This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2999198, IEEE Access

Digital Object Identifier

A Novel Multi-Scale Adversarial Networks for Precise Segmentation of X-ray Breast Mass

JUAN CHEN[1 2], LIANGYONG CHEN[1 2], SHENGSHENG WANG[1 2], PENG CHEN[3]
[1]College of Computer Science and Technology, Jilin University, Changchun 130012, China
[2]Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun 130012, China
[3]The Second Hospital of Jilin University, Changchun 130041, China

Corresponding author: Shengsheng Wang (e-mail: wss@jlu.edu.cn).

This work was supported in part by the Science and Technology Development Project of Jilin Province, China, under Grant 20190302117GX and Grant 20180101334JC, and in part by the Innovation Capacity Construction Project of Jilin Province Development and Reform Commission under Grant 2019C053-3.

ABSTRACT With the constant changes of people's lifestyle and living environment, the morbidity of breast cancer is increasing year by year. It is highly imperative to develop an effective breast mass segmentation method for early breast cancer diagnosis. However, segmenting breast masses in mammograms is still one hot issue with enormous challenges because of masses' irregular shapes and various sizes. In this study, we propose multi-adversarial learning to capture multi-scale image information for accurate breast mass segmentation. To effectively reinforce higher-order consistency in the segmentation results, the proposed network introduces the idea of adversarial networks, mainly consisting of a segmentation network and a discrimination network. An improved U-Net is introduced as the segmentation network to generate masks of the suspicious regions, while the discrimination network combines three convolutional critic networks that operate at different scales to discriminate the input masks. To weaken the unbalanced class problem and produce fine-grained segmentation results, weighted cross entropy loss and Earth-Mover distance are jointly used as an integrated loss function to guide the optimization process. Furthermore, the spectral normalization is adopted to the critics to alleviate the instability of training. The effectiveness of the proposed method is evaluated on two public datasets (INbreast and CBIS-DDSM). Experimental results empirically demonstrate that our method outperforms FCNs-based methods and the state-of-the-art method with dice of 81.64% for INbreast and 82.16% for CBIS-DDSM.

INDEX TERMS Deep learning, semantic segmentation, adversarial learning, breast mass, mammography

I. INTRODUCTION

Breast carcinoma is regarded as one of the most frequent malignant tumors among females across the globe, and it seriously threatens women's physical as well as mental health. According to the prediction of 2019, breast cancer, in the United States, is expected to alone account for 30% of all new cancer diagnoses and 15% of deaths in women [1]. Up to date, the exact cause of breast cancer is not clear enough, and there are not any obvious symptoms in the early stage. Early detection remains the most effective and reliable way to improve the survival rates of breast cancer patients [2]. As a typical screening tool for early diagnosis of breast cancer, mammography has high spatial resolution and is capable of detecting early breast abnormalities like microcalcification, architectural distortion, and breast masses. Due to the complicated biological structures and changeable morphological features of mammary tissues, it is time-consuming and labor-intensive for experts to review the breast pathological images and may even bring about misdiagnosis. In order to increase the diagnostic efficiency and suppress the number of false positives, computer-aided diagnosis (CAD) systems have been established to support the diagnosis of medical experts [3]–[5].

Breast mass is a common clinical symptom of breast cancer, which can be captured by mammography. Judging the nature of masses is able to provide a crucial reference for the follow-up treatment. In general, the more irregular a mass is, the more likely the lesion is malignant [6]. Therefore, the mass segmentation in mammography screening plays an indispensable role in breast cancer CAD systems and is the

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2999198, IEEE Access

key to further qualitative analysis of breast cancer. Nevertheless, since mammogram is two-dimensional imaging of the three-dimensional breast tissue, there may be other normal tissues similar to masses in images, which will interfere with the segmentation of the actual masses. What is worse, the considerable variability of masses in size and shape brings lots of difficulties to the robust and accurate segmentation.

Over the past few decades, scholars have extensively studied image segmentation of breast masses and proposed numerous breast mass segmentation algorithms based on conventional methodologies, typically including region-based algorithms [7], [8], contour-based algorithms [9], [10] and clustering algorithms [11]. Hsu [12] proposed an improved watershed transform based on intrinsic prior information to segment the mammogram image that was first processed by canny edge detector. Specific to the characteristics of breast mass segmentation, Mencattini et al. [13] modified region growing algorithm to reduce the computational overhead of this application. Song et al. [14] applied dynamic programming technique to search the contour of masses from the edge candidate points derived by the plane fitting method. A new CAD system was implemented by Xie et al. [15], in which the level set model was used for the segmentation of preprocessed images. Gu et al. [16] achieved comparatively high precision on breast mass segmentation by means of a superpixel generation and curve evolution method. Given the region of interest selected by an expert, Saleck et al. [17] developed a semi-automatic method of stepwise boundaries refinement in three stages for breast mass segmentation. Unfortunately, the aforementioned methods are mostly reliant on certain low-level image features extracted from the mass images, e.g. grayscale, texture or gradient information, which restricts the overall accuracy owing to lack of the representation for high-level information.

Replacing artificial design with autonomous features learning, deep learning implements function approximation with deep neural networks and directly learns the most essential features of images from raw data. Deep learning approaches based on fully convolutional networks (FCNs) [18], especially U-Net [19], have achieved impressive success in medical image segmentation [20]–[23], attributing to their excellent hierarchical feature representation. In the field of breast mass segmentation, Zhu et al. [24] combined multi-scale FCN with conditional random field (CRF) to improve mass segmentation on INbreast and DDSM-BCRP dataset. Al-Antari et al. [25] attempted an FCN, namely FrCN, for segmenting the breast mass and obtained very good performance in their integrated CAD system. Wang et al. [26] employed a modified U-Net structure to exploit global information for refining the mass coarsely localized by YOLO. Li et al. [27] presented conditional residual U-Net to boost the segmentation performance via incorporating probabilistic graphic models and deep residual learning. Wang et al. [28] integrated contextual information, low-level detail information and high-level semantic information to perform breast mass segmentation with their multi-level nested pyramid network. Hai et al. [29] introduced multi-scale image information into the fully convolutional dense network architecture to promote the improvement of breast tumor segmentation. Despite the promising results these FCNs-based methods have yielded, they are subject to a common property that all the pixel labels are independently predicted, which leads to spatial discontinuity in the output label maps [30]. This also explains why FCNs-based network architectures are often followed by a refinement model, such as CRF, even if its computational cost is rather expensive.

Inspired by generative multi-adversarial networks [31], this work explores a multi-scale adversarial network for X-ray breast mass segmentation. Specifically, an improved U-Net is trained to segment the mass in mammographic images, along with three critic networks that distinguish the ground truth and the output from segmentation network at multi-scale. In this multi-adversarial training, the model can readily reinforce higher-order consistencies between adjacent pixels without redundant complexity of the segmentation network. Additionally, weighted cross entropy loss is adopted to handle pixel class imbalance, and the spectral normalization technique is applied to all the critics to improve the training dynamics. Experiments have suggested that this network structure achieves competitive performance by driving the segmentation network to produce precise segmentation masks indiscernible to the discrimination network.

The main highlights of our work are as follows:

- We design a novel multi-scale adversarial network for breast mass segmentation in whole mammograms. Multi-adversarial learning not only enforces higher-order consistency in output results but also enables the introduction of multi-scale information.
- We formulate an integrated loss function that elaborately merges together Earth-Mover distance and weighted cross entropy, which mitigates pixel class imbalance in mammographic images while ensuring the quality of segmentation.
- Experiments of the proposed method on two public mammographic datasets: INbreast [32] and CBIS-DDSM [33] demonstrate that our method provides an effective and robust way to detect breast mass with high precision.

The subsequent sections in this paper are organized as follows: the details of our method are explicated in section 2, including overall network architecture, object function and training mode. Section 3 mainly provides the experimental results of the proposed network and comparisons with the established models. A discussion and some conclusions are presented in section 4 and 5 respectively.

II. METHODOLOGY

A. NETWORK ARCHITECTURE

GANs [34] has recently shown its outstanding advantages in image generative modeling. No need for trying to approximate intractable probabilistic computations, GANs is

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2999198, IEEE Access

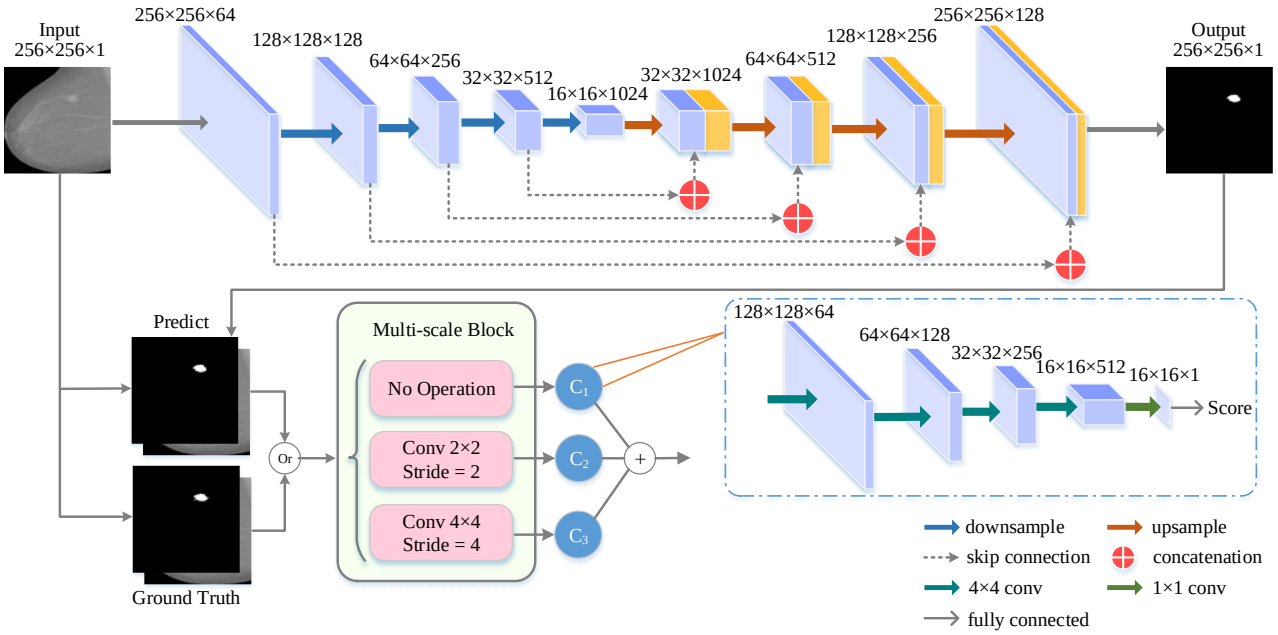


FIGURE 1. Proposed network architecture for breast mass segmentation. Three critics with identical structures discriminate against the masks from ground truths or generated by improved U-Net.

competent in generating realistic pictures with only a simple noise input. Besides, adding certain condition information, cGAN [35] has the ability to perform specific image-to-image translation between different image styles. In fact, the semantic segmentation of an image can be seen as a kind of special image-to-image transformation process in which original images are translated into the corresponding binary segmentation masks. The traditional image segmentation methods based on FCNs usually calculate a pixel-level loss function at the last layer of their network structure, which leads to the segmentation network unable to effectively retain the local and global context information between pixels in the output label map. Therefore, we migrate the GANs network framework to image semantic segmentation to enforce spatial continuity on the output mask. As shown in Figure 1, an improved U-Net acts as the generator in our network architecture, and the discrimination network is made up of deep convolutional neural networks. Considering the importance of multi-scale information for accurate segmentation of breast mass, we designed a generative multi-adversarial network, where three critic networks operate on three different image scales.

As the generator, we adopt the classic U-Net architecture that consists of a contracting path and an expansive path. The contracting path is mainly used to learn the context information in input images, and the expansive path is to accurately locate the position of the region to be segmented. On the contracting path, every encoder block is composed of two successive convolutional layers and one downsampling layer, as shown in Figure 2(a). After feature maps flow through each block, the number of their channels is doubled, except for the first block. In every step of the contracting path, we double the size of the feature maps in both length and width. On the other hand, the typical structure of the decoder blocks on the expansive path is two convolutional layers, followed by an upsampling layer, as shown in Figure 2(b) (note that all activation functions on the expansive path are Rectified Linear Unit (ReLU), except for the last one that is tanh activation function). When feature maps pass by the expansive path, the changes of their sizes and channels are opposite to those when flowing through the contracting path. To make the most of abundant details in the low-level feature maps, layers of the contracting section are incorporated into the corresponding layers on the expansive path by way of skip connection. Compared with the original U-Net network, we add a Batch Normalization layer after all convolution operations on the two paths to mitigate the risk of gradient disappearance. Meanwhile, since max-pooling operation is going to lose the partial position information of the pixels while increasing the receptive field, all downsampling operations are conducted by stride convolutions in our improved U-Net network.

For image segmentation tasks, both detailed texture and overall contour information are necessary to generate good segmentation results for multi-scale targets. For this sake, we do not input the segmentation results directly to the discrimination network but obtain their downsampled images at three different scales (original, one-half and one-quarter size) through a multi-scale block, and then feed them into three critics whose structures are topologically identical to each other. For a single critic, we simply employ a convolutional neural network containing five convolutional layers and a

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2999198, IEEE Access

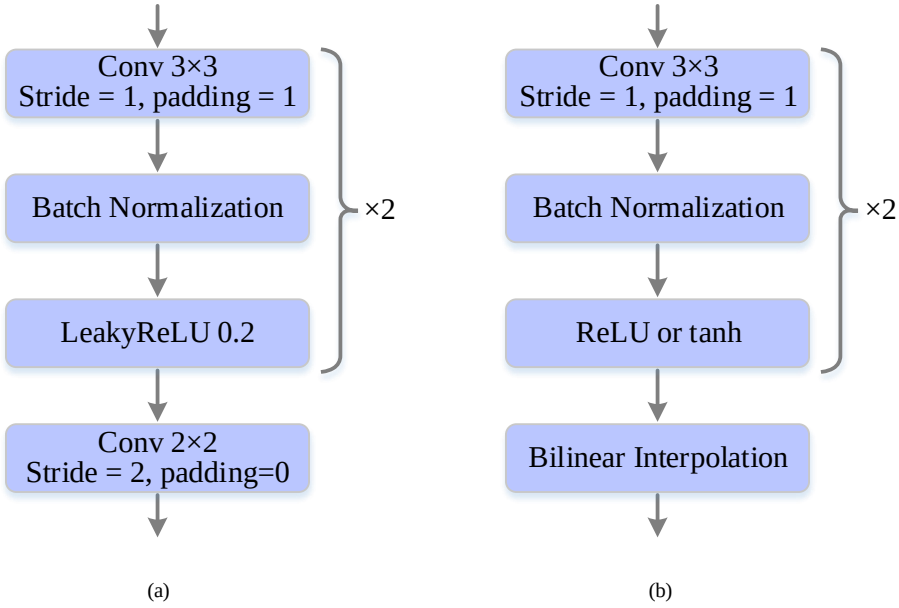


FIGURE 2. Typical structures of blocks in improved U-Net. (a) the structure of encoder block; (b) the structure of decoder block.

fully connected layer. Among the convolutions, the basic structure of the first four layers is a 4×4 convolution with stride 2 and padding 1, followed by a Leaky Rectified Linear Unit (LeakyReLU) with slope 0.2. The kernel size in the last convolutional layer is 1×1, for reducing the number of channels of feature maps. It is worth noting that, in spite of the same structure, all three critics jointly complete a coarse-to-fine discriminating process due to their different receptive fields. Figure 1 just illustrates the details of feature maps in the original-size critic, and the remaining two critics are analogous.

B. OBJECTIVE FUNCTION

In the vanilla GANs framework, the generator G is a mapping from the randomly sampled noise vector z to the image sample output $G(z)$. Then, the discriminator D classifies its input into the generated sample or the real one. The min-max game of GAN for image generation is given by the following expression:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1-D(G(z)))] \quad (1)$$

The objective of generator is to deceive the discriminator by pushing the generated distribution p_g as close as possible to the real data distribution p_{data}, while the goal of discriminator is to make correct judgement on the source of a sample and detect the fake samples. Though this adversarial training, the optimal generator is enabled to produce samples that the discriminator cannot distinguish.

In the proposed multi-scale adversarial network, the discrimination network still differentiates ground truth and synthesized data, but the task of segmentation network is changed. It not only needs to fool all critics, but also predicts the class of each pixel of input images. So we design the objective loss function involving two terms: adversarial loss and segmentation loss.

It is widely acknowledged that the training of GANs can be very unstable. In order to alleviate this training instability, in the adversarial loss term, we take advantages of the earth-mover distance in Wasserstein GAN(WGAN) [36] to measure how close the synthesis distribution and the real distribution rather than Jensen-Shannon divergence in conventional GANs. Then, the first term in our loss function can be formulated as:

$$\mathcal{L}_{GAN}(S,C_k) = \mathbb{E}_{x,y\sim p_{data}(x,y)}[C_k(x\circ y)] - \mathbb{E}_{x\sim p_{data}(x)}[C_k(x\circ S(x))] \quad (2)$$

In this adversarial loss function, x is the image to be segmented, and y is its corresponding ground truth mask. The binary operator $\circ$ donates pixel-wise multiplication between an original image and its mask (ground truth or segmentation result). S and C_k are function maps of the segmentation network and the kth critic, respectively.

Cross entropy loss is the most widely used loss function in traditional image semantic segmentation. This loss function examines each pixel separately, evaluates the class prediction for pixels, and then averages all predictions of them. In the breast mass image, the foreground regions account for extremely small proportions in the whole image, which leads to the training being biased to the background class with more pixels. As a result, it is difficult for the network to characterize the mass features, and its validity will be greatly reduced. To handle the problem arising from the serious class imbalance between masses and background, we take the

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2999198, IEEE Access

class frequencies into consideration, and employ a weighted cross entropy as the segmentation loss term as shown in the equation:

$$\mathcal{L}_{SEG}(S) = \mathbb{E}_{x,y\sim p_{data}(x,y)}[-\sum_{i=1}^{n}\frac{1}{f_{y_i}}(y_i\log(S(x)_i) + (1-y_i)\log(1-S(x)_i))] \quad (3)$$

where $S(x)_i$ and y_i represent the class label of pixel i in prediction mask $S(x)$ and in corresponding ground truth map y, respectively; f_{y_i} represents the class frequency of pixels whose ground truth label is equal to y_i, and n is the number of pixels in a label map.

We apply the mode of adversarial loss in WGAN to the three critics and endow their importance equally. To this end, our complete objective function is:

$$S^* = \arg\min_S(\max_{\|C_1,C_2,...,C_K\|_L\leq 1}\sum_{k=1}^{K}\mathcal{L}_{GAN}(S,C_k) + \lambda\mathcal{L}_{SEG}(S)) \quad (4)$$

where the Lipschitz constants of critics are all constrained to 1. K is the number of critics, and the coefficient λ balances the two loss terms.

C. ADVERSARIAL TRAINING

The segmentation network and discrimination network in our proposed adversarial network are trained in an alternating fashion. We first fix the segmentation network and simultaneously train three critics in the discrimination network to identify the source of its input by scoring. High scores mean the input label map is from the ground truths with a high probability, on the contrary, low scores represent it is more likely to be generated. Three critics operate at different image scales, and independently score the input segmentation mask. The critic operating at the coarsest scale has the largest receptive field while the one that operates at the finest scale is specialized in catching more image details. When the segmentation network is trained to fool the fixed discrimination network, the scores from the three critics are summed to give a comprehensive evaluation for the label map predicted by the segmentation network. The resulting gradient passes to the segmentation network, along with the segmentation loss to guide the update of the segmentation network parameters. In this min-max game, as the critics' recognition capacities are enhanced, the segmentation network has to learn the ability to yield segmentation results that are indistinguishable from the real labels.

Not like the technique of weight clipping in WGAN, we use spectral normalization in the critics to stabilize the training of our network. The spectral normalization limits the intensity of the function change by constraining the spectral norm of weight matrix at every layer of the critics to enforce the 1-Lipschitz continuity. In practice, it is unrealistic to do singular value decomposition for each layer in every training iteration, especially when the dimension of network weights is high. Thereby, we adopt a method called power iteration [37] to derive the approximate solution of singular values. In addition, to train the segmentation network and the discrimination network with 1:1 balanced updates for improving convergence speed, we follow the two-timescale update rule (TTUR) [38] to utilize an imbalanced learning rate for their updates. The training procedure is described in Algorithm 1.

Algorithm 1 Mini-batch gradient back-propagation training of our proposed multi-scale adversarial network.

Require: α_c, the learning rate for the critics. α_s, the learning rate for the segmentation network. m, the batch size. w_k, the kth critic's parameters. θ, the parameters of the segmentation network. K, the number of critics. L, the number of layers in one critic network.

while θ has not converged **do**
  Sample$(x^i, y^i)_{i=1}^m \sim p_{data}(x,y)$ a batch from the real data
  for $k = 1, ..., K$ **do**
    $g_{w_k} \leftarrow \nabla_{w_k}[\frac{1}{m}\sum_{i=1}^{m} C_{w_k}(x^i \circ y^i) - \frac{1}{m}\sum_{i=1}^{m} C_{w_k}(x^i \circ S_\theta(x^i))]$
    $w_k \leftarrow w_k + \alpha_c \cdot Adam(w_k, g_{w_k})$
    for $l = 1, ..., L$ **do**
      $\tilde{v}_k^l \leftarrow (w_k^l)^{\mathrm{T}}\tilde{u}_k^l / \|(w_k^l)^{\mathrm{T}}\tilde{u}_k^l\|_2$
      $\tilde{u}_k^l \leftarrow w_k^l\tilde{v}_k^l / \|w_k^l\tilde{v}_k^l\|_2$
      $w_k^l \leftarrow w_k^l / ((\tilde{u}_k^l)^{\mathrm{T}} w_k^l \tilde{v}_k^l)$
    end for
  end for
  Sample$(x^i, y^i)_{i=1}^m \sim p_{data}(x,y)$ a batch from the real data distribution
  $g_{\theta_a} \leftarrow -\nabla_\theta[\frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{K} C_{w_k}(x^i \circ S(x^i))]$
  $g_{\theta_s} \leftarrow -\nabla_\theta[\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{1}{f_{y_j}}(y_j^i\log(S(x)_j^i) + (1-y_j^i)\log(1-S(x)_j^i))]$
  $\theta \leftarrow \theta - \alpha_s \cdot Adam(\theta, g_{\theta_a} + \lambda \cdot g_{\theta_s})$
end while

III. EXPERIMENT

A. DATASETS AND DATA PREPROCESSING

We test the proposed method on two public and authoritative datasets, INbreast and CBIS-DDSM. The former is created by the Breast Research Group, INESC Porto, Portugal and contains 107 mammographic images with accurate segmentation masks. For CBIS-DDSM, we use a total of 762 representative images in this study, including both craniocaudal and mediolateral oblique views.

We do some data preprocessing on the original images instead of directly extracting the mass-centered image patch like most existing methods. For the whole mammography images, we remove the irrelevant background area in images and the corresponding region in their label maps (remove

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2999198, IEEE Access

the black background and the complete breast image is preserved), and then resize them to 256×256 pixels. Training a model with great quantities of annotated images is often one of the most direct and effective means of avoiding the over-fitting of deep neural networks, which inversely poses a challenge to those medical datasets with a limited volume. Thus, the data augmentation method is additionally performed on the smaller INbreast dataset, and the size of this dataset is expanded to 4 times via random rotation and mirroring transformation. At the same time, the arrangement order of images in INbreast is randomly scrambled to ensure that images with similar texture features (mainly new images transformed from the original images) are not all adjacent. Finally, for both CBIS-DDSM and INbreast dataset, we randomly divide experimental images into three parts: training set, validation set and test set in a ratio of 8:1:1.

B. IMPLEMENTATION DETAILS

Our network model is implemented in the Python programming language using PyTorch and performed on one NVIDIA Titan X Pascal GPU (12G) with a batch size of 8. The learning rates of critics and segmentation network are respectively initialized to 5e-3 and 1e-3, and they will be adaptively reduced by using the Adam optimizer in the training process. As a trade-off between the computation cost and the precision, the number of critics in the discrimination network $K = 3$, and the balance factor λ is set to 10.

C. EVALUATION METRICS

To evaluate the effectiveness of our proposed network, overall accuracy, sensitivity, specificity and dice similarity coefficient are used to quantitatively characterize the segmentation performance on the two datasets. The definitions of all these metrics are given as:

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \times 100\% \quad (5)$$

$$Sensitivity = \frac{TP}{TP + FN} \times 100\% \quad (6)$$

$$Specificity = \frac{TN}{TN + FP} \times 100\% \quad (7)$$

$$Dice = \frac{2 \times TP}{2 \times TP + FP + FN} \times 100\% \quad (8)$$

where TP, TN, FP, FN respectively donate the number of true positive, true negative, false positive and false negative pixel samples.

D. RESULTS ON INBREAST DATASET

We construct a comparison between the proposed model and several well-known segmentation methods on the INbreast dataset. In the current experiments, all the models are trained from scratch and Table 1 lists the experimental results.

It can be observed from Table 1 that, even very simple FCNs architecture such as FCN-8s can achieve a not too bad dice of 67.24% on INbreast dataset. With the addition of skip connection, U-Net provides 6.56% higher values than FCN-8s on dice metric, which indicates the significance of the low-level information for image semantic segmentation. This also is why we choose U-Net structure as the network backbone of our segmentation network. Benefiting from the adversarial training mode, cGAN outstrips the two aforementioned methods with an obvious improvement in terms of sensitivity and dice similarity coefficient. Although Sun et al. [39] got the state of the art result by incorporating attention mechanism, our proposed model gives better metric values, whether its dice or sensitivity. An interesting point is that no matter which model the results come from, both specificity and accuracy maintain at a very high level with values of over 98%. We infer the reason that these two metrics are susceptible to the number of background pixels that overwhelmingly outnumber mass pixels in the whole mammographic images.

TABLE 1. Segmentation performance of different methods on INbreast.

Methods	Dice	Sensitivity	Specificity	Accuracy
FCN-8s	0.6724	0.7169	0.9937	0.9852
U-Net	0.7380	0.7514	0.9935	0.9883
cGAN	0.7830	0.7960	**0.9965**	0.9916
Sun et al. [39]	0.7910	0.8080	-	-
Ours	**0.8164**	**0.8272**	0.9956	**0.9943**

In addition to the quantitative measurement, we also perform visual analysis on INbreast dataset, illustrating some randomly chosen segmentation results.

Figure 3 presents several breast mass segmentation results generated by the FCN-8s, U-Net and cGAN for qualitative comparisons on INbreast dataset. Obviously, cGAN produces more realistic and less noisy segmentation maps than FCN-8s and U-Net, because its training fashion enforces spatial contiguity in the output. When compared with other established models, the masks generated from our network are the closest to the ground truths, which suggests that our model is more powerful in edge preservation of breast masses in various sizes and shapes.

Additionally, we use box plots to depict data distributions of the models' metrics on INbreast dataset, as shown in Figure 4. In contrast to other models, our method has the highest medians with the lowest variances in terms of sensitivity and dice coefficient similarity. On the rest two metrics, our method still provides smaller value ranges when the median values of all models exceed 98%. Box plots show the effectiveness and robustness of the proposed model.

E. RESULTS ON CBIS-DDSM DATASET

Next, we provide another performance comparison on CBIS-DDSM dataset to confirm the proposed method's generalization for X-ray breast mass segmentation, and the results are listed in Table 2.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2999198, IEEE Access

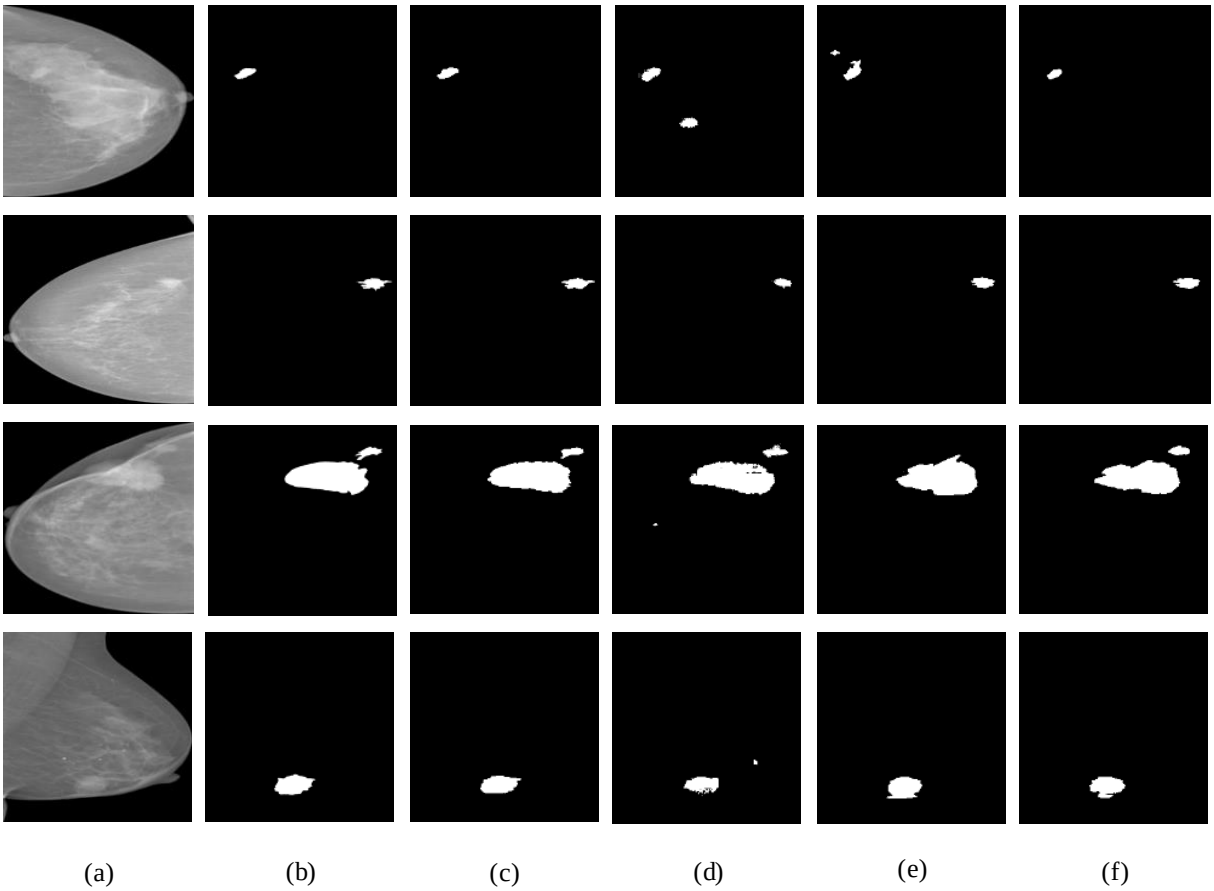


FIGURE 3. Segmentation results of four samples on INbreast dataset. From left to right, the network inputs (a), real label maps (b), and the segmentation results from the proposed model (c), FCN-8s (d), U-Net (e) and cGAN (f) are in order.

TABLE 2. Segmentation performance of different methods on CBIS-DDSM.

Methods	Dice	Sensitivity	Specificity	Accuracy
FCN-8s	0.6886	0.7387	0.9970	0.9852
U-Net	0.7219	0.7645	0.9973	0.9876
cGAN	0.7829	0.8220	0.9982	0.9977
Sun et al. [39]	0.8180	0.8490	-	-
Ours	**0.8216**	**0.8523**	**0.9986**	**0.9981**

As shown in Table 2, our proposed model outperforms the rest on all computed metrics, with very remarkable dice, sensitivity, specificity and accuracy (82.16%, 85.23%, 99.86% and 99.81%, respectively). Specifically, our model performs better than cGAN in terms of dice and sensitivity (+3.87% and +3.03%, respectively), which indicates that our method is more generalized in feature representation of breast mass. In addition, similar performance is obtained between this dataset and INbreast dataset with FCN-8s, U-Net, and cGAN. It is further proved that the encoder-decoder structure with skip-connection exhibits superior performance compared to FCN-8s, and adversarial training indeed promotes the quality of breast mass segmentation. On the other hand, metric values from the state-of-the-art breast mass segmentation model are also shown in Table 2. It can be seen that both sensitivity and dice in our model are higher than the method proposed by Sun et al. [39].

From a more intuitive view, Figure 5 lists some randomly chosen breast mass segmentation on CBIS-DDSM test set from our proposed model, FCN-8s, U-Net, and cGAN. It can be clearly seen that our method provides the best segmentation masks in various sizes and shapes on the CBIS-DDSM dataset. In particular, for the third listed sample (Row 3), when different degrees noises appear in the results of other methods, our method still accurately extracts the contour of the breast mass.

F. ABLATION STUDY

We perform ablation experiments on INbreast to validate the effectiveness of two key components in our model: the multi-scale block and weighted cross entropy loss. For multi-scale block, we ablate it to degrade our network into the cGAN architecture with only one critic (Ablation 1); for segmentation loss term, we replace weighted cross entropy with cross entropy (Ablation 2).

As a result, we obtain the middle 2 rows in Table 3

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2999198, IEEE Access

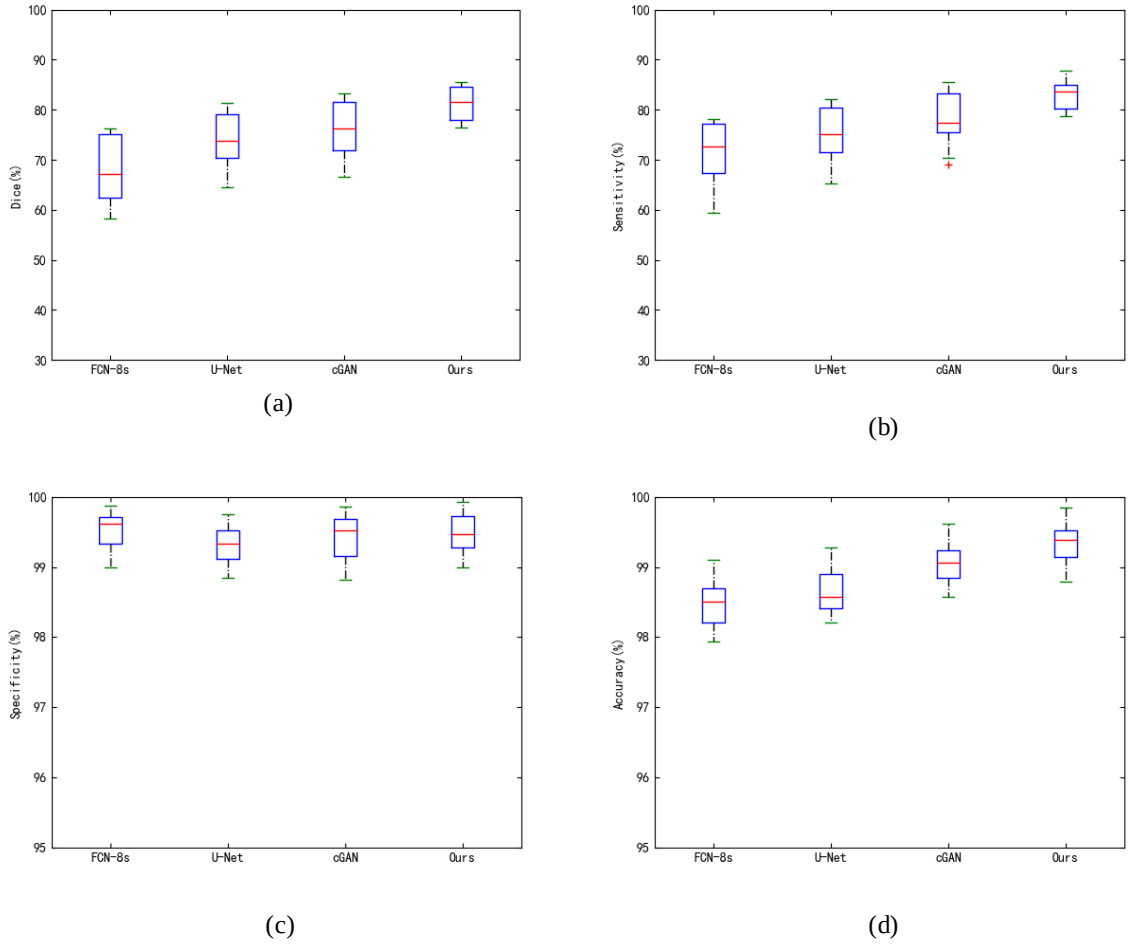


FIGURE 4. Box plots of (a) dice, (b) sensitivity, (c) specificity and (d) overall accuracy metrics of FCN-8s, U-Net, cGAN and our proposed model.

TABLE 3. Segmentation performance of ablation experiments on INbreast.

Methods	Dice	Sensitivity	Specificity	Accuracy
cGAN	0.7830	0.7960	**0.9965**	0.9916
Ablation 1	0.7992	0.8076	0.9943	0.9924
Ablation 2	0.8073	0.8130	0.9962	0.9933
Ours	**0.8164**	**0.8272**	0.9956	**0.9943**

(Ablation 1 and 2) which correspond to the two ablation models in turn. Except for specificity, two ablation models defeat cGAN on all metrics, but their overall performances are still inferior to our complete model. Furthermore, we find that the metrics of Ablation 2 entirely outstrip those of Ablation 1. Therefore, it is not hard to draw a conclusion that although both the multi-scale block and weighted cross entropy loss play positive roles, the introduction of multi-scale information may have a bigger effect than handling class imbalance in our proposed method.

IV. DISCUSSION

In recent years, the development of deep learning has been rapid, and FCNs-based methods represented by U-Net has almost dominated the vast majority of medical image segmentation tasks. However, when it comes to mass segmentation in the whole mammograms, as demonstrated in [30], the way these methods independently make one prediction for each pixel cannot explicitly capture the interactions between pixels. In this work, we transfer the adversarial learning framework to the image segmentation network, which is experimentally proven suitable for solving the problem of spatial discontinuity in label maps. Furthermore, we pay attention to two representative points when segmenting the suspicious regions: (1) breast masses present in various shapes and sizes in mammograms; (2) breast masses occupy only a comparatively small area in the whole images. To address these issues, multiple critics are introduced to enable the network to adapt breast masses of varying sizes and shapes, and weighted cross entropy loss is employed to relieve the class imbalance. Finally, we have testified the validity of the

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2999198, IEEE Access

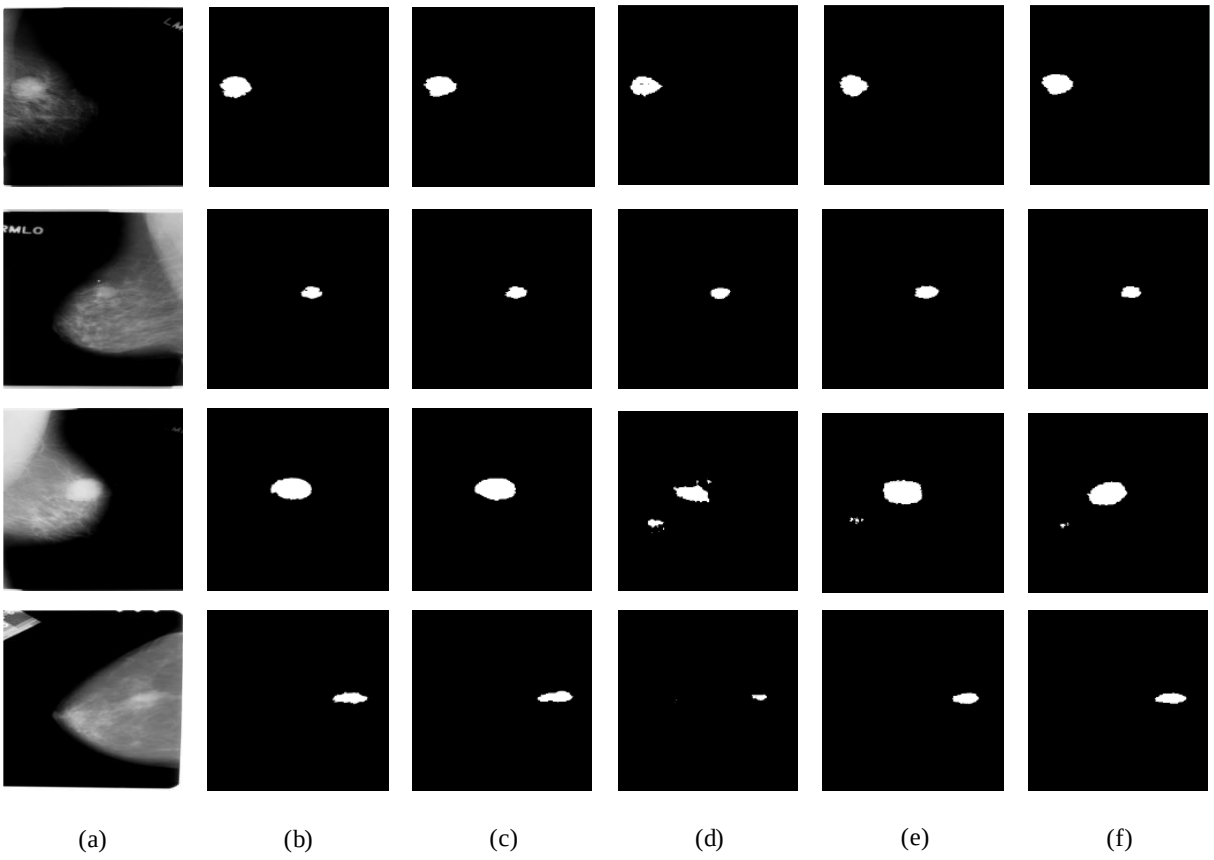


FIGURE 5. Segmentation results on four samples of CBIS-DDSM dataset. From left to right, the network inputs (a), real label maps (b), and the segmentation results from the proposed model (c), FCN-8s (d), U-Net (e) and cGAN (f) are in order.

two components in our ablation study, and both of them reach satisfactory results.

In one min-max game, the capability of segmentation network strengthens as the recognition of discrimination network enhances. By the light of this, we increase the number of critics to boost the discriminating ability while introducing multi-scale information. In our proposed method, we independently train each critic, and simply add the output scores of all the critics with the same weight as the final basis for discrimination. Although this has yielded competitive results, there will be a chance to exploit the maximum potential of individual critic in the discrimination network if some strategies in ensemble learning are applied to the combination of critics. We believe this will be a further improvement direction for multi-scale adversarial networks.

During the experiments, we find the proposed network performs better on CBIS-DDSM even if with additional data augmentations to INbreast dataset, which reflects a common problem of deep neural networks, namely dependence on a large amount of raw data. In addition, due to another three critics compared with conventional FCNs-based networks, our network will take up relatively large memory in training. Fortunately, a simple network is sufficient for the task of judging, so it is within acceptable limits.

V. CONCLUSION

In this paper, we propose a precise breast mass segmentation method via multi-scale adversarial network. Based on the existing research of FCNs in medical image segmentation, the original U-Net network is improved to make it more suitable for x-ray breast mass. The framework of GANs is applied to breast mass segmentation so as to reinforce higher-order consistency in final output masks and three critics correct the segmentation results at multiple scales. With regard to the loss function, the segmentation loss uses the weighted cross entropy loss to deal with the problem of class imbalance between masses and background, and the Wasserstein distance is functioned as the adversarial term to stabilize the training of GANs. Finally, spectral normalization is employed to restrict Lipschitz constants of each layer in the critic networks. The proposed method is verified on INbreast and CBIS-DDSM datasets, and it achieves better segmentation results than those existing methods. We hope that our work will make a difference to CAD systems and inspire the subsequent studies in medical image analysis.

REFERENCES

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," CA: a cancer journal for clinicians, vol. 69, no. 1, pp. 7–34, 2019.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2999198, IEEE Access

[2] C. E. DeSantis, J. Ma, A. Goding Sauer, L. A. Newman, and A. Jemal, "Breast cancer statistics, 2017, racial disparity in mortality by state," CA: a cancer journal for clinicians, vol. 67, no. 6, pp. 439–448, 2017.

[3] M. Al-antari, M. Al-masni, and Y. Kadah, "Hybrid model of computer-aided breast cancer diagnosis from digital mammograms," Journal of Scientific and Engineering Research (JSAER), vol. 4, no. 02, pp. 114–126, 2017.

[4] M. A. Al-antari, M. A. Al-masni, S.-U. Park, J. Park, M. K. Metwally, Y. M. Kadah, S.-M. Han, and T.-S. Kim, "An automatic computer-aided diagnosis system for breast cancer in digital mammograms via deep belief network," Journal of Medical and Biological Engineering, vol. 38, no. 3, pp. 443–456, 2018.

[5] F. S. S. de Oliveira, A. O. de Carvalho Filho, A. C. Silva, A. C. de Paiva, and M. Gattass, "Classification of breast regions as mass and non-mass based on digital mammograms using taxonomic indexes and svm," Computers in biology and medicine, vol. 57, pp. 42–53, 2015.

[6] S. Li, M. Dong, G. Du, and X. Mu, "Attention dense-u-net for automatic breast mass segmentation in digital mammogram," IEEE Access, vol. 7, pp. 59 037–59 047, 2019.

[7] D. Guliato, R. M. Rangayyan, W. A. Carnielli, J. A. Zuffo, J. L. Desautels et al., "Segmentation of breast tumors in mammograms using fuzzy sets," Journal of Electronic Imaging, vol. 12, no. 3, pp. 369–378, 2003.

[8] L. Kinnard, S.-C. Lo, P. Wang, M. T. Freedman, and M. Chouikha, "Automatic segmentation of mammographic masses using fuzzy shadow and maximum-likelihood analysis," in Proceedings IEEE International Symposium on Biomedical Imaging. IEEE, 2002, pp. 241–244.

[9] P. Rahmati, A. Adler, and G. Hamarneh, "Mammography segmentation with maximum likelihood active contours," Medical image analysis, vol. 16, no. 6, pp. 1167–1186, 2012.

[10] S. Timp and N. Karssemeijer, "A new 2d segmentation method based on dynamic programming applied to computer aided detection in mammography," Medical physics, vol. 31, no. 5, pp. 958–971, 2004.

[11] A. Cao, Q. Song, X. Yang, and S. Liu, "Breast mass segmentation on digital mammograms by a combined deterministic annealing method," in 2004 2nd IEEE International Symposium on Biomedical Imaging: Nano to Macro (IEEE Cat No. 04EX821). IEEE, 2004, pp. 1303–1306.

[12] W.-Y. Hsu, "Improved watershed transform for tumor segmentation: application to mammogram image compression," Expert systems with Applications, vol. 39, no. 4, pp. 3950–3955, 2012.

[13] A. Mencattini, G. Rabottino, M. Salmeri, R. Lojacono, and E. Colini, "Breast mass segmentation in mammographic images by an effective region growing algorithm," in International Conference on Advanced Concepts for Intelligent Vision Systems. Springer, 2008, pp. 948–957.

[14] E. Song, L. Jiang, R. Jin, L. Zhang, Y. Yuan, and Q. Li, "Breast mass segmentation in mammography using plane fitting and dynamic programming," Academic radiology, vol. 16, no. 7, pp. 826–835, 2009.

[15] W. Xie, Y. Li, and Y. Ma, "Breast mass classification in digital mammography based on extreme learning machine," Neurocomputing, vol. 173, pp. 930–941, 2016.

[16] S. Gu, Y. Chen, F. Sheng, T. Zhan, and Y. Chen, "A novel method for breast mass segmentation: from superpixel to subpixel segmentation," Machine Vision and Applications, vol. 30, no. 7-8, pp. 1111–1122, 2019.

[17] M. M. Saleck, A. El Moutaouakkil, and M. Rmili, "Semi-automatic segmentation of breast masses in mammogram images," in Proceedings of the International Conference on Pattern Recognition and Artificial Intelligence, 2018, pp. 59–62.

[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] Y. Cao, X. Ban, Z. Han, and B. Shen, "A new method for retinal image semantic segmentation based on fully convolution network," in National Conference of Theoretical Computer Science. Springer, 2018, pp. 27–45.

[21] X. Feng, C. Wang, S. Cheng, and L. Guo, "Automatic liver and tumor segmentation of ct based on cascaded u-net," in Proceedings of 2018 Chinese Intelligent Systems Conference. Springer, 2019, pp. 155–164.

[22] Y.-J. Huang, Q. Dou, Z.-X. Wang, L.-Z. Liu, L.-S. Wang, H. Chen, P.-A. Heng, and R.-H. Xu, "Hl-fcn: Hybrid loss guided fcn for colorectal cancer segmentation," in 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, 2018, pp. 195–198.

[23] V. Shreyas and V. Pankajakshan, "A deep learning architecture for brain tumor segmentation in mri images," in 2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP). IEEE, 2017, pp. 1–6.

[24] W. Zhu, X. Xiang, T. D. Tran, G. D. Hager, and X. Xie, "Adversarial deep structured nets for mass segmentation from mammograms," in 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, 2018, pp. 847–850.

[25] M. A. Al-antari, M. A. Al-masni, M.-T. Choi, S.-M. Han, and T.-S. Kim, "A fully integrated computer-aided diagnosis system for digital x-ray mammograms via deep learning detection, segmentation, and classification," International journal of medical informatics, vol. 117, pp. 44–54, 2018.

[26] J. Wang, C. Gou, T. Shen, and F.-Y. Wang, "Global segmentation-aided local masses detection in x-ray breast images," in 2018 Chinese Automation Congress (CAC). IEEE, pp. 3655–3660.

[27] H. Li, D. Chen, W. H. Nailon, M. E. Davies, and D. Laurenson, "Improved breast mass segmentation in mammograms with conditional residual u-net," in Image Analysis for Moving Organ, Breast, and Thoracic Images. Springer, 2018, pp. 81–89.

[28] R. Wang, Y. Ma, W. Sun, Y. Guo, W. Wang, Y. Qi, and X. Gong, "Multi-level nested pyramid network for mass segmentation in mammograms," Neurocomputing, vol. 363, pp. 313–320, 2019.

[29] J. Hai, K. Qiao, J. Chen, H. Tan, J. Xu, L. Zeng, D. Shi, and B. Yan, "Fully convolutional densenet with multiscale context for automated breast tumor segmentation," Journal of healthcare engineering, vol. 2019, 2019.

[30] P. Luc, C. Couprie, S. Chintala, and J. Verbeek, "Semantic segmentation using adversarial networks," arXiv preprint arXiv:1611.08408, 2016.

[31] I. Durugkar, I. Gemp, and S. Mahadevan, "Generative multi-adversarial networks," arXiv preprint arXiv:1611.01673, 2016.

[32] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "Inbreast: toward a full-field digital mammographic database," Academic radiology, vol. 19, no. 2, pp. 236–248, 2012.

[33] R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy, and D. L. Rubin, "A curated mammography data set for use in computer-aided detection and diagnosis research," Scientific data, vol. 4, p. 170177, 2017.

[34] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in Advances in neural information processing systems, 2014, pp. 2672–2680.

[35] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1125–1134.

[36] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein gan. arxiv preprint arxiv: 170107875," 2017.

[37] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," arXiv preprint arXiv:1802.05957, 2018.

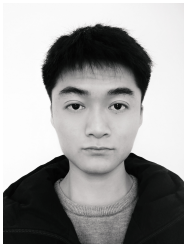
[38] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," in Advances in neural information processing systems, 2017, pp. 6626–6637.

[39] H. Sun, C. Li, B. Liu, Z. Liu, M. Wang, H. Zheng, D. D. Feng, and S. Wang, "Aunet: Attention-guided dense-upsampling networks for breast mass segmentation in whole mammograms," Physics in Medicine & Biology, 2019.

JUAN CHEN received her B.S., M.S., and Ph.D. degrees in computer science from Jilin University in 2000, 2003, and 2007, respectively. She is currently an associate professor in the College of Computer Science and Technology, Jilin University. Her current research interests are in the areas of deep learning, data mining and computing education.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2999198, IEEE Access



LIANGYONG CHEN received the B.S. degree from the College of Computer Science and Technology, Jilin University, in 2018, where he is currently pursuing a M.S degree. His current research interests include deep learning, adversarial learning, and medical image segmentation.

SHENGSHENG WANG received his B.S., M.S., and Ph.D. degrees in computer science from Jilin University in 1997, 2000, and 2003, respectively. He is currently a professor in the College of Computer Science and Technology, Jilin University. His current research interests are in the areas of computer vision, deep learning, and data mining.

PENG CHEN was born in 1976. She received the Ph.D. degree in medicine from Jilin University. She is currently a professor in The Second Hospital of Jilin University.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.