which is listed in Algorithm 1. The input consists of two parts, labeled images $X^L = \{x_1^l, x_2^l, \cdots, x_n^l\}$ and ground truth masks $Y = \{y_1, y_2, \cdots, y_n\}$. The output includes predicted masks $Z = \{z_1, z_2, \cdots, z_n\}$ and confidence coefficient $R = \{r_1, r_2, \cdots, r_n\}$. Here, z_i^1, r_i^1 are the i-th result from G_1 in the batch and z_i^2, r_i^2 are from G_2. A labeled image is input into G_1, G_2 at the same time, and two different predicted masks are generated. At the training time, we optimize generators by using the cross entropy loss L_{ce} to calculate the results of these two generators separately. $G_{1,2}$ means G_1 and G_2 in *Eq.* (1).

$$L_{ce} = -\sum\sum Y_L \log\left(G_{1,2}\left(X^L\right)\right) \tag{1}$$

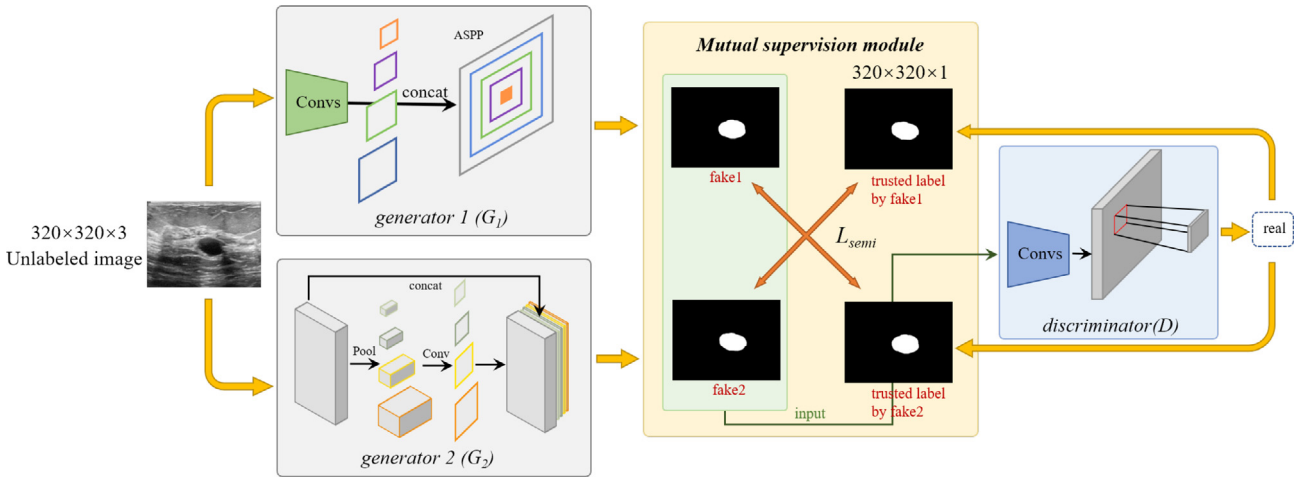

Fig. 3. Semi-supervised flow with unlabeled images. The trusted labels are considered to be real segmentation masks for the discriminators. D judges whether the masks predicted by G_1 and G_2 are real. The trusted label is used as the other generator's supervision signal.

Then, these two results z_i^1, z_i^2 are fed into the discriminator D. The output of D is the evaluation score r_i^1, r_i^2 of z_i^1, z_i^2. They represent the truthfulness of the results. At the same time, the adversarial loss L_{adv} provided by discriminator D is fed back to G_1, G_2. X^L is a labeled image and X^U is an unlabeled image. Such a step aims to reduce the difference between the predicted masks and the ground truth.

$$L_{adv}(G_{1,2}, D) = \log D(Y) + \log\left(1 - D\left(G_{1,2}\left(X^U, X^L\right)\right)\right) \tag{2}$$

Here, L_{adv} is similar to the adversarial loss used in [6], which uses the cross-entropy loss to help the generators output more accurate segmentation masks. To train D, we minimize the cross-entropy loss function L_D of the predicted and real masks.

$$L_D = -\sum(1-y)\log\left(1 - D\left(G_{1,2}\left(X^L, X^U\right)\right)\right) + y\log\left(D(Y_L)\right) \tag{3}$$

where $y = 0$ if the sample is created by the generators, and $y = 1$ if the sample is from the ground truth labels. In addition, $D(G_{1,2}(X^L, X^U))$ is the probability of the predicted masks of G_1, G_2, and $D(Y)$ is the probability of the ground truth labels. G_1, G_2 learn the distribution of the true segmentation masks through fooling the discriminator. For the labeled images, G_1, G_2 are optimized by L_{ce} and L_{adv}. Note that G_1, G_2 input the same images during training, but they do not share weights.

The second is the semi-supervised step, where we propose a mutual supervised module to add unlabeled data for training. Fig. 3 shows the implementation process of the semi-supervised process in ASSGAN, and the pseudocode is described in Algorithm 2. The inputs consist of the unlabeled images $X^U = \{x_1^u, x_2^u, \cdots, x_n^u\}$ and confidence coefficient threshold γ, and the outputs are the predicted masks Z. Q is the total number of epochs. The predicted masks are input into the discriminator. The discriminator network recognizes whether the two predicted masks are real or fake. If the discriminator recognizes a predicted mask as a real label (called a trusted label), then this trusted label is used as the ground truth to supervise the other generator. Otherwise, the results of this segmentation are discarded and the generator parameters are not updated. The semi-supervised loss function is defined as:

Algorithm 2: Semi-Supervised Process

Input:
Unlabeled data X^U;
Confidence Coefficient threshold γ;
Output:
Predicted masks Z;
Confidence coefficient R;
Initialization:
Generator G_1, Generator G_2 and Discriminator D;
$Q = 700$
for epoch = N to Q **do**
  shuffle (unlabeled data)
  divide unlabeled data into small batches with size M
  for i = 1 to M **do**
    Estimate $z_i^1 \Leftarrow G_1(x_i^u)$
    Estimate $z_i^2 \Leftarrow G_2(x_i^u)$
    Optimize G_1, G_2 via Adam
    Estimate $r_i^1, r_i^2 \Leftarrow D(z_i^1, z_i^2)$ via (3), (4)
    if $r_1^i > \gamma$ **then**
      $L_{semi}(z_i^1, z_i^2)$ via (2)
    end if
    if $r_2^i > \gamma$ **then**
      $L_{semi}(z_i^2, z_i^1)$ via (2)
    end if
    update G_1, G_2 using Adam
  end for
end for

$$L_{\text{semi}} = -\sum\sum\Theta\left(D\left(G_1\left(X^U\right)\right) > \gamma\right)\log\left(G_2\left(X^L\right)\right) \tag{4}$$

In the above expression, $\Theta(\cdot)$ is the decision function $\Theta(x) = \begin{cases} 0, x < \gamma \\ 1, x \geqslant \gamma \end{cases}$. In *Eq.* (4). We assume that the discriminator recognizes the prediction result of G_1 as the supervised label and uses that label to supervise G_2. In this way, unlabeled images are utilized. In other words, this method is equivalent to increasing the amount of data. The segmentation network can learn more feature information and improve the segmentation performance.The objective function is:

$$L_{\text{total}}(G_1, G_2, D) = \arg\min_{G_{1,2}}\max_{D}\lambda L_{GAN} + L_{ce} + L_{\text{semi}} \tag{5}$$

3.2. Network Architecture

3.2.1. Segmentation Network

We use DeepLabV3+ [29] based on resnet-101 as the generator G_1, and pretrain on the ImageNet [31] and MSCOCO [32] datasets. During the sampling process, the *conv3* and *conv4* layers are reconstructed to enlarge the receptive field. The *conv3* and *conv4* layers use dilated convolutions at steps 1 and 2 with a dilation rate of 2. After sampling the last layer, we use the Atrous Spatial Pyramid Pooling (ASPP) module to further extract the spatial information of different scales. Since the segmentation of BUS lesions is a binary classification problem, we finally use a sigmoid as the output function. In addition, we use PSPNet [30] as the generator G_2, which is also pretrained on the ImageNet dataset and the MSCOCO dataset. The sigmoid activation function is also used in the last layer of upsampling. The purpose of this step is to make the two networks start training from the same origin, and speed up the first stage of training with initialization parameters.

The global contextual information pyramid pooling module in PSPNet [30] can effectively obtain high-quality results in scene semantic analysis. DeepLabV3+ [29] uses spatial pyramid pooling combined with a simple and effective decoder module to correct the segmentation results, especially for boundary correction. In addition, the DeepLabV3+ [29] decoder uses a feature fusion strategy to retain more low-level information. In general, PSPNet [30] and DeepLabV3+ [29] both use deep and low-level feature fusion, dilated convolution, and pyramid pooling strategies. These excellent strategies make the two networks have a good segmentation performance. But PSPNet [30] pays more attention to scene-level global information to reduce the probability of mis-segmentation, while DeepLabV3+ [29] pays more attention to boundary correction. Therefore, the mutual supervision of two networks can learn the characteristics of each other and play a complementary role.

3.2.2. Discriminator Network

We use the Markov Discriminator (PatchGAN) proposed in [6] as the discriminator network, D, which attempts to classify each $N \times N$ patch in an image as real or fake. This approach is faster than

a full image discriminator; it has fewer parameters and is compatible to images of varying sizes.

3.3. Training Details

Throughout the training process, these two generators have equal positions, which is that no one is superior to the other. At the end, we can obtain two segmentation networks with comparable performance. We use minibatch *SGD* and apply the Adam solver [33], with a learning rate of 0.0001. The first phase consists of 200 epochs and the second phase consists of 500 epochs, making 700 epochs in total. At the first step, the two generators are trained independently. They are both supervised by the ground truth and L_{adv} from the discriminator. According to our experiments, the second step starts when the first step reaches 200 epochs. We used Pytorch to conduct the ASSGAN. The models were trained by using NVIDIA GeForce GTX 1080Ti (12 GB).

In our method, G_1 and G_2 that are really required at last. We did not make too many changes to the structure of the generators during training process. Therefore, our method does not significantly increase the model size and inference time of the developed network. In our experiment, the model size of DeepLabV3+ [29] is 59 MB and PSPNet [30] is 70 MB. It takes 2.8593s to split 50 pictures with NVIDIA GeForce GTX 1080ti, with an average of 0.0572s per picture.

4. Experiments

4.1. Datasets

We used four BUSI datasets to evaluate our methods. To focus on the lesion area and reduce the interference of multiple lesions in one image, we cropped these images. Thus only one lesion area in each image is ensured. DBUI [5] is a publicly available BUSI dataset, which was collected in 2018. It contains BUS images of 600 women between the ages of 25 and 75. There are 437 benign tumors, 210 malignant tumors (647 in total) and the corresponding segmentation masks. The average image size was 500 × 500 pixels. Fig. 4 shows some samples of BUS images and segmentation masks of benign and malignant tumors in DBUI. SPDBUI [34] normalized the size of all of the BUSI to 128 × 128, and the labels of the tumor regions were manually delineated by a well-trained radiologist with over ten years of experience. The dataset is composed of

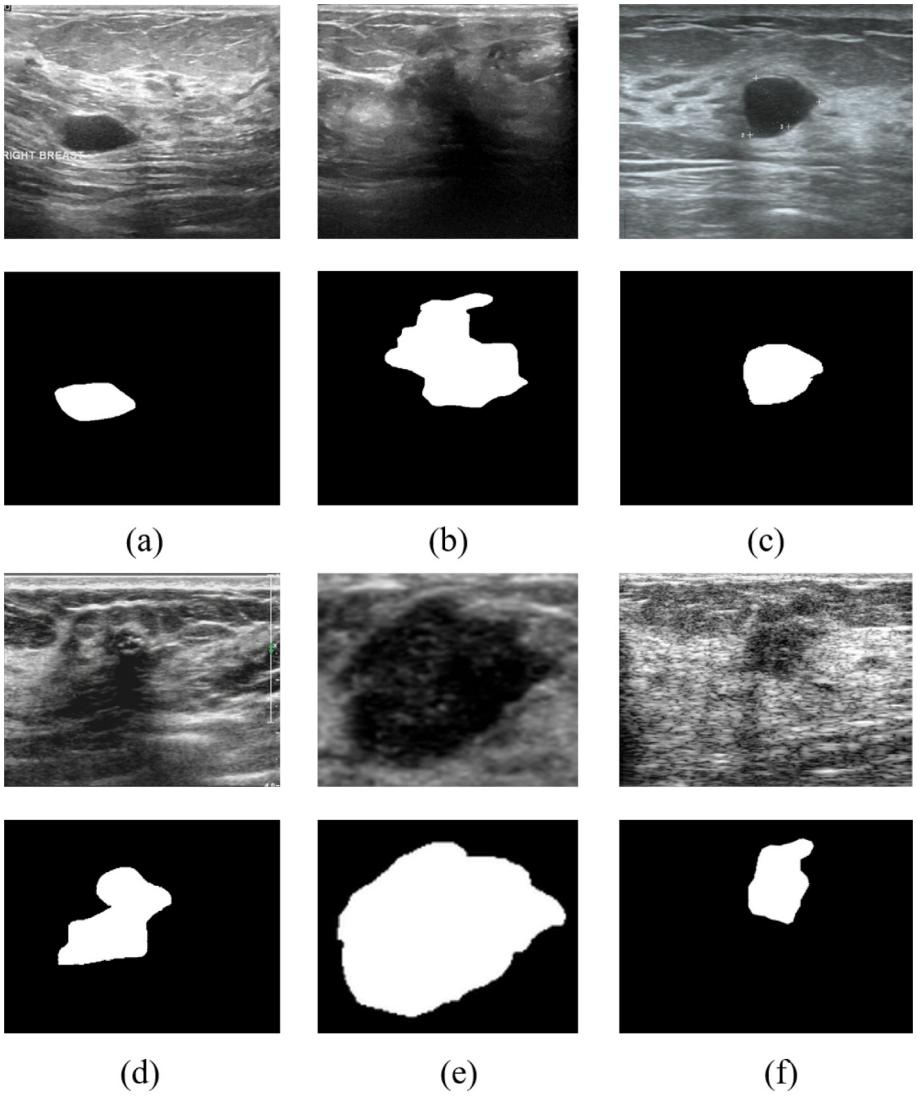

Fig. 4. Some samples of DBUI and SDBUI, including the ultrasound images and segmentation masks. ((a); (b)) are from the DBUI, (a) is a benign sample, and (b) is a malignant sample. ((c); (d)) are from the SDBUI, where (c) is a benign sample, and (d) is a malignant sample. ((e); (f)) are from SPDBUSI and ADBUSI, respectively. The second row is the segmentation mask that corresponds to each image.

160 BUS images with benign tumors and 160 BUS images with malignant tumors. ADBUI [35] includes scans from 52 malignant and 48 benign breast lesions recorded in a group of 78 women. For each lesion, two individual orthogonal scans from the pathological region were acquired with the Ultrasonix SonixTouch Research ultrasound scanner using the L14-5/38 linear array transducer.

The data volume of the above three datasets is relatively small. To further discuss the performance of ASSGAN in the case of different labeling ratios, we built the SDBUI (self-built dataset of breast ultrasound images), which consists of 1805 BUS images provided by the Ultrasound Department of a hospital in Sichuan Province, China. It was taken from an iU Elite System (Philips Medical Systems) with an L12-5 at an imaging frequency of 5.0 MHz. We normalized the sizes of all of the BUS images to 320×320 and the labels of the tumor regions were manually delineated by a well-trained radiologist. We used 1305 images as the training set. In Section *C*, we divided the labeled and unlabeled images in the training set into different proportions for evaluation. To simulate the situation of having inadequate labeled images, we randomly selected some samples from the four data sets as unmarked images. We aliased the datasets that contain partial labels as follows: PDBUI (DBUI), PSPDBUI (SPDBUI), PADBUI (ADBUI), and PSDBUI (SDBUI). The specific distribution of data is shown in Table 1.

Unless otherwise mentioned, the labeled data in SDBUI took 15% of the training set. There are 1110 unlabeled images and 195 labeled images for training. Among the remaining 500 images, 350 were used as the test set, and 150 were used for validation.

Table 2
The number of utilized unlabeled images on different datasets.

Dataset	PDBUI	PSPDBUI	PADBUI	PSDBUI
Unlabeled images	447	170	110	1110
Utilized unlabeled images	243	106	62	667

4.2. Evaluation Metrics

In this paper, the evaluation metrics include intersection over union(*IoU*), pixel accuracy(*Acc*), and dice coefficient(*Dice*). These three evaluation metrics are defined as follows:

$$IoU = \frac{TP}{TP + FP + FN} \tag{6}$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

$$Dice = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{8}$$

where *TP*, *FP*, and *FN* are *true positive*, and *false positive*, and *false negative* pixels, respectively. *IoU* represents the coincidence rate between the predicted area and the real area. *Acc* represents the percentage of pixels of the correct prediction compared to the total pixels. The *Dice* coefficient measures the degree of overlap between the predicted result and the real segmentation mask.

Table 1
The data distribution on different datasets.

Dataset	PDBUI (DBUI)	PSPDBUI (SPDBUI)	PADBUI (ADBUI)	PSDBUI (SDBUI)
Training	100 labeled (18%) 447unlabeled	50 labeled (23%) 170 unlabeled	30 labeled (21%) 110 unlabeled	195 labeled (15%) 1110 unlabeled
Validation	50 labeled	50 labeled	30 labeled	150 labeled
Test	50 labeled	50 labeled	30 labeled	350 labeled

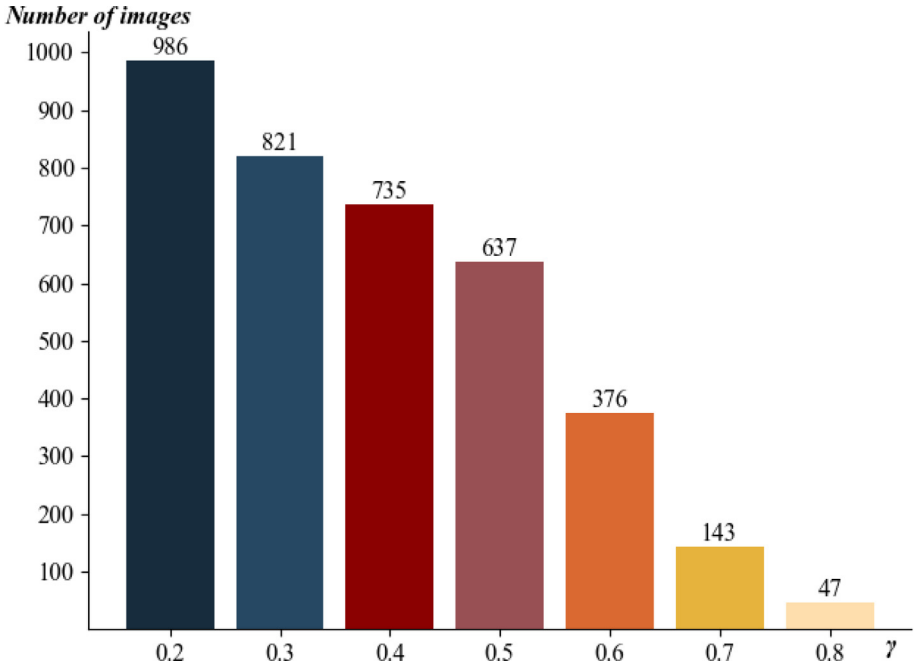

Fig. 5. The number of unlabeled images is used with different γ.

4.3. Analysis of Hyper-parameter γ

The proposed algorithm is influenced by a hyperparameter: γ in *Eq.* (4). γ is a threshold for evaluating whether the image is trusted at the semi-supervised training step. If the output from the discriminator for a predicted mask is greater than γ, we would use that mask as a supervision signal. We can filter (by *Eq.* (4)) the predicted mask by setting different γ.

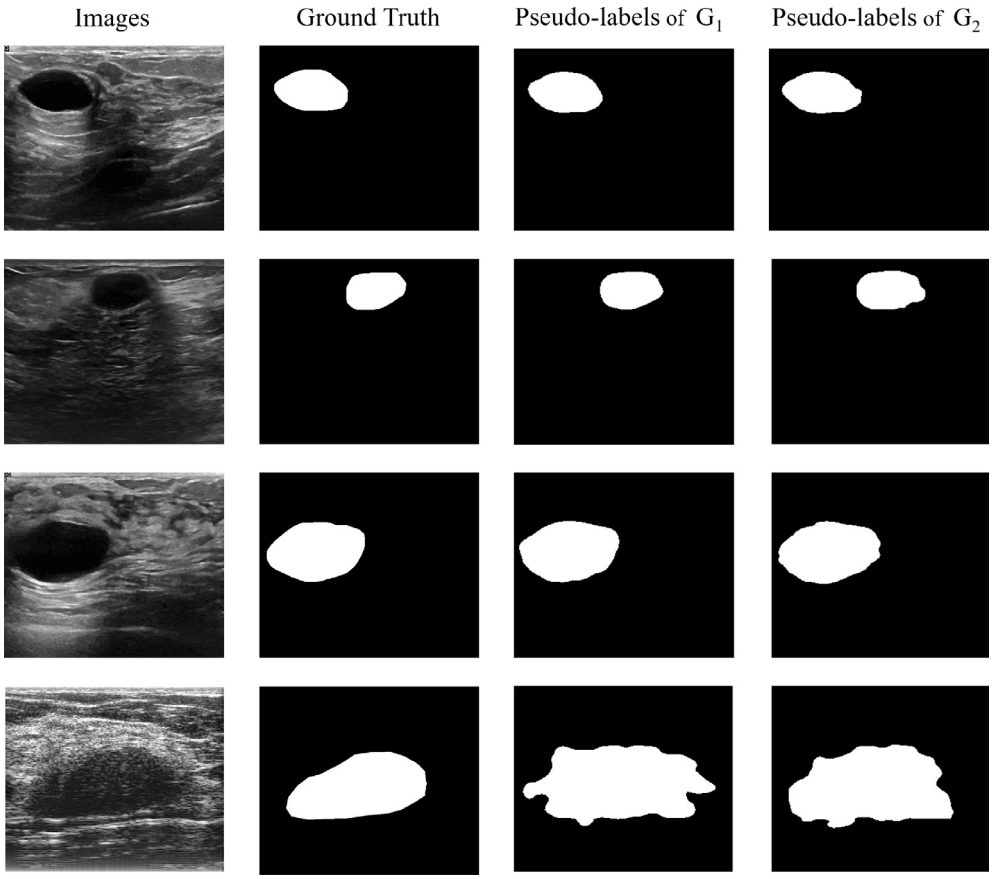

Fig. 6. Some pseudo-labels generated by G_1 and G_2.

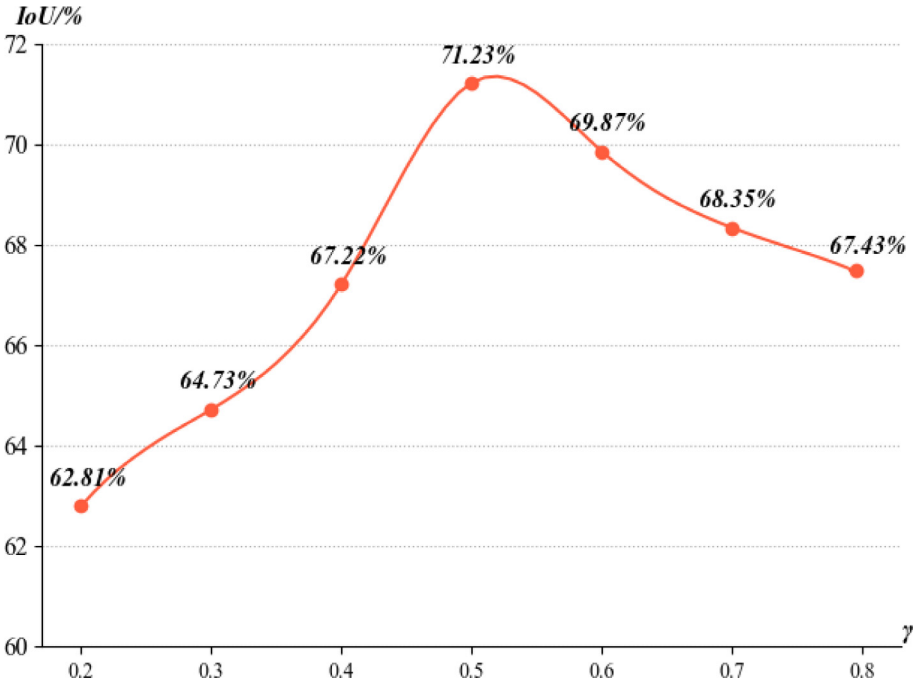

Fig. 7. The mean *IoU* of G_1 and G_2 with respect to different γ.

Table 3
The *IOU* of the four datasets on the five methods under 100% labeled.

Method	DBUI	SPDBUI	ADBUI	SDBUI
U-Net [8]	0.7642	0.8799	0.6362	0.7134
U-Net++ [36]	0.7621	0.8824	0.6235	0.7189
AttenU-Net [37]	0.7404	0.8883	0.6209	0.7234
DeepLabV3+ [29]	**0.7891**	**0.8912**	0.6359	**0.7344**
PSPNet [30]	0.7781	0.8904	**0.6421**	0.7319

The bold values in the table are the maximum values for this column.

To explore the effects of the proposed semi-supervised adversarial algorithm on unlabeled data, we list the number of trusted images under different γ, as shown in Fig. 5. The number of trusted images decreases as γ increases, which is from 986 to 47. Fig. 7 shows the segmentation performance evaluation results of the two generators under different thresholds. Here we take the average *IoU* of G_1 and G_2. It can be found that when γ is set to be too high, it is difficult for the predicated masks to pass the screening from the discriminator. Accordingly, the number of predicted

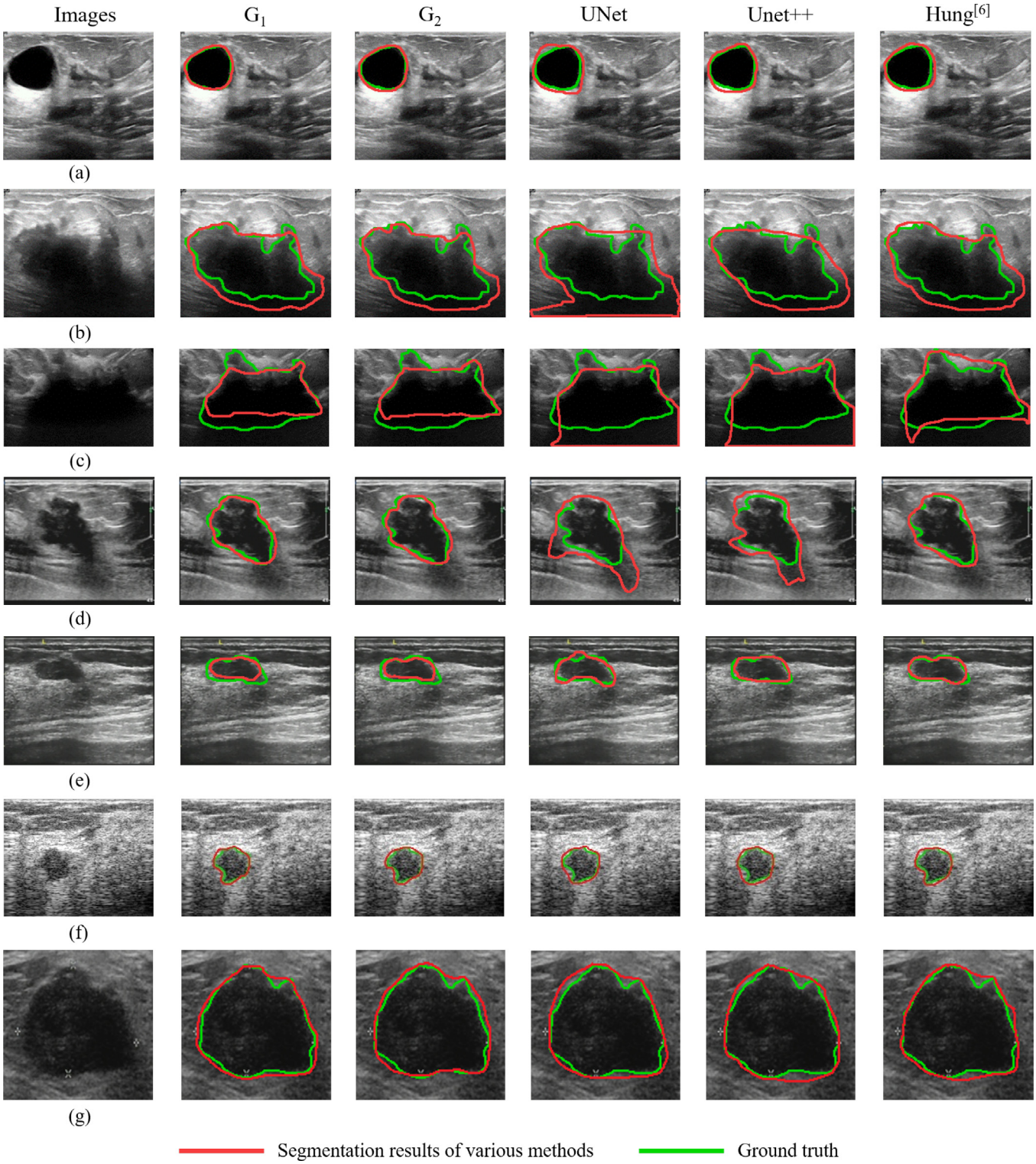

Fig. 8. Some segmentation results of different methods. ((a); (b); (c)) are from PDBUI, ((d); (e))) are from PSDBUI. (f) is from PADBUI. (g) is from PSPDBUI.

masks for the available unlabeled images is reduced. In this case, some good segmentation results could be wrongly discarded. Hence the performance is negatively impacted. Specifically, when γ is 0.8, it can be seen that only 47 segmentation predicted masks were used. The *IoU* is only 67.43%, which is as low as the results without semi-supervision. When the γ is set too low, some poor segmentation results are considered to be valid masks. This circumstance is equivalent to using poor segmentation results as supervision signals, which has poor effects on the overall performance. For example, when $\gamma = 0.2$, 986 segmentation predicted masks are used, and the *IoU* is decreased to 62.81%. Compared with $\gamma = 0.5$, the *IoU* is decreased by 8.42%. Finally, this experiment reveals that the segmentation predicted mask can be best used when $\gamma = 0.5$.

In order to show the promotion effect of our method on each dataset, we have calculated that under the condition that $\gamma = 0.5$, the number of utilized unlabeled images on each dataset. As shown in Table 2. In addition, some unlabeled images and corresponding pseudo-labels used in the training process are shown in Fig. 6. These pseudo-labels increase the amount of training data and improve the precision of segmentation model.

4.4. Segmentation Performance

To compare with the subsequent semi-supervised method and partially labeled data. We first obtained the segmentation performance of the four datasets under 100% labeling, as shown in Table 3. Fig. 8 (a), (b), (c) shows the segmentation results of G_1, G_2 on PDBUI. Tables 4–6, respectively, report the pixel-level performance evaluation results of the 6 methods for the BUS image segmentation task on PDBUI, PSPDBUI and PADBUI. Each evaluation metric is the average of G_1 and G_2. The results show that the performance of each network decreases significantly with a few labels. The performance of PADBUI decreases the most. Its *IoU* decreases by 22.35% compared with ADBUI on U-Net [8] and decreases by 15.66% on DeepLabV3+ [29]. The ASSGAN method proposed in this paper improves the segmentation performance compared with five fully supervised methods: U-Net [8], U-Net++ [36], AttenU-Net [37], DeepLabV3+ [29], and PSPNet [30], with a small number of labeled images. Compared with DeepLabV3+ [29], which has the best performance, *IoU* increased by 4.16%∼13.94%, *Acc* increased by 4.53% at most, and the *Dice* coefficient increased by 2.39%∼11.64%. The performance of our method is close to the effect of having 100% labeled training on the above datasets, with a gap of only 0.6%∼2.34% in the *IoU*.

Due to the limited scale of the above three datasets, the datasets are very limited with regard to verifying the effectiveness of our method. To better evaluate the performance of ASSGAN, we use the PSDBUI for tests, which contains 1805 images. Fig. 8 (d), (e) shows the segmentation results of G_1, G_2 on PSDBUI. In Fig. 8, our method does not have a good segmentation result for image (e). In fact, for some samples with extremely blurred boundaries, *i.e.*, (e) in Fig. 8, the masks generated by the generators are often poor. It is difficult for these masks to pass the filter of the discriminator, and the generators lack training for this type of data. Therefore, the segmentation results are not accurate. However, due to the constraint of L_{adv}, our method has a better segmentation effect than the other methods, which are without adversarial learning. Table 7 reports the segmentation performance results of the above six methods on PSDBUI (195 labeled images and 1110 unlabeled

Table 4
Performance of segmentation on PDBUI.

Method	IoU	Acc	Dice
U-Net [8]	0.5527	0.8549	0.7119
U-Net++ [36]	0.5861	0.8837	0.7391
AttenU-Net [37]	0.5974	0.9048	0.7480
DeepLabV3+ [29]	0.6385	0.9307	0.7794
PSPNet [30]	0.6293	0.9163	0.7725
Ours	**0.7683**	**0.9760**	**0.8690**

The bold values in the table are the maximum values for this column.

Table 5
Performance of segmentation on PSPDBUI.

Method	IoU	Acc	Dice
U-Net [8]	0.8683	0.9401	0.9295
U-Net++ [36]	0.8587	0.9365	0.9240
AttenU-Net [37]	0.8669	0.9387	0.9287
DeepLabV3+ [29]	0.8436	0.9243	0.9152
PSPNet [30]	0.8654	0.9358	0.9278
Ours	**0.8852**	**0.9508**	**0.9391**

The bold values in the table are the maximum values for this column.

Table 6
Performance of segmentation on PADBUI.

Method	IoU	Acc	Dice
U-Net [8]	0.4127	0.9492	0.5843
U-Net++ [36]	0.4496	0.9526	0.6203
AttenU-Net [37]	0.4212	0.9503	0.5927
DeepLabV3+ [29]	0.4793	0.9589	0.6480
PSPNet [30]	0.4673	0.9562	0.6370
Ours	**0.6187**	**0.9605**	**0.7644**

The bold values in the table are the maximum values for this column.

Table 7
Comparison on PSDBUI.

Method	IoU	Acc	Dice
U-Net [8]	0.5879	0.8462	0.7405
U-Net++ [36]	0.6172	0.9137	0.7633
AttenU-Net [37]	0.6258	0.9097	0.7698
DeepLabV3+ [29]	0.6742	0.9474	0.8054
PSPNet [30]	0.6689	0.9429	0.8016
Hung [6]	0.6831	0.9497	0.8117
Ours	**0.7123**	**0.9589**	**0.8319**

The bold values in the table are the maximum values for this column.

Table 8
Segmentation performance under different labeling proportions (ranging from 5% ∼ 50%) on SDBUI.

Method	Labeled(%)	IoU	Acc	Dice
DeepLabV3+ [29]		**0.7344**	**0.9672**	**0.8468**
PSPNet [30]	100%	0.7319	0.9668	0.8452
U-Net [8]		0.7134	0.9595	0.8327
Ours	50%	0.7147	0.9603	0.8336
	30%	0.7131	0.9593	0.8325
	15%	0.7123	0.9589	0.8319
	5%	0.6785	0.9478	0.8085

The bold values in the table are the maximum values for this column.

Table 9
Ablation study on the PSDBUI with 15% labeled data.

Method	IoU	Acc	Dice
1 generator	0.6742	0.9474	0.8054
1 generator + GAN	0.6853	0.9482	0.8133
2 generators	0.6204	0.9053	0.7657
2 generators + GAN(Ours)	**0.7123**	**0.9589**	**0.8319**

The bold values in the table are the maximum values for this column.

images). It can be seen that on PSDBUI, our method has achieved considerable improvement in segmentation performance. Specifically, compared to the method DeepLabV3+, the *IoU* increased by 3.81%, *Acc* increased by 1.15%, and *Dice* coefficient increased by 2.65%. Based on the analysis of the above results, the advantage of fully supervised method is not obvious in the case of the small amount of labeled data. The semi-supervised adversarial algorithm proposed in this paper can make full use of unlabeled images to help the network train.

4.5. Ablation Experiments

Furthermore, we also make a comparison with the method of Hung et al. [6]. In the experiment, we used the same PSDBUI with 15% labeled. In Hung's method, we use the pretrained DeepLabV3+ [29] segmentation network as the generator. Under the same condition of the dataset and the segmentation network, we obtain the data as shown in Table 7. Our method improves the *IoU* by 2.92%, and the *Dice* improves by 2.02%.

To compare the performance gap between the fully supervised methods and ASSGAN. And observe the performance of ASSGAN when the amount of labeled data is extremely small, we performed comparative experiments. Table 8 reports the performance results of ASSGAN under different labeling ratios (5%∼50%). With the fully-supervised method, we used 1,305 labeled images. Next, we reduced the proportion of labeled images. It can be seen that when the labeled images are extremely small (5%), the proposed method still performs well, with the *IoU* reaching 67.85%. When the labeled data reached 15%, the segmentation effect was equivalent to that of U-Net under full supervision, with the *IoU* reaching 71.23%. As the proportion of labeled images increases, the improvement in the segmentation performance becomes less obvious. As can be seen in Table 8, when the amount of labeled data increases to 50%, the *IoU* increases by only 0.24% compared with 15%. The performance of ASSGAN is close to the effect of fully-supervised training with 1305 images, with a gap of only 2.21% in the *IoU*.

To further verify the role of GAN and the mutual supervision module, we conducted an ablation experiment. We use DeepLabV3+ [29] as a single generator, and DeepLabV3+ [29] and PSPNet [30] as two generators. We compared the performance of a single generator, a single generator + GAN, two generators, and two generators + GAN with 15% labeled data.

As shown in Table 9,a single generator + GAN cannot use unlabeled data, so there is little performance improvement compared to a single generator. As shown in the table, *IoU* has promoted 1.11%. Two generators can form a mutual supervision mode. However, since there is no discriminator to filter the generated pseudo-labels, those poor-quality pseudo-labels are noise to the model and will seriously affect model training. This is an unusable strategy, which is much worse than using a single generator. Therefore, the segmentation performance of the model in the second stage deteriorates. Table 9 shows the results obtained in the first stage of DeepLabV3+ [29]. The experimental results prove that the proposed method uses a mutual supervision module and a discriminator to filter pseudo labels, which is better than a single generator + GAN on *IoU* increased 2.7%.

4.6. Failure Case Analysis

Our method still has room for improvement, and the segmentation performance of images with very blurry boundaries is insufficient. As shown in the Fig. 9, we show some failure cases in the network training process. For these images, both generators have lost the boundary information of the lesion area. This leads to the lack of boundary information even if the pseudo-labels used for mutual supervision pass the screening of the discriminator. Therefore, it is difficult for two generators to obtain these boundary information through mutual supervision.

5. Conclusion

In this work, we proposed a semi-supervised segmentation model ASS-GAN based on the GAN architecture. ASSGAN is a new GAN architecture model with two generators and one discriminator. Two generators form a mutual supervision module to use unlabeled data to improve the accuracy and robustness of BUS images. ASSGAN is evaluated on three public datasets and one dataset we collected. The experimental results show that in the absence of labeled data, our network can make full use of unlabeled data to improve segmentation performance. In addition, the proposed segmentation network can be used in other similar medical image segmentation tasks, such as thyroid, prostate, and so on. Our method still cannot adapt well to some extremely difficult samples. Although our approach is proven to make full use of unlabeled data, there is still some room for improvement. Breast tumors also have complex morphological and multi-scale characteristics. Next, we will study how to learn various morphological features of tumors from a small number of labeled images.

CRediT authorship contribution statement

Donghai Zhai: Conceptualization, Methodology, Software, Validation. **Bijie Hu:** Writing - original draft, Writing - review & editing. **Xun Gong:** Investigation, Supervision, Project administration. **Haipeng Zou:** Visualization, Formal analysis. **Jun Luo:** Resources, Data curation.

Funding information

This work is partially supported by the National Natural ScienceFoundation of China (61876158), Fundamental Research-Funds for the Central Universities (2682021ZTPY030).

Declaration of Competing Interest

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

References

[1] Freddie, Bray, Jacques, et al, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, in CA: a cancer journal for clinicians, 2018..
[2] W.-K. Chen, The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2016.
[3] H. Kaiming, G. Georgia, D. Piotr, et al, Mask R-CNN, IEEE Transactions on Pattern Analysis & Machine Intelligence, pp:1-1,2017.
[4] H. Zhao, J. Shi, X. Qi, et al, Pyramid Scene Parsing Network, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE 2017.
[5] W. Al-Dhabyani, M. Gomaa, H. Khaled, A. Fahmy, Dataset of breast ultrasound images, Data in Brief 28 (2020), https://doi.org/10.1016/j.dib.2019.104863 104863.
[6] Hung, Wei-Chih & Tsai, Yi-Hsuan & Liou, Yan-Ting & Lin, Yen-Yu & Yang, Ming-Hsuan, Adversarial Learning for Semi-Supervised Semantic Segmentation,, 2018.
[7] M.H. Yap et al., Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks, IEEE J. Biomed. Health Inform. 22 (4) (July 2018) 1218–1226, https://doi.org/10.1109/JBHI.2017.2731873.
[8] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical

image computing and computer-assisted intervention. Springer, p. 234–241,2015.
[9] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324, https://doi.org/10.1109/5.726791.
[10] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, in: Proceedings of the 25th International Conference on Neural Information Processing Systems, ACM:, Red Hook, 2012, pp. 1097–1105.
[11] K. Huang, H.D. Cheng, Y. Zhang, et al., Medical Knowledge Constrained Semantic Breast Ultrasound Image Segmentation, in: 24th International Conference on Pattern Recognition (ICPR), Beijing, 2018, pp. 1193–1198, https://doi.org/10.1109/ICPR.2018.8545272.2018.
[12] Z. Zhuang, N. Li, A.N.J. Raj, *et al.*, An RDAU-NET model for lesion segmentation in breast ultrasound images, PLoS ONE, 14(8):e0221535,2019.
[13] B. Shareef, M. Xian, A. Vakanski, Stan: Small tumor-aware network for breast ultrasound image segmentation, in IEEE.2020.
[14] X.A. Cheng, Z.B. Lei, C Hf, et al., Global Guidance Network for Breast Lesion Segmentation in Ultrasound Images, in Medical Image Analysis, vol 70, pages 101989, doi: 10.1016/j.media.2021.101989.
[15] L. Zhu, R. Chen, H. Fu, et al., A Second-Order Subregion Pooling Network for Breast Lesion Segmentation in Ultrasound, doi:10.1007/978-3-030-59725-2_16.2020.
[16] Q. Huang, Y. Huang, Y. Luo, et al., Segmentation of breast ultrasound image with semantic classification of superpixels, Medical Image Analysis vol 61 (2020), pages 101657.
[17] Huang Q, Miao Z, Zhou S, et al. Dense Prediction and Local Fusion of Superpixels: A Framework for Breast Anatomy Segmentation in Ultrasound Image with Scarce Data, in IEEE Transactions on Instrumentation and Measurement, PP(99):1-1, 2021.
[18] Zhang Z, Liu Q, Wang Y, Road extraction by deep residual u-net, IEEE Geoscience and Remote Sensing Letters; 15(5):749–753. doi: 10.1109/LGRS.2018.2802944.2018.
[19] X. Li, L. Yu, H. Chen, C.-W. Fu, L. Xing and P.-A. Heng, Transformation-Consistent Self-Ensembling Model for Semisupervised Medical Image Segmentation, in IEEE Transactions on Neural Networks and Learning Systems, doi: 10.1109/TNNLS.2020.2995319.2020.
[20] Dai, Chengliang & Mo, Yuanhan & Angelini, Elsa & Guo, Yike & Bai, Wenjia, Transfer Learning from Partial Annotations for Whole Brain Segmentation, doi:10.1007/978-3-030-33391-1_23.2019.
[21] Z. Chen, L. Zhu, L. Wan, S. Wang, W. Feng, P.-A. Heng, A Multi-task Mean Teacher for Semi-supervised Shadow Detection, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2020), 2020, pp. 5610–5619, https://doi.org/10.1109/CVPR42600.2020.00565.
[22] Y. Liu, L. Zhu, S. Pei, H. Fu, J. Qin, Q. Zhang, L. Wan, W. Feng, From Synthetic to Real: Image Dehazing Collaborating with Unlabeled Real Data, in: Proceedings of the 29th ACM International Conference on Multimedia, 2021, pp. 50–58, https://doi.org/10.1145/3474085.3475331.
[23] X. Wang, S. You, X. Li and H. Ma, Weakly-Supervised Semantic Segmentation by Iteratively Mining Common Object Features, in IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 2018, pp. 1354-1362, doi: 10.1109/CVPR.2018.00147.2018.
[24] X. Wang, S. You, X. Li and H. Ma, Weakly-Supervised Semantic Segmentation by Iteratively Mining Common Object Features, in IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 2018, pp. 1354-1362, doi: 10.1109/CVPR.2018.00147.2018.
[25] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, B. Schiele, The cityscapes dataset for semantic urban scene understanding, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 3213–3223.
[26] Li Z, Wang Y, Yu J, Brain Tum or Segmentation Using an Adversarial Network, in Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. 2018.
[27] A. Lahiri, V. Jain, A. Mondal, et al., Retinal Vessel Segmentation under Extreme Low Annotation, A Generative Adversarial Network Approach (2018).
[28] Moeskops P., Veta M., Lafarge M.W., Eppenhof K.A.J., Pluim J.P.W., Adversarial Training and Dilated Convolutions for Brain MRI Segmentation, in: Cardoso M. et al. (eds) Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. DLMIA 2017, ML-CDS 2017. Lecture Notes in Computer Science, vol 10553. Springer, Cham. 2017, doi: 10.1007/978-3-319-67558-9_7.
[29] Chen LC., Zhu Y., Papandreou G., Schroff F., Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation, in: Ferrari V., Hebert M., Sminchisescu C., Weiss Y. (eds) Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science, vol 11211. Springer, Cham. 2018, doi: 10.1007/978-3-030-01234-2_49.
[30] Zhao, Hengshuang Shi, Jianping Qi, Xiaojuan Wang, Xiaogang Jia, Jiaya, Pyramid Scene Parsing Network,2016.
[31] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, Imagenet: A large-scale hierarchical image database, in CVPR, 2009.
[32] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick, Microsoft coco: Common objects in context, in ECCV, 2014.
[33] D. Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv: 1412.6980, 2014.
[34] Q. Huang, Y. Huang, Y. Luo, et al., Segmentation of breast ultrasound image with semantic classification of superpixels, Medical Image Analysis 61 (101657) (2020).
[35] Hanna Piotrzkowska-Wróblewska, Katarzyna Dobruch-Sobczak, Byra M, et al., Open access database of raw ultrasonic signals acquired from malignant and benign breast lesions, in Medical Physics, 2017, 44(11).
[36] Zhou Z., Rahman Siddiquee M.M., Tajbakhsh N., Liang J., UNet++: A Nested U-Net Architecture for Medical Image Segmentation, in: Stoyanov D. et al. (eds) Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. DLMIA 2018, ML-CDS 2018. Lecture Notes in Computer Science, vol 11045. Springer, Cham. doi: 10.1007/978-3-030-00889-5_1.2018..
[37] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Attention U-Net: Learning Where to Look for the Pancreas, in CVPR.2018.

Donghai Zhai received his B.E. degree in hydrogeology and engineering geology, M.S. degree in computer science and technology, and Ph.D. degree in traffic information engineering and control from Southwest Jiaotong University, China, in 1997, 2000, and 2003 respectively. From 2003 to 2005 he was employed at IBM China Research Laboratory performing system analyses and architecture design. Since 2006 he has been associated with Southwest Jiaotong University in the department of software engineering. He is also visiting scholar at Louisiana State University (LSU), Baton Rouge, USA. His research interests include digital image processing, computer vision, and pattern recognition.

Bijie Hu received the B.S degree in computer science and technology from the Chengdu Technological University (CDTU, China) in 2019. Now he is a master student in computer science and technology from Southwest Jiaotong University (SWJTU, in China). His research interests include computer vision, medical image processing and deep learning.

Xun Gong received the B.S degree in computer science and technology from the Beijing Technology and Business University in 2003. He received the PhD degree in computer science and technology from Southwest Jiaotong University (SWJTU, China) in 2008. He was a visiting scholar of Alberta University, Canada (2015), Louisiana State University, U.S. (2018.7–2019.2). Now he is a professor in the School of Computing and Artificial Intelligence, Southwest Jiaotong University. His research interests include pattern recognition, computer vision, medical image processing and deep learning.

Haipeng Zou received the B.S degree in computer science and technology from the Sichuan Normal University in 2018. Now he is a master student in computer science and technology from Southwest Jiaotong University (SWJYU, in China). His research interests include medical image processing and computer vision.

Jun Luo received the B.S degree in medical science from the Medical Sciences of Chongqing University in 2003. He received the Master's degree in medical science from Sichuan University in 2008. He was a visiting scholar in Italy and Germany from 2014–2015). Now he is an associate professor in School of Medicine UESTC, Sichuan Academy of Medical Sciences, Sichuan Provincial People's Hospital. His research interests include interventional and contrast enhanced ultrasound using in thyroid, breast and liver disease.