Ultrasound in Med. & Biol., Vol. 46, No. 10, pp. 2819−2833, 2020

0301-5629/$ - see front matter

https://doi.org/10.1016/j.ultrasmedbio.2020.06.015

● *Original Contribution*

ATTENTION-ENRICHED DEEP LEARNING MODEL FOR BREAST TUMOR SEGMENTATION IN ULTRASOUND IMAGES

Aleksandar Vakanski,* Min Xian,* and Phoebe E. Freer[†]
* Department of Computer Science, University of Idaho, Idaho Falls, Idaho, USA; and [†] University of Utah School of Medicine, Salt Lake City, Utah, USA

(*Received* 16 *September* 2019; *revised* 12 *June* 2020; *in final from* 19 *June* 2020)

Abstract—Incorporating human domain knowledge for breast tumor diagnosis is challenging because shape, boundary, curvature, intensity or other common medical priors vary significantly across patients and cannot be employed. This work proposes a new approach to integrating visual saliency into a deep learning model for breast tumor segmentation in ultrasound images. Visual saliency refers to image maps containing regions that are more likely to attract radiologists' visual attention. The proposed approach introduces attention blocks into a U-Net architecture and learns feature representations that prioritize spatial regions with high saliency levels. The validation results indicate increased accuracy for tumor segmentation relative to models without salient attention layers. The approach achieved a Dice similarity coefficient (DSC) of 90.5% on a data set of 510 images. The salient attention model has the potential to enhance accuracy and robustness in processing medical images of other organs, by providing a means to incorporate task-specific knowledge into deep learning architectures. (E-mail: vakanski@uidaho.edu) © 2020 World Federation for Ultrasound in Medicine & Biology. All rights reserved.

Key Words: **Breast ultrasound, Medical image segmentation, Visual saliency, Domain knowledge-enriched learning.**

INTRODUCTION

Computer-aided image analysis can assist radiologists' interpretation and diagnosis and reduce error rates, as well as the level of stress regarding erroneous diagnosis (Cheng et al. 2010; Moon et al. 2011; Inoue et al. 2017; Jalalian et al. 2017; Wu et al. 2019). For instance, 3%−6% of all radiologists' image interpretations contain clinically important errors, and significant variability in the inter- and intra-observer image interpretation is often reported (Elmore et al. 1994; Waite et al. 2016; Langlotz et al. 2019).

The emphasis in this work is on automated computer-aided diagnosis of tumors in breast ultrasound (BUS) images (Xian et al. 2018b). A large body of research work employed conventional and deep learning approaches to address tasks related to automated lesion localization, segmentation and classification (Moon et al. 2011; Inoue et al. 2017; Jalalian et al. 2017; Litjens et al. 2017; Wu et al. 2019). Despite this progress, existing methods lack robustness and consistency when processing images taken with different imaging equipment, where the variations in image intensity, contrast and density often result in the degraded performance of models that otherwise perform well on custom-built data sets.

An important way to improve the performance of data-driven models is by incorporating prior domain-specific knowledge (Nosrati and Hamarneh 2016). On the other hand, incorporating prior knowledge in deep models for breast cancer detection is challenging, because unlike other medical organs—such as the kidney or the heart, whose features naturally lend themselves to the application of shape or boundary priors—breast tumors have a large variability in shape and boundaries from case to case. Extracting other priors in the form of curvature, texture, intensity or number of regions for breast tumors is also not an option.

Our proposed approach incorporates topological and anatomic prior information into a deep learning model for image segmentation. More specifically, maps of visual saliency are employed for integrating image topology knowledge (Xu et al. 2016, 2018). The model

Address correspondence to: Aleksandar Vakanski, 1776 Science Center Drive, Idaho Falls, ID 83402, USA. E-mail: vakanski@uidaho.edu

for visual saliency estimation is formulated as a quadratic optimization problem, and it is based on calculations of neutro-connectedness between regions in the image (Xian et al. 2016; Xian 2017). Anatomic prior knowledge is integrated by decomposing the tissue layers into skin, fat, mammary and muscle layers (Xu et al. 2019) and applying higher weights to the salient regions in images belonging to the mammary layer.

In this article, we propose a novel approach to integrating domain knowledge into a deep neural network model by using the attention mechanism (Simonyan et al. 2013). A U-Net architecture (Ronneberger et al. 2015) is selected for incorporating the prior knowledge in the form of a pyramid of visual saliency maps. Attention blocks are integrated with the layers of the encoder to force the network to learn feature representations that place spatial attention on target regions with high saliency values. Unlike similar deep learning models that introduce attention blocks by merging internal feature representations from different layers (Chen et al. 2016; Jetley et al. 2018; Oktay et al. 2018b), the proposed approach employs external auxiliary inputs in the form of visual saliency maps for training the model parameters.

The main contributions of this article are (i) an attention-enriched deep learning model for integrating prior knowledge of tumor saliency and (ii) a confidence level calculation for visual saliency maps.

The article is organized as follows. The next section overviews related works in the literature. The Methods section describes the used image data set, the proposed network architecture, attention blocks and visual saliency maps. Experimental validation is provided in the Results. The Discussion section presents the findings of the experiments, and the Conclusions section summarizes the work.

RELATED WORKS

Computer-aided segmentation in medical imaging has been an important research topic for several decades, and it encompasses a vast body of work in the published literature. Recent advances in deep learning models (LeCun et al. 2015; Goodfellow et al. 2016) have led to great improvements in semantic image segmentation (He et al. 2015; Long et al. 2015; Ronneberger et al. 2015; Badrinarayanan et al. 2017; Lin et al. 2017; Zhao et al. 2017; Chen et al. 2018a, 2018b). Consequently, significant efforts have been devoted toward the implementation and design of deep neural networks for a wide range of medical applications, including segmentation of tumors and lesions (*e.g.,* brain tumor [Kamnitsas et al. 2017], skin lesions [González-Díaz 2017], histopathology images [Chen et al. 2017; Kumar et al. 2017; Graham et al. 2018; Lin et al. 2018; Naylor et al. 2019]) and segmentation of organs (*e.g.,* pancreas [Oktay et al. 2018b], lung [Hu et al. 2019], heart [Oktay et al. 2018a]) or head and neck anatomy (Zhu et al. 2019).

Likewise, the implementation of deep models for *breast tumor segmentation* has spurred interest in the research community in recent years (Xian et al. 2018b). Whereas the most popular image modality for this task has been ultrasound images (Huang et al. 2018; Yap et al. 2018; Abraham and Khan 2019; Chiang et al. 2019) and digital mammography images (Dhungel et al. 2015; Akselrod-Ballin et al. 2017; Kooi et al. 2017; Ribli et al. 2017; Jung et al. 2018; Moor et al. 2018), a body of literature used magnetic resonance images (Jaeger et al. 2018) and histology images (Lin et al. 2018). U-Net (Ronneberger et al. 2015) and its numerous variants and modifications have been the most commonly used architecture for this problem to date. Despite this progress, breast tumor segmentation is still an open research topic because of the challenges related to the inherent presence of noise and low contrast of images; the sensitivity of current methods to the used image-acquisition method, equipment and settings; and the lack of large open data sets of annotated images for training purposes.

Priors in medical image segmentation

Incorporating prior task-specific knowledge for medical image segmentation is important for improved model performance (Nosrati and Hamarneh 2016), and it can be crucial in tasks with small data sets of annotated medical images (*i.e.,* most medical tasks at the present time). Prior knowledge can generally be in the form of shape, boundary, curvature, appearance (*e.g.,* intensity, texture), topology (*e.g.,* connectivity), anatomic information/atlas (structure of tissues or organs), user information (seed points or bounding boxes), moments (size, area, volume), distance (between organs and structures) and other forms. Although recent deep learning-based models have caused a leap in performance of image segmentation over conventional methods based on thresholding, region-growing, graph-based approaches and deformable models (Xiao et al. 2002; Liu et al. 2010; Cai and Wang 2013; Rodrigues et al. 2015; Gómez-Flores and Ruiz-Ortega 2016; Huang et al. 2017), incorporating prior knowledge into deep neural networks has proven to be a difficult task and, consequently, has not been widely investigated. Namely, semantic image segmentation using deep networks typically relies on loss functions that optimize the model predictions at a pixel level, without taking into consideration interpixel interactions and semantic correlations among regions at the image level. To integrate prior knowledge into segmentation models, several works have proposed *custom loss functions* that enforce learning feature representations compatible with the priors. For instance, a loss function that penalizes both geometric priors (boundary smoothness) and topological priors (containment or exclusion of lumen in epithelium and stroma) was devised for histology gland segmentation (BenTaieb and Hamarneh 2016). Likewise, loss functions in fully convolutional networks (FCNs) that encode a shape prior were proposed for kidney segmentation (Ravishankar et al. 2017), cardiac segmentation (Oktay et al. 2018a) and segmentation of star shapes in skin lesions (Mirikharaji and Hamarneh 2018). The disadvantage of this approach is that the related models are task specific and cannot be re-purposed for segmentation of other objects of interest in medical images. Another line of research introduces a *post-processing step* with conditional random fields, where the segmentation predictions by a deep learning network are improved through assignment of class labels to regions with similar topological properties (Havaei et al. 2017; Chen et al. 2018a; Huang et al. 2018). However, these methods increase the processing complexity and computational expense, and have been mostly replaced in recent years with end-to-end training models. Furthermore, a body of work has been proposed to incorporate shape priors by *redesigning the network architecture*. For example, Li et al. (2016) employed an FCN with a VGG-16 base model where shape priors are learned by a consecutive concatenation of the original images with the obtained segmentation maps during several iterations of the procedure. Gonzalez-Diaz (2017) created probability maps based on the knowledge of the patterns of skin lesions (*e.g.,* dots, globules, streaks or vascular structures) and merged them with extracted feature maps in a ResNet-based architecture. Furthermore, a boundary prior was incorporated into a deep learning model called Deep contour-aware network (DCAN) that has two subnetworks for learning concurrently shapes and contour boundaries in histology images (Chen et al. 2017). Yet another class of methods utilizes *generative models* for introducing prior knowledge. For example, in several early pre-FCN image segmentation models, Boltzmann machine networks were employed for learning shape priors (Chen et al. 2013; Eslami et al. 2014). In more recent research, variational Bayes autoencoders have been used to incorporate prior anatomic knowledge of the brain geometry in segmentation of magnetic resonance images (Dalca et al. 2018).

Despite the potential of the above-described research work, to the best of our knowledge, there are no previous studies on the incorporation of prior knowledge into deep models for breast cancer detection. The challenge stems from the fact that unlike other medical organs (*e.g.,* kidney, heart) where shape or boundary priors can be applied, such constraints are not applicable to breast cancer detection, because of the wide difference in the geometry of breast tumors. Analogously, it is difficult to extract generalized prior knowledge regarding curvature, moments, appearance, intensity or number of regions for breast tumors. In this work, we introduce prior topology information in a deep learning segmentation model in the form of region connectivity and visual saliency. Such prior information is combined with anatomic prior knowledge of the tissue layers in breast images, as explained in the subsequent sections.

Attention mechanism in deep learning. Attention mechanism is an approach in deep network layer design where the goal is to recognize discriminative features in the inner activation maps and to utilize this knowledge toward enhanced task-specific data representation and improved model performance (Simonyan et al. 2013). This mechanism contributes to suppressing less relevant features and emphasizing more important features for a considered task; for example, in image classification, important features lie in salient spatial locations in the images.

Attention mechanism has been integrated into various deep learning models designed for image captioning (Xu et al. 2015; Li et al. 2018), language translation (Bahdanau et al. 2015) and image classification (Wang et al. 2017; Jetley et al. 2018). In general, attention in deep neural networks is traditionally implemented in two main forms known as hard and soft attention. The implementation of hard (or stochastic) attention is non-differentiable, the training procedure is based on a sampling technique, and as a consequence, the models are difficult to optimize (Mnih et al. 2014; Stollenga et al. 2014; Cao et al. 2015). Soft (or deterministic) attention models are differentiable and trained with backpropagation; because of these properties, they have been the preferred form of implementation (Jaderberg et al. 2015; Chen et al. 2016; Wang et al. 2017). In image processing, the attention mechanism produces a probabilistic map of spatial locations in images, where the parameters of the attention map are learned in end-to-end training. Furthermore, the introduced architecture designs in image processing typically comprise multiple attention maps with different resolutions, thereby capturing salient features across multiple levels of feature abstraction. For instance, Jetley et al. (2018) introduced attention gates at three intermediate layers in a VGG network, and a weighted combination of the attention maps is used in the last layer for image classification. Chen et al. (2018a) introduced attention blocks in the initial DeepLab model for image segmentation, where attention weights are learned at different scales of a pyramidal feature representation.

Similar attention gates were introduced in a U-Net architecture (Oktay et al. 2018b) and were employed in medical image processing for segmentation of the pancreas (Oktay et al. 2018b) and for breast tumor and skin lesion segmentation (Abraham and Khan 2019). This type of model uses the extracted feature maps in the encoder path of the network for calculation of the attention maps, which are afterward merged with the upsampled feature maps in the decoder network, typically *via* elementwise multiplication. Such design forces the model to encode the locations and shapes of salient regions in extracted representations that are relevant for segmentation of the objects of interest. In the work by Tomita et al. (2018), an attention module was implemented in a 3-D residual convolutional neural network to dynamically identify regions of interest (ROIs) for processing high-resolution microscopy images, thus replacing the commonly used approach of sliding window ROI selection and alleviating the computational burden in processing microscopy images. In a work related to the proposed approach, AttentionNet was designed on top of a ResNeXt encoder-decoder architecture and applied both spatial and channel attention blocks for segmentation of the anatomic tissue layers in BUS images (Li et al. 2019). Conversely to our method, Li et al. (2019) did not apply AttenionNet to breast tumor detection; they also used activation maps of the intermediate layers of the network in the attention blocks.

METHODS

The proposed approach is validated on a data set of 510 BUS images (Xian et al. 2018a). The data set was collected from three hospitals: the Second Affiliated Hospital of Harbin Medical University, the Affiliated Hospital of Qingdao University and the Second Hospital of Hebei Medical University. All images in the data set were de-identified, and informed consent to the protocol was obtained from all involved patients. Different types of imaging ultrasound devices were employed for acquiring the images, including the GE VIVID 7 (General Electric Healthcare, Chicago, IL, USA), GE LOGIQ E9 (General Electric Healthcare), Hitachi EUB-6500 (Hitachi Medical Systems, Chiyoda, Japan), Philips iU22 (Philips Healthcare, Amsterdam, Netherlands) and Siemens ACUSON S2000 (Siemens Healthineers Global, Munich, Germany). The GE VIVID 7 and Hitachi EUB-6500 were used to collect ultrasound images at Harbin Medical University, the GE LOGIQ E9 and Philips iU22 were used at Qingdao University and the Siemens ACUSON S2000 was used at Hebei Medical University. Image annotation related to the segmentation and delineation of tumors in images was initially performed by three experienced radiologists, followed by voting and creation of a single segmentation mask per image on which all three medical professionals agreed. Afterward, the annotations were reviewed by a senior radiologist expert, who either approved or, if needed, applied corrections and amendments to the segmentation boundaries (Xian et al. 2018a).

Network architecture

The proposed network is based on the well-known U-Net architecture (Ronneberger et al. 2015), which consists of fully convolutional encoder and decoder subnetworks with skip connections. The layers in the encoder employ a cascade of convolutional and max-pooling layers, which reduce the resolution of input images and extract increasingly abstract features. The decoder comprises convolutional and upsampling layers that provide an expanding path for recovering the spatial resolution of the extracted feature maps to the initial level of the input images. A unique characteristic of the U-Net architecture is the presence of skip connections from the feature maps in the encoder's contracting path to the corresponding layers in the decoder. The features from the respective encoder's and decoder's layers are merged *via* concatenation, which allows recovery of the spatial accuracy of the objects in images and improves the resulting segmentation masks. Namely, although the central layer of the network offers high-level features with semantic-rich data representation and a large receptive field, it also has low level of spatial context detail because of the downsampling max-pooling layers along the contracting path and affects the localization accuracy around the object boundaries in the predictions. The skip connections provide a means to transmit low-level feature information from the initial high-resolution layers in the encoder to the reconstructing layers in the decoder, thereby restoring the local spatial information in predicted segmentations. Despite the introduction of deeper and more powerful models for image segmentation in recent years, the U-Net architecture has remained popular especially in medical image segmentation, where data sets are small and large models can overfit on the available sets.

Figure 1 is a graphical representation of the proposed model is presented. In addition to the main input consisting of BUS images, the network has an auxiliary input consisting of the corresponding salient maps. Attention blocks introduce salient maps of reduced scale in all layers on the contracting path of the encoder in the form of an image pyramid. This enforces the network to focus the attention on regions in the saliency maps with

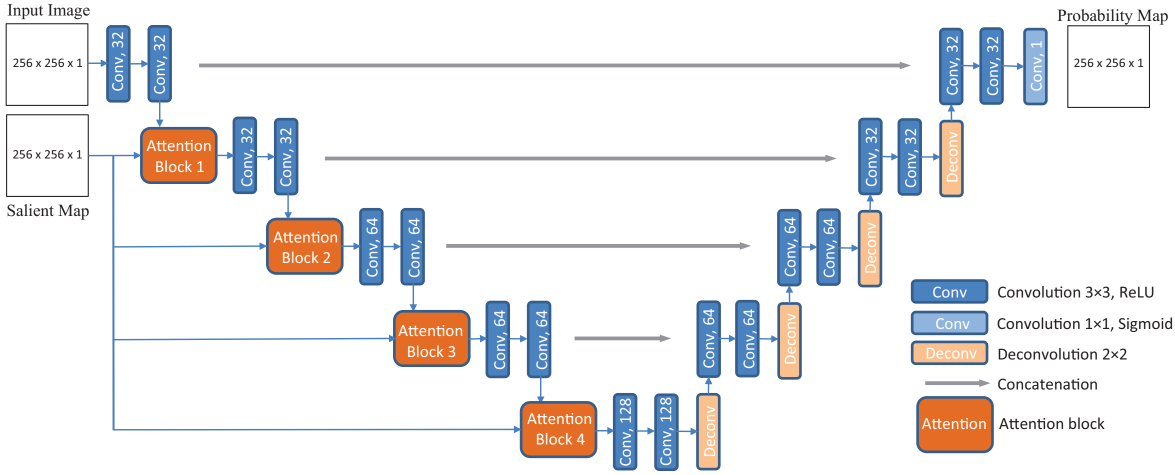

Fig. 1. Architecture of the proposed U-Net model with salient attention. The model uses breast ultrasound images and saliency maps as inputs and produces segmentation probability maps as outputs. Conv = convolution; Deconv = deconvolution.

high intensity values. More specifically, the introduced attention blocks put more weights on areas in the extracted feature maps at each layer that have higher levels of saliency in the salient maps. Thus, the topology of the salient maps influences the learned feature representations.

The images and saliency maps are gray-scale 8-bit data resampled into floating points with normalization. Resized images and saliency maps to 256 × 256 pixels are used as inputs to the model. The number of convolutional filters per layer in the network is (32, 32, 64, 64, 128), which is reduced in comparison to the original U-Net, to account for the relatively small data set. The output segmentation probability maps have the same spatial dimensions as the inputs. The proposed network is trained in an end-to-end fashion; however, the saliency maps are pre-computed and used at both training and inference.

Attention blocks. A block diagram of attention block n is depicted in Figure 2. The input feature maps to the attention block are denoted $F_n = \{f_1, f_2, \ldots, f_{k_n}\}$, where each feature map has horizontal and vertical spatial dimensions of $256/2^{(n-1)} \times 256/2^{(n-1)}$ pixels for the block in the layer level $n \in \{1, 2, 3, 4\}$. The symbol k_n is the channel dimension of the feature maps in block n, that is, $k_n \in \{32, 32, 64, 64\}$. For example, the input feature maps in Figure 2 related to the output activations of the convolutional "Conv 64" layer entering attention block 4 in Figure 1 have dimensions of $32 \times 32 \times 64$ (*i.e.*, for $n = 4$, the size of the feature maps F_4 is $256/2^3 \times 256/2^3 \times k_4 = 32 \times 32 \times 64$).

The input salient map in Figure 2 is denoted S, and it is downsampled through a max-pooling layer, resulting in S_n, which matches the spatial dimension of the input feature maps F_n in attention block n. Next, 1 × 1 convolutions followed by rectified linear unit (ReLU) activation functions are used to increase the number of channels of the saliency map S_n to 128. An elementwise sum block performs addition of F_n and S_n, producing intermediate maps I_n of size $256/2^n \times 256/2^n \times 128$. The intermediate maps I_n are further refined through a series of linear $128 \times 3 \times 3$ and $1 \times 1 \times 1$ convolutions, followed by non-linear ReLU activations. A sigmoid activation function normalizes the values of the activation maps into the [0, 1] range. The produced output is the attention map $A = (\alpha_i)$ with a spatial size of $256/2^n \times 256/2^n \times 1$, where the attention coefficients α_i have scalar values for each pixel i. Next, soft attention is applied *via* elementwise multiplication of the attention map A with the max-pooled features P_n, that is, $O_n = A * P_n$. The activation maps O_n with size $256/2^n \times 256/2^n \times k_n$ are the output of attention block n, and they are further propagated to the next layer, as depicted in Figure 1.

The design of the attention block was inspired by the attention gates in Oktay et al. (2018b) and Jetley et al. (2018). Differently from these two works, where the attention blocks employ activation maps from the intermediate layers in the model as saliency maps to enhance the discriminative characteristics of extracted intermediary features, the proposed attention block in this work utilizes pre-computed saliency maps that point out to target spatial regions. If the attention block in this work applies directly to the self-attention blocks described in Abraham and Khan (2019), Jetley et al. (2018) and Oktay et al. (2018b), the segmentation performance of the model would not improve. The

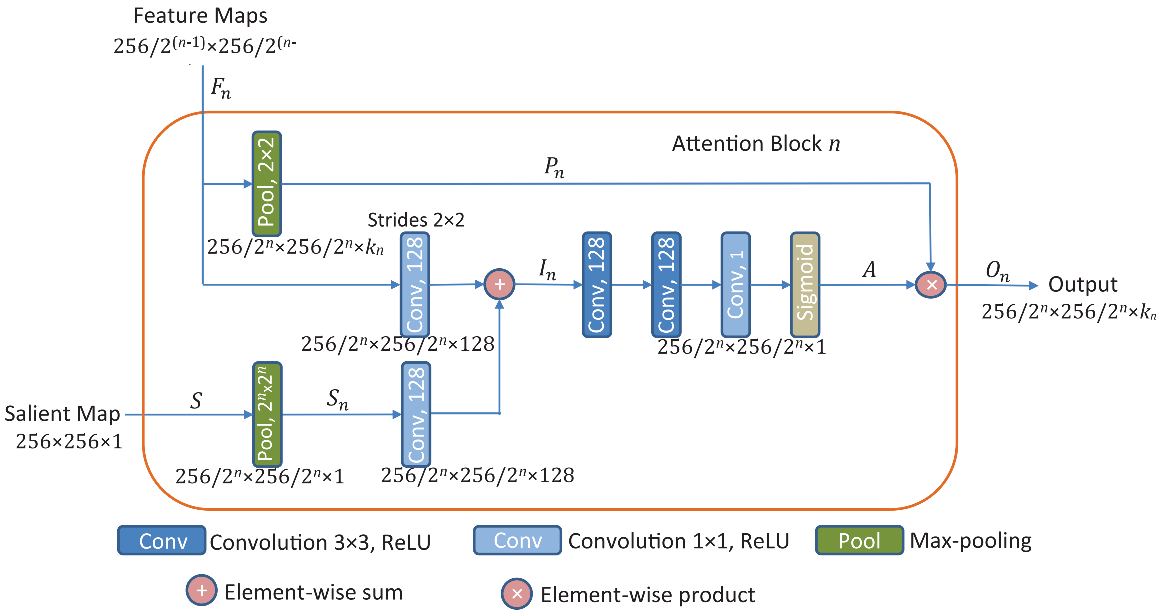

Fig. 2. Attention block n for $n \in \{1, 2, 3, 4\}$. Inputs to the block are feature maps from layer n, with spatial dimensions $256/2^{(n-1)} \times 256/2^{(n-1)}$ with k_n number of channels, and a salient map, whereas the output is downsampled weighted maps with spatial dimensions $256/2^n \times 256/2^n$ and k_n number of channels.

reason for that lies in the distribution of salient regions in the maps used, as in many images, background non-tumor areas have a certain level of saliency in the salient maps; consequently, placing equal attention weights on all salient regions leads to higher levels of false-positive errors and degraded performance. The introduction of additional 3 × 3 and 1 × 1 convolutional layers for feature refinement in the proposed salient attention block was conducive to improved segmentation outputs, which was confirmed *via* empirical validation of the proposed layer design.

Saliency maps. Visual saliency estimation is an important paradigm for automatic tumor diagnosis in BUS images, where the aim is to model the level of saliency of image regions in correspondence to the capacity to attract radiologists' visual attention (Shao et al. 2015; Xie et al. 2017). For an input image, the output of such models is a visual saliency map with saliency values in the [0, 1] range assigned to every image pixel. A high saliency value indicates a high probability that the pixel belongs to a tumor.

The approach adopted for generating saliency maps of BUS images is based on our previous work (Xian et al. 2016, 2017; Xu et al. 2016, 2018, 2019). In particular, the task of visual saliency estimation is formulated as a quadratic programming optimization that integrates high-level image information and low-level saliency assumptions. The model assigns a saliency value s_i to each superpixel region i in an image. The objective function of the model optimizes several terms, as follows. First, one term is a function of a foreground map that calculates the probability that the ith image region belongs to a tumor and the distance between the ith region and the center of the foreground map of the image. Second, another term defines the cost of assigning zero saliency to an image region, and it employs the connectedness to the boundary regions to calculate the probability of the ith region belonging to a non-tumor image background. A third term applies a penalty if similar regions in the image have different saliency values. The formulation of the above functions is based on our neutro-connectedness approach (Xian et al. 2016; Xian 2017), which exploits the information of the degree of connectedness and confidence of connectedness between the image regions. The complete set of formulas for derivation of the optimization model can be found in Xu et al. (2019, 2018).

Our most recent work on this topic (Xu et al. 2019) introduced additional constraints in the model related to breast anatomy by decomposing the images into four anatomic layers: skin, fat, mammary and muscle layers. The four layers have different appearances in BUS images, and the fact that tumors are present predominantly in the mammary layer is used in our framework as an anatomic prior for saliency estimation. Two low-level saliency assumptions are utilized in the framework as well: (i) the adaptive-center bias assumption forces the regions nearer the adaptive center to have higher saliency values; (ii) the region-correlation assumption forces the similar regions to have similar saliency values. The extensive experiments in Xu et al. (2019) indicated that the new model with anatomic knowledge resulted in improved performance compared with models in related works on the data set (Xian 2018b). Another advantage of the approach proposed in Xu et al. (2019) is the capability to interpret images without tumors, whereas many related approaches assume the presence of tumors in each image. Full implementation details can be found in the respective publications.

Examples of breast images and corresponding saliency maps are presented in Figure 3. In the top row of Figure 3 are five BUS images, and in the middle row are the ground truth segmentation masks provided by radiologists. The saliency maps for the images are in the bottom row. It can be noted that the saliency maps assign a value to every pixel regarding the probability of belonging to a tumor, and differently from the ground truth masks, saliency values are assigned to background regions in images as well. Furthermore, the saliency maps are generated in an unsupervised manner; that is, the information of the ground truth is not used by the saliency estimation model.

The incorporation of saliency maps into a deep learning model as complementary prior information is based on an assumption that the areas in images with high saliency values correspond to a high probability of tumor presence. Therefore, it is important that the saliency maps are of adequate quality and provide reliable information regarding the tumor locations. Otherwise, poor-quality saliency maps can degrade model performance.

The five selected examples of saliency maps depicted in Figure 3 have different levels of quality. More specifically, a map is considered of satisfactory quality when the location and intensity of the tumor region are clearly discernable in the saliency map. The example in the middle column in Figure 3 with moderate quality indicates the tumor location correctly, but the tumor shape and boundary do not match very well the ground truth, which may cause errors in the edge segmentation when applied to a deep network. For the case with low quality in Figure 3, there are several regions with similar area and saliency values, and it is not clear which of these regions may be tumors. Lastly, the saliency map with poor quality in Figure 3 assigns zero saliency values to the tumor region and completely misses the tumor.

To account for the cases with lower quality of saliency maps, we devised an algorithm that calculates the level of confidence in the saliency maps and, subsequently, eliminates the maps with low confidence. The approach is based on the following parameters: contour area $A_c = \sum_j p_j$ is the number of pixels of a contour c in an image with a saliency value per pixel p_j greater than a threshold value; cumulative intensity $I_c = \sum_j S(p_j)$ calculates the sum of the saliency values for the pixels in contour c; and mean intensity $M_c = \sum_j S(p_j)/A_c$ of a contour c is calculated as the ratio of the cumulative intensity and the area. The first rule in Algorithm 1 states that if the contour with the largest cumulative intensity

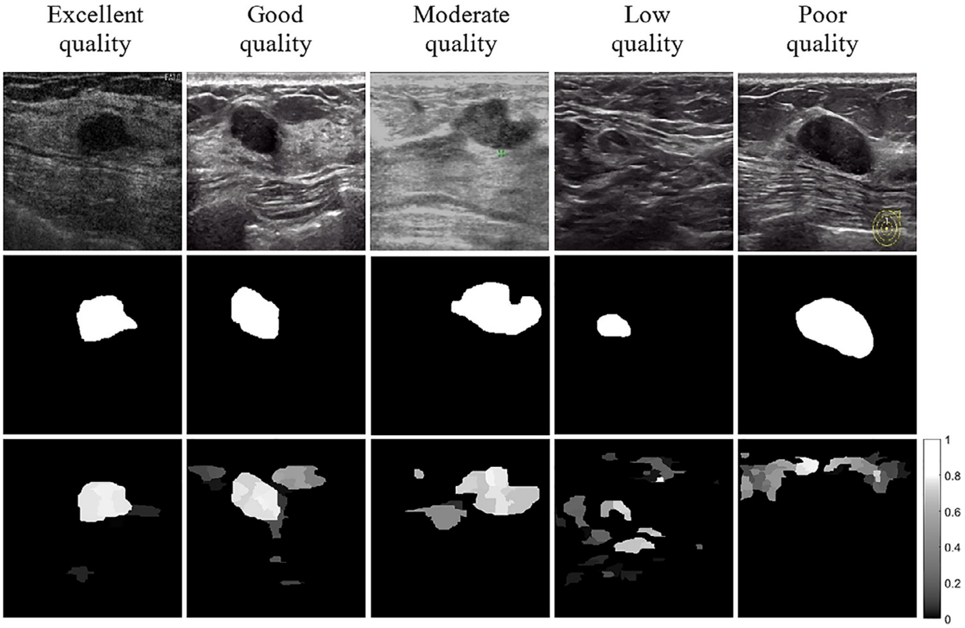

Fig. 3. Examples of saliency maps with varying levels of quality. Top row: Original breast ultrasound image. Middle row: Ground truth mask. Bottom row: Saliency map.

Algorithm 1. Confidence level calculation for saliency maps

```
For saliency map $i = 1 : N$
  Find all fully connected contours with threshold $> 0.3$
  For contour $c = 1 : C$
    if $I_{\text{argmax}(I_{1:c})} < a_1 I_{\text{argmax}(I_{1:c})-1}$ and $M_{\text{argmax}(I_{1:c})} < M_{\text{argmax}(M_{1:c})}$
    or if $I_{\text{argmax}(I_{1:c})} < a_2 I_{\text{argmax}(I_{1:c})-1}$ and
    $M_{\text{argmax}(I_{1:c})} + a_3 < M_{\text{argmax}(M_{1:c})}$
    or if $M_{\text{argmax}(M_{1:c})} > a_4$ and $\text{argmax}(M_{1:\,c}) \neq \text{argmax}(I_{1:\,c})$
    Remove saliency map $i$ from the set
```

$I_{\text{argmax}(I_{1:c})}$ has cumulative intensity similar to that of the second largest contour and its mean intensity $M_{\text{argmax}(I_{1:c})}$ is not the highest of all contours (see Fig. 4, left column), then eliminate the saliency map from the set. The second rule is similar to the first rule and takes into account cases with larger ambiguities in the cumulative intensity and mean intensity of contours (Fig. 4, middle column). The third rule considers the cases when a contour has high mean saliency intensity but smaller cumulative intensity than other contours in the image (see Fig. 4, right column). The parameters in the algorithm are empirically set to $a_1 = 2$, $a_2 = 3$, $a_3 = 0.2$ and $a_4 = 0.55$. In total, 52 saliency maps satisfied the given conditions and were removed from the original set of 562 images, resulting in a reduced set of 510 images. That is, approximately 91% of the saliency maps are with high level of confidence. Having a low level of confidence for a saliency map does not necessarily mean that the saliency map is not correct (*e. g.,* one can argue that the saliency for the example in the middle column in Figure 4 is correct). Rather, the proposed algorithm is designed to identify saliency maps with ambiguities regarding the spatial regions for tumor existence. The algorithm takes as inputs only the saliency maps, and it does not use the knowledge of the ground truth in estimating the level of confidence.

Evaluation metrics

We used the DSC, Jaccard index (JI), true-positive ratio (TPR), false-positive ratio (FPR) and global accuracy (ACC) to evaluate performance of the model:

$$\text{DSC} = \frac{2|A_g \cap A_p|}{|A_g| + |A_p|} \tag{1}$$

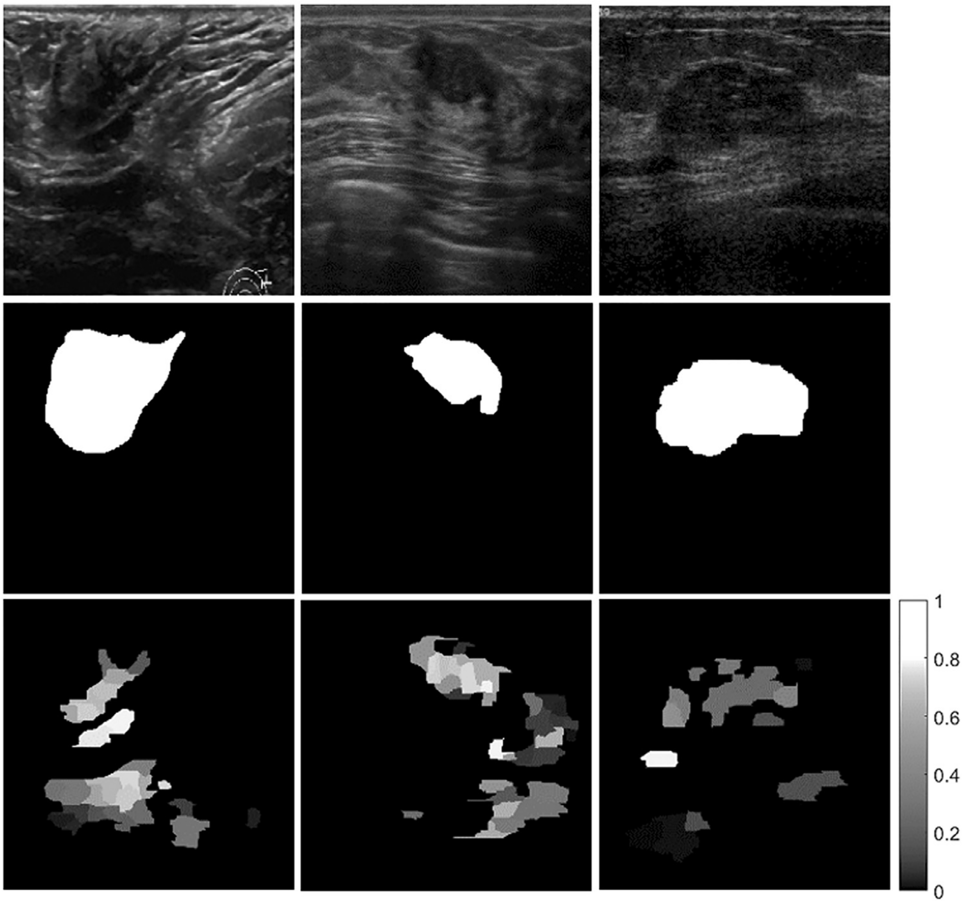

Fig. 4. Examples of eliminated saliency maps from the original data set. Top row: Original breast ultrasound image. Middle row: Ground truth mask. Bottom row: Saliency map.

$$\mathrm{JI}=\frac{|A_{\mathrm{g}}\cap A_{\mathrm{p}}|}{|A_{\mathrm{g}}\quad \text{x222A; } A_{\mathrm{p}}|} \tag{2}$$

$$\mathrm{TPR}=\frac{|A_{\mathrm{g}}\cap A_{\mathrm{p}}|}{|A_{\mathrm{g}}|} \tag{3}$$

$$\mathrm{FPR}=\frac{|A_{\mathrm{g}}\cup \mathrm{A_p}-\mathrm{A_g}|}{|A_{\mathrm{g}}|} \tag{4}$$

$$\mathrm{ACC}=\frac{|A_{\mathrm{g}}\cap A_{\mathrm{p}}|+|\ A_{\mathrm{g}}-A_{\mathrm{g}}\cup \mathrm{A_p}|}{|A_{\mathrm{g}}|+|\overline{A}_{\mathrm{g}}|} \tag{5}$$

In eqns (1)–(5), A_{g} is the set of pixels that belong to a tumor region in the ground truth segmented images, $\overline{A}_{\mathrm{g}}$ is the set of pixels that belong to the background region without tumors in the ground truth segmented images and A_{p} is the corresponding set of pixels that are predicted to belong to a tumor region by the segmentation method. It is important to note that FPR is calculated as the ratio divided by the number of positives (*i.e.,* pixels in tumor regions in the ground truth masks), as opposed to a ratio divided by the number of negatives (*i.e.,* pixels in the background regions in the ground truth masks), as it is often defined in related tasks. Because the positive regions are smaller in BUS tumor segmentation, the selected formulation for FPR is more descriptive for this task. Additional metrics that we used for performance evaluation are the area under the curve of the receiver operating characteristic score (AUC-ROC), Hausdorff distance (HD) and mean distance (MD). For most of these metrics, the values are in the [0, 1] range, where higher values indicate improved performance (except for FPR, HD and MD, where low values are preferred).

The differences in the values of the metrics obtained by different models are evaluated with paired-comparison statistical hypothesis testing. A null hypothesis assumes that the metrics values are drawn from the same distribution and have a median value equal to zero.

Implementation details

The proposed approach was validated on the described data set of BUS images. We used fivefold cross-validation, where fourfolds (80% of images) are used for training, and onefold (20% of images) is used for testing. Validation during training is performed on 20% of the training set of images. All images in the data set are first resized to a 256 × 256-pixel resolution. Because we focused on understanding the impact of the introduced salient attention on model performance, we did not apply image augmentation.

The proposed model is trained with randomly initialized weights using Xavier normal initialization (Glorot and Bengio 2010). The Dice loss function was used for training, defined as

$$\mathcal{L}\ =\ 1-\mathrm{DSC}\ =\ 1-\frac{2|A_{\mathrm{g}}\cap A_{\mathrm{p}}|}{|A_{\mathrm{g}}|+|A_{\mathrm{p}}|} \tag{6}$$

where the same notation is preserved (*i.e.,* A_{g} and A_{p} denote the ground truth and predicted masks, respectively).

The models were implemented using TensorFlow (Google, Menlo Park, CA, USA) and Keras (Francois Chollet, Menlo Park, CA, USA) libraries on the Google Colaboratory cloud computing services, which employ Tesla K80 GPUs. The network was trained by using the adaptive moment estimation optimizer (Adam) with a learning rate of 10^{-4} and a batch size of four images. The training was stopped when the loss of the validation set did not improve for 20 epochs.

RESULTS

Evaluation and comparative analysis

Experimental validation of the proposed approach is based on a comparative analysis of three models:

1. U-Net
2. U-Net-SA, which applied the proposed salient attention approach
3. U-Net-SA-C, which is a model with salient attention applied to a modified version, where only one contour with the highest saliency is extracted in each salient map.

Examples of input BUS images, ground truth masks, saliency maps and output segmentation maps by the models are provided in Figure 5. The values of the performance metrics are listed in Table 1. For the BUS images displayed in Figure 5, the segmentation outputs by the U-Net model are inferior in comparison to the predicted masks produced by the models with salient attention U-Net-SA and U-Net-SA-C. One particular aspect of improved performance entails the false-positive predictions by U-Net (see rows A–G in Fig. 5). In these cases, U-Net produces positive predictions of tumor presence for image regions that do not belong to a tumor. The attention models U-Net-SA and U-Net-SA-C benefited from the information in the salient maps, which led to a reduced rate of false-positive predictions in A–G. This is especially noticeable in rows B, E and G, which have high-quality salient maps, resulting in great improvement over the predictions by the basic U-Net model.

Furthermore, improved performance with respect to the true-positive predictions by U-Net is illustrated for rows H and I in Figure 5. The provision of salient maps for these two cases helps the model to focus on target regions with high saliency, leading to higher true-positive rates of the segmentation masks by U-Net-SA over the basic U-Net model. In addition, rows J and K provide examples where the geometry of the salient regions in the saliency maps contributes to more accurate predictions of the proposed models in comparison to U-Net. Cases C and I are instances of BUS images with small tumors, where the salient attention models successfully located the tumor regions. As explained earlier, the U-Net-SA-C model employs salient maps with one contour with the highest saliency intensity, and in many images it further improves the segmentation outputs. This is noticeable in row A in Figure 5, where the false positives in the segmentation are reduced in comparison to U-Net-SA. However, the U-Net-SA-C model is based on an assumption that there is only one tumor in the images, which may not always be the case.

Table 1 lists the averages (standard deviations) per fold in the fivefold cross-validation procedure for the three deep models. The values obtained indicate that the models with salient attention U-Net-SA and U-Net-SA-C outperform the basic U-Net network without attention blocks for all performance metrics. The model U-Net-SA-C trained on the data set with a single contour in the salient maps produced improved segmentation performance in comparison to U-Net-SA. The average training time per fold for the basic U-Net model was 7.58 min, whereas the corresponding times for training the salient attention models U-Net-SA and U-Net-SA-C were 8.54 and 8.08 min, respectively. Segmentation of the testing set of images with a trained model took 1.09, 1.26 and 1.37 s per fold (*i.e.,* 102 images) for U-Net, U-Net-SA and U-Net-SA-C, respectively. This translates to processing times of 12 ms per image for U-Net-SA and 13 ms per image for U-Net-SA-C.

A Wilcoxon signed rank test was adopted for statistical analysis, based on the distribution of the metrics values. The hypothesis testing results are summarized in Table 2. The cells with asterisks indicate rejection of the null hypothesis at a p value <0.05. ACC and AUC-ROC metrics are not included in the test because their values are calculated per a fold of 20% of the images, and not

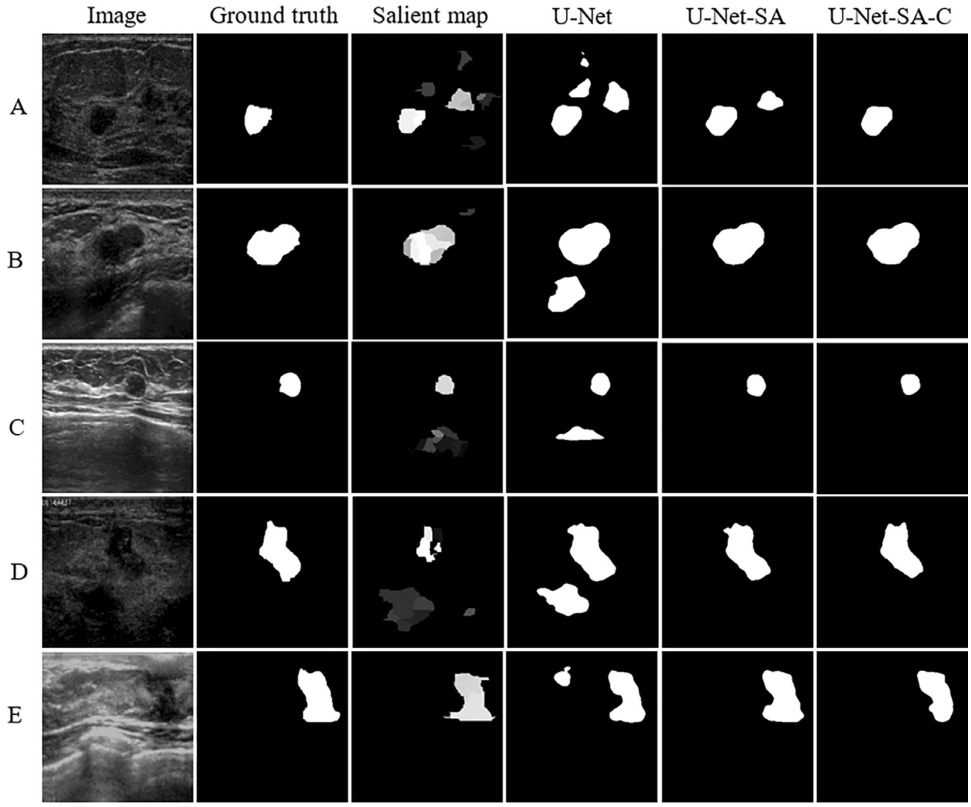

Fig. 5. Segmentation results. First column: original breast ultrasound image. Second column: Ground truth mask. Third column: Saliency map. Fourth column: Segmentation mask produced by U-Net. Fifth column: Segmentation mask produced by U-Net-SA. Sixth column: Segmentation mask produced by U-Net-SA-C.

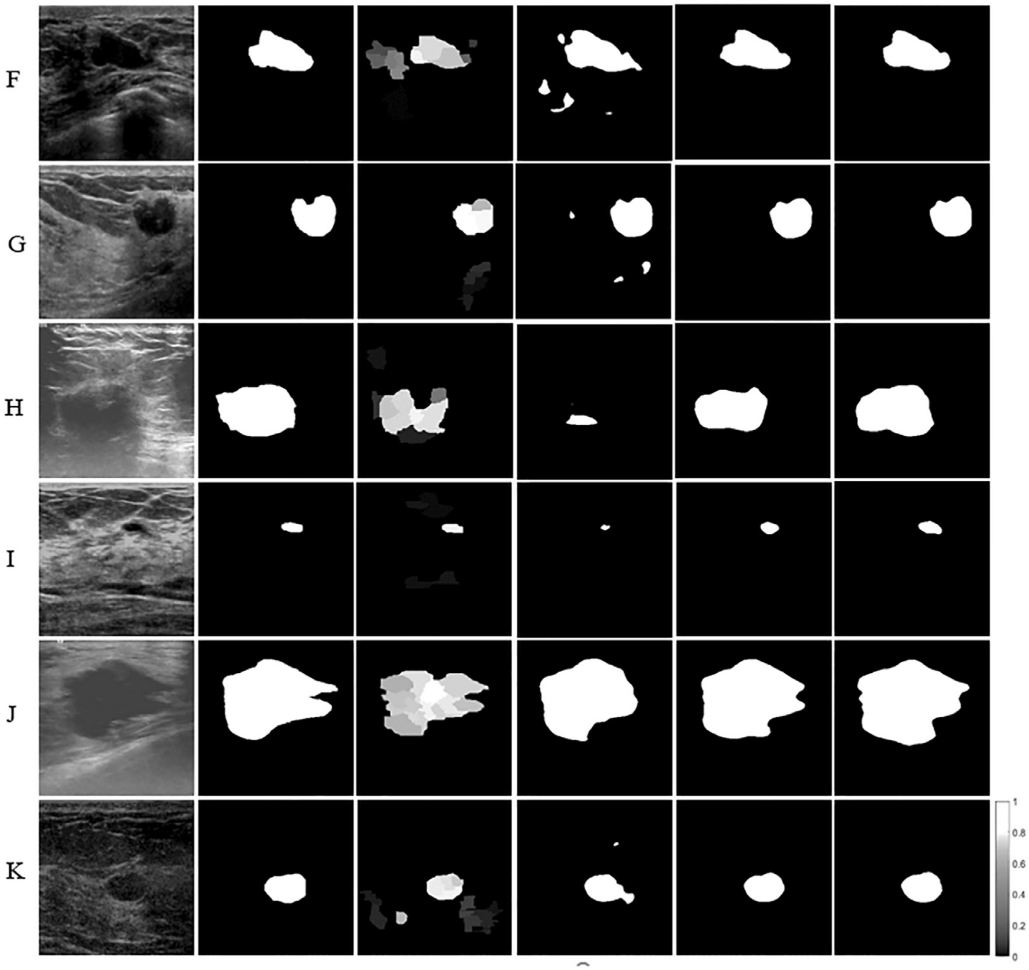

Fig. 5. Continued

Table 1. Performance evaluation metrics for models without and with salient attention*

Model	DSC	JI (IOU)	TPR	FPR	ACC	AUC-ROC	HD	MD
U-Net	0.894 (±0.013)	0.821 (±0.017)	0.903 (±0.011)	0.107 (±0.019)	0.978 (±0.002)	0.951 (±0.006)	4.346 (±1.377)	0.224 (±0.240)
U-Net-SA	0.901 (±0.013)	0.832 (±0.014)	0.904 (±0.016)	0.092 (±0.008)	0.979 (±0.001)	0.955 (±0.002)	4.326 (±1.360)	0.209 (±0.234)
U-Net-SA-C	**0.905** (±0.013)	**0.838** (±0.014)	**0.910** (±0.011)	**0.089** (±0.012)	**0.980** (±0.001)	**0.957** (±0.004)	**4.271** (±1.326)	**0.201** (±0.218)

ACC = global accuracy; AUC-ROC = area under the receiver operating characteristic curve; DSC = Dice similarity coefficient; FPR = false-positive ratio; HD = Hausdorff distance; JI (IOU) = Jaccard index (intersection over union); MD = mean distance; TPR = true-positive ratio.

* Values are the average (±standard deviation) per fold in fivefold cross-validation.

The values in bold font indicate the best performance for each metric

Table 2. Wilcoxon signed rank test of the performance metrics per image

Model	*p* value					
	DSC	JI (IOU)	TPR	FPR	HD	MD
U-Net and U-Net-SA	0.0011*	<0.0001*	0.5822	<0.001*	0.2592	<0.001*
U-Net and U-Net-SA-C	<0.000*	<0.0001*	0.0052*	0.0098*	0.0345*	<0.00*

DSC = Dice similarity coefficient; FPR = false-positive ratio; HD = Hausdorff distance; JI (IOU) = Jaccard index (intersection over union); MD = mean distance; TPR = true-positive ratio.

* Statistically significant difference, *p* value <0.05.

The values in bold font indicate the best performance for each metric

Table 3. Values of the performance metrics for tumor segmentation by different models*

Model	Training setting	DSC	JI (IOU)	TPR	FPR	ACC	AUC- ROC
Seg-Net	LR = 8 · 10^{-4}, decreased by 0.5 after 10 epochs until 1 · 10^{-4}	0.889 (±0.011)	0.811 (±0.015)	0.877 (±0.019)	**0.088** (±0.014)	0.977 (±0.002)	0.957 (±0.004)
DenseNet-26	LR = 1 · 10^{-3}, decreased by 0.1 after 10 epochs until 1 · 10^{-4}	0.888 (±0.016)	0.818 (±0.017)	0.886 (±0.019)	0.093 (±0.025)	0.978 (±0.002)	**0.958** (±0.005)
PSPNet-ResNet18	LR = 1 · 10^{-4}, decreased by 0.5 after 10 epochs until 5 · 10^{-5}, image size of 384 × 384 pixels	0.886 (±0.008)	0.808 (±0.008)	0.884 (±0.014)	0.107 (±0.016)	0.976 (±0.002)	0.953 (±0.005)
Ours: U-Net-SA	LR = 1 · 10^{-4}	**0.901** (±0.013)	**0.832** (±0.014)	**0.904** (±0.016)	0.092 (±0.008)	**0.979** (±0.001)	0.955 (±0.002)

ACC = global accuracy; AUC-ROC = area under the receiver operating characteristic curve; DSC = Dice similarity coefficient; FPR = false-positive ratio; JI (IOU) = Jaccard index (intersection over union); LR = learning rate used for training the model; TPR = true-positive ratio.

* Values are the average (±standard deviation) per fold in fivefold cross-validation.

Table 4. Performance evaluation metrics for the models on the original data set of 562 images*

Model	DSC	JI (IOU)	TPR	FPR	ACC	AUC- ROC
U-Net	0.891 (±0.005)	0.817 (±0.008)	0.900 (±0.009)	0.120 (±0.027)	0.977 (±0.002)	0.950 (±0.006)
U-Net-SA	0.894 (±0.006)	0.824 (±0.008)	**0.901** (±0.017)	0.111 (±0.032)	0.978 (±0.002)	0.952 (±0.012)
U-Net-SA-C	**0.896** (±0.007)	**0.825** (±0.010)	0.899 (±0.020)	**0.106** (±0.025)	**0.978** (±0.002)	**0.955** (±0.010)

ACC = global accuracy; AUC-ROC = area under the receiver operating characteristic curve; DSC = Dice similarity coefficient; FPR = false-positive ratio; JI (IOU) = Jaccard index (intersection over union); TPR = true-positive ratio.

* Values are the average (±standard deviation) per fold in fivefold cross-validation.

per individual images. Accordingly, for almost all metrics there is a statistically significant difference in the median values by the proposed models in comparison to U-Net. The exceptions are the TPR and HD values between U-Net and U-Net-SA, for which there is no statistically significant difference.

Next, a comparison of our salient attention model for tumor segmentation U-Net-SA and three respective deep models for image segmentation is provided in Table 3. The data set with 510 images is used for training the models. For a fair comparison, all models are trained in the same manner as the proposed architecture, that is, fivefold cross-validation, batch size of 4, Xavier normal weights initialization, Dice loss, Adam optimizer and a stopping criterion of 20 epochs of non-improved validation loss. Because of the relatively small size of the data set, for the comparison we selected smaller versions of the models. For instance, DenseNet is based on a network with 26 layers, and for PSPNet (that requires a base model), the small residual model ResNet18 is employed. The learning rate is fine-tuned for the different models, where an initial learning rate is selected, and when the validation loss does not improve for 10 epochs, the learning rate is reduced by a certain step size. The procedure is repeated until a pre-set value for the learning rate is reached. The details regarding the used learning rates for the different models are provided in Table 3. Our proposed U-Net-SA model, listed last in the table, outperformed the other deep learning networks for image segmentation on most of the performance metrics employed.

Table 4 provides the values of the performance metrics for the models on the original data set of 562 images. In comparison to the values presented in Table 1 on the reduced data set of 510 images, the results in Table 4 indicate that the performance of the proposed attention-enriched models U-Net-SA and U-NET-SA-C is reduced on the original data set. Moreover, the basic U-Net model without salient attention also has reduced performance on the data set of 562 images, which implies that the subset of 52 images that were removed from the original data set contains breast tumors that are more challenging for segmentation in general. In conclusion, the algorithm for determining the level of confidence of the saliency maps contributed to improved performance on the reduced data set of 510 images, by ensuring that the model predictions are not inhibited by poor data.

DISCUSSION

On the basis of the evaluation results presented in Table 1, the models with attention blocks outperformed the basic U-Net model. In addition, if only one contour with the highest saliency is extracted in the saliency maps (the U-Net-SA-C model), the performance improves further. This can be explained by the increasing spatial attention to a single salient region in the maps, resulting in reduced false positives in the outputs. As we mentioned earlier, this is based on an assumption that there is only one tumor in the images, which may not always be the case.

The design of the attention blocks has an impact on the segmentation output; therefore, we investigated several alternatives for the block layers and their parameters. Compared with similar attention blocks in deep models (Chen et al. 2016; Jetley et al. 2018; Oktay et al. 2018b), the block used in this work requires additional feature refinement by using convolutional 3 × 3 and 1 × 1 layers. The refinement layers balance the impact of inaccurate boundaries of the regions in salient maps on the learned features. In other words, the saliency maps do not provide accurate local information on the edges and boundaries of tumors in images, but rather, they provide global information on the spatial probability regarding the presence of tumors. Larger values of the attention coefficients place more emphasis on the edges and boundaries in salient maps and can reduce the segmentation outputs. The use of additional refinement layers lessens the values of the attention coefficients and results in improved tumor segmentation.

The fact that the ultrasound images for validation of the approach were collected with various imaging systems is a strength of the work described here because it makes the data set suitable for training data-driven models with enhanced robustness to variations across images from different sources.

One limitation of the presented approach is that it relies on the quality of saliency maps. Using low-quality maps can at best not improve the results or can result in degraded performance. To deal with this shortcoming, we proposed an algorithm that calculates a confidence score and eliminates the saliency maps with low confidence in their level of quality. Whereas visual saliency estimation is not the focus of this work, improvements in the models for visual saliency estimation can lead to improved segmentation by the proposed approach.

Avenues for future work include investigation of custom loss functions in deep learning models for encoding prior information and working with medical partners to obtain annotated images with breast tissue layers and afterward integrating such anatomic priors with salient maps in a unified segmentation model.

CONCLUSIONS

In this article, a novel deep learning architecture that incorporates radiologists' visual attention for breast tumor segmentation is proposed. The proposed architecture consists of a variant of the basic U-Net model with attention blocks integrated along the contracting path in the layers of the encoder. The proposed attention blocks allow the deep learning model to suppress spatial regions with low saliency values and, respectively, to focus on regions with high saliency values. The attention blocks use multiscaled versions of the saliency maps. The approach is validated on a data set of 510 images, and the results indicate improved segmentation performance. The importance of this work stems from the difficulties in incorporating priors into deep learning models for medical image processing, and in particular for segmentation of BUS images, where most of the traditionally used prior forms cannot be applied.

Acknowledgments—This work was supported by the Center for Modeling Complex Interactions (CMCI) at the University of Idaho through National Institutes of Health Award P20GM104420. We thank Fei Xu for providing the visual saliency maps from her latest research and for her constructive feedback and review of the article.

REFERENCES

Abraham N, Khan NM. A novel focal Tversky loss function with improved attention U-Net for lesion segmentation. 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy. Piscataway, NJ. : IEEE; 2019. p. 683–687.

Akselrod-Ballin A, Karlinsky L, Hazan A, Bakalo R, Horesh AB, Shoshan Y, Barkan E. Deep learning for automatic detection of abnormal findings in breast mammography. In: Proceedings, Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support—Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, held in conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, 2017. Cham. : Springer; 2017. p. 321–329.

Badrinarayanan V, Kendall A, Cipolla R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell 2017;39:2481–2495.

Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. In: Bengio Y, LeCun Y, (eds). 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings. .

BenTaieb A, Hamarneh G. Topology aware fully convolutional networks for histology gland segmentation. In: Ourselin S, Joskowicz L, Sabuncu MR, Ünal GB, Wells W, (eds). Medical Image Computing and Computer-Assisted Intervention, MICCAI 2016, 19th International Conference, Athens, Greece, October 17–21, 2016, Proceedings, Part II. Cham. : Springer; 2016. p. 460–468.

Cai L, Wang Y. A phase-based active contour model for segmentation of breast ultrasound images. 6th International Conference on Biomedical Engineering and Informatics, BMEI 2013, Hangzhou, China, December 16–18. 91–95.

Cao C, Liu X, Yang Y, Yu Y, Wang J, Wang Z, Huang Y, Wang L, Huang C, Xu W, Ramanan D, Huang TS. Look and think twice: Capturing top-down visual attention with feedback convolutional neural networks. 2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7–13, 2015 IEEE Computer Society. 2956–2964.

Chen F, Yu H, Hu R, Zeng X. Deep learning shape priors for object segmentation. 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, June 23–28. 2013 IEEE Computer Society. 1870–1877.

Chen LC, Yang Y, Wang J, Xu W, Yuille AL. Attention to scale: Scale-aware semantic image segmentation. 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las

Vegas, NV, USA, June 27−30; 2016 IEEE Computer Society. 3640–3649.

Chen H, Qi X, Yu L, Dou Q, Qin J, Heng PA. DCAN: Deep contour-aware networks for object instance segmentation from histology images. Med Image Anal 2017;36:135–146.

Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans Pattern Anal Mach Intell 2018a;40:834–848.

Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder−decoder with atrous separable convolution for semantic image segmentation. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y, (eds). Computer Vision—ECCV 2018—15th European Conference, Munich, Germany, September 8−14, 2018, Proceedings, Part VII. Cham. : Springer; 2018b. p. 833–851.

Cheng HD, Shan J, Ju W, Guo Y, Zhang L. Automated breast cancer detection and classification using ultrasound images: A survey. Pattern Recognition 2010;43:299–317.

Chiang TC, Huang YS, Chen RT, Huang CS, Chang RF. Tumor detection in automated breast ultrasound using 3-D CNN and prioritized candidate aggregation. IEEE Trans Med Imaging 2019;38:240–249.

Dalca AV, Guttag JV, Sabuncu MR. Anatomical priors in convolutional networks for unsupervised biomedical segmentation. 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18−22, 2018. IEEE Computer Society. 9290–9299.

Dhungel N, Carneiro G, Bradley AP. Deep learning and structured prediction for the segmentation of mass in mammograms. In: Navab N, Hornegger JIII, Wells WM, Frangi AF, (eds). Medical Image Computing and Computer-Assisted Intervention, MICCAI 2015, 18th International Conference, Munich, Germany, October 5−9, 2015, Proceedings, Part I. Cham. : Springer; 2015. p. 605–612.

Elmore JG, Wells CK, Lee CH, Howard DH, Feinstein AR. Variability in radiologists' interpretations of mammograms. N Engl J Med 1994;331:1493–1499.

Eslami SMA, Heess N, Williams CKI, Winn JM. The shape Boltzmann machine: A strong model of object shape. Int J Computer Vision 2014;107:155–176.

Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: The YW, Titterington M, (eds). AISTATS 2010—Thirteenth International Conference on Artificial Intelligence and Statistics, Chia Laguna Resort, Sardinia, Italy, May 13−15, 2010. 9, 249–256. Proc Mach Learn Res.

Gómez-Flores W, Ruiz-Ortega BA. New fully automated method for segmentation of breast lesions on ultrasound based on texture analysis. Ultrasound Med Biol 2016;42:1637–1650.

González-Díaz I. Incorporating the knowledge of dermatologists to convolutional neural networks for the diagnosis of skin lesions. CoRR 2017; abs/1703.01976.

Goodfellow I, Bengio Y, Courville A. Deep learning. Cambridge, MA: MIT Press; 2016.

Graham S, Vu QD, Raza SEA, Kwak JT, Rajpoot NM. XY network for nuclear segmentation in multi-tissue histology images. CoRR 2018; abs/1812.06499.

Havaei M, Davy A, Warde-Farley D, Biard A, Courville AC, Bengio Y, Pal C, Jodoin PM, Larochelle H. Brain tumor segmentation with deep neural networks. Med Image Anal 2017;35:18–31.

He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. arXiv:151203385 [cs]2015.

Hu S, Worrall DE, Knegt S, Veeling BS, Huisman H, Welling M. Supervised uncertainty quantification for segmentation with multiple annotations. CoRR 2019; abs/1907.01949.

Huang Q, Luo Y, Zhang Q. Breast ultrasound image segmentation: a survey. Int J Comput Assist Radiol Surg 2017;12:493–507.

Huang K, Cheng HD, Zhang Y, Zhang B, Xing P, Ning C. Medical knowledge constrained semantic breast ultrasound image segmentation. 24th International Conference on Pattern Recognition, ICPR 2018, Beijing, China, August 20−24, 2018. : IEEE Computer Society; 2018. p. 1193–1198.

Inoue K, Yamanaka C, Kawasaki A, Koshimizu K, Sasaki T, Doi T. Computer aided detection of breast cancer on ultrasound imaging using deep learning. Ultrasound Med Biol 2017;43:S19.

Jaderberg M, Simonyan K, Zisserman A, Kavukcuoglu K. Spatial transformer networks. In: Cortes C, Lawrence ND, Lee DD, Sugiyama M, Garnett R, (eds). Annual conference on neural information processing systems, December 7−12, 2015, Montreal, QC, Canada. 28, 2017–2025. Adv Neural Inf Process Syst.

Jaeger PF, Kohl SAA, Bickelhaupt S, Isensee F, Kuder TA, Schlemmer HP, Maier-Hein KH. Retina U-Net: Embarrassingly simple exploitation of segmentation supervision for medical object detection. CoRR 2018; abs/1811.08661.

Jalalian A, Mashohor S, Mahmud R, Karasfi B, Saripan MIB, Ramli ARB. Foundation and methodologies in computer-aided diagnosis systems for breast cancer detection. EXCLI J 2017;16:113–137.

Jetley S, Lord NA, Lee N, Torr PHS. Learn to pay attention. 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30−May 3, 2018, Conference Track Proceedings, OpenReview.net.

Jung H, Kim B, Lee I, Yoo M, Lee J, Ham S, Woo O, Kang J. Detection of masses in mammograms using a one-stage object detector based on a deep convolutional neural network. PLOS One 2018;13: e0203355.

Kamnitsas K, Bai W, Ferrante E, McDonagh SG, Sinclair M, Pawlowski N, Rajchl M, Lee M, Kainz B, Rueckert D, Glocker B. Ensembles of multiple models and architectures for robust brain tumour segmentation. In: Crimi A, Bakas S, Kuijf HJ, Menze BH, Reyes M, (eds). Brain lesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries—Third International Workshop, BrainLes 2017, held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 14, 2017, Revised Selected Papers. Cham. : Springer; 2017. p. 450–462.

Kooi T, Litjens GJS, van Ginneken B, Gubern-Mérida A, Sánchez CI, Mann R, den Heeten A, Karssemeijer N. Large scale deep learning for computer aided detection of mammographic lesions. Med Image Anal 2017;35:303–312.

Kumar N, Verma R, Sharma S, Bhargava S, Vahadane A, Sethi A. A dataset and a technique for generalized nuclear segmentation for computational pathology. IEEE Trans Med Imaging 2017;36:1550–1560.

Langlotz CP, Allen B, Erickson BJ, Kalpathy-Cramer J, Bigelow K, Cook TS, Flanders AE, Lungren MP, Mendelson DS, Rudie JD, Wang G, Kandarpa K. A roadmap for foundational research on artificial intelligence in medical imaging: From the 2018 NIH/RSNA/ACR/The Academy Workshop. Radiology 2019;291:781–791.

LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–444.

Li K, Hariharan B, Malik J. Iterative Instance Segmentation. 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27−30, 2016. : IEEE Computer Society; 2016. p. 3659–3667.

Li L, Tang S, Zhang Y, Deng L, Tian Q. GLA: Global−local attention for image description. IEEE Trans Multimedia 2018;20:726–737.

Li H, Cheng JZ, Chou YH, Qin J, Huang S, Lei B. AttentionNet: Learning where to focus via attention mechanism for anatomical segmentation of whole breast ultrasound images. 16th IEEE International Symposium on Biomedical Imaging, ISBI 2019, Venice, Italy, April 8−11, 2019. Piscataway, NJ. : IEEE; 2019. p. 1078–1081.

Lin G, Milan A, Shen C, Reid ID. RefineNet: Multi-path Refinement Networks for High-Resolution Semantic Segmentation. 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017 IEEE Computer Society. 5168–5177.

Lin H, Chen H, Dou Q, Wang L, Qin J, Heng P-A. ScanNet: A fast and dense scanning framework for metastatic breast cancer detection from whole-slide image. 2018 IEEE Winter Conference on Applications of Computer Vision, WACV 2018, Lake Tahoe, NV, USA, March 12−15, 2018. : IEEE Computer Society; 2018. p. 539–546.

Litjens GJS, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JAWM, van Ginneken B, Sánchez CI. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60–88.

Liu B, Cheng HD, Huang J, Tian J, Tang X, Liu J. Fully automatic and segmentation-robust classification of breast tumors based on local texture analysis of ultrasound images. Pattern Recognition 2010;43:280–298.

Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7−12, 2015. IEEE Computer Society; 2015. p. 3431–3440.

Mirikharaji Z, Hamarneh G. Star Shape Prior in Fully Convolutional Networks for Skin Lesion Segmentation. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, (eds). Medical Image Computing and Computer Assisted Intervention, MICCAI 2018, 21st International Conference, Granada, Spain, September 16−20, 2018, Proceedings, Part IV. Cham. : Springer; 2018. p. 737–745.

Mnih V, Heess N, Graves A, Kavukcuoglu K. Recurrent Models of Visual Attention. In: Ghahramani Z, Welling M, Cortes C, Lawrence ND, Weinberger KQ, (eds). Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8−13 2014. Montreal, Quebec, Canada2204–2212.

Moon WK, Shen YW, Huang CS, Chiang LR, Chang RF. Computer-aided diagnosis for the classification of breast masses in automated whole breast ultrasound images. Ultrasound Med Biol 2011;37:539–548.

de Moor T, Rodríguez-Ruiz A, Mann RM, Teuwen J. Automated soft tissue lesion detection and segmentation in digital mammography using a U-net deep learning network. ArXiv 2018; abs/1802.06865.

Naylor P, Lae M, Reyal F, Walter T. Segmentation of nuclei in histopathology images by deep regression of the distance map. IEEE Trans Med Imaging 2019;38:448–459.

Nosrati MS, Hamarneh G. Incorporating prior knowledge in medical image segmentation: a survey. CoRR 2016; abs/1607.01092.

Oktay O, Ferrante E, Kamnitsas K, Heinrich MP, Bai W, Caballero J, Cook SA, de Marvao A, Dawes T, O'Regan DP, Kainz B, Glocker B, Rueckert D. Anatomically constrained neural networks (ACNNs): Application to cardiac image enhancement and segmentation. IEEE Trans Med Imaging 2018a;37:384–395.

Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B, Glocker B, Rueckert D. Attention U-Net: Learning where to look for the pancreas. 1st Conference on Medical Imaging with Deep Learning (MIDL). Amsterdam, The Netherlands1–10.

Ravishankar H, Venkataramani R, Thiruvenkadam S, Sudhakar P, Vaidya V. Learning and incorporating shape models for semantic segmentation. In: Descoteaux M, Maier-Hein L, Franz AM, Jannin P, Collins DL, Duchesne S, (eds). Medical Image Computing and Computer Assisted Intervention, MICCAI 2017, 20th International Conference, Quebec City, QC, Canada, September 11−13, 2017, Proceedings, Part I. Cham. : Springer; 2017. p. 203–211.

Ribli D, Horváth A, Unger Z, Pollner P, Csabai I. Detecting and classifying lesions in mammograms with Deep Learning. CoRR 2017; abs/1707.08401.

Rodrigues R, Braz R, Pereira M, Moutinho J, Pinheiro AMG. A two-step segmentation method for breast ultrasound masses based on multi-resolution analysis. Ultrasound Med Biol 2015;41:1737–1748.

Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, III WMW, Frangi AF, (eds). Medical Image Computing and Computer-Assisted Intervention, MICCAI 2015, 18th International Conference, Munich, Germany, October 5−9, 2015, Proceedings, Part III. Cham. : Springer; 2015. p. 234–241.

Shao H, Zhang Y, Xian M, Cheng H-D, Xu F, Ding J. A saliency model for automated tumor detection in breast ultrasound images. 2015 IEEE International Conference on Image Processing, ICIP 2015, Quebec City, QC, Canada, September 27−30. 1424–1428.

Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv:13126034 [cs]2013.

Stollenga MF, Masci J, Gomez FJ, Schmidhuber J. Deep networks with internal selective attention through feedback connections. In: Ghahramani Z, Welling M, Cortes C, Lawrence ND, Weinberger KQ, (eds). Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8−13 2014. Montreal, Quebec, Canada3545–3553.

Tomita N, Abdollahi B, Wei J, Ren B, Suriawinata AA, Hassanpour S. Finding a needle in the haystack: Attention-based classification of high resolution microscopy images. CoRR 2018; abs/1811.08513.

Waite S, Scott J, Gale B, Fuchs T, Kolla S, Reede D. Interpretive error in radiology. Am J Roentgenol 2016;208:739–749.

Wang F, Jiang M, Qian C, Yang S, Li C, Zhang H, Wang X, Tang X. Residual attention network for image classification. 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21−26, 2017 IEEE Computer Society. 6450–6458.

Wu N, Phang J, Park J, Shen Y, Huang Z, Zorin M, Jastrzebski S, Févry T, Katsnelson J, Kim E, Wolfson S, Parikh U, Gaddam S, Lin LLY, Ho K, Weinstein JD, Reig B, Gao Y, Toth H, Pysarenko K, Lewin A, Lee J, Airola K, Mema E, Chung S, Hwang E, Samreen N, Kim SG, Heacock L, Moy L, Cho K, Geras KJ. Deep neural networks improve radiologists' performance in breast cancer screening. CoRR 2019; abs/1903.08297.

Xian M. Neutro-connectedness theory, algorithms and applications. PhD thesis. Utah State University; 2017.

Xian M, Zhang Y, Cheng H, Xu F, Ding J. Neutro-Connectedness Cut. IEEE Trans Image Process 2016;25:4691–4703.

Xian M, Zhang Y, Cheng HD, Xu F, Huang K, Zhang B, Ding J, Ning C, Wang Y. A benchmark for breast ultrasound image Segmentation (BUSIS). CoRR 2018; abs/1801.03182.

Xian M, Zhang Y, Cheng HD, Xu F, Zhang B, Ding J. Automatic breast ultrasound image segmentation: A survey. Pattern Recognition 2018b;79:340–355.

Xiao G, Brady M, Noble JA, Zhang Y. Segmentation of ultrasound B-mode images with intensity inhomogeneity correction. IEEE Trans Med Imaging 2002;21:48–57.

Xie Y, Chen K, Lin J. An automatic localization algorithm for ultrasound breast tumors based on human visual mechanism. Sensors 2017;17:1101.

Xu K, Ba JL, Kiros R, Cho K, Courville A, Salakhutdinov R, Zemel RS, Bengio Y. Show, attend and tell: Neural image caption generation with visual attention. In: Proceedings of the 32nd International Conference on International Conference on Machine Learning. 37, 2048–2057. J Mach Learn Res.

Xu F, Xian M, Cheng HD, Ding J, Zhang Y. Unsupervised saliency estimation based on robust hypotheses. 2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016. Lake Placid, NY, USA; 2016. p. 1–6.

Xu F, Xian M, Zhang Y, Huang K, Cheng HD, Zhang B, Ding J, Ning C, Wang Y. A hybrid framework for tumor saliency estimation. 24th International Conference on Pattern Recognition, ICPR 2018, Beijing, China, August 20−24. Cham. Springer; 2018. p. 3935–3940.

Xu F, Zhang Y, Xian M, Cheng HD, Zhang B, Ding J, Ning C, Wang Y. Tumor saliency estimation for breast ultrasound images via breast anatomy modeling. CoRR 2019; abs/1906.07760.

Yap MH, Pons G, Martí J, Ganau S, Sentís M, Zwiggelaar R, Davison AK, Marti R. Automated breast ultrasound lesions detection using convolutional neural networks. IEEE J Biomed Health Inform 2018;22:1218–1226.

Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid scene parsing network. 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21−26. 2017 IEEE Computer Society. 6230–6239.

Zhu W, Huang Y, Zeng L, Chen X, Liu Y, Qian Z, Du N, Fan W, Xie X. AnatomyNet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. Med Phys 2019;46:576–589.