IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 2, 2019, accepted April 23, 2019, date of publication May 7, 2019, date of current version May 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2914873

Attention Dense-U-Net for Automatic Breast Mass Segmentation in Digital Mammogram

SHUYI LI[1,2], MIN DONG[1], GUANGMING DU[1,2], AND XIAOMIN MU[1]
[1]School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China
[2]Institute of Industrial Technology, Zhengzhou University, Zhengzhou 450001, China

Corresponding author: Min Dong (dm880612dm@163.com)

This work was supported in part by the Key Research Projects of Henan Higher Education Institutions under Grant 18A510017, and in part by the Henan Postdoctoral Research Project under Grant 001701004.

ABSTRACT Breast mass is one of the most distinctive signs for the diagnosis of breast cancer, and the accurate segmentation of masses is critical for improving the accuracy of breast cancer detection and reducing the mortality rate. It is time-consuming for a physician to review the film. Besides, traditional medical segmentation techniques often require prior knowledge or manual extraction of features, which often lead to a subjective diagnosis. Therefore, developing an automatic image segmentation method is important for clinical application. In this paper, a fully automatic method based on deep learning for breast mass segmentation is proposed, which combines densely connected U-Net with attention gates (AGs). It contains an encoder and a decoder. The encoder is a densely connected convolutional network and the decoder is the decoder of U-Net integrated with AGs. The proposed method is tested on the public and authoritative database—Digital Database for Screening Mammography (DDSM) database. F1-score, mean intersection over union, sensitivity, specificity, and overall accuracy are used to evaluate the effectiveness of the proposed method. The experimental results show that dense U-Net integrated AGs achieve better segmentation results than U-Net, attention U-Net, DenseNet, and state-of-the-art methods.

INDEX TERMS Breast masses segmentation, deep learning, biomedical image processing, attention gates, densely connected convolutional network.

I. INTRODUCTION

Breast cancer is one of the most common malignancies in female. National Institute of Cancer (NCI) of the China shows that the incidence of breast cancer ranks first among female cancer and it is still rising. New cases of breast cancer are 1.38 million and 1.67 million, at the same time 458,000 and 522,000 people died from it in 2008 and 2012 [1]. If the breast cancer can be early detected, the 5-year relative survival rate is 99%; On the contrary if the cancer has spread to a distant part of the body, the 5-year survival rate is only 27% [2]. Therefore, early diagnosis of breast cancer is particularly important. As well known, the most reliable and effective technique for breast cancer early detection is digital mammography [3]. Because the structure of the breast is very complicated, medical experts sometimes cannot correctly determine the location of lesions, and the heavy reading task also lead to missed diagnosis and misdiagnosis. In order to improve the accuracy of breast disease examination, Computer Aided Diagnosis (CAD) technology come into being, which provide convenience for breast cancer screening, and reduce the workload of doctors. It has brought technical breakthroughs in medical diagnosis [4].

Breast mass is one of the most distinctive signs for diagnosis of breast cancer, and its marginal information reflects the growth pattern and biological characteristics. Generally speaking, benign masses are regular in shape, and masses with irregular margins are often malignant. In other words, the accuracy of masses segmentation will affect the benign or malignant classification of the masses. Therefore, masses segmentation is a very important process in breast cancer CAD, which can help doctors diagnose and treat breast cancer early. However, breast masses have diverse features such as different size, shape, and ill-defined boundaries, so the accurate segmentation has become a difficult and hot issue in CAD technology [5], [6].

In the past few years, lots of algorithms for breast mass segmentation have been widely studied [7], such as

The associate editor coordinating the review of this manuscript and approving it for publication was Cristian A. Linte.

Active Contour Model (ACM) and Chan-Vese methods *et al.* [8]–[13]. Hu *et al.* [8] combined adaptive global threshold and adaptive local threshold method to achieve mass segmentation on Mammographic Image Analysis Society (MIAS) database. Dominguez and Nandi [9] employed dynamic-programming-based method and constrained region-growing method to segment the mass. Kozegar *et al.* [10] used the seed position as the only priori information and incorporated shape information in training masses, this proposed method outperformed the adaptive region growing and regularized level set evolution method in terms of Dice measure. Pereira *et al.* [11] employed multiple threshold, wavelet transform and genetic algorithm to segment the mass randomly selected from DDSM. Gupta and Tiwari [12] proposed histogram modified gray relational analysis method to enhance mammogram image and then segmented the breast mass, absolute mean brightness error, structural similarity index measure and Peak Signal to Noise Ratio (PSNR) were used to evaluate the effectiveness. Kashyap *et al.* [13] presented variation level set function based approaches in mass segmentation.

However, it is necessary to set some parameters (such as thresholds) and extract effective features manually in these traditional segmentation methods, which may based on experience. Consequently, the results of these techniques are not stable enough. In addition, due to the presence of noise and artifacts in mammograms, the background region is complex and the size of the mass is too small relative to background region. The segmentation results of the traditional methods are often not ideal [7]. These existing traditional techniques have difficulty in achieving automatic end-to-end precise segmentation of breast masses.

Recently, due to the ever-increasing computing power and the ever-increasing amount of data available, deep learning has made significant progress in the field of medical image [14]–[20]. Especially, Convolutional Neural Network (CNN) can capture the nonlinear mapping between input and output, and automatically learn local area features and high-level abstract features through multi-layer network structures which are usually better than manual extraction and predefined feature sets. CNN has been used to address a variety of biomedical image processing issues, including medical image segmentation [21]–[23] and achieved good segmentation results on public authoritative datasets.

However, the mass segmentation based on CNN uses an image block around a pixel as the network input to classify, which not only causes large storage overhead but also makes low calculation efficiency because of repeated convolution calculation during training and prediction [24]. In addition, the inputs of large image blocks require more pooling layers which will reduce the segmentation accuracy. Conversely, using small image blocks means that the size of receptive fields is reduced. Small receptive fields only extract local features and the accuracy of segmentation is limited.

In order to solve the problem of traditional CNN, Long *et al.* [25] proposed Fully Convolutional Networks (FCN) for natural image segmentation which replaced all the fully connected layers in the traditional CNN. Although the FCN-8s is much better than the FCN-32s, the up-sampling results are still blurry and smooth, and the network is not sensitive enough for image details. On the other hand, the final segmentation results only use deep feature mapping and ignore the spatial domain information extracted by shallow layer of the network. As well known, compared with natural images segmentation, breast masses segmentation is more complicated and requires more precision. Consequently, Ronneberger *et al.* [26] proposed a more suitable medical image segmentation method utilizing U-Net architecture based on FCN. Medical images generally have blurred boundaries and complex gradients, and they often need more high-resolution information for detection and identification. U-Net concatenates the features extracted from the shallow layer and the deep layer by skip connections. It uses high-resolution information of breast masses obtained by the shallow layer supply the missing detailed spatial information during the up-sampling process, and obtain better segmentation results. However, the performance of single U-Net segmentation method is not so excellent when the target organ or tissue differs in shape and size greatly; it often relies on multi-stage cascaded CNN [27]. The cascaded network first locates the approximate location of the lung tumor through a network and then uses another network to achieve accurate segmentation in this area. But the cascading multiple networks lead to redundant use of computational resources and model parameters. In addition, breast masses usually vary in shape and size; the scale between training samples varies widely. In order to ensure the network achieving a better segmentation performance when training on a limited set for breast masses with different shapes and sizes without increasing the convolution kernel, the network needs to extract as many features as possible. U-Net is slightly insufficient at this point.

Oktay *et al.* [28] proposed Attention Gates (AGs) and used the network structure integrated with AGs to segment pancreas without adding extra network, which got better segmentation results than single U-Net. Sometime, AGs are also applied to language processing [29] and natural image analysis [30]. Trainable attention is categorized as hard-attention and soft-attention. Hard-attention is general non-different and depends on reinforcement learning to update parameters which will make training procedure more difficult. On the contrary, soft-attention is probabilistic and uses standard back-propagation to update parameters. Bahdanau *et al.* [31] introduced additive soft attention in the encoder-decoder neural machine translation model and achieved a translation performance comparable to the existing state-of-the-art system. Jetley *et al.* [32] proposed self-attention to eliminate the dependency on exterior gating information.

In this paper, an improved U-Net architecture for breast masses segmentation to solve the above problems is proposed. The proposed network is a U-type network structure including an encoder and a decoder. The encoder is

a densely-connected CNN structure, which is the feature extractor of the network. This connection method proposed in [33] makes each feature extractor layer accepting the features obtained by all the previous layers as input. These feature maps are concatenated directly. It realizes reuse of features, so that as many breast masses features with various sizes and shapes as possible can be extracted without adding more parameters. The decoder is a CNN integrated with AGs. The addition of the AGs could help get more accurate segmentation results than a single U-Net without cascading more. At train time, AGs learn to focus on target without additional supervision automatically; at test time, AGs generate soft region proposals implicitly and highlight marked features for breast masses segmentation task. Experiments have shown that this network structure yields a promising breast masses segmentation results than the basic U-Net and state-of-the-art methods, which demonstrate the effectiveness of the combined use of dense connections and attention mechanisms on breast masses segmentation.

To provide a better understanding of the proposed algorithms, the rest of the article is organized in the following order. In Section II, the network architecture and some of the theoretical knowledge are illustrated. Experimental results of segmentation are explained in Section III, and finally in Section IV the conclusion is drawn.

II. MATERIALS AND METHODS

The proposed breast mass segmentation network can be schematically described in Fig. 1. We can see that our network is encoder-decoder architecture: the encoder is a densely-connected CNN and the decoder is a CNN integrated with AGs.

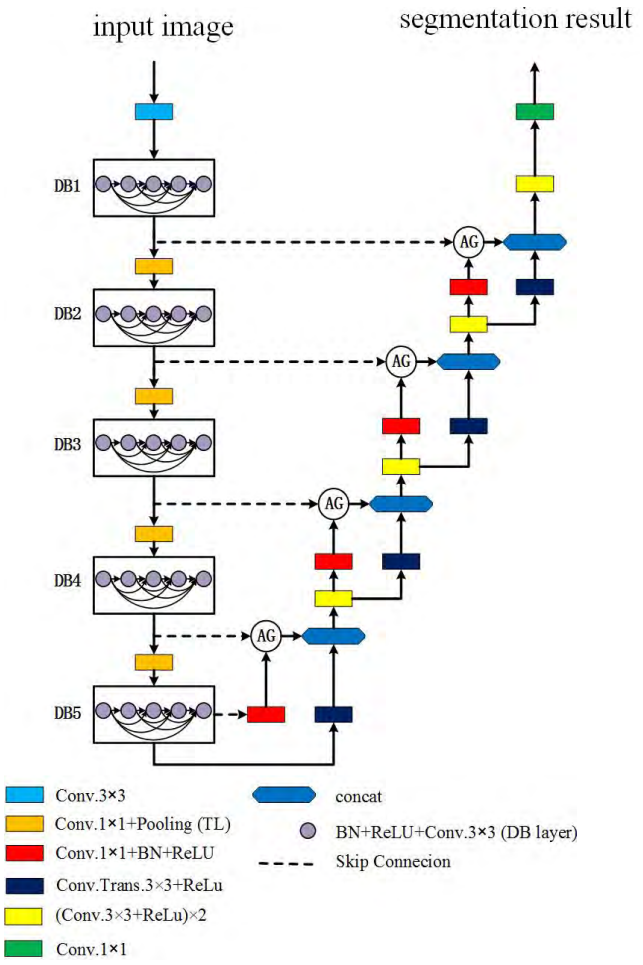

FIGURE 1. The architecture of Attention Dense U-Net.

A. DATASET AND DATA AUGMENTATION

In the study, the DDSM [34] of X-ray mammography from the University of Florida is used to train, test and evaluate the proposed network. The database includes approximately 2,500 cases and every case contains two views of each breast, as well as some associated patient information (age, breast density rating, rating for abnormalities and keyword description of abnormalities) and image information (scanner, spatial resolution and so on). Images containing suspicious areas have associated pixel-level ''ground truth'' information about the locations and types of suspicious regions. The lossless encoded ''.LJPEG'' format images can be downloaded from the official website. In our experiment, 400 representative mammography X-ray images with masses of the DDSM database are selected and they are randomly divided into three parts: training set, verification set and test set in a ratio of 4:1:1. All these chosen images have been annotated by experienced radiologists and the segmentation performance is evaluated by comparing the difference between the prediction by networks and the contours annotated by experts. In order to standardize the dataset, the average of image is subtracted from each pixel, and then the standard deviation of the image is divided.

The mammograms in the DDSM database are generally composed of the mammary region and the background region containing label. The segmented target region (breast masses) usually appears in the breast region, and the background region generally contains noise information. The presence of these noises can interfere with the training of the network and affect the results of the breast masses segmentation. Therefore, before the image is sent to the network, these mammograms need to be pre-processed to extract the breast region and remove the label from the background region in the image. First, we use gamma transform to achieve gray adjustment of the original image to enhance the breast region and its contour, but it also enhances the label and other interference information in the background area; then the gray-adjusted image is binaries; in order to remove the interference information, it is necessary to obtain the maximum connected region in binary image. Finally, the original image is dot multiplied by the maximum connected area to obtain the unlabeled image. The result is as in Fig. 2.

One of the dominating concerns when employing deep learning is over fitting, especially for CNN. In order to train deep learning networks without over fitting and achieve great performance, large amounts of annotated images are

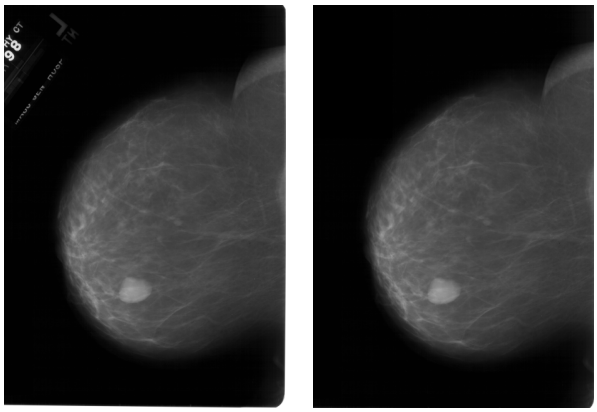

FIGURE 2. The result of removing the label. (a) Original image. (b) Breast region.

usually needed. However, medical datasets generally have a quite small number of images because of limited patient volume. The little size of breast image datasets available puts forward a challenge for the training task at present. Recently, data augmentation is proposed to handle this challenge. It can better extract features and avoid over fitting by generating new data from the original data. There are many existing data augmentation methods. In this study, in order to preserve the features of breast images, combination of random rotation and random mirroring techniques are used. Each original image is mirrored horizontally. Then, the original and reflected images are rotated in the range of -10 to 10 degrees. Such a data augmentation method allows masses to be displayed in different orientations. In this way, it can improve the generalization ability and robustness of the model. Last but not least, in order to adapt the network, experimental images are resized to 512×512.

B. PROPOSED METHOD

1) FULLY CONVOLUTIONAL NETWORK (FCN)

In order to solve the problems using traditional CNN in image segmentation, Long *et al.* [25] proposed FCN for image segmentation which replaced all the fully connected layers in the traditional CNN with convolutional layers. And they used deconvolutional layers to up-sample the output feature map of the last convolutional layer to restore the size of original images. The network generates a prediction for every pixel and remains the spatial information of original input. Finally, the pixel-by-pixel classification is performed on the output feature map of the deconvolutional layer to achieve pixel-level classification. Compared to the segmentation of CNN, FCN can accept input images of any size because there is no fully connected layer. In addition, because it does not use image blocks around a pixel as input when training and predicting, segmentation accuracy is improved and storage overhead is reduced. However, since the final segmentation results of FCN only utilize deep feature map, the segmentation results are not good enough to satisfy the accuracy of medical images segmentation.

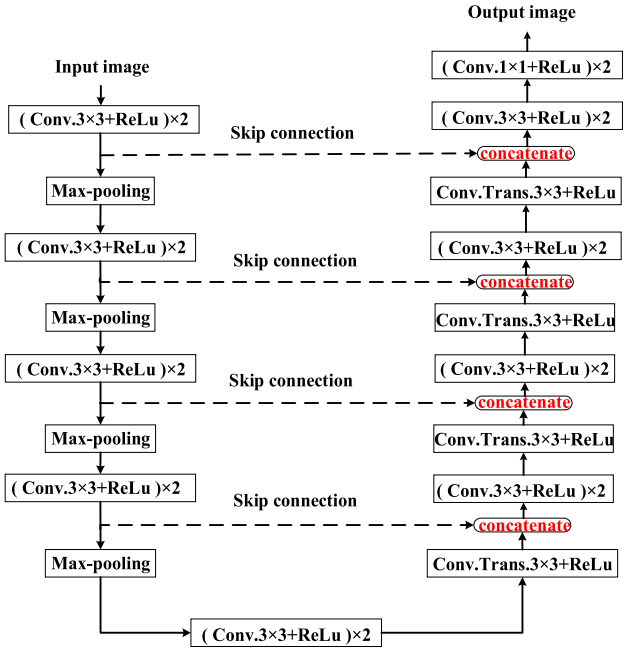

FIGURE 3. U-Net architecture.

2) U-NET

In order to make the network more suitable for segmenting medical images, Ronneberger *et al.* [26] proposed a U-Net segmentation architecture based on FCN which had been confirmed that it was more suitable for medical images segmentation than FCN. The U-Net architecture is as shown in Fig. 3. It composes of an encoder (a contracting path on the left side) and a decoder (an expansive path on the right side). The encoder uses the representative architecture of a CNN, which is responsible for down-sampling and is the feature extractor of the network. Each down-sampling module includes two convolutional layers (3×3, without padding) and a maximum pool layer (2×2, stride $=2$). For each down-sampling, the number of feature channels is doubled and the size of feature maps is reduced by half. The decoder is responsible for up-sampling. Each up-sampling module contains a deconvolutional layer (2×2) and two convolution layers (3×3, without padding). The feature maps obtained by the shallow layer are concatenated with the features maps output from deep layer by skip connections, as is shown in the Fig. 3. For each up-sampling, the number of feature channels is reduced by half and the size of images is doubled. The last convolutional layer (1×1) map 64-channels feature vector to the desired classification result, making prediction for each pixel. Rectified Linear Unit (ReLU) activation functions are used after each convolutional layer. Because it concatenates the output of the shallow layer and the output of the deep layer, the network can simultaneously consider the contribution of shallow information and deep information at the

final output, resulting in better segmentation results on the segmentation of medical images than FCN.

3) ATTENTION DENSE U-NET

For the characteristics of the dataset used in this work: a). the size and shape of the different masses vary greatly; b). the size of mass is too small relative to the background region. According to these features, encoder-decoder architecture with AGs and dense connections is proposed to segment breast masses. The following is a detailed introduction about the network.

The performance of a single U-Net segmentation is not so excellent when the small target organ or tissue differs in shape and size greatly, for example in the field of medical. For the sake of improving segmentation accuracy, multi-stage cascaded CNNs framework is proposed. The architecture uses an additional localization network to determine the approximate location of the ROI and then uses another network to perform segmentation on the ROI. However, this method results in redundant use of computational resources and model parameters.

Oktay [28] proposed AGs and used it for the segmentation of the pancreas. The visual attention mechanism is a unique brain signal processing mechanism to human vision. Human quickly scan the entire image to find the target area that needs to be focused on, which is commonly referred to as the focus of attention, and then invest more attention resources in this area to obtain more detail information. And at the same time, suppress useless information and interference information. This is a means for human to quickly screen out high-value information from a large amount of information using limited attention resources. It is a survival mechanism of human beings formed in long-term evolution. The human visual attention mechanism greatly improves the efficiency and accuracy of visual information processing. The attention mechanism in deep learning is essentially similar to the human visual attention mechanism. Its core goal is also to find the most critical information for the current task (mass segmentation) from a large amount of information. The attention mechanism improves the performance of the network by suppressing feature activations that are not related to the task. In this way, there is no need to cascade additional positioning network at the same time, it can get higher prediction accuracy than a single network. The architecture of AGs is shown in Fig. 4.

The x_l is the feature map of the output of layer l (low-level features). The g_i is the gating signal vector which is collected from a coarser scale and used for every pixel to decide focus region. α is the attention coefficient which reserves activations related to target task by suppressing unrelated feature responses. The output of AGs is element-wise multiplication of the two.

$$x_{lout} = x_l \cdot \alpha_i \tag{1}$$

Learning multi-dimensional attention coefficients is recommended in the situation of multiple semantic classes.

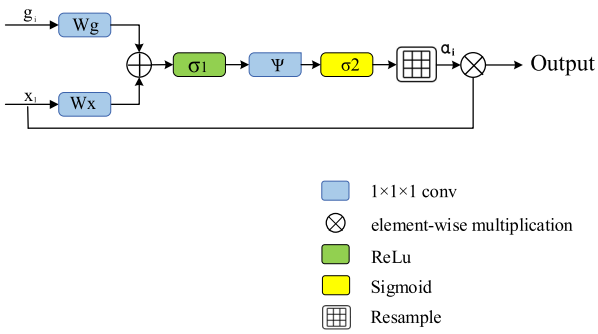

FIGURE 4. Schematic of additive AGs.

We use additive attention instead of multiplicative attention to obtain the gating coefficient. Although additive attention increases computational burden, it can obtain higher accuracy. The formulation for attention coefficient is as follows:

$$\alpha_i = \sigma_2(\psi^T(\sigma_1(W_x^T x_l + W_g^T g_i + b_g)) + b_\psi) \tag{2}$$

σ_1 is often chosen as ReLU function: $\sigma_1(m) = \max(0, m)$, and σ_2 is the Sigmoid function: $\sigma_2(m) = \frac{1}{1+e^{-m}}$. W_x, W_g and ψ are linear transformations, b_ψ and b_g are bias terms. These linear transformations are computed using channel-wise $1\times1\times1$ convolutions for the feature map x_l and gating vector g_i. Parameters of AGs use a normal distribution initialization method and are updated according to the back-propagation principle.

On the other hand, breast masses usually vary in shape and size and the number of images available are limited. For the sake of obtaining good performance on segmenting breast masses with different shapes and sizes without additional convolution kernel on small dataset, the network needs to make the best use of the extracted features. G. Huang *et al.* [33] proposed dense connection. This manner let each layer connect to every other layer in a feed-forward pattern. It means that each layer receives additional inputs from its all preceding layers and passes on its feature maps to all following layers. This connection method realizes feature reuse, which allows network to make better use of features and achieve higher segmentation accuracy on small objects with large scale changes. Traditional CNN only connects the output feature map of l^{th} layer as input to $(l+1)^{th}$ layer, which causes to the layer transition: $x_l = H_l(x_{l-1})$. While l^{th} layer of densely connected networks obtains the feature maps of all forming layers, which gives rise to the layer transition:

$$x_l = H_l([x_0, x_1, \ldots, x_{l-1}]) \tag{3}$$

where $[x_0, x_1, \ldots, x_{l-1}]$ represents the concatenation of the feature maps generates in layers $0, \ldots, l-1$. And $H_l(\cdot)$ is defined as a composite function of three continuous operations: Batch Normalization (BN), a ReLU and a 3×3 convolution (Conv).

To make the network structure suitable for the characteristics of the breast masses dataset: a. the shape and size of different masses are quite various; b. the mass is very small

relative to the background area of the breast, we add attention mechanism and dense connections to traditional U-Net. This architecture can achieve feature reuse and obtain better results on segmenting breast masses without cascading additional positioning networks and adding more convolution kernels. The network is an encoder-decoder architecture which is composed of an encoder (a contracting path on the left side) and a decoder (an expansive path on the right side). The proposed network is shown as Fig. 1.

The encoder is a densely-connected network. Firstly, the input image is subjected to a 2×2 convolution calculation, and then the output is sent to the first dense block containing five densely-connected convolutional layers (a BN operation, a ReLU activation function and a 3×3 convolution). The input of every layer in the dense block is the feature maps output from all the previous layers of this dense block. There are five dense blocks in the proposed network. Each dense block contains five densely connected convolutional layers. Convolution operation changes image size and concatenation operation in the decoder is not viable when the size changes. In order to make concatenation successful, there is a transition layer between two dense blocks to recover the size of the feature maps. It is made up of a 1×1 convolution and a 2×2 average pooling operation. This is also why the network is divided into multiple dense blocks. Besides, every layer connects to the gradients from loss function and original input directly, resulting in alleviating the phenomenon of gradient disappearance. The dense connections have a regularizing effect, which can reduce over fitting on tasks using smaller training set.

The decoder is a CNN integrated with AGs which can highlight salient features. Using the feature maps extracted from down-sampling stage in the model as the gating signal. So the output of the encoder is firstly transformed into the gating signal vector which contains contextual information to eliminate unrelated and noisy responses. The gating signal and the feature map obtained by the fifth dense block are computed by the AGs and its output is concatenated with feature map obtained by the fourth dense block (through skip connection). Because during backward pass, gradients obtained by background regions are down-weighted, the parameters of lower layer are updated mostly based on the masses regions. There are four AGs architectures in the decoder.

Fig. 5 is a flow chart of the entire algorithm of our proposed algorithm. The inputs are the breast images to be segmented. Firstly, performing data augmentation and then sending images to the proposed network. The output images of the network are the corresponding segmentation results.

C. PERFORM EVALUATION

Sensitivity, specificity, overall accuracy, F1-score or Dice similarity coefficient and Receiver Operating Characteristic (ROC) curve are used to evaluate the mass segmentation performance of the networks (our proposed network, U-Net [26], Attention U-Net [28] and DenseNet [33]) in an objective perspective. All of the metrics are defined as follows,

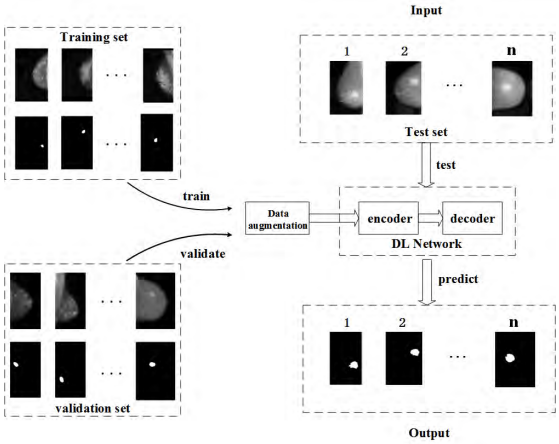

FIGURE 5. The flow chart of the proposed algorithm.

TABLE 1. The concept of confusion matrix.

Predict result	Actual test	
	Positive	Negative
Positive	TP	FP
Negative	FN	TN

$$\text{Sensitivity(True Positive Rate,TPR)} = \frac{\text{TP}}{\text{TP+FN}} \quad (4)$$

$$\text{Specificity(True Negative Rate, TNR)} = \frac{\text{TN}}{\text{TN+FP}} \quad (5)$$

$$\text{Overall accuracy} = \frac{\text{TP+TN}}{\text{TP+FN+TN+FP}} \quad (6)$$

$$\text{F1-score(Dice Coefficient)} = \frac{2\times\text{TP}}{2\times\text{TP+FP+FN}} \quad (7)$$

where TP, FP, TN, FN respectively indicate the number of pixels in the true positive, false positive, true negative, and false negative regions, as defined in Table 1. Sensitivity represents the ability of the mass region annotated by experts (ground truth) to be predicted by the model as mass area. Specificity represents the ability of background region annotated by experts to be predicted by the model as background area. F1-score represents the degree of similarity between mass region predicted by model and mass region in the ground truth. All parameters range between 0 and 1. And values closer to 1 are better.

At same time, we use ROC curve combining sensitivity and specificity in a graphical manner to evaluate segmentation performance comprehensively. It is plotted with sensitivity as ordinate and 1-specificity as abscissa. AUC is the area under the ROC curve, is also used to evaluate the segmentation network. The closer the curve is to the upper left corner or the larger the value of AUC, the better the test results are.

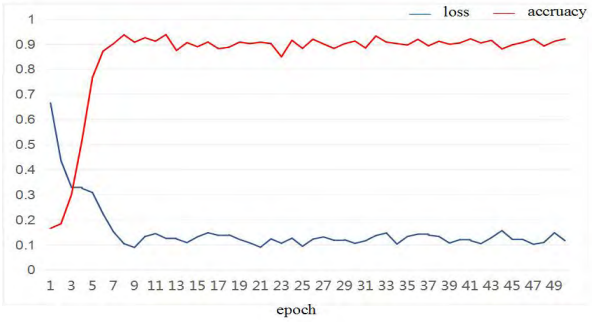

FIGURE 6. The curves of the loss and accuracy of the proposed model as the number of iterations increases during training process (blue curve -> loss, red curve -> accuracy).

TABLE 2. F1-score, sensitivity, specificity, overall accuracy from DDSM database.

Method	F1-score (%)	Sensitivity (%)	Specificity (%)	Overall accuracy (%)
Proposed	**82.24 ±0.06**	**77.89 ±0.08**	84.69 ±0.09	**78.38 ±0.04**
U-Net	75.71 ±0.12	66.92 ±0.13	**87.62 ±0.09**	74.37 ±0.10
Attention U-Net	77.32 ±0.10	68.01 ±0.11	86.24 ±0.09	74.83 ±0.09
DenseNet	81.36 ±0.08	76.19 ±0.11	80.82 ±0.11	77.93 ±0.09

III. EXPERIMENTAL STUDY

A. IMPLEMENTATION

The network architecture is implemented in the Python programming language using the Keras interface to Tensorflow, and performed on one NVIDIA Titan X Graphics Processing Unit (GPU) with 8GB of memory. The initial learning rate is set to 0.0001, and it is adaptively reduced during the training process using the Adam optimizer [15]. The momentum is set to 0.9. The cross entropy loss function is used to calculate the difference between the predicted and the real, and the loss is transmitted to the layers of the network through the back propagation algorithm to guide the update of the weights and biases of each layer of the network.

B. RESULTS AND DISCUSSION

Under normal conditions, judging a deep learning model reach a convergence state or not is according to the change of loss and accuracy during the training process. Fig. 6 shows the curve of the loss and accuracy of the proposed model as the number of iterations increases during training process. As is shown in Fig. 6, the loss is in a steady state close to 0, and the accuracy is in a steady state close to 1. It can be seen that the network has reached a stable convergence state.

The segmentation performances of the proposed network architecture, traditional U-Net, Attention U-Net, and DenseNet are presented in Table 2 by using objective evaluation indicators. As is shown in the table, the results consist of two parts: the mean and standard deviation. The mean reflects the overall segmentation performance of the network.

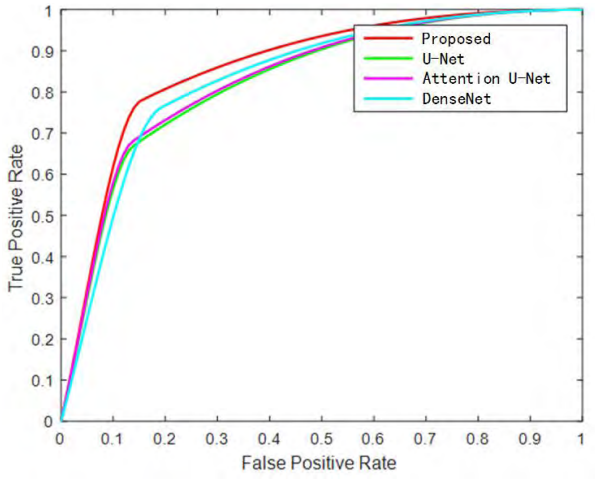

FIGURE 7. ROC curve analysis of the various segmentation methods.

TABLE 3. ROC area indicators of the various segmentation methods.

	Proposed	U-Net	Attention U-Net	DenseNet
AUC	**0.8605**	0.8277	0.8326	0.8336

The standard deviation is used to measure the degree of dispersion of the segmentation results. The smaller the standard deviation is, the stronger the generalization ability of the network is. Our proposed network outperforms the other methods with F1-socre of 82.24%, sensitivity of 77.89% and overall accuracy of 78.38%. Besides, the standard deviation is also smaller, indicating that our proposed network has stronger generalization ability.

It can be seen that the specificity of U-Net and Attention U-Net exceeds the proposed method. High specificity represents a low rate of misdiagnosis. On occasions, higher sensitivity values are at the expense of lower specificity values. Ideally we hope both sensitivity and specificity to be high, but in reality we generally look for a balance between sensitivity and specificity, which can be represented by the ROC curve, as shown in Fig. 7.

Table 3 shows the AUC of the four methods. As is obviously shown in Table 3 and Fig. 7, the ROC curve of the proposed method is to the upper left corner and the AUC is larger than other three networks (0.8605).

In order to further confirm the effectiveness of our network, the segmentation results of representative images in test set are shown using qualitative methods in Fig. 8 (contours obtained by four networks) and quantitative evaluation indicator in Table 4 and Table 5 (F1-score).

It can be seen that from Fig. 8, Table 4 and Table 5 whether the masses are benign or malignant, the proposed network has achieved better segmentation results on masses with various sizes and shapes than the other three networks.

Table 6 shows the time spent on training the four networks and average testing time on an image. Because the proposed network structure in the paper is more complicated, it takes more time to update network parameters in training process. Besides, it takes a little more time to segment the mass in

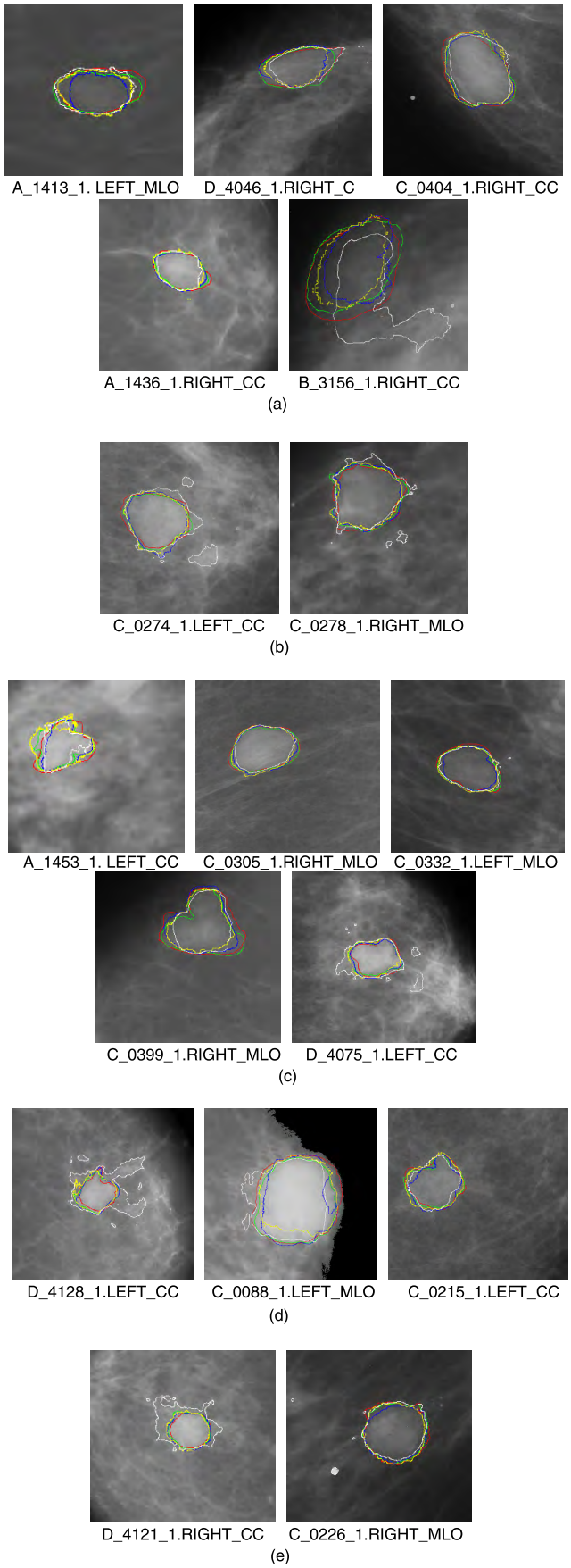

FIGURE 8. The manual annotation and segmentation results obtained by four networks. Ground truth → red, Proposed Method → green, Attention U-Net → blue, DenseNet → yellow and U-Net →w. (a) Benign masses with oval outlines. (b)Benign masses with round outlines. (c) Benign masses with lobulated outlines. (d) Malignant masses with oval outlines. (e) Malignant masses with round outlines. (f) Malignant masses with lobulated outlines. (g) Malignant masses with irregular outlines.

an image than the other three networks which is within acceptable limits. Breast mass segmentation often requires higher accuracy, so we choose increase the appropriate time cost in exchange for the improvement of the accuracy of breast mass segmentation.

Results from state-of-the-art breast mass segmentation models are summarized in Table 7 for comparison. Though the sensitivity obtained by our method is not as high as that of [35], F1-score is higher than that of [35]. Both sensitivity and F1-socre are higher than method proposed in [36]. Sensitivity is higher than [37].

In this paper, Attention Dense U-Net segmentation network for breast masses is proposed based on the features of breast mass dataset. A single U-Net segmentation method is not so excellent when the masses differ in shape and size greatly. Cascaded CNN can get higher accuracy while resulting in redundant use of computational resources and model parameters. The introduction of AGs improves the

TABLE 4. The F1-socre for benign masses segmentation obtained by four methods.

Name	Shape	Margins	F1-score			
			Proposed	U-Net	Attention U-Net	DenseNet
A_1413_1. LEFT_MLO	oval	ill_defined	0.91	0.87	0.76	0.88
D_4046_1.RIGHT_CC	oval	obscured	0.92	0.87	0.87	0.85
C_0404_1.RIGHT_CC	oval	circumscribed	0.94	0.89	0.91	0.91
A_1436_1. RIGHT_CC	oval	circumscribed	0.95	0.89	0.91	0.92
B_3156_1.RIGHT_CC	oval	circumscribed	0.86	0.51	0.76	0.76
C_0274_1.LEFT_CC	round	circumscribed	0.92	0.79	0.89	0.89
C_0278_1.RIGHT_MLO	round	micircumscribed	0.93	0.87	0.90	0.90
A_1453_1. LEFT_CC	lobulated	obscured	0.87	0.77	0.84	0.80
C_0305_1.RIGHT_MLO	lobulated	circumscribed	0.95	0.91	0.92	0.92
C_0332_1.LEFT_MLO	lobulated	circumscribed	0.94	0.91	0.87	0.90
C_0399_1.RIGHT_MLO	lobulated	circumscribed	0.90	0.84	0.88	0.84
D_4075_1.LEFT_CC	lobulated	circumscribed	0.92	0.77	0.90	0.88

TABLE 5. The F1-socre for malignant masses segmentation obtained by four methods.

Name	Shape	Margins	F1-score			
			Proposed	U-Net	Attention U-Net	DenseNet
D_4128_1.LEFT_C	oval	ill_defined	0.88	0.54	0.86	0.86
C_0088_1.LEFT_MLO	oval	micircumscribed	0.92	0.85	0.88	0.87
C_0215_1.LEFT_CC	oval	spiculated	0.92	0.88	0.87	0.90
D_4121_1.RIGHT_CC	round	spiculated	0.94	0.65	0.92	0.90
C_0226_1.RIGHT_MLO	round	spiculated	0.95	0.91	0.91	0.92
C_0011_1.RIGHT_CC	lobulated	circumscribed	0.89	0.87	0.79	0.86
C_0360_1.LEFT_CC	lobulated	microlobulated	0.92	0.88	0.87	0.89
C_0023_1.LEFT_CC	lobulated	microlobulated	0.90	0.54	0.76	0.85
C_0051_1.RIGHT_CC	lobulated	microlobulated	0.96	0.85	0.90	0.92
C_0086_1.RIGHT_MLO	lobulated	microlobulated	0.91	0.76	0.87	0.84
C_0176_1.LEFT_CC	lobulated	microlobulated	0.94	0.87	0.86	0.90
C_0457_1.RIGHT_CC	irregular	spiculated	0.90	0.84	0.80	0.80
C_0028_1.LEFT_CC	irregular	spiculated	0.94	0.90	0.91	0.92
C_0031_1.RIGHT_CC	irregular	spiculated	0.94	0.77	0.89	0.90
C_0050_1.LEFT_CC	irregular	spiculated	0.93	0.91	0.87	0.91
C_0102_1.RIGHT_MLO	irregular	spiculated	0.92	0.90	0.85	0.88
C_0208_1.RIGHT_CC	irregular	microlobulated	0.95	0.86	0.91	0.92

TABLE 6. Training time and testing time on an image of the various segmentation methods.

Method	Training time (s)	Testing time on an image (s)
Proposed	12987	1.53
U-Net	9732	1.06
Attention U-Net	12574	1.40
DenseNet	11458	1.26

TABLE 7. Comparison of the proposed and state-of-the-art methods.

Method	F1-score (%)	Sensitivity (%)
Proposed	**82.24±0.06**	77.89±0.08
Hai et al. [36]	76.97	79.83
Le et al. [37]	74.00	69.00
Dhungel et al. [38]	-	75.00

performance of the network by suppressing feature activations that is not related to the task. Also, the amount of mammograms available is little, and the masses to be segmented are too small compared to the background area, and the required segmentation accuracy is high. So, dense-connected CNN are chosen as feature extractor. Dense connections realize feature reuse, and can get higher segmentation accuracy without training more parameters. In addition, since each layer in the dense block is directly connected to input and loss, vanishing gradient problem can be alleviated. The design of the proposed network structure is based on the specific application (breast masses segmentation), which proves that AGs and dense connections are effective on this specific application. Qualitative and quantitative comparisons demonstrate that the proposed algorithm can achieve automatic breast mass segmentation and has high segmentation precision for various size and shapes of breast tumor.

IV. CONCLUSION

In this paper, a deep learning network is employed to segment suspicious breast masses in mammograms. The proposed Attention Dense U-Net algorithm can segment breast masses in real time without manually extracting features and selecting certain parameters like traditional segmentation algorithms, and performs on par with other popular deep learning segmentation networks (U-Net, Attention U-Net, DenseNet). The increment of segmentation accuracy is due to combined

contribution of AGs and dense connections. Advances in suspicious masses segmentation provide potential benefits for masses classification.

As network is trained in more unique and more diverse situations, that is, by increasing the data set, the performance of our proposed algorithm can increase. Besides, in clinical diagnosis, medical experts typically observe and segment masses based on a number of adjacent slices along the z-axis. 2D convolutions cannot fully utilize spatial information along the 3D. As a result, to further improve the performance of the network, we will try to use a bigger data-set and 3D convolution in the future.

ACKNOWLEDGMENT

The authors would like to sincerely thank all the reviewers for the tremendous efforts towards the success of this paper. They would also like to thank the Editor-in-Chief, Associate Editor and the Administrator for their help.

REFERENCES

[1] W. Chen *et al.*, "Cancer incidence and mortality in China in 2013: An analysis based on urbanization level," *Chin. J. Cancer Res.*, vol. 29, no. 1, pp. 1–10, Feb. 2017.
[2] C. E. Desantis, J. Ma, A. G. Sauer, L. A. Newman, and A. Jemal, "Breast cancer statistics, 2017, racial disparity in mortality by state," *CA: A Cancer J. Clinicians*, vol. 67, no. 6, p. 439, Oct. 2017.
[3] O. Akin *et al.*, "Advances in oncologic imaging: Update on 5 common cancers," *CA: A Cancer J. Clinicians*, vol. 62, no. 6, pp. 364–393, Oct. 2012.
[4] M. Dong, X. Lu, Y. Ma, Y. Guo, Y. Ma, and K. Wang, "An efficient approach for automated mass segmentation and classification in mammograms," *J. Digit. Imag.*, vol. 28, no. 5, pp. 613–625, Mar. 2015.
[5] M. Dong, Z. Wang, C. Dong, X. Mu, and Y. Ma, "Classification of region of interest in mammograms using dual contourlet transform and improved KNN," *J. Sensors*, vol. 2017, Nov. 2017, Art. no. 3213680.
[6] H. Min, S. S. Chandra, N. Dhungel, S. Crozier, and A. P. Bradley, "Multi-scale mass segmentation for mammograms via cascaded random forests," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Melbourne, VIC, Australia, Apr. 2017, pp. 113–117,.
[7] A. Oliver *et al.*, "A review of automatic mass detection and segmentation in mammographic images," *Med. Image Anal.*, vol. 14, no. 2, pp. 87–110, 2010.
[8] K. Hu, X. Gao, and F. Li, "Detection of suspicious lesions by adaptive thresholding based on multiresolution analysis in mammograms," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 2, pp. 462–472, Feb. 2011.
[9] A. R. Domínguez and A. K. Nandi, "Toward breast cancer diagnosis based on automated segmentation of masses in mammograms," *Pattern Recognit.*, vol. 42, no. 6, pp. 1138–1148, Jun. 2009.
[10] E. Kozegar, M. Soryani, H. Behnam, M. Salamati, and T. Tan, "Mass segmentation in automated 3-D breast ultrasound using adaptive region growing and supervised edge-based deformable model," *IEEE Trans. Med. Imag.*, vol. 37, no. 4, pp. 918–928, Apr. s2018.
[11] D. C. Pereira, R. P. Ramos, and M. Z. do Nascimento, "Segmentation and detection of breast cancer in mammograms combining wavelet analysis and genetic algorithm," *Comput. Methods Programs Biomed.*, vol. 114, no. 1, pp. 88–101, Apr. 2014.
[12] B. Gupta and M. Tiwari, "A tool supported approach for brightness preserving contrast enhancement and mass segmentation of mammogram images using histogram modified grey relational analysis," *Multidimensional Syst. Signal Process.*, vol. 28, no. 4, pp. 1549–1567, Oct. 2017.
[13] K. L. Kashyap, M. K. Bajpai, and P. Khanna, "Globally supported radial basis function based collocation method for evolution of level set in mass segmentation using mammograms," *Comput. Biol. Med.*, vol. 87, pp. 22–37, Aug. 2017.
[14] N. Dhungel, G. Carneiro, and A. P. Bradley, "Deep learning and structured prediction for the segmentation of mass in mammograms," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Berlin, Germany, 2015, pp. 605–612.
[15] Y. Yuan, M. Chao, and Y.-C. Lo, "Automatic skin lesion segmentation using deep fully convolutional networks with Jaccard distance," *IEEE Trans. Med. Imag.*, vol. 36, no. 9, pp. 1876–1886, Sep. 2017.
[16] X. Li *et al.*, "3D multi-scale FCN with random modality voxel dropout learning for intervertebral disc localization and segmentation from multi-modality MR images," *Med. Image Anal.*, vol. 45, pp. 41–54, Apr. 2018.
[17] K. Umehara, J. Ota, and T. Ishida, "Application of super-resolution convolutional neural network for enhancing image resolution in chest CT," *J. Digit. Imag.*, vol. 31, no. 4, pp. 441–450, Aug. 2017.
[18] A. A. Mohamed, W. A. Berg, H. Peng, Y. Luo, R. C. Jankowitz and S. Wu, "A deep learning method for classifying mammographic breast density categories," *Med. Phys.*, vol. 45, no. 1, pp. 314–321, Jan. 2017.
[19] G. González *et al.*, "Disease staging and prognosis in smokers using deep learning in chest computed tomography," *Amer. J. Respiratory Crit. Care Med.*, vol. 197, no. 2, pp. 194–203, Jan. 2018.
[20] H. Chougrad, H. Zouaki, and O. Alheyane, "Deep convolutional neural networks for breast cancer screening," *Comput. Methods Programs Biomed.*, vol. 157, pp. 19–30, Apr. 2018.
[21] B. He, D. Xiao, Q. Hu, and F. Jia, "Automatic magnetic resonance image prostate segmentation based on adaptive feature learning probability boosting tree initialization and CNN-ASM refinement," *IEEE Access*, vol. 6, pp. 2005–2015, Dec. 2017.
[22] T. Kooi *et al.*, "Large scale deep learning for computer aided detection of mammographic lesions," *Med. Image Anal.*, vol. 35, pp. 303–312, Jan. 2017.
[23] Y. Li and L. Shen, "cC-GAN: A robust transfer-learning framework for HEp-2 specimen image segmentation," *IEEE Access*, vol. 6, pp. 14048–14058, 2018.
[24] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, Lake Tahoe, NV, USA, 2012, pp. 2843–2851.
[25] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, Nov. 2017, pp. 3431–3440.
[26] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (IEEE MICCAI)*, Munich, Germany, 2015, pp. 234–241.
[27] J. Jiang *et al.*, "Multiple resolution residually connected feature streams for automatic lung tumor segmentation from CT images," *IEEE Trans. Med. Imag.*, vol. 38, no. 1, pp. 134–144, Jul. 2018.
[28] O. Oktay *et al.* (Jul. 2018). "Attention U-Net: Learning where to look for the pancreas." [Online]. Available: https://arxiv.org/abs/1804.03999
[29] T. Shen, T, Zhou, G. Long, J. Jiang, S. Pan, and C. Zhang. (Sep. 2017). "DiSAN: Directional self-attention network for RNN/CNN-free language understanding." [Online]. Available: https://arxiv.org/abs/1709.04696
[30] F. Wang *et al.*, "Residual attention network for image classification," in *Proc. IEEE CVPR*, Honolulu, HI, USA, 2017, pp. 3156–3164.
[31] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, 2015, pp. 1–15. [Online]. Available: https://arxiv.org/abs/1409.0473
[32] S. Jetley, N. A. Lord, N. Lee, and P. H. S. Torr, "Learn to pay attention," in *Proc. Int. Conf. Learn. Represent.*, Vancouver, BC, Canada, 2018, pp. 1–14. [Online]. Available: https://arxiv.org/abs/1804.02391
[33] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, 2017, pp. 4700–4708.
[34] M. Heath, K. Bowyer, D. Kopans, R. Moore, and P. Kegelmeyer, "The digital database for screening mammography," in *Proc. 5th Int. Workshop Digit. Mammogr.*, Toronto, ON, Canada, 2017, pp. 212–218.
[35] J. Hai *et al.*, "Fully convolutional densenet with multiscale context for automated breast tumor segmentation," *J. Healthcare Eng.*, vol. 2019, Jan. 2019, Art. no. 8415485.
[36] L. Sun *et al.*, "An image segmentation framework for extracting tumors from breast magnetic resonance images," *J. Innov. Opt. Health Sci.*, vol. 11, pp. 1850014-1–1850014-15, Jul. 2018.
[37] N. Dhungel, G. Carneiro, and A. P. Bradley, "Automated Mass Detection from Mammograms using Deep Learning and Random Forest," in *Proc. Breast Image Anal. Workshop(BIA), 18th Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, Munich, Germany, 2015, pp. 1–8.

SHUYI LI received the B.Sc. degree from the School of Information Science and Engineering, Hunan University. She is currently pursuing the master's degree with the Institute of Industrial Technology, Zhengzhou University. Her current research interests include medical image processing and analysis, deep learning, and pattern recognition.

MIN DONG received the M.A.Sc. and Ph.D. degrees from the School of Information Science and Engineering, Lanzhou University. She is currently a Lecturer with the School of Information Engineering, Zhengzhou University. Her current research interests include image analysis and processing, pattern recognition, and signal detection and estimation.

GUANGMING DU received the B.Sc. degree from the School of Electrical and Computer Engineering, Jilin Jianzhu University. He is currently pursuing the master's degree with the Institute of Industrial Technology, Zhengzhou University. His current research interests include medical image processing and analysis, pattern recognition, and signal detection and estimation.

XIAOMIN MU received the B.E. degree from the Beijing Institute of Technology, Beijing, China, in 1982. She is currently a Full Professor with the School of Information Engineering, Zhengzhou University. She has authored many papers in the field of signal processing and coauthored two books. Her research interests include signal processing in communication systems, wireless communications, and cognitive radio.