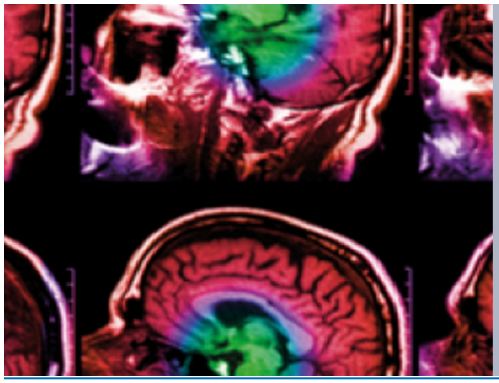
PAPER

AUNet: attention-guided dense-upsampling networks for breast mass segmentation in whole mammograms

To cite this article: Hui Sun *et al* 2020 *Phys. Med. Biol.* **65** 055005



IOP Publishing Phys. Med. Biol. 65 (2020) 055005 (17pp) https://doi.org/10.1088/1361-6560/ab5745

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

PAPER

RECEIVED
25 March 2019

REVISED
14 October 2019

ACCEPTED FOR PUBLICATION
13 November 2019

PUBLISHED
28 February 2020

AUNet: attention-guided dense-upsampling networks for breast mass segmentation in whole mammograms

Hui Sun[1,2,6], **Cheng Li**[1,6], **Boqiang Liu**[2], **Zaiyi Liu**[3], **Meiyun Wang**[4], **Hairong Zheng**[1], **David Dagan Feng**[5] and **Shanshan Wang**[1,7]

[1] Paul C. Lauterbur Research Center for Biomedical Imaging, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong 518055, People's Republic of China
[2] School of Control Science and Engineering, Shandong University, Jinan, Shandong 250100, People's Republic of China
[3] Department of Radiology, Guangdong General Hospital, Guangdong Academy of Medical Sciences, Guangzhou, Guangdong 510080, People's Republic of China
[4] Department of Radiology, Henan Provincial People's Hospital, Zhengzhou, Henan 450003, People's Republic of China
[5] The Biomedical and Multimedia Information Technology Research Group, School of Information Technologies, The University of Sydney, Sydney, NSW 2006, Australia
[6] These authors contribute equally to this paper.
[7] Author to whom any correspondence should be addressed.

E-mail: sophiasswang@hotmail.com

Keywords: breast cancer, mammogram, segmentation, deep learning

Supplementary material for this article is available online

Abstract

Mammography is one of the most commonly applied tools for early breast cancer screening. Automatic segmentation of breast masses in mammograms is essential but challenging due to the low signal-to-noise ratio and the wide variety of mass shapes and sizes. Existing methods deal with these challenges mainly by extracting mass-centered image patches manually or automatically. However, manual patch extraction is time-consuming and automatic patch extraction brings errors that could not be compensated in the following segmentation step. In this study, we propose a novel attention-guided dense-upsampling network (AUNet) for accurate breast mass segmentation in whole mammograms directly. In AUNet, we employ an asymmetrical encoder–decoder structure and propose an effective upsampling block, attention-guided dense-upsampling block (AU block). Especially, the AU block is designed to have three merits. Firstly, it compensates the information loss of bilinear upsampling by dense upsampling. Secondly, it designs a more effective method to fuse high- and low-level features. Thirdly, it includes a channel-attention function to highlight rich-information channels. We evaluated the proposed method on two publicly available datasets, CBIS-DDSM and INbreast. Compared to three state-of-the-art fully convolutional networks, AUNet achieved the best performances with an average Dice similarity coefficient of 81.8% for CBIS-DDSM and 79.1% for INbreast.

1. Introduction

Latest investigations demonstrate that breast cancer persists as one of the most threatening cancer types to female, accounting for 29% of cancer incidence and 15% of cancer mortality in women (Siegel *et al* 2017). Early diagnosis of breast cancer is vital for the survival of patients. Mammography is one of the most effective and efficient breast cancer screening tools. However, analyzing mammograms by radiologists is tedious and the interpretations are subject to substantial inter- and intra-observer variations, which may lead to missed cancers as well as overdiagnosis (Birdwell *et al* 2001, Løberg *et al* 2015). Therefore, a computer-aided detection/diagnosis (CAD) system that can work as a second reader is important and necessary.

Various types of abnormalities may show in mammograms, such as asymmetrical breast tissues, adenopathy, density, microcalcifications, and masses. Among them, breast masses are believed to contribute significantly to

breast cancers (Giger *et al* 2013). Currently, the majority of breast mass studies concentrated on image-level lesion detection and patch-level mass classification or segmentation (Wei *et al* 2006, Dhungel *et al* 2015b, 2017, Jiang *et al* 2016, Han *et al* 2017, Kim *et al* 2018). However, image-level lesion detection can only give the bounding box of the mass without the boundary information, which has been identified as an important indicator of its malignancy (Guliato *et al* 2008). And patch extraction around the mass before segmentation is a tedious and difficult work for radiologists. Therefore, mass segmentation of whole mammograms is of high application value for breast cancer detection and diagnosis. Specifically, our focus in this study is the automatic breast mass segmentation in whole mammograms, i.e. the segmentation in full fields of view (FOVs) of input mammograms rather than extracted regions of interest (ROIs).

Recently, deep learning models, especially convolutional neural networks (CNNs), have seen great successes in computer vision and medical imaging (Greenspan *et al* 2016, Litjens *et al* 2017, Hamidinekoo *et al* 2018). In respect of medical image segmentation, the most well-known network is UNet (Ronneberger *et al* 2015) and UNet-like architectures are frequently investigated (Balagopal *et al* 2018, Li *et al* 2019b). However, most deep learning-based models developed for mammographic mass segmentation focus on extracted patches instead of the original whole mammograms (Hai *et al* 2019). In addition to the limited studies, existing deep learning-based studies conduct whole mammographic mass segmentation by simply combing classic models with some effective network modules developed for natural image processing. Atrous spatial pyramid pooling and attention gates have been introduced to FCDenseNet and Dense-U-Net to enhance the segmentation capacity (Hai *et al* 2019, Li *et al* 2019a). In these studies, both the network architecture and the added modules were not specifically optimized for the breast mass segmentation purpose. There is still a large gap to be filled and a lot of work to be done. Besides, although CAD systems have been widely developed to assist radiologists in identifying suspicious regions, their performance can still be improved since contradictory conclusions exist regarding their effectiveness in mammogram interpretation (Lehman *et al* 2015, Kooi *et al* 2017). Therefore, we feel motivated to investigate the whole mammographic mass segmentation project, which is expected to be a significant add-on to the current CAD system for mammographic diagnosis.

In this paper, we propose a new model, attention-guided dense-upsampling network (AUNet), for the segmentation of mammographic masses. Different from the classical symmetric encoder–decoder architecture of UNet (Ronneberger *et al* 2015), AUNet employs an asymmetrical structure—different encoder and decoder blocks—through the implementation of residual connections. Furthermore, we design a novel upsampling module, attention-guided dense-upsampling block (AU block), to compensate the information loss caused by bilinear upsampling, effectively fuse the high- and low-level features, and at the same time, highlight the rich-information channels. The performance of the proposed network was evaluated on two public mammographic datasets, CBIS-DDSM and INbreast. With AUNet, we achieved an average Dice score of 81.8% for CBIS-DDSM and 79.1% for INbreast. Both improved the segmentation results of UNet by more than 8.0%. Our major contributions are: (1) A more effective asymmetric encoder–decoder network architecture is introduced; (2) We propose a new block, AU block, that can effectively extract important information from both high- and low-level features; (3) AU block can serve as a universal decoder module that is compatible with any encoder–decoder segmentation network; (4) Implementing both AU block and the asymmetrical structure, our proposed network, AUNet, is able to accurately segment masses in whole mammograms without the need of ROI extraction; (5) Better breast mass segmentation performances were achieved by AUNet compared to commonly utilized fully convolutional networks (FCNs) in medical imaging. Our code is available at https://github.com/lich0031/AUNet.

2. Related works

In this section, we review the related works on deep learning models for image segmentation and existing methods for mammographic mass segmentation.

2.1. Segmentation networks

Since the introduction of FCNs in 2015 (Long *et al* 2015), most segmentation models follow a similar encoder–decoder network backbone design. The encoder pathway first extracts high dimensional and high abstract feature maps from the inputs, usually with severely decreased resolutions, and then the decoder pathway is responsible for the recovery of image resolution and generation of the segmentation results. However, due to the information loss during the encoding process by pooling or convolution with strides, the reconstructed segmentation results are usually not satisfactory. To solve this issue, works have been done to include conditional random fields as a post processing method, which has shown a significant improvement (Kamnitsas *et al* 2017, Chen *et al* 2018). Another direction is the application of dilated convolution (Yu *et al* 2017). Dilated convolution can increase the receptive field and, in the meantime, keep the image resolution unchanged. Nevertheless, limited by the current available computing power, dilated convolution at high image resolutions is hard to achieve if not impossible

(Yu *et al* 2017). UNet proposed another solution to the problem (Ronneberger *et al* 2015). The main idea of UNet is to fuse high-level feature maps that are rich in semantic information with low-level feature maps that are rich in location information. By fusing feature maps from different layers, UNet is capable of generating accurate segmentation maps for small datasets. However, the feature fusion of UNet is done through simple concatenation, which is not effective enough and improvement is necessary for different applications (Lin *et al* 2017, Zhang *et al* 2018).

2.2. Upsampling approaches

Different methods have been adopted in literature to upsample the low-resolution feature maps. Bilinear interpolation is a simple and efficient method that has been commonly used (Zhao *et al* 2017, Chen *et al* 2018). The output of bilinear interpolation is fixed and not learnable, which may cause information loss (Wang *et al* 2018a). Deconvolution was first proposed along with FCNs (Long *et al* 2015) and adopted in later works. Deconvolution could be realized in two ways. One is through the reverse operation of convolution (Long *et al* 2015). The other is through unpooling, where the low-resolution feature maps are first upsampled to high-resolution feature maps using the stored max pooling indices and then the sparse feature maps are densified by convolutions (Badrinarayanan *et al* 2017). Both methods result in learnable upsampling procedure but require zero padding at the first step. The last method is dense upsampling convolution (DUC) (Wang *et al* 2018a), derived from the sub-pixel convolution method originally developed for image super resolution task (Shi *et al* 2016). DUC is also learnable. In addition, different from deconvolution, no zero padding is required for DUC.

2.3. Attention mechanism

Attention mechanism in neural networks has attracted a lot of attention recently. It is proposed in accordance with the human visual attention that human beings always focus on a certain part of a given image after quickly glimpsing through it. Attention could be viewed as a tool to force the network focusing on the most informative part of the inputs or features (Mnih *et al* 2014). It has been widely applied in natural language processing and image captioning (Chen *et al* 2017, Vaswani *et al* 2017). Studies also found that CNNs could learn implicitly to localize the most important regions of the input images (Zhou *et al* 2016), which could be treated as a kind of attention. To improve image classification accuracies, both spatial and channel-wise attention modules have been proposed in literatures (Hu *et al* 2018, Roy *et al* 2018). Attention has also been explicitly used for image segmentation (Mirikharaji and Hamarneh 2018, Nie *et al* 2018). Different from these works, which utilize attention mechanism to focus on regions of inputs, our proposed AU block implements attention to select important channels for breast mass segmentation.

2.4. Segmentation of mammographic mass

Automatic mammographic mass segmentation methods could be divided into unsupervised and supervised methods. Unsupervised methods include region-based (Gulsrud *et al* 2005, Wei *et al* 2006), contour-based (Shi *et al* 2008, Rahmati *et al* 2012), and clustering models (Ball *et al* 2004, Abdel-Dayem and EI-Sakka 2005). These models encounter various problems when applied to mammographic mass segmentation (Oliver *et al* 2010). Region-based models rely on region homogeneity and prior information is usually needed, such as the locations of seeding points and shape information (Kupinski and Giger 1998). Contour-based models are based on edge detection whereas it is challenging to extract the boundary between masses and normal breast tissues (Sahiner *et al* 2001). Hierarchical clustering models are computational expansive while partitional clustering models need to know the number of regions in advance (Li *et al* 2002). Supervised methods have a training and testing procedure. Pattern matching is widely used for segmentation and detection (Freixenet *et al* 2008, Song *et al* 2010). Nonetheless, mammographic masses can be in a wide variety of shapes, which hinders the usage of pattern matching approaches (Oliver *et al* 2010). Deep learning models belong to supervised methods. Deep structured models have been successfully applied to segmenting masses from ROIs rather than whole mammograms (Dhungel *et al* 2015b, 2015c, 2017). And using manually extracted ROIs could improve the segmentation performance compared to automatically detected bounding boxes generated by detection models (Dhungel *et al* 2017), which indicates that the segmentation results depend on the patch extraction process and it is difficult to achieve fully automatic mammographic mass segmentation employing this approach. Very few attempts on mass segmentation of whole mammograms could be found probably caused by the previously discussed difficulties (Hai *et al* 2019). These studies mainly combined famous segmentation models with some special network modules developed for natural image analysis. For example, atrous spatial pyramid pooling and attention gates have been introduced to FCDenseNet and Dense-U-Net to enhance the segmentation capacity (Hai *et al* 2019, Li *et al* 2019a). Considering the gap between medical and natural image domains, these models may not be perfectly suitable for the breast mass segmentation task. Moreover, the experiments were not comprehensive, and the models were not publicly available. Aiming to address these challenges, our AUNet

is designed specifically for fully automatic mammographic mass segmentation. Two public datasets have been tested and the models are publicly available.

3. Methodology

In this section, we first describe the datasets used in the study. Then, the proposed network architecture, including the asymmetrical encoder–decoder backbone and the AU block, is presented. After that, loss function selection is discussed. Finally, quantitative evaluation metrics are listed.

3.1. Datasets

We instantiated our proposed network with two publicly available datasets, CBIS-DDSM (Heath *et al* 2000, Lee *et al* 2017) and INbreast (Moreira *et al* 2012). For CBIS-DDSM, a total of 858 images were used in the current study with 690 images for training and 168 for validation. The INbreast dataset contains 107 images with accurate mass segmentation masks. A 5-fold cross-validation experiment was conducted for INbreast.

All the images along with the masks were first processed to remove the irrelevant background regions (rows and columns have negligible maximum intensities) and then resized to 256×256, followed by an intensity normalization. Before inputting into the networks, the gray images were changed to RGB images by copying the pixel values to the other two channels. The importance of this step will be discussed later. No further data processing or augmentation was applied.

Figure 1(a) shows representative images from the two datasets. It could be observed that mammographic masses are in a wide variety of shapes and sizes, which increases the difficulty of training the segmentation network. Figures 1(b) and (h) give the area ratio distributions of the two datasets. Both indicate that most masses only occupy very small regions of the whole mammograms. Results confirm more than 81.8% masses occupy less than 1.0% area of the whole mammograms for CBIS-DDSM. For INbreast, more than 81.0% masses occupy less than 4.0% area of the whole mammograms. Therefore, it is much more difficult to train a network capable of accurately segmenting masses in whole mammograms than in mass-centered mammographic patches. Other available important information including subtlety, mass shape and margin, BIRADS category, and pathology are also plotted in figure 1 to comprehensively describe the datasets.

3.2. Asymmetrical network backbone

Our proposed network employs an encoder–decoder architecture backbone (figure 2(a)). The encoder pathway contains five encoder blocks with the first four followed by max pooling. Thus, the downsampling ratio is 16 in total. The decoder pathway is composed of four alternating upsampling and decoder blocks. The upsampling block will be discussed in the next section. The classic UNet employs symmetrical encoder and decoder pathways, where the basic unit (figure 2(b)) is implemented for both the encoder and decoder blocks (Ronneberger *et al* 2015). Although this simple design contributes to the efficiency of the network, the effectiveness needs to be explored. Inspired by the recently wide spread usage of ResNet (He *et al* 2016), we investigated the feasibility of another two configurations, deep unit (figure 2(c)) and res unit (figure 2(d)).

For the three different units, we have the respective outputs as follows:

$$y_{basic}(x) = \delta(W_{b2} * \delta(W_{b1} * x + b_{b1}) + b_{b2}) \tag{1}$$

$$y_{deep}(x) = \delta(W_{d3} * \delta(W_{d2} * \delta(W_{d1} * x + b_{d1}) + b_{d2}) + b_{d3}) \tag{2}$$

$$y_{res}(x) = \delta((W_{r3} * \delta(W_{r2} * \delta(W_{r1} * x + b_{r1}) + b_{r2}) + b_{r3}) + (W_{r1} * x + b_{r1})) \tag{3}$$

where y is the respective output of the different units and x is the corresponding input. δ refers to the ReLU function. W and b refer to the weights and bias of the different convolution layers. $*$ is the convolution operation.

Moreover, we also evaluated the different combinations of applying the three units as the encoder/decoder block. In the results section, we will show that constructing an asymmetrical network backbone by applying the res unit as the encoder block and the basic unit as the decoder block, the network can achieve the best segmentation performance.

3.3. Attention-guided dense-upsampling block

Our major novelty regarding the network design lies in the upsampling block, where we introduce our proposed AU block (figure 3(b)). The original UNet used deconvolution to upsample the feature maps (Ronneberger *et al* 2015). However, our preliminary experiments found that deconvolution was not as effective as bilinear upsampling for our application (supplementary file table S1 (stacks.iop.org/PMB/65/055005/mmedia)), and thus, bilinear upsampling was utilized throughout the study.

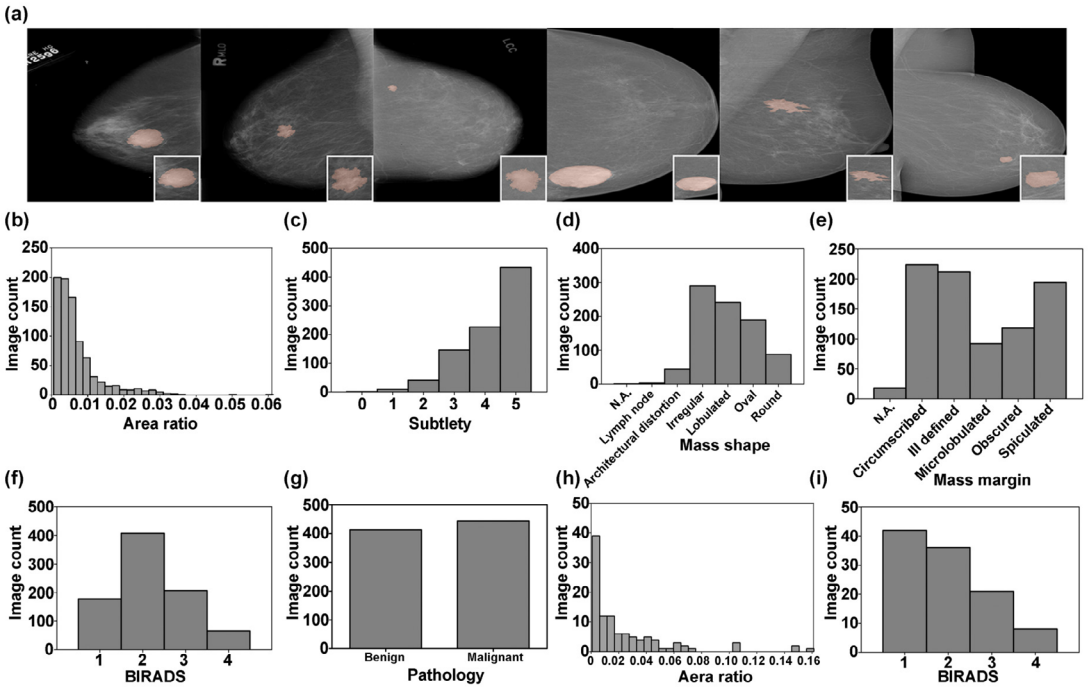


Figure 1. Representative mammographic images (a) and distributions of the area ratios (area ratio = area of the mass/area of the whole mammogram) of masses, the subtlety, the mass shape, the mass margin, the BIRADS category, and the pathology diognosis for the CBIS-DDSM dataset (b)–(g) and area ratios of masses and the BIRADS category for the INbreast dataset (c). The left three images in (a) are from the CBIS-DDSM dataset and the right three from the INbreast. Pink color regions indicate the masses. Insets are enlarged patches containing the masses. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The bilinear upsampling block (BU block) of UNet is shown in figure 3(a), where the high-level features are simply upsampled and concatenated with the low-level features after passing a convolution layer. The goal of the proposed AU block (figure 3(b)) is to extract all important information from both high- and low-level features. The high-level low-resolution features (F_{high}) are firstly upsampled using two different methods. One is dense upsampling convolution (F_{duc}), and the other is bilinear upsampling followed by a convolution layer (F_{buc}). The convolution layer is always followed by batch normalization and ReLU activation unless otherwise specified. Then, F_{duc} is combined with the low-level features (F_{low}) by summation (F_{sum}). A convolution layer is applied before F_{sum} is concatenated with F_{buc} (F_{concat}) to smooth the concatenation process. In this way, we expect that F_{concat} contains all the information from both F_{high} and F_{low}.

The next step is to select the important information from F_{concat}. Motivated by the squeeze-and-excitation networks (Hu *et al* 2018), we adopt a channel-wise attention. Firstly, global average pooling is applied to obtain a channel-wise descriptor Z_c:

$$Z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} (F_{concat,c}(i,j)) \tag{4}$$

where $F_{concat,c}$ is the (c^{th}) channel of F_{concat}. H and W refer to the height and width of $F_{concat,c}$. Z_c passes through two fully connected layers (FC layers), one with ReLU and one without, and a Sigmoid function to get the channel-wise weights S:

$$S = \sigma(W_2 * \delta(W_1 * Z + b_1) + b_2) \tag{5}$$

where σ refers to the Sigmoid function. $W_1 \in R^{2n/r \times 2n}$, $W_2 \in R^{2n \times 2n/r}$, $b_1 \in R^{2n/r}$, and $b_2 \in R^{2n}$ are the weights and bias of the FC layers, respectively. r is a reduction ratio. The output of the AU block is:

$$\tilde{F}_{concat,c} = S_c \cdot Z_c. \tag{6}$$

After that, $\tilde{F}_{concat,c}$ goes through a basic unit (figure 2(b)), which is composed of two convolution layers, and then, is treated as the high-level feature input to the next AU block.

3.4. Loss function

The commonly used cross-entropy loss function for two-class segmentation task is defined as:

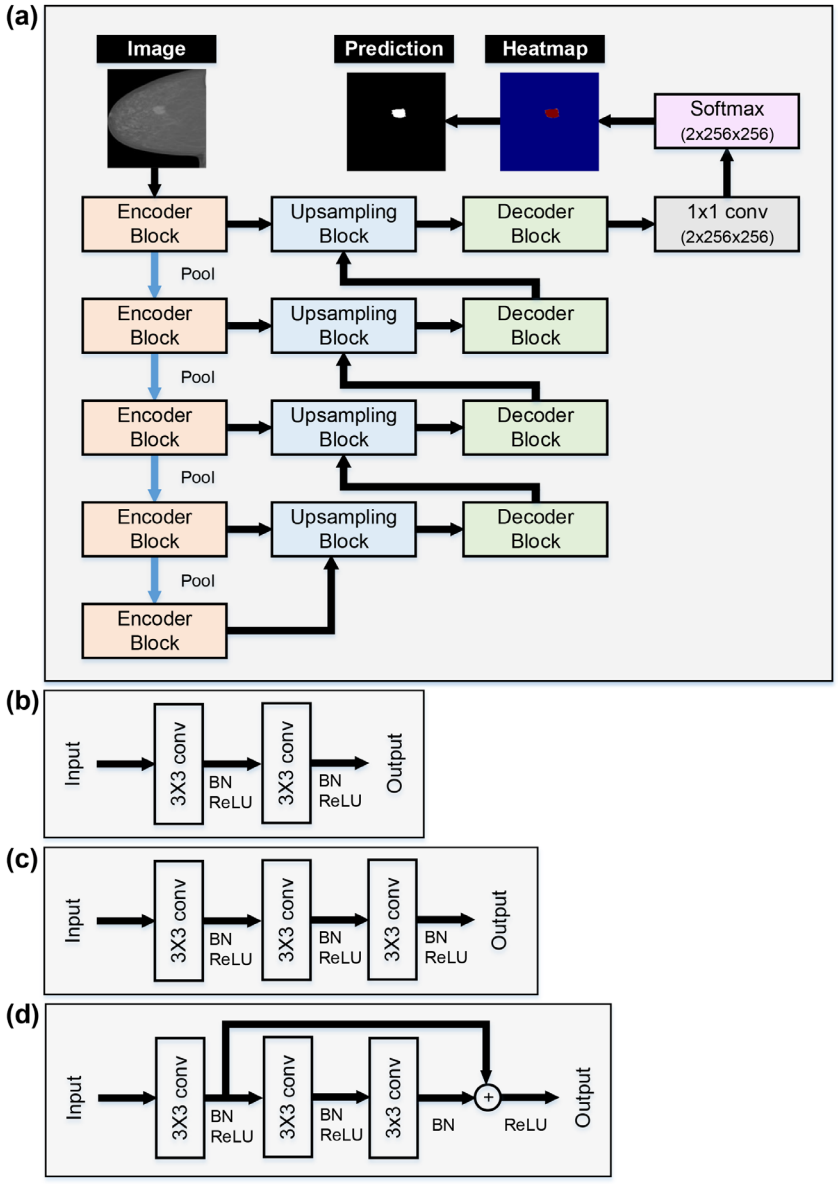


Figure 2. Network architecture and different blocks. (a) Architecture backbone. (b)–(d) Different blocks named as basic unit (b), deep unit (c), and res unit (d).

$$L_{CE} = -\frac{1}{N}(y_i \sum_{i=1}^{N} p_i + (1 - y_i) \sum_{i=1}^{N} (1 - p_i)). \tag{7}$$

For 2D inputs, N is the total number of pixels in the image. $y_i \in \{0, 1\}$ is the ground truth label of the i^{th} pixel with 0 refers to the background and 1 refers to foreground. $p_i \in [0, 1]$ is the corresponding predicted probability of the pixel belonging to the foreground class.

From the definition, positive and negative pixels contribute equally to the cross-entropy loss. However, from figure 1, we know a severe class imbalance problem exists for both datasets that masses only occupy small regions of the whole mammograms. Minimization of the cross-entropy loss function may bias the model towards correctly predicting the negative class. To solve this issue, we introduced another loss function, the Dice loss. The Dice loss in our situation is defined as:

$$L_{Dice} = 1 - \frac{2\sum_{i=1}^{N} p_i y_i + \varepsilon}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} y_i + \varepsilon} \tag{8}$$

where ε is a constant to keep numerical stability. It has been reported that applying only the Dice loss makes the optimization process unstable (Zhu *et al* 2019). Therefore, we use a combined loss function for our model, which is defined as:

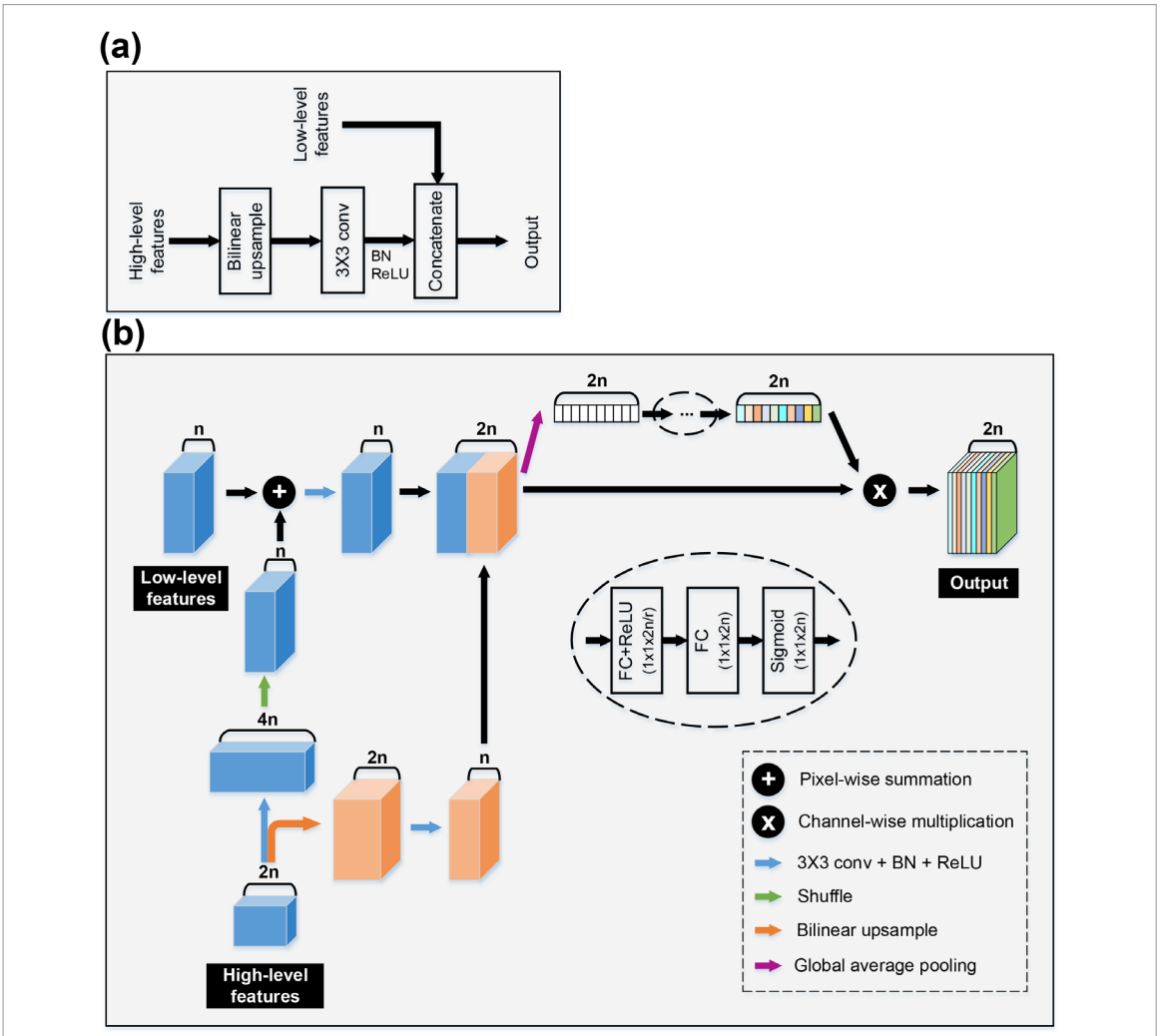


Figure 3. Upsampling block. (a) The bilinear upsampling block (BU block). (b) The attention-guided dense-upsampling block (AU block). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$L = L_{Dice} + \alpha L_{CE} \tag{9}$$

where α is a weight constant to control the trade-off between the cross-entropy loss and the Dice loss.

3.5. Evaluation metrics

To quantitatively evaluate the proposed model, we use Dice similarity coefficient (*DSC*), sensitivity (*SEN*), relative area difference (ΔA), and Hausdorff distance (*HAU*) to characterize the performances of the methods on the test datasets. We use the overall average metrics to select the best model during the network architecture optimization. To comprehensively compare our final model to the existing networks, in addition to the overall average metrics, we also evaluate the results with respect to the image properties for the CBIS-DDSM dataset (figures 1(c)–(g)). *DSC*, *SEN*, ΔA, and *HAU* are defined as:

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{10}$$

$$SEN = \frac{TP}{TP + FN} \tag{11}$$

$$\Delta A = \frac{|A_{pred} - A_{GT}|}{A_{GT}} = \frac{|(TP + FP) - (TP + FN)|}{TP + FN} \tag{12}$$

$$HAU = max(h(pred, GT), h(GT, pred)) \tag{13}$$

where *pred* refers to network predictions and *GT* referes to ground truth segmentations. A_{pred} refers to the predicted mass area and A_{GT} refers to the ground-truth mass area. *TP*, *FP*, and *FN* refer to true positives, false

positives, and false negatives. $h(A,B) = max(a \in A)(b \in B)||a-b||$ and $||\cdot||$ refers to the L2 distance between the two points.

Differences between the different models were evaluated by Wilcoxon signed-rank test with a significance threshold of $p < 0.05$.

3.6. Experimental set-up

Our proposed network as well as the comparison models were implemented with PyTorch (Paszke *et al* 2017). Network training and testing were run on a NVIDIA GeForce GTX 1080Ti GPU (11GB) with batch size of 4. We used ADAM with the AMSGRAD optimization method (Reddi *et al* 2017). The learning rate was initially set to 1×10^{-4}, and step decay policy was applied, specifically with [40, 30, 30, 20] epochs at the learning rate of $[1 \times 10^{-4}, 5 \times 10^{-5}, 1 \times 10^{-5}, 1 \times 10^{-6}]$. The INbreast dataset contains 107 images, which may limit the proper training of a deep neural network. Therefore, we tried to fine-tune the models pretrained on the CBIS-DDSM dataset. We set the respective hyper-parameters in (8) and (9) empirically to $\varepsilon = 1.0$ and $\alpha = 1.0$. We have tested α with different values (0.5, 1.0 and 2) and found that 1.0 achieved the best segmentation performance (supplementary file table S2). The determination of the reduction ratio r will be discussed in the results section.

To validate the effectiveness of our proposed AUNet, we conducted ablation experiments. Specifically, to select the best network backbone, we have tried to substitute the encoder and decoder blocks in figure 2(a) with the deep unit (figure 2(c); Deep-UNet) or res unit (figure 2(d); Res-UNet) but keep the BU block (figure 3(a)) unchanged. In addition, different combinations of the encoder and decoder units have been tested to check the feasibility of symmetric and asymmetric structures. Finally, we compare the segmentation results of the proposed AUNet with three established FCNs, UNet (Ronneberger *et al* 2015), FusionNet (Quan *et al* 2016), and FCDenseNet (Jégou *et al* 2017). The original UNet utilizes deconvolution for upsampling. However, experimental results demonstrated that bilinear upsampling is more effective for our application (supplementary file table S1). We adopted bilinear upsampling for all the networks. FusionNet introduces residual connections to UNet and increases the network depth by adding more convolution layers in each unit (5 convolutions per unit). FCDenseNet103 extends the recently published architecture DenseNet to fully convolutional networks for image segmentation task. Similarly, all the networks were trained from scratch for the CBIS-DDSM dataset and fine-tuning was investigated on the INbreast dataset. We show that although FusionNet and FCDenseNet103 are much deeper than AUNet, AUNet can still generate better segmentation results, which highlights the effectiveness of the proposed AU block. Three independent experiments were done for each network and the results are presented as (*mean* $\pm$ *s.d.*).

4. Experimental results

In this section, we present the results on the two public datasets, CBIS-DDSM and INbreast, and compare the results of the proposed AUNet to other FCNs.

4.1. Results on CBIS-DDSM dataset

In this section, we firstly discuss the choice of the different encoder/decoder blocks. Then the determination of the reduction ratio r is demonstrated. Finally, we compare the results of the optimized AUNet to the three FCNs.

4.1.1. Optimization of the network backbone

Results of networks employing different encoder and decoder blocks are presented in table 1. The model names indicate the units applied with the first word referring to the encoder block and the second referring to the decoder block. For example, the model Basic-Deep-UNet means we utilized the basic unit (figure 2(b)) for the encoder pathway and the deep unit (figure 2(c)) for the decoder pathway. From table 1, two general conclusions could be made: (a) Deeper networks generally achieve better performances with higher *DSC*, higher *SEN*, lower ΔA, and lower *HAU* (compare UNet to Deep–Deep-UNet); (b) Models with asymmetric structures, especially those employing the basic unit in only one pathway, perform better than models with symmetric structures (compare Res-Basic-UNet to Res–Res-UNet and Res-Deep-UNet).

By taking all the four evaluation parameters into consideration, we selected the model 'Res-Basic-UNet' as our network backbone since it achieves the highest average *DSC* (0.775 ± 0.002) among all the models and, in the meantime, comparable *SEN* (0.823 ± 0.015 versus 0.825 ± 0.008), ΔA (0.352 ± 0.036 versus 0.347 ± 0.009), and *HAU* (3.18 ± 0.04 versus 3.16 ± 0.03) to the respective best results.

4.1.2. Performance enhancement by the AU block

The introduction of the AU block (figure 3(b)) to our network backbone brings an obvious performance increment shown by all the four evaluation characteristics (table 2). The reduction ratio r is very important for the capacity and computational cost of the proposed AUNet. Therefore, we have conducted experiments to

Table 1. Ablation experiments employing different encoder–decoder blocks.

Models	*DSC* (%)	*SEN* (%)	ΔA (%)	*HAU*
UNet (Basic–Basic)	73.6 ± 0.2	79.4 ± 1.3	42.7 ± 3.1	3.38 ± 0.04
Basic-Deep-UNet	74.3 ± 0.1	78.8 ± 0.4	37.7 ± 1.2	3.28 ± 0.05
Basic-Res-UNet	74.6 ± 0.3	80.0 ± 0.7	42.0 ± 1.6	3.31 ± 0.07
Deep-Basic-UNet	77.3 ± 0.5	**82.5 ± 0.8**	36.6 ± 0.2	**3.16 ± 0.03**
Deep–Deep-UNet	77.3 ± 0.4	81.3 ± 1.2	35.7 ± 1.8	3.19 ± 0.04
Deep-Res-UNet	76.9 ± 0.7	81.6 ± 1.1	36.4 ± 1.9	**3.16 ± 0.05**
Res-Basic-UNet	**77.5 ± 0.2**	**82.3 ± 1.5**	**35.2 ± 3.6**	**3.18 ± 0.04**
Res-Deep-UNet	76.3 ± 0.2	81.9 ± 1.5	38.4 ± 3.8	3.23 ± 0.04
Res–Res-UNet	76.2 ± 0.1	80.4 ± 0.2	**34.7 ± 0.9**	3.19 ± 0.02

Table 2. Investigating the influence of the reduction ratio *r*.

Reduction ratio	*DSC* (%)	*SEN* (%)	ΔA (%)	*HAU*
2	80.6 ± 0.2	83.7 ± 0.9	29.1 ± 1.2	3.02 ± 0.03
4	80.8 ± 0.2	84.6 ± 0.4	28.5 ± 0.8	2.97 ± 0.02
8	81.0 ± 0.3	84.4 ± 0.3	29.1 ± 2.4	2.97 ± 0.06
16	**81.8 ± 0.0**	**84.9 ± 0.3**	**26.9 ± 0.3**	**2.96 ± 0.03**
32	80.8 ± 0.0	84.1 ± 0.5	28.5 ± 1.4	2.98 ± 0.01

finalize the selection. A wide range of r has been tested from 2 to 32. Results indicate that with $r = 16$, the best model performance could be achieved (table 2). Besides, it could also be observed that regardless of the choice of r, the proposed AU Block could always enhance the segmentation performance compared to the selected network backbone (Res-Basic-UNet), which demonstrates the general effectiveness of the proposed block. For all the following experiments, $r = 16$ is applied unless otherwise specified.

4.1.3. Comparison to established FCNs

Our proposed AUNet achieves better segmentation metrics when compared to established FCNs (table 3). Comparing among the three established models, FusionNet gives the highest *DSC*, the lowest ΔA, and the lowest *HAU*, whereas FCDenseNet103 presents the highest *SEN*. This indicates that FCDenseNet103 increases its capability of finding the mass locations by generating more false positives. Since FCDenseNet103 is much deeper than the other networks, it suggests that very deep networks perform worse on the mammographic datasets probably caused by overfitting. On the other hand, our proposed AUNet achieves the highest *DSC*, the highest *SEN*, the lowest ΔA, and the lowest *HAU*, which demonstrates the suitability of our proposed network for our whole mammographic mass segmentation task. Our model shows an average *DSC* increase of at least 2.0% (statistically significant with $p < 0.05$ by Wilcoxon signed-rank test) compared to the best performed FCNs.

Considering the inherent differences among the images having different categories (subtlety, BIRADS, mass shape, mass margin, and pathology), the segmentation performances of the different networks are also presented with regards to these properties. Combining the different categories (21 in total: 5 subtlety groups, 4 BIRADS categories, 5 shape groups, 5 margin categories, and 2 pathology groups) with the different evaluation metrics (*DSC*, *SEN*, ΔA, and *HAU*), there are 84 cases (detalied results in supplementary file tables S3–S6). Overall, our AUNet still achieves the best results, ranking the 1^{st} in 56 cases (16 for *DSC*, 11 for *SEN*, 14 for ΔA, and 15 for *HAU*). FusionNet and FCDenseNet103 obtain the best results in 15 and 10 cases, respectively. UNet performs the worst in this aspect with only 3 1^{st} cases.

To directly compare the performances of the different networks, the empirical cumulative distributions of *DSC* were plotted (figure 4). The closer the distribution line to the lower right position in the figure, the more images are segmented with high *DSC* values by the corresponding network. Thus, we could conclude that for the CBIS-DDSM dataset, AUNet achieves a relatively better mass segmentation performance, followed by FusionNet, FCDenseNet, and UNet.

4.2. Results on INbreast dataset

The INbreast dataset is smaller than the CBIS-DDSM dataset. As such, we tried to re-use the CBIS-DDSM trained models and fine-tuned those models using the INbreast dataset. Moreover, 5-fold cross-validation experiments were conducted to generate meaningful and convincing results.

The segmentation results of the proposed AUNet and the three established models with/without pretraining on CBIS-DDSM are listed in table 4. It could be observed that with or without the pretraining step, AUNet

Table 3. Segmentation performance of different FCNs on the CBIS-DDSM dataset.

Models	*DSC* (%)	*SEN* (%)	ΔA (%)	*HAU*
UNet	73.6 ± 0.2	79.4 ± 1.3	42.7 ± 3.1	3.38 ± 0.04
FusionNet	79.8 ± 0.5	83.9 ± 0.8	31.3 ± 0.5	3.01 ± 0.03
FCDenseNet103	78.2 ± 0.1	84.2 ± 0.6	40.2 ± 0.3	3.13 ± 0.04
AUNet	**81.8 ± 0.0**	**84.9 ± 0.3**	**26.9 ± 0.3**	**2.96 ± 0.03**

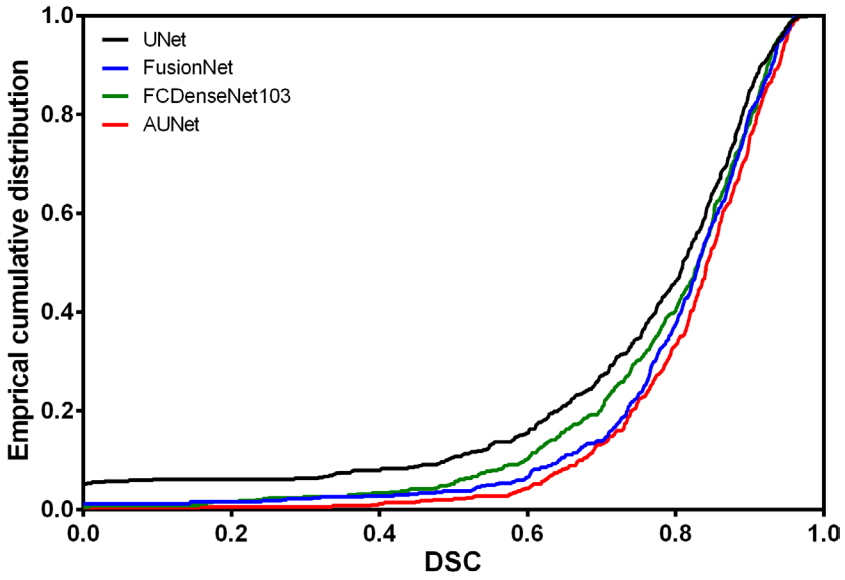


Figure 4. Empirical cumulative distribution of *DSC* on CBIS-DDSM. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4. Segmentation performance of different FCNs on the INbreast dataset.

Models	*DSC* (%)	*SEN* (%)	ΔA (%)	*HAU*
UNet (w/o[a])	62.3 ± 3.7	62.7 ± 4.0	54.3 ± 19.7	4.73 ± 0.26
UNet (w/[b])	69.3 ± 6.8	70.4 ± 8.8	44.0 ± 13.3	4.54 ± 0.42
FusionNet (w/o)	62.1 ± 5.8	65.1 ± 5.4	62.7 ± 30.9	4.80 ± 0.33
FusionNet (w/)	73.2 ± 5.8	74.6 ± 5.4	69.8 ± 33.8	4.33 ± 0.34
FCDenseNet103 (w/o)	42.9 ± 8.5	52.8 ± 11.9	149.5 ± 71.8	6.20 ± 0.52
FCDenseNet103 (w/)	76.1 ± 4.6	77.9 ± 4.7	47.1 ± 17.3	4.35 ± 0.35
AUNet (w/o)	64.0 ± 7.6	66.0 ± 7.4	51.6 ± 21.0	4.66 ± 0.43
AUNet (w/)	**79.1 ± 6.0**	**80.8 ± 7.1**	**37.6 ± 15.4**	**4.04 ± 0.33**

[a] w/o—Without pretraining on CBIS-DDSM
[b] w/—With pretraining on CBIS-DDSM

always generates better segmentation metrics and pretraining improves the segmentation performance of all the methods significantly. With pretraining on CBIS-DDSM, the results of the three established models present a different pattern from the CBIS-DDSM dataset. Among the three established models, FCDenseNet103 generates the highest *DSC* and *SEN* value, UNet shows the lowest ΔA, and FusionNet gives the lowest *HAU*. It is interesting that FusionNet shows much worse performance on INbreast than that on CBIS-DDSM. On the other hand, compared to the three models, our proposed AUNet still generates the highest *DSC*, the highest *SEN*, the lowest ΔA, and the lowest *HAU*. AUNet shows an average *DSC* increase of at least 3.0% (statistically significant with $p < 0.05$ by Wilcoxon signed-rank test) and *HAU* decrease of 0.29 (statistically significant with $p < 0.05$ by Wilcoxon signed-rank test). Similarly, the empirical cumulative distribution plot indicates that for INbreast, AUNet still achieves a relatively better segmentation performance, followed by FCDenseNet, FusionNet, and UNet (figure 5).

4.3. Qualitative results

Figure 6 presents several segmentation results generated by the different networks for qualitative comparisons. We can see, overall, our proposed AUNet performs better than the other three FCNs for our whole mammographic

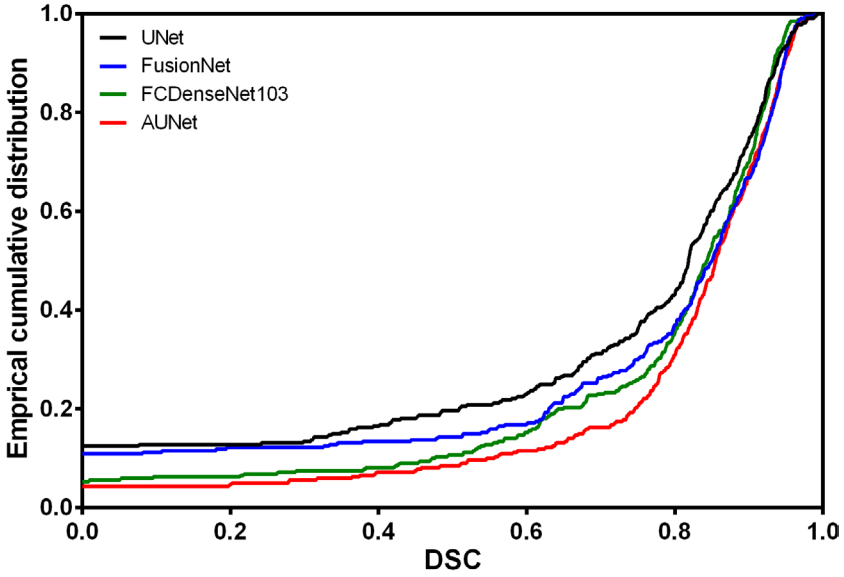


Figure 5. Empirical cumulative distribution of *DSC* on INbreast with pretraining. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

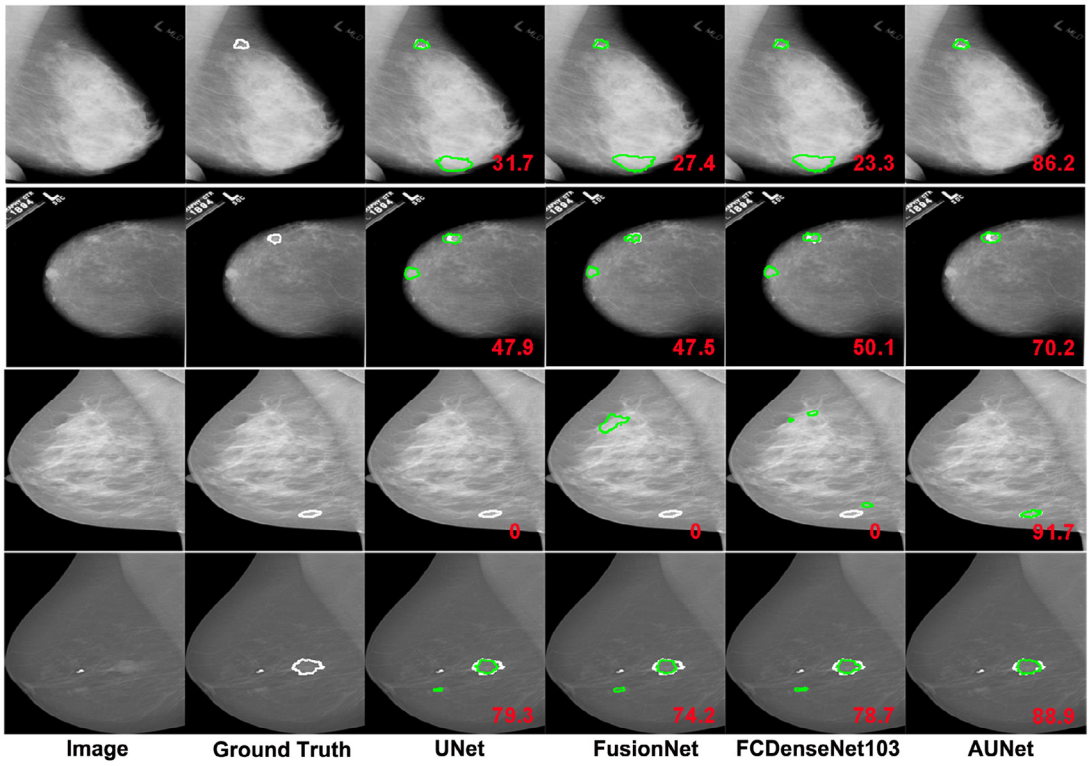


Figure 6. Segmentation results of different networks. From left to right, the columns correspond to the input images, the ground truth labels, the segmentation results of UNet, FusionNet, FCDenseNet103, and our proposed AUNet, respectively. The white circles indicate the boundaries of the labels and the green circles indicate the boundaries of the segmentation results. The red number on the right bottom of each image is the *DSC* value of the segmentation result. The first two rows are from the CBIS-DDSM dataset and the last two rows are from the INbreast dataset. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

mass segmentation task. In addition, it could be observed that AUNet displays an impressive ability to suppress the false positive results of UNet without increasing the number of false negatives, whereas both FusionNet and FCDenseNet103 are not effective in this aspect or even make the situation worse (figure 6; the first, second, and last rows). This observation is consistent with the quantitative results discussed before. Lastly, our AUNet could give accurate segmented masses for difficult samples when the other three networks could barely find the targeted regions at all, such as the third example in figure 6.

Table 5. Segmentation performance on mass-centered image patches for INbreast dataset.

Models	*DSC* (%)	*SEN* (%)	ΔA (%)	*HAU*
Cardoso *et al* (2015)	$0.88 \times 100\%$	—	—	—
Dhungel *et al* (2015b)	$(0.90 \pm 0.06) \times 100\%$	—	—	—
Dhungel *et al* (2017)[a]	$(0.85 \pm 0.02) \times 100\%$	—	—	—
UNet	92.0 ± 0.8	93.1 ± 1.2	8.3 ± 2.5	6.88 ± 0.18
FusionNet	92.0 ± 0.8	92.7 ± 1.0	8.1 ± 2.9	6.94 ± 0.19
FCDenseNet103	89.5 ± 0.8	89.6 ± 2.0	11.7 ± 2.2	7.23 ± 0.14
AUNet	$\mathbf{92.4 \pm 0.9}$	$\mathbf{93.7 \pm 0.9}$	$\mathbf{7.5 \pm 2.6}$	$\mathbf{6.85 \pm 0.28}$

[a] Patches were extracted based on detection results.

Table 6. Computational complexities of different networks.

Models	UNet	FusionNet	FCDenseNet103	AUNet ($R = 16$)
Convolutional and FC layers	23	50	103	44
Parameters (million)	34.5	78.5	13.9	75.5
FPS (with 256×256 inputs)	59	36	27	32

4.4. Results on extracted image patches

In order to compare the performance of proposed network directly to the literature on breast mass segmentation, we also conducted experiments on extracted mass-centered image patches for the INbreast dataset. For each mammogram, we first found the smallest rectangular that could accommodate the mass. Then, the mass-centered image patch was extracted through enlarging the rectangular by 20% in area with an equal elongation ratio in width and height of $\sqrt{1.2}$. Similar to the whole mammogram situation, 5-fold cross-validation experiments with three replicates were done. Results in table 5 confirms that our proposed AUNet could also achieve better segmentation metrics on mass-centered image patches compared to both the three FCNs and the literature reported results.

4.5. Model complexity

Table 6 lists the total number of convolutional and FC layers, the optimizable parameters, and the inference time in terms of frames per second (FPS) with input resized to 256×256. Obviously, UNet is the simplest and fastest model, and the other three models (FusionNet, FCDenseNet103, and AUNet) have similar inference speeds with AUNet achieves the best segmentation performance for our task.

5. Discussion

Segmentation of mammographic masses is a challenging task as mammograms have low signal-to-noise ratio and breast masses may vary in shapes and sizes. An easy alternative is to segment masses from extracted ROIs. However, manual extraction of ROIs is a tedious task. Automatic detection algorithms still subject to high false positives and specially designed post processing methods are required to achieve expected performance (Dhungel *et al* 2015a). Therefore, automatic breast mass segmentation in whole mammograms is of great clinical value. There are several reports targeting at developing deep learning models for whole mammographic mass segmentation, such as the ASPP-FC-DenseNet and the Attention Dense-U-Net (Hai *et al* 2019, Li *et al* 2019a). ASPP-FC-DenseNet achieved a *DSC* of 76.97% on the private dataset and Attention Dense-U-Net achieved a sensitivity of 77.89% on the selected DDSM dataset. Both are much smaller than the results achieved in this study, which confirms the suitability of our proposed AUNet for the task. Figure 7 presents a few segmentation results of AUNet. We admit that compared to inputs with irregular or small masses, AUNet performs slightly better for inputs with large and regular masses. However, figure 6 indicates that AUNet still performs better than the three FCNs for inputs with small and irregular masses. Overall, figures 6 and 7 conclude that for inputs with different mass shapes and sizes, our AUNet could always give very accurate segmentation results.

Mammograms are taken with high resolutions. Images from CBIS-DDSM dataset have a width ranging from 1786 to 5431 pixels and a height ranging from 3920 to 6931 pixels. Images form INbreast have either 3328×4084 or 2560×3328 pixels. To facilitate the training and testing of deep neural networks, necessary image preprocessing steps are required, such as image patch extraction or resizing. Although patch extraction method can preserve all the original image information and researchers have developed elegant approaches to extract informative image patches (Qin *et al* 2018), we adopted resizing in this study. On one hand, it has been suggested in the computer vision field that global contextual information is important for accurate image segmentation (Wang *et al*

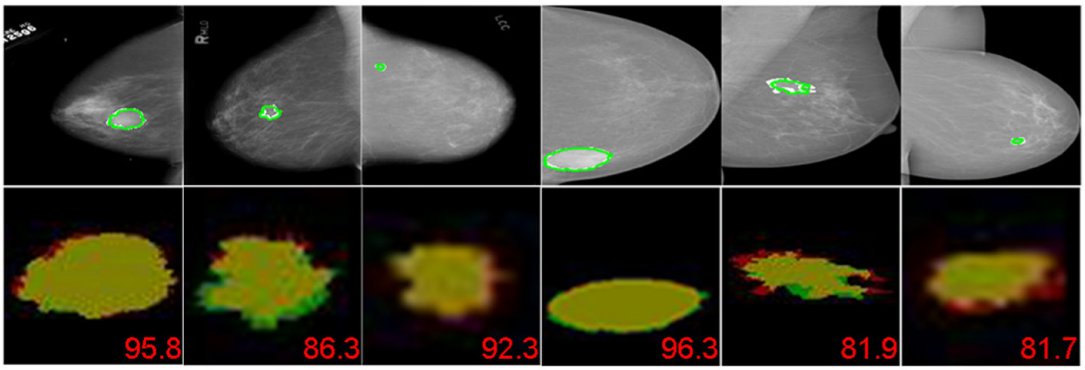


Figure 7. Segmentation results of AUNet when input images contain masses of different shapes and sizes. The left three columns are images from the CBIS-DDSM dataset and the right three columns are from the INbreast dataset. The white circles indicate the boundaries of the labels and the green circles indicate the boundaries of the segmentation results. Red color regions indicate the ground-truth masses and green indicate the segmentation results. Yellow color regions indicate the overlap between the segmentation and the ground-truth. The red number on the right bottom of each image is the *DSC* value of the segmentation result. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 7. Segmentation performance of AUNet with different input settings on CBIS-DDSM dataset.

Inputs	*DSC* (%)	*SEN* (%)	ΔA (%)	*HAU*
Gray input (resize 256 × 256)	80.9 ± 0.4	84.5 ± 0.5	29.9 ± 0.2	2.97 ± 0.01
RGB input (resize 256 × 256)	**81.8 ± 0.0**	**84.9 ± 0.3**	**26.9 ± 0.3**	**2.96 ± 0.03**
RGB input (resize 512 × 512)	78.7 ± 0.5	81.1 ± 0.1	28.6 ± 1.0	4.30 ± 0.02
RGB input (pad & resize 256 × 256)	78.9 ± 0.2	83.3 ± 0.2	30.4 ± 1.1	**2.20 ± 0.01**

2018b). Patch extraction restricts the field of view of the network, which may influence the segmentation performance. Therefore, the correlations between the patches need to be carefully considered, which we will investigate in the following work. On the other hand, after resizing, most masses still occupy hundreds to thousands of pixels. We believe these downsampled masses are large enough to preserve the overall mass information. Moreover, different input settings have been tested with gray or RGB inputs, with different resolutions (256 × 256 inputs or 512 × 512 inputs), and with fixed aspect ratios by zero padding the images before resizing (table 7). Although different inputs show influence on the final segmentation results, our proposed AUNet always achieves the best performance (more results in supplementary file tables S7–S9). Thus, it can be anticipated that our method should also be able to achieve the best segmentation performance if the full resolution inputs are utilized. With detailed inspection, the results show that RGB inputs could improve the segmentation performance. Even though it was not investigated in the current study, RGB inputs can also facilitate the direct transfer learning of networks trained on natural images. Resizing to the higher resolution (512 × 512 pixels) showed negative effects on the segmentation performance, which was also observed for the three established FCNs (supplementary file table S8). This weakened performance might be caused by two reasons. One is due to the GPU memory limitation, batch size of 2 was applied for inputs with 512 × 512 pixels instead of 4 for inputs with 256 × 256 pixels. The other is it is difficult to accurately define the mass boundaries in mammograms. At higher resolutions, the images are more sensitive to manual label errors. Zero padding brings large regions of background to the inputs and hinders the segmentation process. In this study, our experiments were done with RGB inputs resized to 256 × 256 pixels to maximize the segmentation performance.

Our AUNet, as well as the three comparison networks, showed severely worse performance on the INbreast dataset compared to that on the CBIS-DDSM dataset when trained from scratch (tables 3 and 4). A major cause could be the large difference in the sample size. Much better results were obtained when the networks were pre-trained on the CBIS-DDSM dataset. But still, the performance is not as good as that on the CBIS-DDSM dataset. Except for the sample size, another observable difference between the two datasets is the different image intensity ranges (figure 8). Although images from both datasets were stored with a 16-bit integer data type, all images from CBIS-DDSM have an intensity range of [0, 65 535], whereas images from INbreast have different intensity ranges with the minimum of [0, 1811] and maximum of [0, 4095]. Even though intensity normalization was conducted before the images were inputted into the networks, the original differences might also affect the results. Besides, as shown in figure 1, the image distributions are also different between the two datasets, which might influence the results a little bit. The difference between digitized film-screen mammograms of CBIS-DDSM and full-fild digital mammograms of INbreast is another possible cause. The observed model performance difference between the two datasets agrees with the commonly accepted condition for the application of trained deep

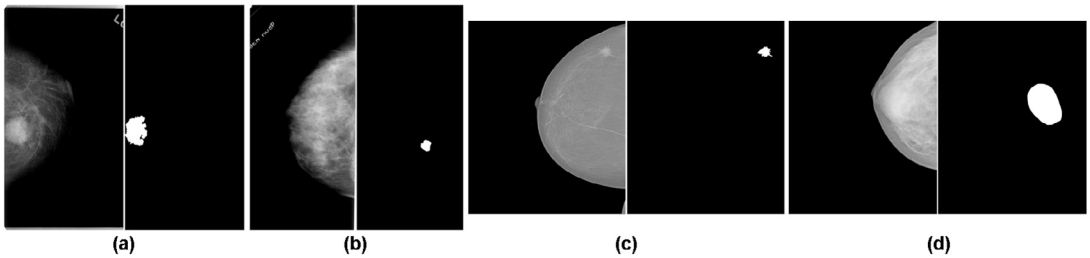


Figure 8. Example images from the two datasets. Image and label from the CBIS-DDSM categorized as BIRADS 1 (a), from the CBIS-DDSM categorized as BIRADS 4 (b), from the INbreast categorized as BIRADS 1 (c), and from the INbreast categorized as BIRADS 4 (d).

learning models that only when the testing dataset has a similar data distribution with the training dataset, the models can be applied directly. When the testing dataset has a different distribution, the models need to be fine-tuned. To apply the optimized models to unknown digital mammograms, there are two possible solutions. One is labeling a small dataset in the target domain and fine-tuning the trained models accordingly. The other is transforming the mammograms to fit the distribution of the CBIS-DDSM/INbreast dataset.

UNet is a very powerful network for biomedical image segmentation (Ronneberger *et al* 2015) and is the template for many following-up studies (Balagopal *et al* 2018, Li *et al* 2019b). Our proposed AUNet adopts a similar encoder–decoder architecture. To enhance the performance, we first investigated the network backbone design. Compared to the basic unit (figure 2(b)) used in both the encoder and decoder pathways of naive UNet, we found that our asymmetrical network backbone Res-Basic-UNet was more suitable for our application. This is reasonable as the res unit (figure 2(d)) in the encoder pathway promotes the information and gradient propagation while the basic unit (figure 2(b)) in the decoder pathway better preserves important semantic information of the high-level features. Our results show that Res-Basic-UNet improves the *DSC* by 3.9% over UNet (statistically significant with $p < 0.05$ by Wilcoxon signed-rank test).

Then, we believe that the simple bilinear upsampling method and the feature fusion through concatenation adopted by UNet are not effective enough. Significant information loss might happen, which could greatly worsen the segmentation results. Therefore, we proposed a new upsampling block, AU block, to solve these problems. AU block utilizes the high-level features in two means. In one way, the high-level features are densely upsampled and fused with the low-level features by summation. In the other, the high-level features are bilinear upsampled and concatenated with the convolution smoothed summation (figure 3(b)). Moreover, in order to select the rich-informative channels, a channel-wise attention component is used after the concatenation. With AU block, our AUNet increases the *DSC* by another 4.3% over Res-Basic-UNet (statistically significant with $p < 0.05$ by Wilcoxon signed-rank test). Besides, AUNet generates better segmentation results than the three widely used FCN segmentation networks and recently published studies for both CBIS-DDSM and INbreast datasets.

False positive and false negative are important issues that need to be considered for CAD systems. False positive is commonly found to be the problem that hinders the application of automatic detection algorithms to medical imaging (Dhungel *et al* 2015a, Samala *et al* 2016b). It can bring huge psychological stress and depression to patients and result in unnecessary biopsies. False negative, on the other hand, is detrimental for clinical applications which can miss early diagnosis. It is important to reduce both false positive and false negative results. The low signal-to-noise ratio of a mammogram makes it difficult to clearly differentiate the masse from the normal breast tissues (figures 1(a) and 6). All the three FCNs show serious false positive segmentation results, which greatly affected the evaluation metrics (figure 6). On the contrary, AUNet is able to effectively reduce the false positive incidences without increasing the false negative results through the information selection by channel-wise attention. Moreover, thanks to the full utilization of the feature map information, AUNet also performs better at decreasing the false negative results (figure 6; the third example).

The proposed whole mammographic mass segmentation method has two major limitations. According to the results of the two datasets, the performance of the network depends on the datasets to some extent. The two datasets we utilized have their own advantages. CBIS-DDSM has a lot more cases than INbreast whereas INbreast has higher precision (Moreira *et al* 2012, Lee *et al* 2017). We lack a comprehensive dataset that can fully validate the effectiveness of the proposed method. The other limitation is our method is developed solely on images, and the clinical parameters, such as the age of patients, were not considered, which we will consider in the following work. On the other hand, breast masses are significant contributors to breast cancers (Giger *et al* 2013). Mass segmentation is an important step for the following disease diagnosis and treatment planning. After the mass segmentation, image features can be extracted from and surrounding the specific regions and different analyses can be conducted. These image features could be used to differentiate breast cancer subtypes (Wu *et al* 2017a).

They were found to be associated with tumor-infiltrating lymphocytes in breast cancer, which is a promising predictive biomarker for the effectiveness of immunotherapy treatment (Wu *et al* 2018b). Some of them were identified as valuable prognostic markers for adjuvant and neoadjuvant chemotherapies (Wu *et al* 2017b, 2018a). As a necessary next step for our current work, we will also study the corresponding image feature extraction methods as well as imaging-based disease diagnosis and treatment plan selection in the future.

Traditional two-dimensional (2D) mammograms are taken when the breast is compressed, which may lead to tissue overlap and influence the subsequent diagnosis. The recent introduction of digital breast tomosynthesis (DBT) technology has been proved to have improved sensitivity and specificity compared with 2D mammography (Heather *et al* 2015, Conant *et al* 2019). There are a number of studies working on the CNN-based analysis of DBT, including the reconstruction of DBT volume (Ayyagari *et al* 2018), detection of breast masses (Samala *et al* 2016a) or microcalcifications (Samala *et al* 2016b), and classification of benign and malignant breast masses (Samala *et al* 2018). Limited datasets to train a deep neural network is always a key problem. For DBT, this issue is even severer (Geras *et al* 2019). Transfer learning is an effective solution, and studies indicated a multi-stage knowledge transfer strategy, consisting of transferring from natural image to mammography and mammography to DBT, achieved better breast mass malignancy classification performance than direct transfer from natural image to DBT (Samala *et al* 2018). Similarly, for our task on breast mass segmentation, our model is expected to serve as a better baseline to be transferred to DBT analysis.

6. Conclusion

In this work, we propose a new network, AUNet, for the mass segmentation in whole mammograms. Specifically, we utilized an asymmetrical encoder–decoder architecture and introduced a new upsampling block, AU block, to boost the segmentation performance. Comprehensive experiments have been conducted. AUNet presented improved segmentation behaviors on both CBIS-DDSM and INbreast datasets compared to existing FCN models, which proves its effectiveness and robustness. In addition, AUNet could greatly reduce both false negative and false positive results. We make our code available, by which we hope our work can attract and inspire more following-up studies in the field.

Acknowledgments

This work was supported by funding from the National Natural Science Foundation of China (61601450, 61871371, and 81830056), Key-Area Research and Development Program of Guangdong Province (2018B010109009), Science and Technology Planning Project of Guangdong Province (2017B020227012), the Basic Research Program of Shenzhen (JCYJ20180507182400762), Youth Innovation Promotion Association Program of Chinese Academy of Sciences (2019351), the National Nature Science Foundation of China (61903227 and 11801313), and Shandong Provincial Natural Science Foundation (ZR2019QA007).

ORCID iDs

Shanshan Wang https://orcid.org/0000-0002-0575-6523

References

Abdel-Dayem A R and EI-Sakka M R 2005 Fuzzy entropy based detection of suspicious masses in digital mammogram images *Proc. IEEE EMBS* pp 4017–22

Ayyagari D, Ramesh N, Yatsenko D, Tasdizen T and Atria C 2018 Image reconstruction using priors from deep learning *Proc. SPIE* **10574** 105740H

Badrinarayanan V, Kendall A and Cipolla R 2017 Segnet: a deep convolutional encoder–decoder architecture for image segmentation *IEEE Trans. Pattern Anal. Mach. Intell.* **39** 2481–95

Balagopal A, Kazemifar S, Nguyen D, Lin M, Hannan R, Owrangi A and Jiang S 2018 Fully automated organ segmentation in male pelvic CT images *Phys. Med. Biol.* **63** 245015

Ball J E, Butler T W and Bruce L M 2004 Towards automated segmentation and classification of masses in digital mammograms *Proc. IEEE EMBS* pp 1814–7

Birdwell R L, Ikeda D M, O'Shaughnessy K F and Sickles E A 2001 Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection *Radiology* **219** 192–202

Cardoso J S, Domingues I and Oliveira H P 2015 Closed shortest path in the original coordinates with an application to breast cancer *Int. J. Pattern Recognit. Artif. Intell* **29** 1555002

Chen L C, Papandreou G, Kokkinos I, Murphy K and Yuille A L 2018 Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs *IEEE Trans. Pattern Anal. Mach. Intell.* **40** 834–48

Chen L, Zhang H, Xiao J, Nie L, Shao J, Liu W and Chua T S 2017 SCA-CNN: spatial and channel-wise attention in convolutional networks for image captioning *IEEE CVPR (Honolulu, HI, 21–26 July 2017)* pp 5659–67

Conant E F *et al* 2019 Association of digital breast tomosynthesis versus digital mammography with cancer detection and recall rates by age and breast density *JAMA Oncol.* **5** 635–42
Dhungel N, Carneiro G and Bradley A P 2015a Automated mass detection in mammograms using cascaded deep learning and random forests *IEEE DICTA (Adelaide, SA, 23–25 November 2015)* pp 1–8
Dhungel N, Carneiro G and Bradley A P 2015b Deep learning and structured prediction for the segmentation of mass in mammograms *MICCAI* pp 2950–4
Dhungel N, Carneiro G and Bradley A P 2015c Deep structured learning for mass segmentation from mammograms *IEEE ICIP (Quebec City, QC, 27–30 September 2015)* pp 2950–4
Dhungel N, Carneiro G and Bradley A P 2017 A deep learning approach for the analysis of masses in mammograms with minimal user intervention *Med. Image Anal.* **37** 114–28
Freixenet J, Oliver A, Marti R and Lladó X 2008 Eigendetection of masses considering false positive reduction and breast density information *Med. Phys.* **35** 1840–53
Geras K J, Mann R M and Moy L 2019 Artificial intelligence for mammography and digital breast tomosynthesis: current concepts and future perspectives *Radiology* **293** 246–59
Giger M L, Karssemeijer N and Schnabel J A 2013 Breast image analysis for risk assessment, detection, diagnosis, and treatment of cancer *Annu. Rev. Biomed. Eng.* **15** 327–57
Greenspan H, Ginneken B V and Summers R M 2016 Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique *IEEE Trans. Med. Imaging* **35** 1153–9
Guliato D, Rangayyan R M, Carvalho J D and Santiago S A 2008 Polygonal modeling of contours of breast tumors with the preservation of spicules *IEEE Trans. Biomed. Eng.* **55** 14–20
Gulsrud T O, Engan K and Hanstveit T 2005 Watershed segmentation of detected masses in digital mammograms *Proc. EMBS* pp 3304–7
Hai J, Qiao K, Chen J, Tan H, Xu J, Zeng L, Shi D and Yan B 2019 Fully convolutional densenet with multiscale context for automated breast tumor segmentation *J. Healthcare Eng.* **2019** 1–11
Hamidinekoo A, Denton E, Rampun A, Honnor K and Zwiggelaar R 2018 Deep learning in mammography and breast histology, an overview and future trends *Med. Image Anal.* **47** 45–67
Han S, Kang H, Jeong J, Park M, Kim W, Bang W and Seong Y 2017 A deep learning framework for supporting the classification of breast lesions in ultrasound images *Phys. Med. Biol.* **62** 7714–28
He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition *IEEE CVPR (Las Vegas, NV, 27– 30 June 2016)* pp 770–8
Heath M, Bowyer K, Kopans D, Moore R and Kegelmeyer J P 2000 The digital database for screening mammography *Proc. 5th Int. Workshop on Digital Mammography* pp 212–8
Heather H R, Nicholson B E, Rochman C M, Merchant J K, Mayo R C III and Harvey J A 2015 Digital breast diagnostic setting: indications and clinical applications *Radiographics* **35** 975–90
Hu J, Shen L and Sun G 2018 Squeeze-and-excitation networks *IEEE CVPR (Salt Lake City, UT, 18–23 June 2018)* pp 7132–41
Jégou S, Drozdzal M, Vazquez D, Romero A and Bengio Y 2017 The one hundred layers tiramisu: fully convolutional densenets for semantic segmentation *IEEE CVPRW (Honolulu, HI, 21–26 July 2017)* pp 1175–83
Jiang M, Zhang S, Zheng Y and Metaxas D N 2016 Mammographic mass segmentation with online learned shape and appearance priors *MICCAI* pp 35–43
Kamnitsas K, Ledig C, Newcombe V F J, Simpson J P, Kane A D, Menon D K, Rueckert D and Glocker B 2017 Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation *Med. Image Anal.* **36** 61–78
Kim S T, Lee J, Lee H and Ro Y M 2018 Visually interpretable deep network for diagnosis of breast masses on mammograms *Phys. Med. Biol.* **63** 235025
Kooi T, Litjens G, Ginneken B V, Gubern-Mérida A, Sánchez C I, Mann R, Heeten A D and Karssemeijer N 2017 Large scale deep learning for computer aided detection of mammographic lesions *Med. Image Anal.* **35** 303–12
Kupinski M A and Giger M L 1998 Automated seeded lesion segmentation on digital mammograms *IEEE. Trans. Med. Imaging* **17** 510–7
Lee R S, Gimenez F, Hoogi A, Miyake K K, Gorovoy M and Rubin D L 2017 Data descriptor: a curated mammography data set for use in computer-aided detection and diagnosis research *Sci. Data* **4** 170177
Lehman C D, Wellman R D, Buist D S M, Kerlikowske K, Tosteson N A, Miglioretti D L and for the Breast Cancer Surveillance Consortium 2015 Diagnostic accuracy of digital screening mammography with and without computer-aided detection *JAMA Intern. Med.* **175** 1828–37
Li L, Clark R A and Thomas J A 2002 Computer-aided diagnosis of masses with full-field digital mammography *Acad. Radiol.* **9** 4–12
Li S, Dong M, Du G and Mu X 2019a Attention dense-u-net for automatic breast mass segmentation in digital mammogram *IEEE Access* **7** 59037–47
Li X, Hong Y, Kong D and Zhang X 2019 Automatic segmentation of levator hiatus from ultrasound images using u-net with dense connections *Phys. Med. Biol.* **64** 075015
Lin G, Milan A, Shen C and Reid I 2017 Refinenet: multi-path refinement networks for high-resolution semantic segmentation *IEEE CVPR (Honolulu, HI, 21–26 July 2017)* pp 1925–34
Litjens G, Kooi T, Bejnordi B E, Setio A A A, Ciompi F, Ghafoorian M, van der Laak J A, van Ginneken B and Sánchez C I 2017 A survey on deep learning in medical image analysis *Med. Image Anal.* **42** 60–88
Long J, Shelhamer E and Darrell T 2015 Fully convolutional networks for semantic segmentation *IEEE CVPR (Boston, MA, 7–12 June 2015)* pp 3431–40
Løberg M, Lousdal M L, Bretthauer M and Kalager M 2015 Benefits and harms of mammography screening *Breast Cancer Res.* **17** 1–12
Mirikharaji Z and Hamarneh G 2018 Star shape prior in fully convolutional networks for skin lesion segmentation *MICCAI* pp 737–45
Mnih V, Heess N, Graves A and Kavukcuoglu K 2014 Recurrent models of visual attention *NIPS* pp 1–9
Moreira I C, Amaral I, Domingues I, Cardoso A, Cardoso M J and Cardoso J S 2012 Inbreast: toward a full-field digital mammographic database *Acad. Radiol.* **19** 236–48
Nie D, Gao Y, Wang L and Shen D 2018 ASDNET: attention based semi-supervised deep networks for medical image segmentation *MICCAI* pp 370–8
Oliver A, Freixenet J, Martí J, Pérez E, Pont J and Denton E R E 2010 A review of automatic mass detection and segmentation in mammographic images *Med. Image Anal.* **14** 87–110
Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, Lin Z, Desmaison A, Antiga L and Lerer A 2017 Automatic differentiation in pytorch. *NIPS* pp 1–4
Qin W, Wu J, Yuan Y, Zhao W, Ibragimov B, Gu J and Xing L 2018 Superpixel-based and boundary-sensitive convolutional neural network for automated liver segmentation *Phys. Med. Biol.* **63** 095017

Quan T M, Hildebrand D G C and Jeong W K 2016 Fusionnet: a deep fully residual convolutional neural network for image segmentation in connectomics (arXiv:1612.05360)
Rahmati P, Adler A and Hamarneh G 2012 Mammography segmentation with maximum likelihood active contours *Med. Image Anal.* **16** 1167–86
Reddi S J, Kale S and Kumar S 2017 On the convergence of Adam and beyond *ICLR* pp 1–23
Ronneberger O, Fischer P and Brox T 2015 U-net: convolutional networks for biomedical image segmentation *MICCAI* pp 234–41
Roy A G, Navab N and Wachinger C 2018 Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks *MICCAI* pp 421–9
Sahiner B, Petrick N, Chan H P, Hadjiiski L M, Paramagul C, Helvie M A and Gurcan M N 2001 Computer-aided characterization of mammographic masses: Accuracy of mass segmentation and its effects on characterization *IEEE Trans. Med. Imaging* **20** 1275–84
Samala R K, Chan H P, Hadjiiski L, Helvie M A, Richter C D and Cha K 2018 Breast cancer diagnosis in digital breast tomosynthesis: effects of training sample size on multi-stage transfer learning using deep neural nets *IEEE Trans. Med. Imaging* **38** 686–96
Samala R K, Chan H P, Hadjiiski L, Helvie M A, Wei J and Cha K 2016a Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography *Med. Phys.* **43** 6654–66
Samala R K, Chan H P, Hadjiiski L M, Cha K and Helvie M A 2016b Deep-learning convolution neural network for computer-aided detection of microcalcifications in digital breast tomosynthesis *Proc. SPIE* **9785** 97850Y
Shi J *et al* 2008 Characterization of mammographic masses based on level set segmentation with new image features and patient information *Med. Phys.* **35** 280–90
Shi W, Caballero J, Huszár F, Totz J, Aitken A P, Bishop R, Rueckert D and Wang Z 2016 Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network *IEEE CVPR (Las Vegas, NV, 27–30 June 2016)* pp 1874–83
Siegel R L, Miller K D and Jemal A 2017 Cancer statistics, 2017 *CA. Cancer J. Clin.* **67** 7–30
Song E, Xu S, Xu X, Zeng J, Lan Y, Zhang S and Hung C C 2010 Hybrid segmentation of mass in mammograms using template matching and dynamic programming *Acad. Radiol.* **17** 1414–24
Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N and Kaise L 2017 Attention is all you need *NIPS* pp 1–11
Wang P, Chen P, Yuan Y, Liu D, Huang Z, Hou X and Cottrell G 2018a Understanding convolution for semantic segmentation *WACV* pp 1451–60
Wang X, Girshick R, Gupta A and He K 2018 Non-local neural networks *CVPR (Salt Lake City, UT, 18–23 June 2018)* pp 7794–803
Wei J, Chan H P, Sahiner B, Hadjiiski L M, Helvie M A, Roubidoux M A, Zhou C and Ge J 2006 Computer-aided detection of breast masses on mammograms: dual system approach with two-view analysis *Med. Phys.* **36** 4157–68
Wu J, Cui Y, Sun X, Cao G, Li B, Ikeda D M, Kurian A W and Li R 2017a Unsupervised clustering of quantitative image phenotypes reveals breast cancer subtypes with distinct prognoses and molecular pathways *Clin. Cancer Res.* **23** 3334–42
Wu J, Li B, Cao G, Rubin D L, Napel S, Ikeda D M, Kurian A W and Li R 2017b Heterogeneous enhancement patterns of tumor-adjacent parenchyma at MR imaging are associated with dysregulated signaling pathways and poor survival in breast cancer *Radiology* **285** 401–13
Wu J, Cao G, Sun X, Lee J, Rubin D L, Napel S, Kurian A W, Daniel B L and Li R 2018a Intratumoral spatial heterogeneity at perfusion MR imaging predicts recurrence-free survival in locally advanced breast cancer treated with neoadjuvant chemotherapy *Radiology* **288** 26–35
Wu J, Li X, Teng X, Rubing D L, Napel S, Daniel B L and Li R 2018b Magnetic resonance imaging and molecular features associated with tumor-infiltrating lymphocytes in breast cancer *Breast Cancer Res.* **20** 101
Yu F, Koltun V and Funkhouser T 2017 Dilated residual networks *IEEE CVPR (Honolulu, HI, 21–26 Jul 2017)* pp 472–80
Zhang Z, Zhang X, Peng C, Cheng D and Sun J 2018 Exfuse: enhancing feature fusion for semantic segmentation *ECCV (Lecture Notes in Computer Science)* vol 11214 (Berlin: Springer) pp 1–16
Zhao H, Shi J, Qi X, Wang X and Jia J 2017 Pyramid scene parsing network *IEEE CVPR (Honolulu, HI, 21–26 July 2017)* pp 2881–90
Zhou B, Khosla A, Lapedriza A, Oliva A and Torralba A 2016 Learning deep features for discriminative localization *IEEE CVPR (Las Vegas, NV, 27–30 June 2016)* pp 2921–9
Zhu W, Huang Y, Zeng L, Chen X, Liu Y, Qian Z, Du N and Fan W 2019 AnatomyNet: deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy *Med. Phys.* **46** 576–89