SAP-cGAN: Adversarial learning for breast mass segmentation in digital mammogram based on superpixel average pooling

Yamei Li and Guohua Zhao
School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China
Collaborative Innovation Center for Internet Healthcare, Zhengzhou University, Zhengzhou 450052, China

Qian Zhang
School of Computer Science, Zhongyuan University of Technology, Zhengzhou 450007, China

Yusong Lin
Collaborative Innovation Center for Internet Healthcare, Zhengzhou University, Zhengzhou 450052, China
School of Software, Zhengzhou University, Zhengzhou 450002, China
Hanwei IoT Institute, Zhengzhou University, Zhengzhou 450002, China

Meiyun Wang[a)]
Collaborative Innovation Center for Internet Healthcare, Zhengzhou University, Zhengzhou 450052, China
Department of Radiology, People's Hospital of Zhengzhou University, Zhengzhou 450003, China

(Received 24 September 2020; revised 11 December 2020; accepted for publication 11 December 2020; published 10 January 2021)

Purpose: Breast mass segmentation is a prerequisite step in the use of computer-aided tools designed for breast cancer diagnosis and treatment planning. However, mass segmentation remains challenging due to the low contrast, irregular shapes, and fuzzy boundaries of masses. In this work, we propose a mammography mass segmentation model for improving segmentation performance.
Methods: We propose a mammography mass segmentation model called SAP-cGAN, which is based on an improved conditional generative adversarial network (cGAN). We introduce a superpixel average pooling layer into the cGAN decoder, which utilizes superpixels as a pooling layout to improve boundary segmentation. In addition, we adopt a multiscale input strategy to enable the network to learn scale-invariant features with increased robustness. The performance of the model is evaluated with two public datasets: CBIS-DDSM and INbreast. Moreover, ablation analysis is conducted to evaluate further the individual contribution of each block to the performance of the network.
Results: Dice and Jaccard scores of 93.37% and 87.57%, respectively, are obtained for the CBIS-DDSM dataset. The Dice and Jaccard scores for the INbreast dataset are 91.54% and 84.40%, respectively. These results indicate that our proposed model outperforms current state-of-the-art breast mass segmentation methods. The superpixel average pooling layer and multiscale input strategy has improved the Dice and Jaccard scores of the original cGAN by 7.8% and 12.79%, respectively.
Conclusions: Adversarial learning with the addition of a superpixel average pooling layer and multiscale input strategy can encourage the Generator network to generate masks with increased realism and improve breast mass segmentation performance through the minimax game between the Generator network and Discriminator network. [https://doi.org/10.1002/mp.14671]

Key words: breast mass segmentation, generative adversarial network, multiscale features, superpixel pooling

1. INTRODUCTION

Breast cancer is one of the leading causes of cancer deaths in the world.[1] Early screening, diagnosis, and treatment are considered to be the main approaches to improving the survival rates of breast cancer. Mammography is currently the most reliable method for diagnosing breast cancer, and it is widely used to detect abnormalities in the breast.

Breast masses, in relation to microcalcifications and architectural distortion, are the most important type of breast cancer abnormalities because they represent a very high possibility of malignant tumors.[2] However, the interpretations of mass by radiologists are subject to substantial inter- and intra-observer variations, which may lead to missed cancers as well as overdiagnosis.[3] Therefore, computer-aided diagnosis (CAD) systems are now widely used to assist radiologists as a second reader in detecting and diagnosing breast masses.[4] An important tasks of breast CAD systems is to accurately segment masses from breast images. However, detecting and segmenting masses are challenging because they have a low contrast, irregular shapes, and fuzzy boundaries.[5]

Mass segmentation is a decisive process in the breast cancer CAD system. The accuracy of mass classification is

highly dependent on automatic segmentation algorithms. In addition, the automatic segmentation of a mass is essential for further quantitative and qualitative analyses. Some loose correlations exist between the mass morphology information obtained from segmentation results and the molecular subtype of breast cancer,[6,7] which is the key to identifying the best oncological treatment for breast cancer.

In this work, we propose an improved conditional generative adversarial network (cGAN) for breast mass segmentation in digital mammograms to solve the above problems. We introduce a superpixel average pooling layer into the Generator (G) network of cGAN, which utilizes superpixels as a new pooling layout to improve boundary segmentation. In addition, we adopt a multiscale input strategy in the G and Discriminative (D) networks to promote the networks to learn scale-invariant features with increased robustness. Experiments show that this network structure can promote G to generate masks with increased realism and improve the mass segmentation performance through the minimax game between G and D.

The major contributions of this work are as follows:

- We propose SAP-cGAN, a novel mammography mass segmentation architecture that utilizes superpixels as a new pooling layout to form a superpixel average pooling layer and joints it to the cGAN. This work also provides the first combination of the superpixel average pooling layer and the cGAN for use in mammogram images and effectively improves the performance of breast mass segmentation.
- We also adopt a multiscale input strategy that promotes the network to learn robust features of different scales to solve the recognition problem caused by the inconsistent sizes of masses.
- We use two public datasets for the quantitative and qualitative evaluation of the performance of the proposed SAP-cGAN architecture, which outperforms the current state-of-the-art breast mass segmentation methods.

2. RELATED WORK

Given the complexity of the morphology and boundary of breast masses, accurate segmentation has become a difficult and hot issue in breast cancer CAD systems.[8] Traditional methods, such as conditional random field (CRF)[9] and active contour model[10] were widely used in early studies on breast mass segmentation. Dhungel et al.[11] adopted CRF and structured support vector machine to explore the use of deep belief networks to segment breast masses in structured prediction models. Kozegar et al.[12] considered the seed position as the only prior condition and proposed a two-stage segmentation method that combines the shape information of a mass. The adaptive region growing algorithm is used to estimate the mass boundary roughly, and a new geometric edge-based deformable model is introduced to refine the mass boundary. However, the setting of the initial parameters of these traditional methods relies excessively on experience. This reliance usually leads to subjective diagnosis. In addition, most of these methods cannot realize the end-to-end automatic segmentation of a mass, and manual intervention is often required during segmentation.

Since the full convolutional network (FCN)[13] was first proposed, semantic segmentation methods have made great progress in deep learning and have been widely used in medical image segmentation. Li et al.[14] combined densely connected U-Net with attention gates and proposed a fully automatic breast mass segmentation method based on deep learning. The method presented better segmentation performance than U-Net, attention U-Net, and DenseNet. Cheng et al.[15] used a spatial enhanced rotation aware network to avoid mass misclassification in the background area to facilitate separating a mass from a complex background. Thus far, different extensions of deep learning have been applied in the detection, segmentation, and classification of breast masses.[16–18] The results obtained by these methods are more accurate than those obtained by traditional methods. However, network training requires additional pixel-level annotation data and training costs.

Dong et al.[19] used the GAN[20] network for medical image segmentation for the first time. The image-to-image method improves the accuracy and computational efficiency of segmenting the liver from three-dimensional (3D) computed tomography (CT) scans. This method can perform powerful learning by using less data than traditional deep learning segmentation networks. Pang et al.[21] proposed a new framework based on GAN, CTumorGAN, which achieved superior segmentation performance on the CT datasets of various diseases. Singh et al.[22] proved the advantages of applying the modified GAN model (cGAN) to the mass segmentation task in digital mammograms.

Superpixels are a popular method that combines spatial priors with various computer vision problems.[23] Superpixel generation is an over-segmentation process that can retain highly detailed boundary information. We have reason to believe that this characteristic is helpful for detecting masses with unclear and spiculated boundaries. Kwak et al.[24] added a superpixel pooling layer to the classification network and achieved remarkable performance in the weakly supervised semantic segmentation task of natural images. In recent years, superpixel pooling has been extensively studied in weakly supervised segmentation or natural image classification.[25,26] As far as we know, however, it has not been applied for the segmentation of breast masses in digital mammograms.

3. MATERIALS AND METHODS

3.A. Problem formulation

Assume that kth mass image I_k is present in digital mammogram datasets, and M_k is the corresponding segmentation ground truth (a binary mask of the mass area). I_k and M_k form the paired training data $\sum_{i=1}^{n}(I_k, M_k)$. The

goal of the GAN is to learn the mapping from the random noise vector z to the output mask image M, that is, $z \rightarrow M$. By contrast, the cGAN learns a mapping from the observed image I (conditional information) and random noise vector z to M,[27] $\{I,z\} \rightarrow M$. The addition of conditional information prompts the network to generate the output $\hat{M}_k$ related to the input image I_k.

The cGAN framework based on the superpixel average pooling proposed in this work, namely, SAP-cGAN, is shown in Fig. 1 : SAP-cGAN is composed of two parts: G and Discriminator D. G is trained to generate a mask image that is similar to the ground truth to confuse D, whereas D tries its best to distinguish the generated image $\hat{M}_k$ from the ground truth M_k. In such a minimax game, our goal is to improve the accuracy of mass segmentation by making the mask image generated by G increasingly similar to the ground truth segmentation mask and making D work hard to identify the "fakes" generated by G.

3.B. Superpixel average pooling layer

We use simple linear iterative clustering (SLIC) to generate superpixels in the mass images. It can iteratively cluster the pixels on the basis of local structural features and the spatial relationship between the pixels and merge the pixels with similar low-level features.[23]

3.B.1.. Superpixel generation

Input mass image $I \in R^{W \times H}$ is denoted, where W and H denote the width and height, respectively. Q cluster center E_i, $i \in \{1, 2, 3,\ldots,Q\}$ is initialized uniformly on the mass image I. E_i is moved to the lowest gradient position that is in a 3×3 neighborhood to prevent the cluster center from falling on the edge of the image. For each pixel p in the $2K \times 2K$ neighborhood of E_i ($K = \sqrt{P/Q}$, P is the total number of pixels in I), the distance D_{E_ip} between E_i and p is calculated by Eq. (1), and the pixel p is classified to the corresponding E_i in accordance with the minimum distance principle.

$$D_{E_ip} = \sqrt{{d_{gray}}^2 + \left(d_{space}/K\right)^2 m^2}, \tag{1}$$

where $d_{gray} = \sqrt{(I_i - I_j)}$ represents the distance in gray space; I_i and I_j are the gray values of the two pixels, respectively; $d_{space} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}$ denotes the Euclidean distance; and m is a tightness parameter that controls the relative relationship between gray-scale similarity and spatial similarity.

Superpixel generation is an oversegmentation process, and the mass images will be divided into several superpixels. The pixels inside each superpixel have similar structural features. Therefore, the superpixel average pooling layer is used to reduce the dimensionality of the feature map inside the superpixel to the same features. After the feature map is subjected to mapping and $\times$ reduction through the superpixel average pooling, internal redundant information can be removed such that the network focuses on the information difference between classes. This approach will be helpful for locating masses with irregular shapes or fuzzy boundaries.

3.B.2.. Superpixel average pooling layer

The process of superpixel average pooling is shown in Fig. 2. After a series of operations such as convolution,

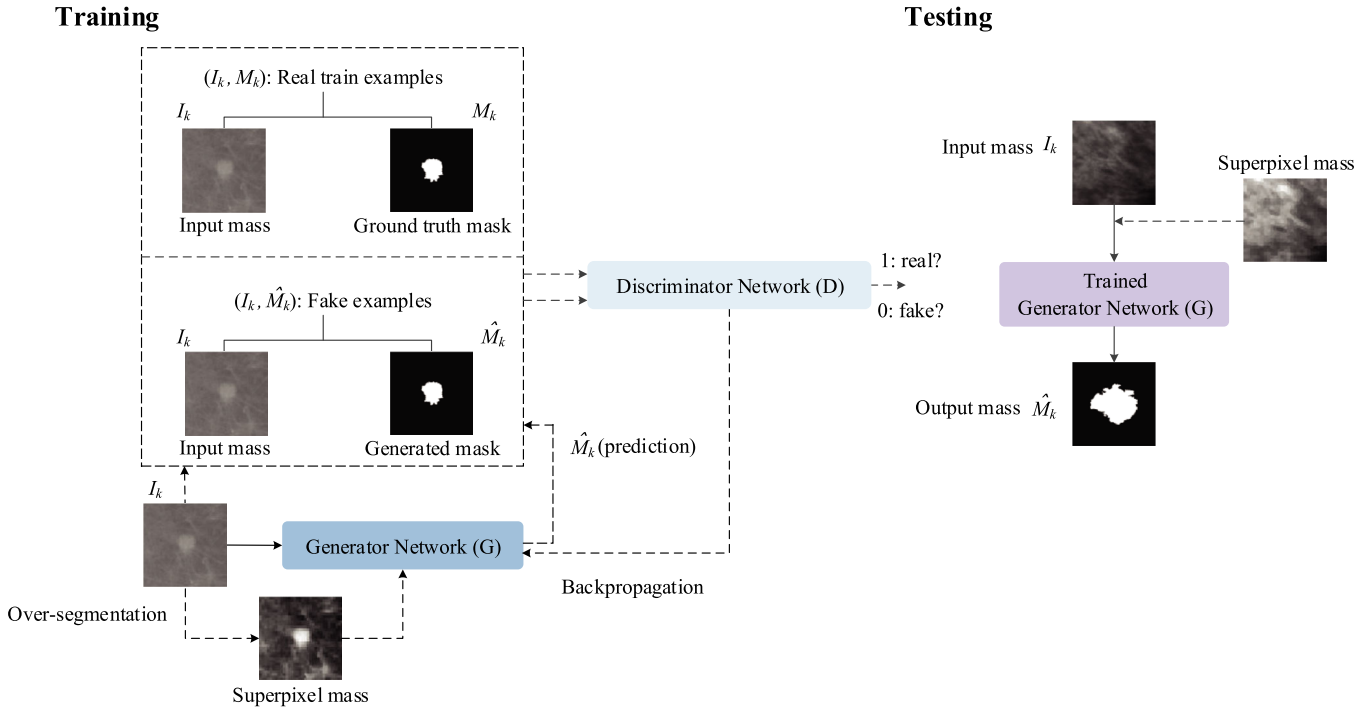

FIG. 1. Proposed SAP-cGAN framework for mass segmentation in digital mammogram. [Color figure can be viewed at wileyonlinelibrary.com]

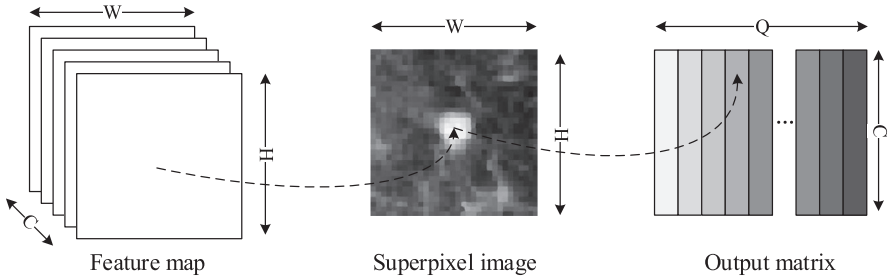

FIG. 2. Illustration of the superpixel pooling layer.

pooling, and upsampling, the mass image I obtains the feature map $X \in R^{W \times H \times C}$, where W and H represent the width and height of the image, and C represents the number of channels. The oversegmentation result generated by the SLIC is a single-channel map $S \in L^{W \times H}$, where $L = [1,Q]$ indicates the label of each superpixel, and Q is the number of superpixels. $S_i = Q$ indicates that the pixel i belongs to the Qth superpixel. As shown in Fig. 2, the output Y of the superpixel pooling layer is a matrix with size $C \times Q : Y \in R^{C \times Q}$ that can be denoted as

$$Y_{c,q} = avg\{X_{c,i} | i : S_i = Q\}, \tag{2}$$

where $avg\{\cdot\}$ is the average pooling function.

In backward propagation, the gradient calculation formula for superpixel average pooling layer is

$$\frac{\delta Y_{c,q}}{\delta X_{c,i}} = \begin{pmatrix} \frac{1}{N(S_i)}, & if \ S_i = Q \ and \ c' = c \\ 0 \ , & othewise \ , \end{pmatrix} \tag{3}$$

where $N(S_i)$ is the number of pixels in the superpixel with label Q.

The superpixel average pooling layer is very similar to the average pooling of the neural network. The difference is that the superpixel average pooling layer is performed in the irregular region that is determined by the superpixel shape. The superpixel map contains the original low-level shape prior information of the mass image, which has a good effect on refining the boundary of the irregular mass.

3.C.. The SAP-cGAN architecture

3.C.1.. Generator network

Figure 3 shows the suggested architectures for G. In G, we improve the cGAN architecture used in the work of Singh et al.[22] For the encoder of G, we adopt a multiscale input strategy[28] to obtain highly effective and scale-invariant feature representation such that the network can exhibit improved performance when locating masses with different sizes. The input images are resized to 1/8, 1/4, and 1/2 of its original size, and a convolution layer with 3×3 kernels is used to extract features from each scale. Given that the feature maps extracted from different scales have different sizes, they must be upsampled to the size of the feature map extracted from the original image before aggregating multiscale features. After all the features are aggregated, they are input into the encoder.

The detailed settings of the encoder and decoder are shown in Fig. 3. In addition to the last decoder layer (DF8), skip connections are added between the encoder and decoder layers such that the decoder can combine the features of the lower layers when upsampling to avoid losing necessary detailed information. In addition, dropout is applied at the first three decoding layers (DF1, DF2, and DF3) to avoid overfitting. The feature map gradually recovers to the size of the original image after multiple upsampling in the decoder. The superpixel average pooling layer described in Section 3.B is combined with the DF8 layer. The DF8 layer is matched with the input superpixel map of the superpixel average pooling layer for calculation. The superpixel average pooling layer aggregates features inside of each superpixel by exploiting an input superpixel map as the pooling layout. According to Eq. (2), the output of the superpixel average pooling layer becomes a $C \times Q$ matrix ($C = 32$ in the current SAP-cGAN architecture). A single $C \times 1$ feature vector is assigned to each superpixel which will be classified by the fully connected classification layer following the superpixel average pooling layer. After calculating the classification score of each superpixel, the tensor of $W \times H \times 1$ is obtained. The tanh activation function is then applied to generate a binary segmentation mask of the breast mass.

3.C.2.. Discriminator network

The architecture of D is shown in Fig. 4. As shown in Fig. 4, D is composed of five encoding layers. The input of D is the original mass image, the ground truth mask, and the "fake" mask generated by G. Similar to the multiscale input of the G, the mass images and the masks are resized into the same three scales. Sigmoid is applied at the last layers, and the output of the network is 0.0–1.0, which represents the degree of complete credibility (0.0 represents completely fake).

In accordance with the symbols presented in Section 3.B, the loss function of G is

$$L_G = E_{I,M,z}(-\log(D(I, G(I,z)))) + \lambda E_{I,M,z}(L_{Dice}), \tag{4}$$

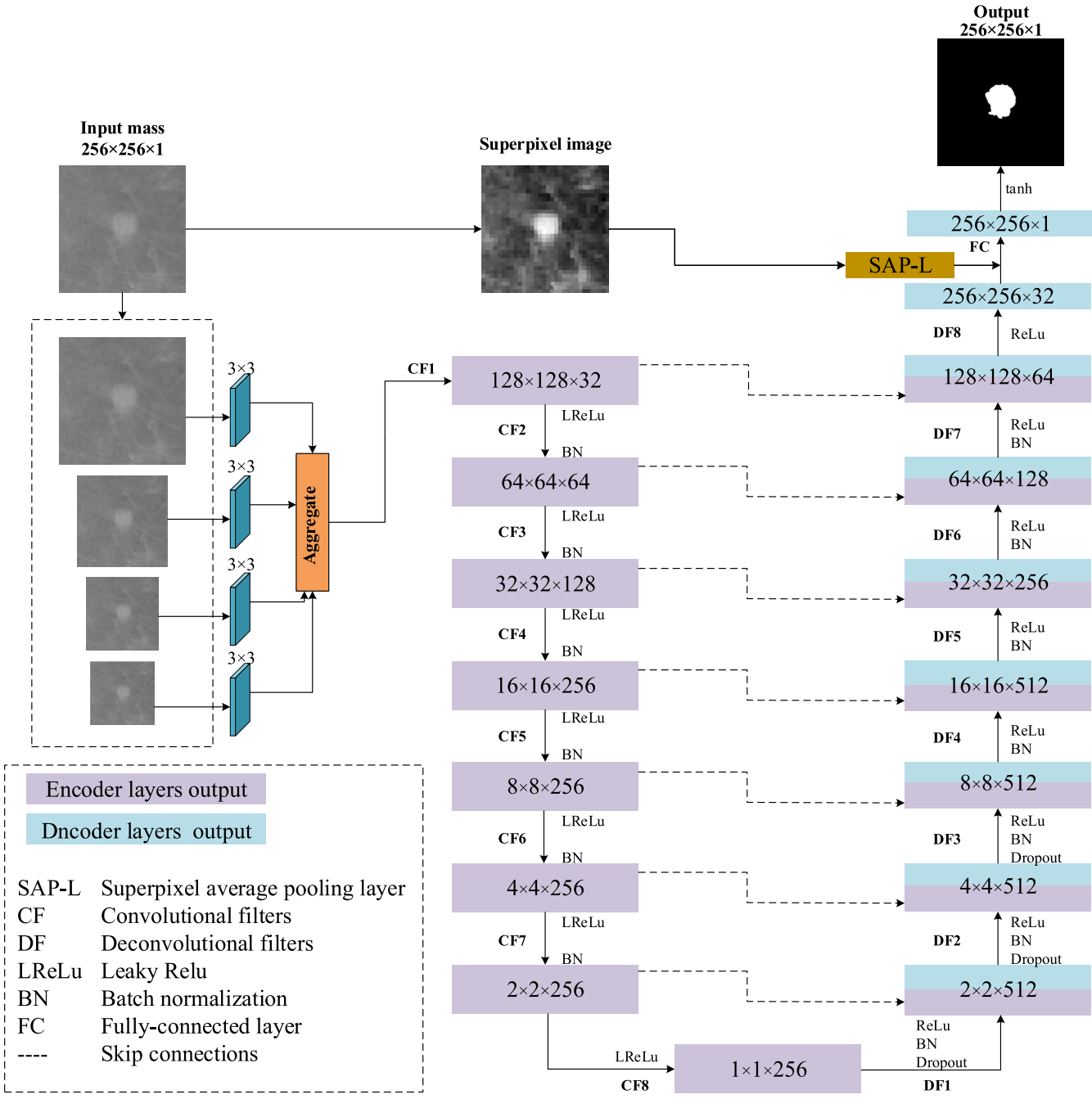

FIG. 3. Generator network architecture of the proposed SAP-cGAN. [Color figure can be viewed at wileyonlinelibrary.com]

where $D(I,G(I,z))$ and $G(I,z)$ are introduced as the output of D and G, λ is an empirical weighting factor, and L_{Dice} is the Dice loss[29] between M and $\hat{M}$.

The loss function of the D is

$$L_D = E_{I,M,z}(-\log(D(I,M))) + E_{I,M,z}(-\log(1 - D(I,G(I,z)))). \quad (5)$$

When optimizing G and D, the optimizer makes G work hard to generate a mask similar to $\hat{M}$, and D attempts to distinguish the "fake" $\hat{M}$ among them. During this adversarial learning process, the network will output a binary mask with increased clarity and realism.

3.D.. Dataset and processing

3.D.1.. CBIS-DDSM

The CBIS-DDSM dataset is the curated breast imaging subset of DDSM,[30] which contains mass and calcification data. It consists of 861 mass cases that including the mediolateral oblique (MLO) and cephalo-caudal (CC) views of the mammograms. Among the images in the dataset, 912 images are marked as benign and the remaining 784 are marked as malignant. The location and type of suspicious areas in the mass image are marked as the ground truth by experienced radiologists. The CBIS-DDSM is divided into training and

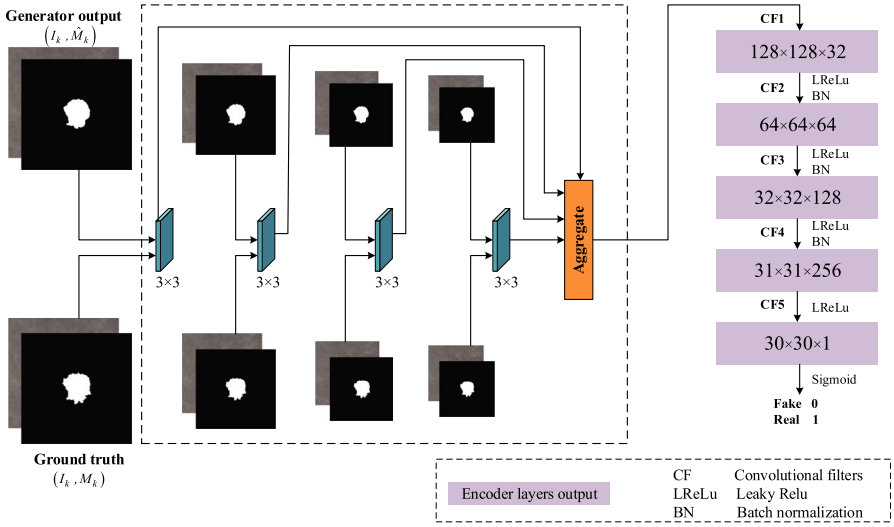

FIG. 4. Discriminator network architecture of the proposed SAP-cGAN. [Color figure can be viewed at wileyonlinelibrary.com]

testing subsets[31]; the former consists of 1318 images, and the latter contains 378 images.

3.D.2.. INbreast

The INbreast dataset was created by the Breast Research Group, INESCPorto, Portugal.[32] It contains images of 115 patients for a total of 410 images, including images of masses, calcifications, and other abnormalities. For mass segmentation, it specifically provides 106 images with corresponding masks. Therefore, we select these 106 images as the external validation data of the proposed model.

Similar to other researchers working on mass segmentation,[33] we extracted 256 × 256 image regions centered at masses in the two datasets to build our dataset. The affine transformation is used to rotate the training images by 90° and 180° and to flip the images along the horizontal axes to augment the dataset to enable the network to learn features with increased richness and avoid overfitting. That is, the training dataset contains 5272 mass images. The two test sets comprise 378 mass images from CBIS-DDSM and 106 mass images from INbreast.

The prepared data are then fed into SAP-cGAN, and a binary mask of the mass is generated during adversarial learning between G and D. We apply morphological operations to postprocess the mask to remove small speckles. As shown in Fig. 5, small speckles surrounded by red boxes in the generated mask are present. These speckles can be filtered out after morphological operations.

3.E.. Experiment and parameter settings

We quantitatively and qualitatively analyze the proposed method to evaluate the effectiveness of our method comprehensively. For the quantitative analysis, FCN,[13] U-Net,[34] SegNet[35] and other two mass segmentation models (called SegModel_CRF[11] and SegModel_cGAN[22]) are compared with the proposed SAP-cGAN method. We implement these models for comparison and retain the experimental setup in their original studies. In addition, we compare our method with the previously described state-of-the-art mass segmentation methods.[36,37] We also perform ablation analysis on SAP-cGAN to verify the effectiveness of the superpixel average pooling layer and multiscale input strategy integrated in SAP-cGAN. On the basis of the original cGAN, different components are sequentially added until SAP-cGAN is reached. Before the components are added, the results of the evaluation metrics are recorded. In qualitative analysis, we visualize and analyze the segmentation results of SAP-cGAN and the above five models.

All experiments are conducted with Tensorflow framework version 1.15. When using SLIC to generate superpixels, the Dice coefficient and Jaccard index are applied to measure the value of Q. The best effect is obtained when $Q = 600$.

FIG. 5. Postprocessing example. [Color figure can be viewed at wileyonlinelibrary.com]

The network is optimized through Adam with $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The batch size is 8, and the initial learning rate is 0.0001. After 120 epochs of training, G and D obtain the best results. For the ablation analysis, we adopt the same initial learning rate and Adam to optimize the network during the training process.

3.F.. Evaluation metrics

Accuracy, the Dice coefficient, the Jaccard index (i.e., intersection over union), sensitivity, and specificity are used as evaluation metrics to verify the effectiveness of the proposed method quantitatively.[38] For mass segmentation, let G be the ground truth mask of the mass image, and R be the mask predicted by the network. Then, the true positive (TP) refers to the mass region in G and R, that is, TP = G∩R, whereas true negative (TN) is TN = $\bar{G}\cap\bar{R}$. The false positive (FP) is defined as FP = $\bar{G}$∩R, which indicates the mass region included in R that does not belong to G. By comparison, false negative (FN) is defined as FN = G∩$\bar{R}$, which denotes the real mass region missed by the model.

3.G.. Statistical analysis

The differences between results are assessed using Student's *t*-tests or Mann-Whitney *U*-tests for continuous variables, as appropriate. The above statistical analyses are performed with SPSS Statistics software, version 18.0. The two-tailed threshold of $P<0.05$ was considered statistically significant.

4. RESULTS

4.A.. Comparison with Other methods

As shown in Table I, the quantitative results of SAP-cGAN with the testing datasets from CBIS-DDSM and INbreast are compared with those of other state-of-the-art models. The results for the CBIS-DDSM show that our method SAP-cGAN outperforms the compared models in terms of all evaluation metrics ($P < 0.05$). The Dice coefficient and Jaccard index, which are two important metrics for evaluating segmentation results, have been obviously improved and are 93.37% and 87.57%, respectively. Our model has also achieved high accuracy, specificity, and sensitivity values of 98.35%, 98.46%, and 97.50%, respectively. This finding shows that the superpixel average pooling layer and the multi-scale input strategy promote the network to learn additional discriminative features in the mass images. Our proposed SAP-cGAN architecture can improve the mass segmentation performance in digital mammograms.

The results of the mass segmentation performance on INbreast show that the performances of U-Net, SegModel_CRF and SAP-cGAN significantly decreased ($P < 0.05$) because the model is trained in CBIS-DDSM and the INbreast dataset is not used to fine-tune the network parameters during testing. Although the performance of SegModel_CRF on CBIS-DDSM are better than that of FCN ($P < 0.05$), it has the lowest Dice and Jaccard scores on INbreast, indicating that this model cannot learn generalized features of masses. However, the decrease in Dice coefficient, Jaccard index, and other metrics of SAP-cGAN for INbreast is significantly greater than that for CBIS-DDSM ($P < 0.05$). Thus, excellent segmentation performance is retained. The U-Net obtains the highest specificity (99.05%), but its sensitivity is lower than the sensitivity of SAP-cGAN (16.92%), indicating that it misses more real mass regions than the proposed SAP-cGAN.

In addition, Table I presents a comparison between our method and the state-of-the-art mass segmentation methods. We directly list the metrics of the methods that were presented in their original articles because their source codes have not yet been published. Our method outperforms other models in terms of accuracy, sensitivity, and specificity.

TABLE I. Comparison of SAP-cGAN with other mammogram mass segmentation methods.

Dataset	Methods	Dice	Jaccard	Accuracy	Specificity	Sensitivity
CBIS-DDSM	FCN	0.8476	0.7356	0.9340	0.9464	0.8865
	U-Net	0.8748	0.7774	0.9151	0.9755	0.8104
	SegNet	0.8435	0.7293	0.9265	0.9187	0.9564
	SegModel_CRF	0.8640	0.7606	0.9496	0.9584	0.9083
	SegModel_cGAN	0.8947	0.8094	0.9608	0.9641	0.9453
	SAP-cGAN	**0.9337**	**0.8757**	**0.9835**	**0.9846**	**0.9750**
INbreast	FCN	0.7758	0.6338	0.9296	0.9553	0.7887
	U-Net	0.8868	0.7967	0.9279	**0.9924**	0.8079
	SegNet	0.8292	0.7083	0.9528	0.9540	0.9442
	SegModel_CRF	0.6813	0.5167	0.8762	0.8797	0.8566
	SegModel_cGAN	0.8935	0.8075	0.9305	0.9769	0.8428
	SAP-cGAN	0.9154	0.8440	**0.9855**	0.9863	**0.9771**
	Tianyu et al.[37]	0.8881	0.8043	0.9145	0.9155	0.9200
	Alantari et al.[36]	**0.9269**	**0.8637**	0.9297	0.9321	0.9272

Bold indicates the highest value under each evaluation metric.

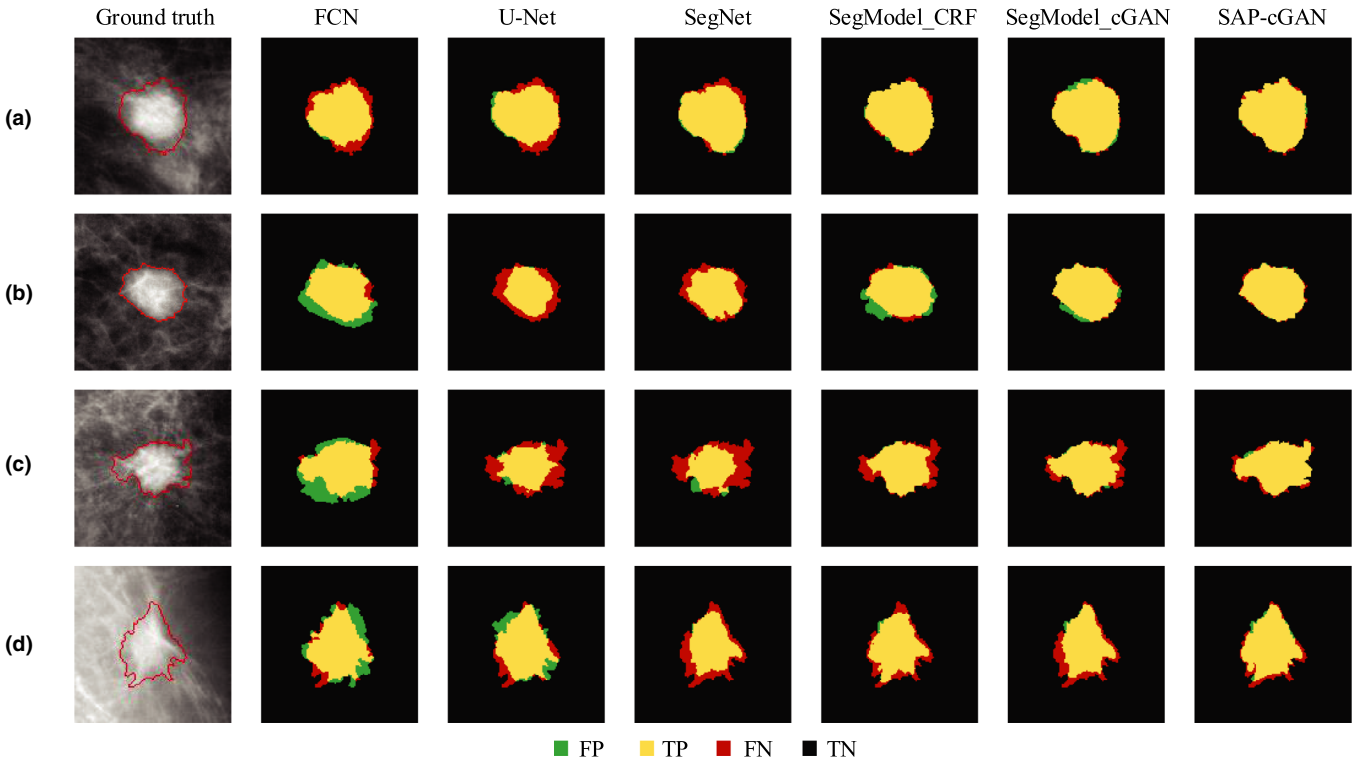

FIG. 6. Segmentation result of six models with the CBIS-DDSM dataset. (a) Round mass, (b) oval mass, (c) irregular mass, (d) lobular mass. From left to right: original images and the results of the FCN, U-Net, SegNet, SegModel_CRF, SegModel_cGAN, and SAP-cGAN models. [Color figure can be viewed at wileyonlinelibrary.com]

However, our model does not provide the highest Dice and Jaccard scores, which are 1.15% and 1.97% lower, respectively, than the scores of the work of Al-antari et al.[36] This difference is attributed to our use of the INbreast dataset as an external testing set, whereas the model in the work of Al-antari et al.[36] is trained on the INbreast dataset. Moreover, the results of the previous model are only slightly better than those of our model, showing that our model has a certain universality.

Figure 6 shows a comparison among the performances of our model and other five segmentation models on the same mass image from the CBIS-DDSM testing set. The round, oval, irregular, and lobular masses shown in Figure 6 are used to reflect the segmentation results of the models for masses with different shapes. The shape label is provided by the CBIS-DDSM. For each model, we compare the segmentation results with the ground truth mask, and mark TP(yellow), FP (green), TN(black), and FN(red) in different colors. As shown in Fig. 6, the FP and FN of our proposed method are obviously lower than those of the other five models. All five models show relatively good performance for regular masses, such as round and oval masses. However, although our model still shows the best performance for irregular and lobular masses, its FN and FP values have increased. The wrong area is mainly located near the border of the mass. This result is due to the rough and fuzzy borders of irregular and lobular masses that complicate accurate segmentation. Nevertheless, compared with other models, our model has evidently reduced FP and FN values. This characteristic also reduces the possibility of the misclassification of the boundary region. This result indicates that our model can achieve high segmentation accuracy for masses with relatively smooth boundaries and can improve the segmentation of irregular and fuzzy masses.

TABLE II. Ablation analysis of the performance of the proposed method on CBIS-DDSM.

Models	Dice	Jaccard	Accuracy	Specificity	Sensitivity
BaseLine	0.8557	0.7478	0.9319	0.8941	0.9676
BaseL + MS	0.8674	0.7659	0.9341	0.9787	0.8751
BaseL + SAP	0.9076	0.8308	0.9276	0.9442	0.8678
SAP-cGAN	**0.9337**	**0.8757**	**0.9835**	**0.9846**	**0.9750**

Bold indicates the highest value under each evaluation metric.

These results demonstrate that SAP-cGAN has great advantages in improving the accuracy and integrity of mass segmentation and can learn features with increased discriminativeness and universality. These characteristics can improve the performance of mass segmentation.

4.B.. Ablation analysis

In Table II, we present the results of the ablation analysis for SAP-cGAN with CBIS-DDSM to demonstrate the effectiveness of our proposed module. We first train an original

cGAN (BaseLine) without adding the superpixel average pooling layer or mutilscale input module. Then, we add the mutilscale input module to the G and D of the original cGAN (BaseL+MS). Furthermore, the superpixel average pooling layer is separately added to original cGAN(BaseL+ SAP). Finally, the proposed model (SAP-cGAN) is constructed by adding the superpixel average pooling layer and mutilscale input module.

Table II presents the results of the BaseLine, BaseL+MS, BaseL+ SAP and SAP-cGAN models. For the dataset of CBIS-DDSM, the BaseLine yields Dice and Jaccard scores of 85.57% and 74.78%, respectively, whereas the BaseL+MS model exhibits a small improvement of 1.17% and 1.81%. This improvement is obviously small because the multiscale input module provides the model with features in different scales such that the network can learn scale-invariant features. Moreover, the BaseL+SAP model gives a Dice of 90.76% and Jaccard scores of 83.08% (5.19% and 8.3% higher than the scores provided by BaseLine). This comparison shows that the superpixel average pooling layer can effectively improve the performance of mass location and segmentation. SAP-cGAN significantly improves the results compared with BaseLine ($P < 0.05$), in terms of Dice score (93.37%), Jaccard score (87.57%), accuracy (98.35%), specificity (98.46%), and sensitivity (97.50%).

The qualitative results of the ablation analysis are shown in Fig. 7. A small mass, a big mass and a mass with spiculated boundaries are used to determine the effects of adding different components on the segmentation results. All four models show good performance in segmenting the small mass partly because of its smooth boundaries. By contrast, BaeLine evidently cannot segment the big mass as well as the small mass. After the multiscale input module is added, the FN value obviously decreased. Base+SAP improves the results of boundary region segmentation of the mass with spiculated boundaries compared with BaseLine. The SAP-cGAN model, which is implemented by adding the superpixel average pooling layer and multiscale input module, shows superior performance in segmentating these types of masses.

5. DISCUSSION

In this study, we propose a novel network architecture for mass image segmentation in digital mammograms. We integrate two modules into the basic cGAN: a superpixel average pooling layer and a multiscale input module. These modules are used to provide prior boundary information and scale-invariant features. The performance of the proposed model is evaluated by using mass images from two public datasets, namely, CBIS-DDSM and INbreast. In terms of Dice and Jaccard scores, accuracy, specificity, and sensitivity, our model outperforms the current state-of-the-art mass segmentation methods, such as FCN, U-Net, SegNet, SegModel_CRF, and SegModel_cGAN. Superpixels can provide the prior shape and boundary information of a mass, and the multiscale input module can provide features with increased robustness. The ablation analysis results show that these two modules can be combined with the abstract semantic information in the cGAN, which helps generate pixel-level classification results with increased accuracy.

Although the SAP-cGAN model proposed in this paper is superior to other segmentation models, its results for masses with complex tissue structures are not ideal. As shown in Fig. 8, the mass in (a) has numerous boundary burrs and intensity inhomogeneity. In this case, our model achieves a Dice score of 85.19% and a Jaccard score of 74.02%. For the mass in (b) with spiculated boundaries and low contrast, our proposed SAP-cGAN obtains a Dice score of 79.50% and a Jaccard score of 65.98%. All the six models have failed to segment accurately the mass with fuzzy and spiculated boundaries, low contrast, and intensity inhomogeneity.

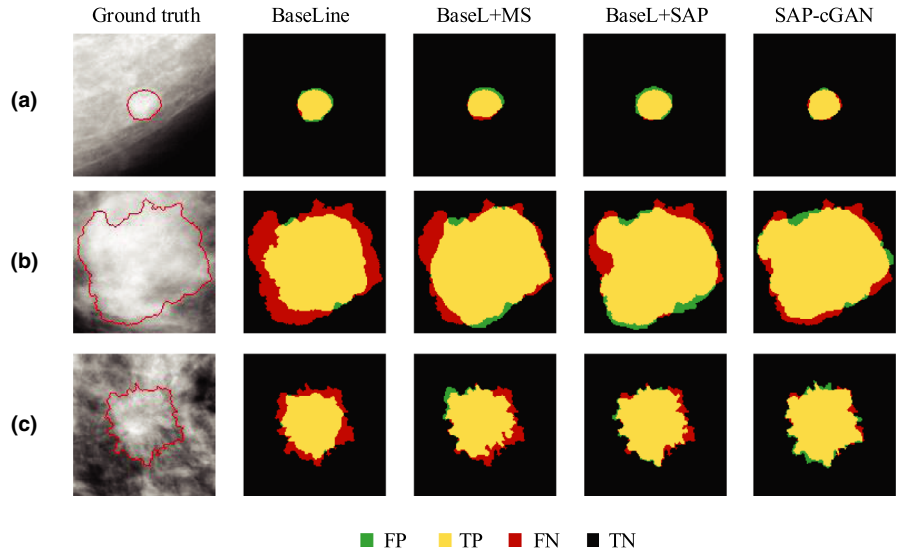

FIG. 7. Segmentation result of ablation analysis with the CBIS-DDSM dataset. (a) Small mass, (b) big mass, (c) mass with spiculated boundaries. From left to right: original images and the results of the BaseLine, BaseL+MS, BaseL+SAP and SAP-cGAN models. [Color figure can be viewed at wileyonlinelibrary.com]

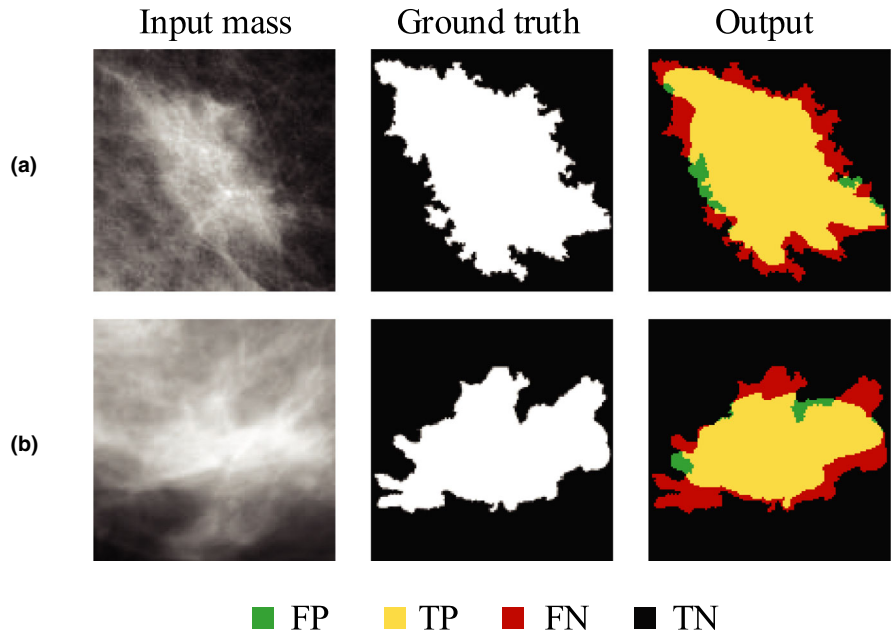

FIG. 8. Examples of masses that were inaccurately segmented by the proposed SAP-cGAN. [Color figure can be viewed at wileyonlinelibrary.com]

Nevertheless, in contrast to the other models, our model can provide segmentation results in such difficult situations. In the future, we intend to optimize the performance for segmenting such masses and increase the accuracy of segmentation.

In addition, the present work has several limitations. In fact, there are significant correlations between the lesions in MLO and CC views. Our model extracts the features of each view, and this correlation may be ignored during feature extraction. However, the issue of how these correlations will affect the segmentation result of breast masses is unclear. In a follow-up study, we will explore the correlations between mass segmentation and MLO/CC views.

6. CONCLUSIONS

We propose a mammography mass segmentation model called SAP-cGAN, which is based on an improved cGAN. We introduce a superpixel average pooling layer into the decoder of cGAN, which utilizes superpixels as a pooling layout to improve boundary segmentation. In addition, we adopt a multiscale input strategy to enable the network to learn scale-invariant features with increased robustness. In mammography mass segmentation with CBIS-DDSM and INbreast datasets, the proposed SAP-cGAN exhibits remarkable improvements qualitatively and quantitatively over the baseline cGAN and state-of-the-art methods. We prove that SAP-cGAN can provide the prior shape, boundary information, and multiscale features of a mass. These pieces of information can be combined with the high-level abstract semantic information in cGAN. Moreover, our model is fully automated and can be easily integrated into breast cancer CAD systems to assist physicians in the detection and diagnosis of breast cancer.

ACKNOWLEDGMENT

This work is supported by National Natural Science Foundation of China under Grant81772009, Scientific and Technological Research Project of Henan Province under Grant182102310162.

CONFLICTS OF INTEREST

The authors have no conflicts to disclose.

[a)]Author to whom correspondence should be addressed. Electronic mail: mywang@ha.edu.cn.

REFERENCES

1. Khazaei Z, Jarrahi AM, Momenabadi V, Ghorat F, Goodarzi E. Global Cancer Statistics 2018: globocan estimates of incidence and mortality worldwide stomach cancers and their relationship with the human development index (HDI). *Adv Hum Biol*. 2019;9(3):245–250.
2. Liberman L, Menell JH. Breast imaging reporting and data system (BI-RADS). *Radiol Clin N Am*. 2002;40:409–430.
3. Bretthauer M, Kalager M. Principles, effectiveness and caveats in screening for cancer. *British Journal of Surgery*. 2013;100(1):55–65.
4. Dong M, Lu X, Ma Y, et al. An efficient approach for automated mass segmentation and classification in mammograms. *J Digit Imaging*. 2015;28:613–625.
5. Oliver A, Freixenet J, Martý J, et al. A review of automatic mass detection and segmentation in mammographic images. *Med Image Anal*. 2010;14:87–110.
6. Tamaki K, Ishida T, Miyashita M, et al. Correlation between mammographic findings and corresponding histopathology: potential predictors for biological characteristics of breast diseases. *Cancer Sci*. 2011;102:2179–2185.
7. Liu S, Wu X-D, Xu W-J, et al. Is there a correlation between the presence of a spiculated mass on mammogram and Luminal a subtype breast cancer? *Korean J Radiol*. 2016;17:846–852.

8. Tang J, Rangayyan RM, Xu J, Naqa IE, Yang Y. Computer-aided detection and diagnosis of breast cancer with mammography: recent advances. *IEEE Trans Inf Technol Biomed*. 2009;13:236–251.
9. Sutton C, McCallum A, Rohanimanesh K. Dynamic conditional random fields: factorized probabilistic models for labeling and segmenting sequence data. *Journal of Machine Learning Research*. 2007;8:693–723.
10. Rahmati P, Adler A, Hamarneh G. Mammography segmentation with maximum likelihood active contours. *Med Image Anal*. 2012;16:1167–1186.
11. Dhungel N, Carneiro G, Bradley AP. Deep learning and structured prediction for the segmentation of mass in mammograms, in International Conference on Medical Image Computing & Computer-assisted Intervention, 2015.
12. Kozegar E, Soryani M, Behnam H, Salamati M, Tan T. Mass segmentation in automated 3-D breast ultrasound using adaptive region growing and supervised edge based deformable model. *IEEE Trans Med Imag*. 2018;37:918–928.
13. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. *IEEE Trans Pattern Anal Mach Intellig*. 2015;39:640–651.
14. Li S, Dong M, Du G, Mu X. Attention dense-U-Net for automatic breast mass segmentation in digital mammogram. *IEEE Access*. 2019;7: 59037–59047.
15. Cheng Y, Gao Y, Xie L, Xie X, Lin WJIA. Spatial enhanced rotation aware network for breast mass segmentation in digital mammogram. *IEEE Access*. 2019:1.
16. Hamidinekoo A, Denton E, Rampun A, Honnor K, Zwiggelaar R. Deep learning in mammography and breast histology, an overview and future trends. *Med Image Anal*. 2018;47:45–67.
17. Yassin NIR, Omran S, Houby EMFE, Allam H. Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: a systematic review. *Comput Meth Progr Biomed*. 2018;156:25–45.
18. Shen T, Gou C, Wang J, Wang FY. Simultaneous segmentation and classification of mass region from mammograms using a mixed-supervision guided deep model. *IEEE Signal Process Lett*. 2020;27:196–200.
19. Dong Y, Daguang X, Zhou SK, et al. Automatic liver segmentation using an adversarial image-to-image network, in Medical Image Computing and Computer Assisted Intervention - MICCAI 2017. 20th International Conference. Proceedings: LNCS 10435, 2017:507–515.
20. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. In Ghahramani Z, Welling M, Cortes C, Lawrence ND, Weinberger KQ, eds. *Advances in Neural Information Processing Systems 27*, Advances in Neural Information Processing Systems, 2014: 2672–2680.
21. Pang S, Du A, Organ MA, et al. CTumorGAN: a unified framework for automatic computed tomography tumor segmentation. *Eur J Nucl Med Mol Imaging*. 2020;47:2248–2268.
22. Singh VK, Rashwan HA, Romani S, et al. Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network. *Expert Syst Appl*. 2020;139:14.
23. Achanta R, Shaji A, Smith K, et al. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans Pattern Anal Mach Intell*. 2012;34:2274–2281.
24. Kwak S, Hong S, Han B. Weakly supervised semantic segmentation using superpixel pooling network, Thirty-First Aaai Conference on Artificial Intelligence, 2017.
25. Chen J, He F, Zhang Y, Sun G, Deng M. SPMF-Net: weakly supervised building segmentation by combining superpixel pooling and multi-scale feature fusion. *Remote Sens*. 2020;12:1049.
26. Yang Y, Wang X. Improved CNN based on super-pixel segmentation, in International Conference on Intelligence Science.
27. Isola P, Zhu J-Y, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks, IEEE Conference on Computer Vision and Pattern Recognition, 2017:5967–5976.
28. Noord NV, Postma E. Learning scale-variant and scale-invariant features for deep image classification. *Pattern Recognit*. 2017;61:583–592.
29. Milletari F, Navab N, Ahmadi S-A. V-Net: fully convolutional neural networks for volumetric medical image segmentation, International Conference on 3D Vision, 2016:565–571.
30. Heath M, Bowyer K, Kopans D, et al. *The Digital Database for Screening Mammography*. Netherlands: Springer; 2001.
31. Lee RS, Gimenez F, Hoogi A, et al. A curated mammography data set for use in computer-aided detection and diagnosis research. *Sci Data*. 2017;4:170177.
32. Moreira IC, Amaral I, Domingues I, et al. INbreast: toward a full-field digital mammographic database. *Acad Radiol*. 2012;19:236–248.
33. Jiang M, Zhang S, Zheng Y, Metaxas DN. Mammographic mass segmentation with online learned shape and appearance priors, in The 19th International Conference on Medical Image Computing and Computer-Assisted Intervention, 2016:35–43.
34. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation, vol. 9351 of Lecture Notes in Computer Science, 2015:234–241.
35. Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell*. 2017;39:2481–2495.
36. Al-antari MA, Al-masni MA, Choi M-T, Han S-M, Kim T-S. A fully integrated computer-aided diagnosis system for digital x-ray mammograms via deep learning detection, segmentation, and classification. *Int J Med Informatics*. 2018;117:44–54.
37. Shen T, Gou C, Wang FYHe ZChen W. Learning from adversarial medical images for x-ray breast mass segmentation. *Comput Meth Progr Biomed*. 2019;180, 105012.
38. Lalande A, Garreau M, Frouin F. *Evaluation of Cardiac Structure Segmentation in Cine Magnetic Resonance Imaging*. Hoboken: Wiley-Blackwell, 2015.