Journal Pre-proof

Multi-Task Learning for Segmentation and Classification of Tumors in 3D Automated Breast Ultrasound Images

Yue Zhou, Houjin Chen, Yanfeng Li, Qin Liu, Xuanang Xu,
Shu Wang, Dinggang Shen, Pew-Thian Yap

Please cite this article as: Yue Zhou, Houjin Chen, Yanfeng Li, Qin Liu, Xuanang Xu, Shu Wang, Dinggang Shen, Pew-Thian Yap, Multi-Task Learning for Segmentation and Classification of Tumors in 3D Automated Breast Ultrasound Images, *Medical Image Analysis* (2020), doi: https://doi.org/10.1016/j.media.2020.101918

128x64x128
Input

Stage 1

Stage 2

Stage 3

Stage 4

Stage 5

Stage 6

Stage 7

Stage 8

Stage 9

argmax

128x64x128
Output

Input Features

Feature Extraction

Maligant/Benign

Probability

Conv + BN + ReLu

Conv + BN + ReLu + Sigmoid

Down-Sampling

Up-Sampling
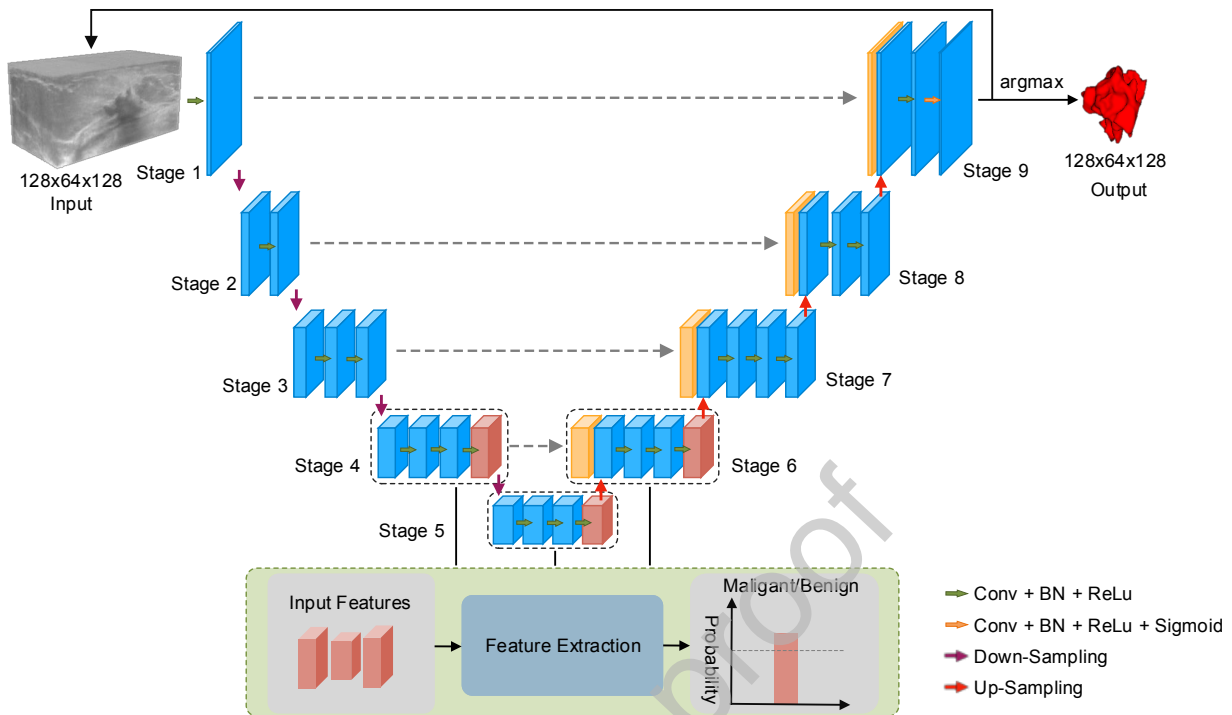
Highlights

Multi-Task Learning for Tumor Segmentation and Classification in 3D Automated Breast Ultra-sound Images

Yue Zhou,Houjin Chen,Yanfeng Li,Qin Liu,Xuanang Xu,Shu Wang,Pew-Thian Yap,Dinggang Shen

- A multi-task learning framework is designed to joint 3D ABUS tumor segmentation and classification.

- A multi-scale feature extraction network is proposed for the classification task.

- An iterative feature-refining mechanism is used for highlighting tumor regions.

- Segmentation and classification promote each other during the training process.

Multi-Task Learning for Segmentation and Classification of Tumors in 3D Automated Breast Ultrasound Images

Yue Zhou[a,b], Houjin Chen[a,*], Yanfeng Li[a], Qin Liu[b], Xuanang Xu[b], Shu Wang[c], Dinggang Shen[b], Pew-Thian Yap[b,*]

[a] School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, 100044, China.
[b] Department of Radiology and BRIC, University of North Carolina, Chapel Hill, NC, 27599, USA.
[c] Peking University People's Hospital, Beijing, 100044, China.

ARTICLE INFO

ABSTRACT

Tumor classification and segmentation are two important tasks for computer-aided diagnosis (CAD) using 3D automated breast ultrasound (ABUS) images. However, they are challenging due to the significant shape variation of breast tumors and the fuzzy nature of ultrasound images (e.g., low contrast and signal to noise ratio). Considering the correlation between tumor classification and segmentation, we argue that learning these two tasks jointly is able to improve the outcomes of both tasks. In this paper, we propose a novel multi-task learning framework for joint segmentation and classification of tumors in ABUS images. The proposed framework consists of two sub-networks: an encoder-decoder network for segmentation and a light-weight multi-scale network for classification. To account for the fuzzy boundaries of tumors in ABUS images, our framework uses an iterative training strategy to refine feature maps with the help of probability maps obtained from previous iterations. Experimental results based on a clinical dataset of 170 3D ABUS volumes collected from 107 patients indicate that the proposed multi-task framework improves tumor segmentation and classification over the single-task learning counterparts.

1. Introduction

Breast cancer is the most commonly diagnosed cancer and is the leading cause of cancer death in females around the world (Bray et al., 2018). Fortunately, with advanced medical imaging technologies, more than 90% cases of breast cancer can be diagnosed early and treated before becoming metastatic (Bleicher et al., 2016). Mammography is the primary imaging modality for breast cancer screening. However, exposure to ionizing radiation from repeated mammography may increase the risk for breast cancer (Skaane et al., 2019). Breast ultrasound, such as 3D automated breast ultrasound (ABUS), is an alternative to mammography due to its safety, speed, and reproducibility (Vourtsis, 2019; Giger et al., 2016; Zanotel et al., 2018). The early detection rate of breast cancer is significantly increased using ABUS in conjunction with mammography for women with dense breasts (Le et al., 2019).

3D ABUS systems provide a large field of view using high-frequency transducers (Chough et al., 2020). In clinical practice, three to five acquisitions in different orientations are acquired to cover the whole breast and each ABUS volume includes hundreds of 2D slices (Brem et al., 2015). Reviewing
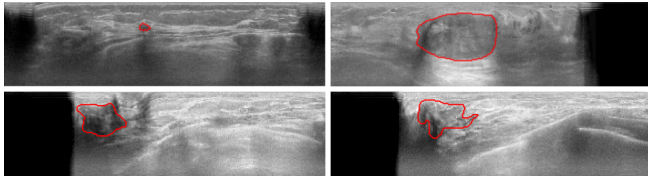
Fig. 1: Tumor examples in ABUS images. Tumors vary significantly in size and shape with irregular and ambiguous boundaries (red lines).

the whole ABUS examination slice-by-slice is extremely time-consuming even for experienced radiologists. Thus effective computer-aided diagnosis (CAD) systems are essential for assisting physicians in specifying treatments and surgical plans. However, designing CAD systems for ABUS is a challenging task, due to the large tumor size and shape variation, irregular and ambiguous tumor boundaries, and low signal-to-noise ratio in ultrasound volumes, as illustrated in Fig. 1.

Tumor segmentation and classification are two basic tasks in CAD systems. Benign and malignant tumors in ABUS images usually have different shape characteristics. Specifically, benign tumors commonly have smooth, round, and oval boundaries, whereas most malignant tumors have irregular and spiculated shapes (Yang et al., 2008). In clinical diagnosis, boundary characteristics of tumors assist radiologists in distinguishing benign tumors from malignant tumors and shape properties are useful for both tumor classification and segmentation. Consequently, training two tasks jointly in one network to encourage feature sharing between breast tumor classification and segmentation is a promising direction to explore.

In this paper, we propose a multi-task learning method to jointly train tumor segmentation and classification in an end-to-end network for 3D ABUS. The proposed method includes an encoder-decoder network for segmentation and a light-weight multi-scale network for classification. We employ VNet as the backbone network for both tumor classification and segmentation (Milletari et al., 2016). The segmentation and classification tasks share features extracted from the encoding path.

To solve the problem of large tumor size variation, many methods attempt to capture multi-scale information at the image-level by image resampling or at the feature-level by combining convolutional layers with different kernel sizes into deep learning networks (Shen et al., 2015; Zhang et al., 2017). Furthermore, many multi-scale convolutional layer connection methods are proposed for single classification task feature extraction, such as feature pyramid network (FPN) (Lin et al., 2017a). Unlike existing methods that extract classification features from a single task, we extract multi-scale features for classification from a multi-task learning network by connecting feature maps from different convolutional neural network (CNN) layers. To suppress noise and further improve the multi-task learning performance, we use an iterative feature-refining strategy inspired by auto-context feature extraction proposed in (Wu et al., 2017). The auto-context mechanism integrates context information by fusing low-level features with context to implicit shape information (Tu and Bai, 2009). In contrast to the existing method (Wu et al., 2017) that cascades three fully convolutional networks (FCNs) to extract features for a single segmentation

task, we train the multi-task network using an iterative training strategy to extract auto-context features for both tasks. We conducted extensive experiments on a clinical dataset of 170 ABUS volumes from 107 patients to evaluate the effectiveness of the proposed method. Experimental results indicate that the proposed method improves the performance of both tasks. Our main contributions are summarized as follows:

1. We design a novel multi-task learning network to simultaneously conduct 3D ABUS tumor segmentation and classification. A multi-scale feature extraction network is added to the end of the encoding path in VNet for classification.

2. We utilize an iterative feature-refining training strategy where the predicted segmentation maps are added as a part of the input to guide feature extraction.

The remaining sections are organized as follows. Related studies of breast ultrasound image classification, segmentation, and multi-task learning are reviewed in Section 2. Section 3 provides a detailed description of the proposed method. Section 4 describes the ABUS dataset used in this study and the experimental details. Section 5 presents the experiment results. Section 6 provides further discussion and Section 7 concludes this work.

2. Related Work

2.1. Ultrasound Image Segmentation

In previous studies, many segmentation methods are proposed for ultrasound images based on conventional image processing approaches, such as watershed transform (Gomez et al., 2010; Chen et al., 2005; Cheng et al., 2010b), region growing (Kozegar et al., 2017; Cheng et al., 2010a), and active contour (Galińska et al., 2017; Lee et al., 2020). Gomez et al. (2010) proposed to employ marker-controlled watershed transformations on contrast-enhanced ultrasound images for mitigating over-segmentation and robust segmentation of objects with closed contours. Kozegar et al. (2017) designed a two-stage segmentation approach. First, an adaptive region growing algorithm produced a rough estimation of the mass boundary. Second, a geometric edge-based deformable model used the rough estimation as the initial contour for further refinement. This two-stage segmentation method heavily relies on the accuracy of the initial seed selection.

In recent years, CNN-based methods have been introduced for ultrasound image segmentation (Litjens et al., 2017; Cao et al., 2020; Wang et al., 2018a; Xing et al., 2019). Wang et al. (2018a) developed a 3D deeply supervised CNN model for ABUS based on 3D U-Net with a threshold loss for balanced low false positive (FP) rate and high sensitivity. However, tumors in the shadows of nipples were missed by this method (Wang et al., 2018a). Xing et al. (2019) proposed a semi-pixel-wise cycle generative adversarial network that took the advantages of a FCN and a generative adversarial network (GAN) for tumor segmentation. In addition to tumor segmentation, the

anatomy segmentation is also an important branch in breast ultrasound image analysis, which is beneficial to remove segmentation FPs (Bian et al., 2017). Lei et al. (2018) used a boundary regularized deep convolutional encoder-decoder network to segment breast anatomical layers in ABUS. They further added a self-co-attention mechanism to improve segmentation consistency between consecutive slices (Lei et al., 2020).

2.2. Ultrasound Image Classification

Traditional methods for ultrasound image classification typically manually extract local texture features, including shape, orientation, margin, lesion boundary, echo pattern, and posterior acoustic features. Moon et al. (2011) and Moon et al. (2012) extracted texture, morphologic, and descriptor features in segmented tumor regions and fed these features to a classifier to categorize the tumors as benign or malignant. Uniyal et al. (2014) generated malignancy maps from the estimated cancer likelihood in an ultrasound radiofrequency (RF) time series to classify malignant breast tumors. Flores et al. (2015) improved the classification accuracy of breast tumors on ultrasonography by analyzing distinct morphological and texture features. The above classification methods are reliant on the accuracy of manually extracted features and cannot be generalized to other classification tasks.

Deep learning based methods with the powerful feature extraction capability are able to overcome the limitations of traditional methods. Zeimarani et al. (2020) used a custombuilt CNN with a few hidden layers and applied regularization techniques to improve the classification performance for 2D ultrasound images. Since 2D slices from different views in ABUS volumes have distinct characteristics, some ABUS classification models focus on tumor classification by fusing 2D multi-view features. Wang et al. (2020) proposed a modified Inception-v3 architecture to extract efficient features from ABUS slices in axial and coronal views (Szegedy et al., 2016). Kong et al. (2018) designed three CNN models to fuse multiview classification information. Although 2D information fusion is useful to classify tumors in ABUS, 3D volumes provide more global contextual features for tumor classification.

2.3. Joint Classification and Segmentation

There are few methods for jointly training classification and segmentation of ABUS images. However, the idea of multi-task learning has already been applied to 2D ultrasound images and other medical applications. For 2D ultrasound images, Wang et al. (2018b) proposed a multi-feature guided CNN architecture based on a modified U-Net. They added a classification branch to the U-Net bottom for simultaneous segmentation and classification of bone surfaces in 2D ultrasound data. Xie et al. (2018) used a two stage multi-task method based on ResNet and Mask R-CNN. A pre-trained ResNet and a transfer learning approach were used to extract and classify candidates. Then, an improved Mask R-CNN model was used for accurate tumor segmentation. Singh et al. (2019) added an atrous convolution layer and a channel-wise weighting block to the conditional GAN (cGAN) model for breast tumor segmentation and extracted statistical features from the predicted mask boundaries for tumor classification.

Multi-task learning is also applied to other medical images. Song et al. (2020) employed a joint training network by combining a region proposal network (RPN) backbone with three sub-networks for detection, classification, and segmentation of skin lesions. Chakravarty and Sivaswamy (2018) performed segmentation and image-level classification for glaucoma by exploiting both image appearance and structural features. This method used a network similar to U-Net for segmentation and down-sampled the segmentation map to the same scale as the feature map from the last down-sample layer to extract features for classification. Chen et al. (2018) designed a multi-task learning network for 3D left atrium images by sharing feature maps from the fourth max-pooling layer in an improved 3D U-Net. Shared features were fed into two fully connected layers for classification. Qu et al. (2019) connected a prediction network and a perceptual loss network for nucleus classification and segmentation in histopathology images. The prediction network produced a segmentation map for every type nucleus. The perceptual loss network further improved segmentation accuracy.

The above multi-task learning methods cannot be directly applied to 3D ABUS data for joint segmentation and classification. Firstly, the tumor boundaries in ABUS images are ill-defined, which makes tumor delineation in ABUS more complicated than other applications. Secondly, significant size and shape variations of breast tumors require the network to cater to tumors with various scales.

3. Method

In this paper, we propose an iterative feature-refining multi-task learning method for ABUS images. The proposed framework integrates classification and segmentation in an end-to-end CNN model which takes ABUS volumes as inputs and produces two outputs, including a 3D tumor segmentation map and a volume-level classification probability.

3.1. Multi-Task Learning Network

The proposed network is shown in Fig. 2. We utilize VNet as the backbone network for multi-task learning due to its great performance in 3D medical image segmentation. The VNet architecture consists of three parts: (i) an encoding path, (ii) a decoding path, and (iii) skip connections between them. As shown in Fig. 2a, VNet includes stacked convolution layers in each stage. Specifically, the encoding path employs four down-sampling operations to extract high-level semantic features. Symmetrically, the decoding path utilizes four up-sampling operations to restore the feature maps to the original input size and interprets the extracted features from the encoding path to predict the segmentation mask of the target. Skip connections connect feature maps from the encoding path to the decoding path to propagate spatial information and refine segmentation outcomes. We set all convolution kernel sizes to 3×3×3 in each convolution layer. Every convolution operation is followed by batch normalization (BN) and rectified linear unit (ReLU). We use convolution operations with 2×2×2 kernel size and 2×2×2
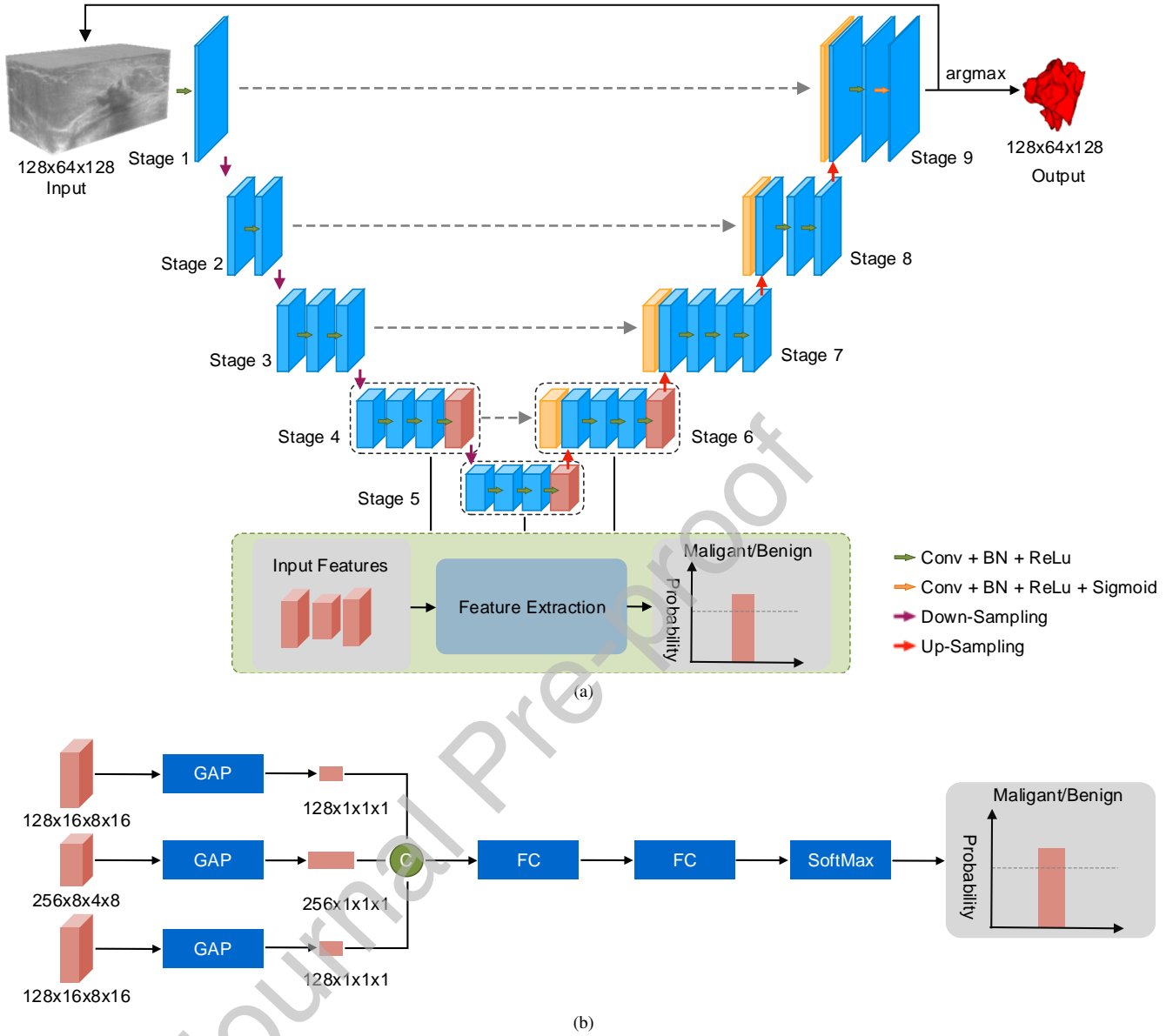
Fig. 2: (a) Overview of the proposed multi-task learning network. (b) Architecture of the multi-scale classification path.

stride to downsample feature maps instead of pooling operations to preserve positional information.

Many classic image classification networks use high-level feature maps from CNN models, such as VGG (Simonyan and Zisserman, 2014) and ResNet (He et al., 2016). Inspired by this observation, we share high-level feature maps from VNet to extract common features for both classification and segmentation in the proposed multi-task learning network. A classification branch is added to the bottom of the VNet, as illustrated in Fig. 2b. Firstly, feature maps from Stage 4, Stage 5, and Stage 6 are fed into the classification network. Then, we fuse these shared feature maps for the classification task. Finally, we input fused features to the classification branch which has two fully connected (FC) layers and one softmax layer to predict the input volume as benign or malignant.

3.2. Multi-Scale Feature Extraction

To analyze features of different scales, we visualize and re-shape 2D axial view feature maps to the original input size from different stages in the proposed network, as shown in Fig. 3. Case 1 is a large benign tumor and Case 2 is a small malignant tumor. We can observe that low-level features mainly capture shape and boundary information, high-level features summarize attributes of different targets and are commonly used in classification tasks. However, capturing features for small objects can be challenging when network depth increases with more convolution and downsampling operations (Liu et al., 2015). To solve this problem, we design a multi-scale feature concatenation model for the classification task, as shown in Fig. 2b. We connect and fuse feature maps from Stage 4 to Stage 6 in VNet as classification features. Since features maps from the different
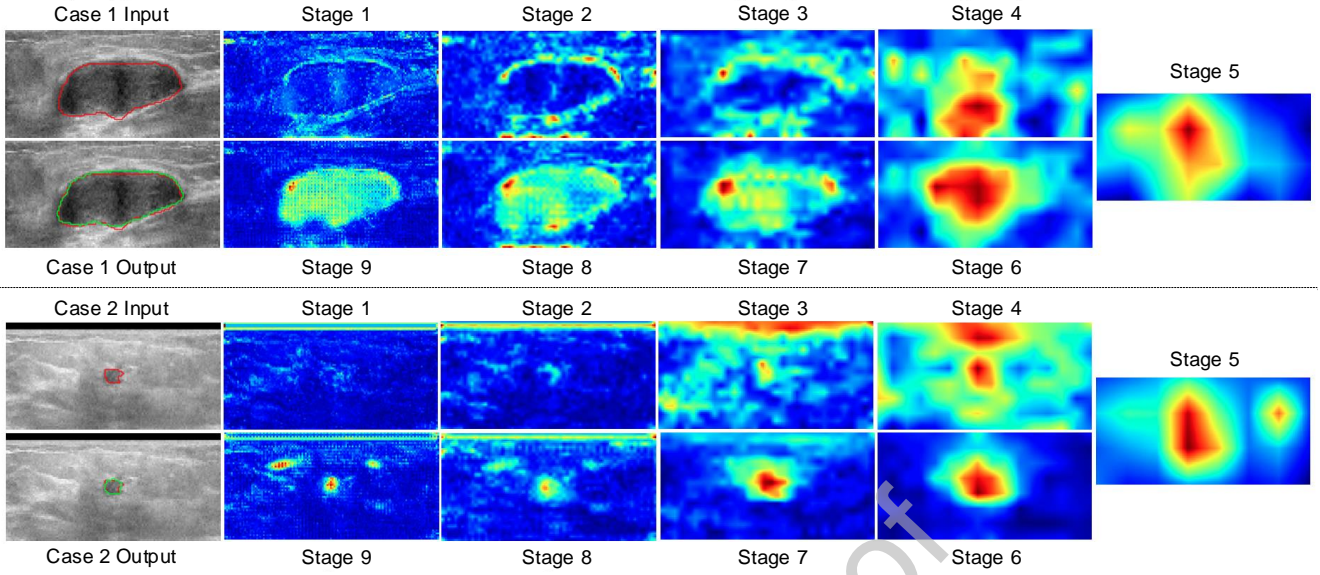
Fig. 3: 2D axial views of feature maps from different stages of the proposed network. Case 1 is a large benign tumor and Case 2 is a small malignant tumor.

stages have different sizes and cannot be connected directly, we use channel-wise global average pooling (GAP) to convert feature maps from different stages to the same size in each channel. Although fully connected (FC) layers and global max pooling (GMP) function similarly as GAP, FC layers increase network parameters and training time and GMP uses the max voxel to represent the whole feature map, neglecting a lot of useful spatial information. In contrast, GAP uses voxel averaging to express spatial information and is more robust to image spatial translation.

Generally, the amplitudes of features from the deeper layers are smaller than the shallow layers. Thus shallow layers may dominate deeper layers in terms of feature magnitude (Zeiler and Fergus, 2014). In light of this, we normalize the features from multiple scales with the ℓ_2 norm before connection:

$$\hat{x}_i = \frac{x_i}{\sqrt{\sum_i^c \|x_i\|^2}}, \tag{1}$$

where c is the channel number, x_i is the global feature from each channel, and $\hat{x}_i$ is the normalized feature.

3.3. Iterative Feature Refinement

ABUS volumes have severe noise and tumors in ABUS have ambiguous boundaries, which complicates tumor segmentation. To suppress noise and improve segmentation performance, we employ an iterative feature-refining strategy during the training process. This strategy updates the features in each iteration using segmentation probability maps from the previous iteration, as illustrated in Algorithm 1. In the initial iteration, the input to the multi-task learning network is a 3D ABUS volume and the output is a 3D segmentation probability map with a classification score. The predicted probability maps contain contextual information and can be used to guide the network to focus on tumor regions. In our case, this is performed on feature maps of Stage 9 in Fig. 3. In subsequent iterations, the input volume is

modulated via addition with the probability map from the last iteration.

Algorithm 1 Iterative Feature Refinement

Require: Original Training volume, V_{orig}
Require: Classification label, y_{cls}
Require: Segmentation label, Y_{seg}
Require: Number of iterations, N
Require: Network, $f_\theta(\text{Input})$
 1: **for** $i = 1$ to N **do**
 2:     **if** $i = 1$ **then**
 3:         Input $= V_{\text{orig}}$
 4:     **else**
 5:         Input $= V_{\text{orig}} + \text{SegProbMap}_{i-1}$
 6:     **end if**
 7:     $\text{SegProbMap}_i, \text{p}_{\text{cls},i} = f_\theta(\text{Input})$
 8:     $\text{P}_{\text{seg},i} = \text{argmax}(\text{SegProbMap}_i)$
 9:     $\text{Loss} = L(\text{P}_{\text{seg},i}, \text{p}_{\text{cls},i}, Y_{\text{seg}}, y_{\text{cls}})$
10:     Update θ using ADAM optimizer
11: **end for**
12: **return** $\text{P}_{\text{seg},N}, \text{p}_{\text{cls},N}$

3.4. Multi-Task Loss Function

Class imbalance is a common challenge in medical image classification. For example, in our AUBS dataset, the number of malignant tumors is roughly twice that of benign tumors. To account for this, we use a modified weighted focal loss (Lin et al., 2017b) as the classification loss function:

$$\begin{aligned} L_{\text{cls}}(p_{\text{cls}}, y_{\text{cls}}) = &-w_m(1 - p_{\text{cls}})^\gamma y_{\text{cls}} \log(p_{\text{cls}}) \\ &- w_n p_{\text{cls}}^\gamma (1 - y_{\text{cls}}) \log(1 - p_{\text{cls}}), \end{aligned} \tag{2}$$

where p_{cls} and y_{cls} are the predicted volume classification probability from the proposed network and the ground truth class of

this volume ($y_{cls} = 0$ for benign and $y_{cls} = 1$ for malignant). The focusing parameter γ is set as 2. w_m and w_n are weights for malignant and benign cases defined as:

$$w_m = \frac{N_n}{N_m + N_n}, \quad w_n = \frac{N_m}{N_m + N_n}, \tag{3}$$

where N_n and N_m are the numbers of benign and malignant volumes, respectively.

For the segmentation task, the imbalance between foreground and background may cause segmentation bias. To account for this problem, a segmentation loss based on the Dice coefficient is utilized to emphasize shape similarity between segmentation maps and ground truth regions. The segmentation loss is defined as:

$$L_{seg}(P_{seg}, Y_{seg}) = 1 - \frac{2P_{seg}Y_{seg} + 1}{P_{seg} + Y_{seg} + 1}, \tag{4}$$

where L_{seg} is the segmentation loss, P_{seg} and Y_{seg} denote the predicted segmentation map from the proposed network and the labeled tumor map.

In our method, classification loss and segmentation loss are linear combined as a multi-task loss by a hyperparameter λ. The multi-task loss is defined as:

$$L_{joint} = \lambda L_{cls} + (1 - \lambda)L_{seg}, \tag{5}$$

where L_{joint} is the multi-task loss, and $\lambda \in [0, 1]$ is the weight of the classification task.

4. Data and Experiments

4.1. Data

ABUS volumes used in this paper were acquired at Peking University People's Hospital, Beijing, China. All samples were collected with a freehand 3D imaging method using an ACUSON 2000 Ultrasound System (Siemens). Each ABUS volume contains 318 2D axial slices. Tumor boundaries were labeled in each 2D axial slice by two experienced radiologists and the tumors were classified as malignant or benign. In order to cover the entire breast, at least two volumes from the lateral view and the medial view were scanned for each breast. A total number of 170 volumes from 107 patients were collected. To evaluate the effectiveness of the proposed method, we employed four-fold cross-validation for all experiments. ABUS volumes in each fold were divided at the patient-level to prevent ABUS volumes from the same patient from appearing in both training and testing sets.

Obvious tumor volume variation among patients is a limitation in the ABUS dataset, as presented in Fig. 4. 95 volumes contain small tumors ($\leq 1 \, cm^3$), and the smallest tumor is only $0.026 \, cm^3$. In contrast, the largest tumor is $18.022 \, cm^3$. We resampled each volume to an isotropic resolution of $0.5 \times 0.5 \times 0.5 \, mm^3$ from $0.5 \times 0.07 \times 0.2 \, mm^3$. Tumor regions were cropped with the size of $64 \times 32 \times 64 \, mm^3$ based on tumor centers.
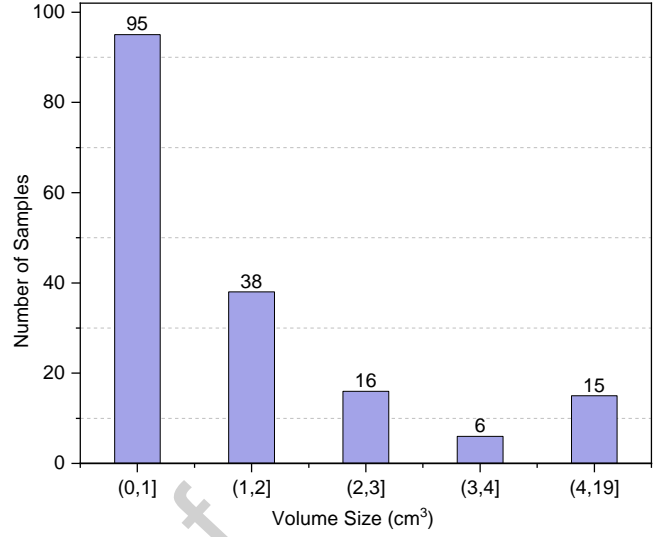

Fig. 4: Tumor volumes in 170 ABUS scans from 117 patients.

Table 1: Ablation models. CLS: classification; SEG: segmentation.

Model	Task	Multi-Scale	Iterative Training
VNet	SEG	×	×
ClsNet	CLS	×	×
CVNet	CLS + SEG	×	×
CVNet$_{Iter}$	CLS + SEG	×	✓
C$_{MS}$VNet	CLS + SEG	✓	×
C$_{MS}$VNet$_{Iter}$	CLS + SEG	✓	✓

4.2. Experiment Details

To evaluate the effectiveness of the proposed method, we conducted several ablation experiments. The differences among ablation models are listed in Table 1. The single-task models ClsNet and VNet were trained as the classification and the segmentation baseline models, respectively. ClsNet uses the encoding path from VNet followed by a GAP layer and two FC layers for classification. For multi-task learning, we firstly added a single-scale classification branch to VNet, which used the feature map from Stage 5 in VNet as the classification feature (denoted as CVNet), and then trained CVNet using the iterative feature-refining strategy with the number of iterations, $N=2$ (denoted as CVNet$_{Iter}$). Moreover, we used the proposed multi-scale classification branch in Fig. 2b instead of the single-scale classification branch for multi-task learning (denoted as C$_{MS}$VNet). Finally, we trained the C$_{MS}$VNet model using the iterative feature-refining strategy with the number of iterations $N=2$ (denoted as C$_{MS}$VNet$_{Iter}$).

We further compared the proposed method with existing multi-task learning methods (see Section 2.3), including Wang et al. (2018b), Chakravarty and Sivaswamy (2018), Xie et al. (2018), and Qu et al. (2019). Specifically, we extended all 2D operations in these methods to 3D operations and changed multi-class classification to two-class classification. For fair comparison, we employed in all comparison methods the same

VNet backbone as the proposed method. To compare segmentation and classification performance, we re-trained all compared models using our ABUS dataset with four-fold cross-validation and used the proposed loss function in Eq. 5 with λ=0.3. The other parameters of the compared models were identical to those in (Wang et al., 2018b; Chakravarty and Sivaswamy, 2018; Xie et al., 2018; Qu et al., 2019).

All models were implemented using Pytorch (Paszke et al., 2019) and were trained and tested on NVIDIA TITAN Xp GPUs (12G) using the ADAM optimizer with a decaying learning rate initialized at 1e-4 (Kingma and Ba, 2014). The batch size was set as 2 for each GPU. Data augmentation (Nalepa et al., 2019; Al-Dhabyani et al., 2019) was carried out using four transformations: (i) rotation of 0 to 360 degrees, (ii) random flipping along axes in axial, coronal, or sagittal planes, (iii) elastic transformation for shape variation, and (iv) scaling of tumors to various sizes.

4.3. Evaluation Metrics

We employed Dice similarity coefficient (DSC), Jaccard index (JI), and 95th percentage of the asymmetric Hausdorff distance (95HD) for quantitative evaluation of tumor segmentation. Due to the sensitive of Hausdorff distance (HD) to outliers, 95HD was utilized instead of the maximum (Jesorsky et al., 2001). These metrics are calculated as follows:

$$DSC(A, B) = \frac{2|A \cap B|}{|A| + |B|} \tag{6}$$

$$JI(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{7}$$

$$95HD(A, B) = \max\{\sup_{x \subset A} \inf_{y \subset B} \|x - y\|, \sup_{y \subset B} \inf_{x \subset A} \|x - y\|\} \tag{8}$$

where A and B are the segmentation results and the ground truth, x and y are the voxels in A and B. DSC and JI are sensitive to tumor areas and 95HD to shape.

We employed receiver operating characteristic (ROC), area under ROC curve (AUC), recall (REC), precision (PRE), accuracy (ACC), false positive rate (FPR), and F1-score (F1) for quantitative evaluation of tumor classification:

$$REC = \frac{TP}{TP + FN} \tag{9}$$

$$PRE = \frac{TP}{TP + FP} \tag{10}$$

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \tag{11}$$

$$FPR = \frac{FP}{FP + TN} \tag{12}$$

$$F1 = \frac{2 \times REC \times PRE}{REC + PRE} \tag{13}$$

where TP, FP, TN, FN are the number of true positives, false positives, true negatives and false negatives, respectively.

Table 2: Segmentation performance (Mean ± SD) of VNet with respect to number of iterations.

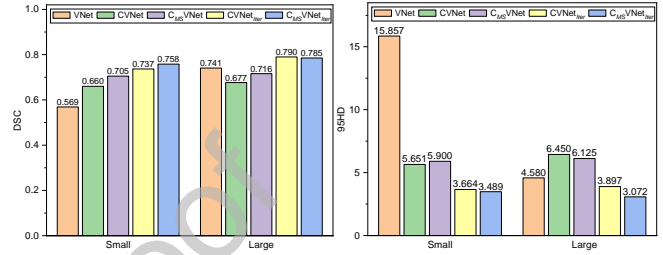| N	DSC	JI	95HD(mm)
0	0.645 ± 0.252	0.520 ± 0.238	10.882 ± 12.571
1	0.742 ± 0.165	0.610 ± 0.172	4.886 ± 6,921
2	**0.754 ± 0.165**	**0.628** ± 0.176	**3.870 ± 6.390**
3	0.751 ± 0.146	0.621 ± 0.159	3.879 ± 6.022
4	0.735 ± 0.166	0.603 ± 0.172	4.428 ± 7.170

Fig. 5: Segmentation results for small ($\leq 1\,cm^3$) and large ($>1\,cm^3$) tumors.

5. Results

5.1. Tumor Segmentation

To achieve the best performance of the iterative feature-refining strategy, we trained VNet using the iterative strategy with different number of iterations, N. As listed in Table 2, when N=2, the iterative training strategy yields the best performance with DSC, JI, and 95HD at 0.754, 0.628, and 3.870mm, respectively. Thus we set N=2 in the following experiments.

Table 3 lists quantitative results of ablation experiments. For the segmentation task, the proposed $C_{MS}VNet_{Iter}$ model outperforms other models in all metrics with DSC, JI, and 95HD at 0.778, 0.650, and 3.303 mm, respectively. By comparing the results from CVNet and VNet, we can observe that the segmentation task achieves better performance in CVNet, which verifies that classification information is helpful for segmentation. In addition, different from CVNet, $C_{MS}VNet$ uses multi-scale feature extraction for classification and achieves better performance on segmentation than both CVNet and VNet. In Table 3, iterative featurere finement improves segmentation performance. For example, $CVNet_{Iter}$ and $C_{MS}VNet_{Iter}$ increase the DSC value by 9.3% and 8.8% compared with CVNet and $C_{MS}VNet$.

To further analyze the robustness of the proposed method on different tumor volumes, we divide tumors into two groups based on tumor volume, including small tumors ($\leq 1\,cm^3$) and large tumors ($>1\,cm^3$). Fig. 5 shows that $C_{MS}VNet_{Iter}$ achieves the best performance for small tumors with DCS at 0.758 and 95HD at 3.489mm and both $C_{MS}VNet_{Iter}$ and $CVNet_{Iter}$ perform better for large tumors.

Fig. 6 displays the segmentation results in 3D renderings generated using visualization toolkit (VTK) (Schroeder et al., 2004) for two large tumors (Case 1: $17.99\,cm^3$ and Case 2: $10.01\,cm^3$) and two small tumors (Case 3: $0.56\,cm^3$ and Case 4:

Table 3: Segmentation and classification performance of ablation models.

Model	Segmentation (Mean ± SD)			Classification				
	DSC	JI	95HD (mm)	ACC	REC	PRE	FPR	F1
VNet	0.645 ± 0.252	0.520 ± 0.238	10.882 ± 12.571	–	–	–	–	–
ClsNet	–	–	–	0.594	0.714	0.708	0.686	0.711
CVNet	0.668 ± 0.198	0.527 ± 0.181	6.033 ± 8.137	0.700	0.840	0.757	0.627	0.797
CVNet$_{Iter}$	0.761 ± 0.159	0.635 ± 0.168	3.766 ± 5.705	0.700	0.840	0.757	0.627	0.797
C$_{MS}$VNet	0.690 ± 0.208	0.550 ± 0.208	9.334 ± 11.743	0.718	**0.849**	0.771	0.588	0.808
C$_{MS}$VNet$_{Iter}$	**0.778 ± 0.145**	**0.650 ± 0.170**	**3.303 ± 4.513**	**0.741**	0.798	**0.826**	**0.392**	**0.811**
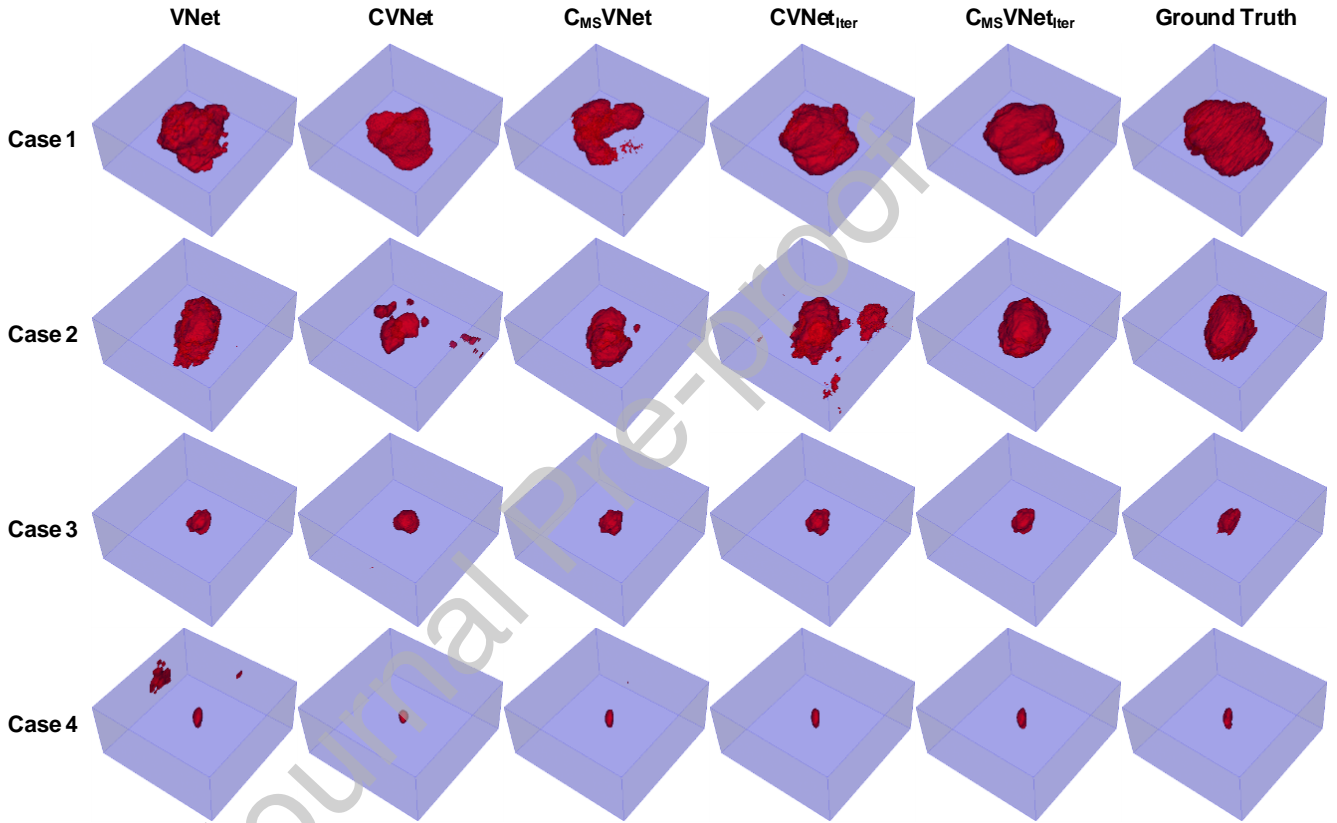
Fig. 6: 3D renderings of segmentation results for large tumor (Case 1: 17.99 cm^3 and Case 2: 10.01 cm^3) and two small tumors (Case 3: 0.56 cm^3 and Case 4: 0.20 cm^3).

0.20 cm^3). Compared with other methods, C$_{MS}$VNet$_{Iter}$ is capable of adapting to tumors of various sizes. Fig. 7 presents corresponding tumor boundaries in 2D axial, coronal, and sagittal views, indicating our method gives results that are consistent with the ground truth.

5.2. Tumor Classification

Table 3 also summarizes quantitative classification results of the single classification model and multi-task learning models. CVNet uses single-scale feature extraction for classification and improves classification results compared with the classification baseline ClsNet, which verifies that segmentation improves classification. With multi-scale feature extraction, C$_{MS}$VNet obtains the best REC at 0.849 and C$_{MS}$VNet$_{Iter}$ achieves the best ACC, PRC, FPR, and F1 score at 0.741, 0.826, 0.392, and

0.811, respectively. Fig. 8 presents ROCs from each ablation model. Compared with CVNet, C$_{MS}$VNet increases AUC from 0.676 to 0.683 and the proposed C$_{MS}$VNet$_{Iter}$ model further increases AUC to 0.787. Quantitative classification results in Table 3 and Fig. 8 indicate that multi-scale feature concatenation improves classification performance.

We also compare the impact of feature map combinations from different scales on the classification task. C$_{MS}$VNet was trained with different feature combinations, as listed in Table 4. C$_{MS}$(i-j)VNet connects feature maps from Stage i to Stage j in the VNet. For example, C$_{MS}$(4-6)VNet model uses feature maps from Stage 4, Stage 5, and Stage 6. We can see that involving too many features with different scales does not necessarily result in better classification performance. In fact, higher-level features summarize tumor attributes and are beneficial for
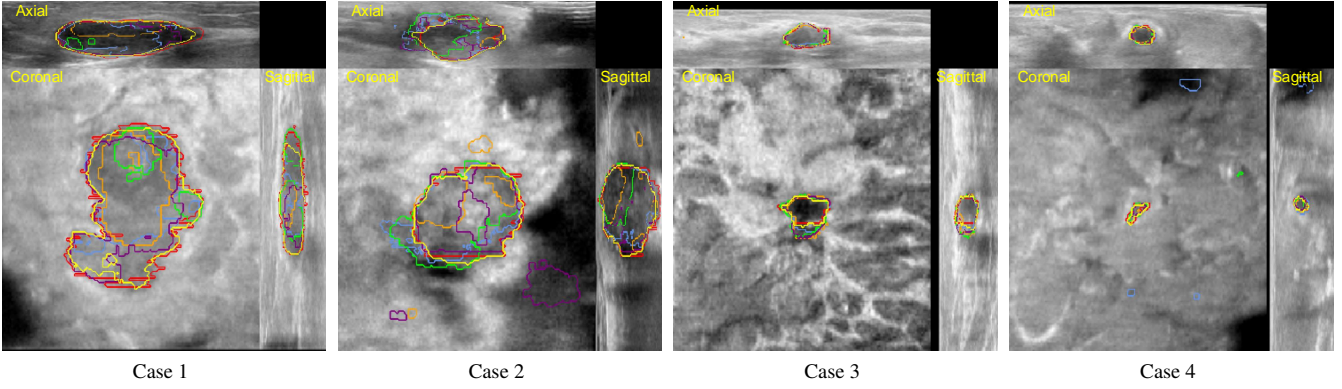
Fig. 7: Segmentation boundaries given by VNet (blue), CVNet (orange), C_{MS}VNet (green), CVNet$_{Iter}$ (purple), C_{MS}VNet$_{Iter}$ (yellow), and ground truth (red).

Table 4: Classification performance using different multi-scale features.

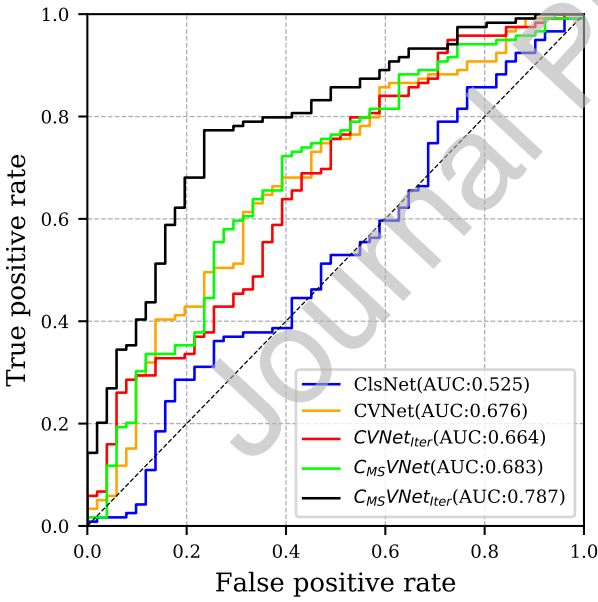| Model	ACC	REC	PRE	FPR	F1
C_{MS}(1-9)VNet	0.682	**0.849**	0.737	0.706	0.789
C_{MS}(2-8)VNet	0.659	0.706	**0.785**	**0.451**	0.743
C_{MS}(3-7)VNet	0.682	0.815	0.752	0.627	0.782
C_{MS}(4-6)VNet	**0.718**	**0.849**	0.771	0.588	**0.808**
C_{MS}(5)VNet	0.700	0.840	0.757	0.627	0.797

Fig. 8: ROCs of ablation models.

ABUS classification.

5.3. Multi-Task Learning

The hyperparameter in Eq. 5 balances classification and segmentation performance during multi-task learning. Classification and segmentation performance with different λ on the proposed C_{MS}VNet$_{Iter}$ model are listed in Table 5. When λ is set

to 0.3, C_{MS}VNet$_{Iter}$ balances classification and segmentation and gives the best performance on DSC, JI, ACC, PRE, FPR, and F1 score. Although classification REC performance is best at 0.848 with $\lambda = 0.9$, segmentation performance is obviously worse than $\lambda = 0.3$. Therefore, we set $\lambda = 0.3$ in the experiments.

We compare the proposed method with other multi-task learning methods, the comparison is summarized in Table 6. We trained four existing methods on our ABUS data (Wang et al., 2018b; Chakravarty and Sivaswamy, 2018; Xie et al., 2018; Qu et al., 2019). Table 6 indicates that the proposed C_{MS}VNet$_{Iter}$ model achieves the best performance for all metrics except REC and C_{MS}VNet yields the second best REC among all methods.

6. Discussion

ABUS is widely used in breast cancer diagnosis, due to its safety, better reproducibility, and short acquisition time (Vourtsis, 2019; Giger et al., 2016; Zanotel et al., 2018). However, ABUS generates multiple 3D large view volumes for one breast, making the reviewing of ABUS volumes a time-consuming task. Thus designing CAD systems for ABUS is necessary. Because of the large tumor size and shape variation, irregular and ambiguous tumor boundaries, and low signal to noise ratio in ultrasound volumes, ABUS tumor segmentation and classification are two challenge tasks. In this paper, we have introduced a multi-task learning framework for segmentation and classification of tumors in 3D ABUS images.

We employ a multi-scale feature concatenation network, realized via GAP layers and feature channel concatenations, to fuse features from different VNet stages for classification. As reported in Table 3, the proposed C_{MS}VNet and C_{MS}VNet$_{Iter}$ networks improve classification performance compared with the single classification model and multi-task models with single-scale feature concatenation. Furthermore, the proposed multi-scale feature connection classification method is able to rectify classification errors in the classification baseline. We show in Fig. 9 a benign tumor and a malignant tumor that are misclassed by the classification baseline due to their similar appearances. However, they are correctly classified by our method when multi-scale features are used (i.e., C_{MS}VNet and C_{MS}VNet$_{Iter}$).

Table 5: The effects of λ on multi-task learning.

	Segmentation (Mean ± SD)			Classification				
λ	DSC	JI	95HD (mm)	ACC	REC	PRE	FPR	F1
0.1	0.768 ± 0.154	0.644 ± 0.166	**2.992 ± 3.955**	0.700	0.773	0.793	0.471	0.783
0.3	**0.778 ± 0.145**	**0.650 ± 0.170**	3.303 ± 4.513	**0.741**	0.798	**0.826**	**0.392**	**0.811**
0.5	0.768 ± 0.143	0.643 ± 0.160	3.288 ± 4.413	0.694	0.773	0.786	0.491	0.780
0.7	0.762 ± 0.148	0.635 ± 0.166	3.683 ± 4.938	0.694	0.781	0.781	0.510	0.782
0.9	0.746 ± 0.163	0.618 ± 0.178	3.876 ± 4.825	0.700	**0.848**	0.753	0.647	0.798

Table 6: Comparison with existing multi-task learning models.

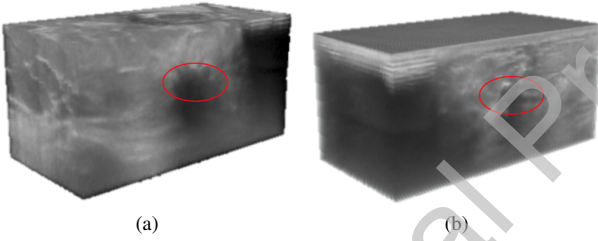|	Segmentation (Mean ± SD)			Classification				
Method	DSC	JI	95HD (mm)	ACC	REC	PRE	FPR	F1
Wang et al. (2018b)	0.651 ± 0.244	0.523 ± 0.232	9.362 ± 11.669	0.676	0.823	0.724	0.667	0.781
Chen et al. (2018)	0.656 ± 0.234	0.527 ± 0.228	10.522 ± 12.363	0.653	0.748	0.754	0.568	0.751
Xie et al. (2018)	0.681 ± 0.193	0.544 ± 0.198	8.512 ± 10.599	0.658	0.832	0.723	0.745	0.773
Qu et al. (2019)	0.620 ± 0.212	0.483 ± 0.213	12.460 ± 11.741	0.700	**0.882**	0.739	0.725	0.804
C_{MS}VNet	0.690 ± 0.208	0.550 ± 0.208	9.334 ± 11.743	0.718	0.849	0.771	0.588	0.808
C_{MS}VNet$_{Iter}$	**0.778 ± 0.145**	**0.650 ± 0.170**	**3.303 ± 4.513**	**0.741**	0.798	**0.826**	**0.392**	**0.811**

Fig. 9: (a) A malignant tumor and (b) a benign tumor misclassed by the classification baseline and but correctly classified by our method.

We have also demonstrated that iterative feature refinement improves segmentation and classification outcomes by progressive focusing on the tumor region. C_{MS}VNet$_{Iter}$ yields more precise segmentation and higher classification accuracy compared with C_{MS}VNet, as reported in Table 3. In Table 2, with the increase of N, the segmentation network achieves more and more accurate outcomes. However, when the network was optimized more times, the performance starts to degrade possibly, such as results of N=3 and N=4. One possible reason is that the network may be guided by wrong prediction results from the beginning during iterative training. Another reason is that overfitting occurs when the iteration increases. Although training VNet with the N=2 has the best segmentation results in a single segmentation task, the proposed C_{MS}VNet$_{Iter}$ model still outperforms the VNet$_{Iter}$ model by incorporation information from tumor classification. Multi-task learning can not only improve the performance of both tasks but also salve training time. We compare the training time of each epoch under same parameters and hardware conditions. C_{MS}VNet$_{Iter}$ costs 6 minutes, which is less than training these two tasks independently (9 muiutes).

Despite improved segmentation and classification perfor-

mance, our method has some limitations. First, due to the imbalance of samples for benign and malignant tumors, the network may be affected classification bias. Although employing a weighted focal loss to mitigate this problem, the FPR for C_{MS}VNet$_{Iter}$ is still high at 0.392, as listed in Table 3. Second, tumors adjacent to nipple regions are difficult to be segmented, as presented in Fig. 10. The proposed method successfully locates such tumors, however, segmentation metrics are not satisfied. The DSC value of cases in Fig. 10a and Fig. 10b are at 0.404 and 0.415, respectively. Third, the multi-task weight λ in Eq. 5 is adjusted manually to balance classification and segmentation performance. This adjusting method is imprecise and time-consuming. Inspired by paper Kendall et al. (2018) and Nair et al. (2020), multi-task weights can be selected automatically based on training performance by considering the homoscedastic uncertainty of each task, which will be researched in our further work.

7. Conclusion

In this paper, we introduce a multi-task learning network for end-to-end segmentation and classification in 3D ABUS volumes by one pass. Specifically, we design a multi-scale feature concatenation network for the classification task and adopt an iterative feature-refining training strategy to highlight tumor regions for both tasks. The effectiveness of this proposed method is evaluated on a database with 170 volumes obtained from 107 patients. Experimental results demonstrate that the proposed multi-task model provides better results on both tumor segmentation and classification in comparison to the individually trained single-task models, achieving state-of-the-art performance in the existing methods.
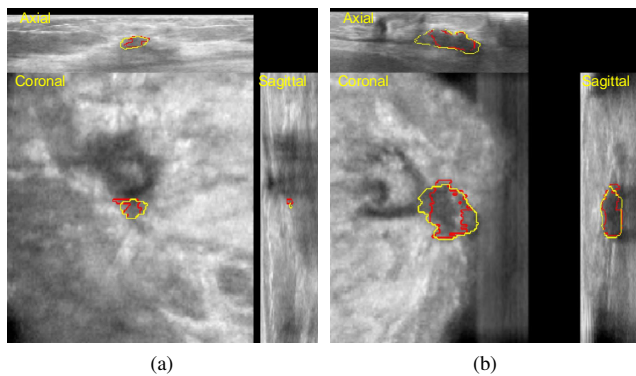
(a)   (b)

Fig. 10: Near-nipple tumors are challenging to segment, as the discrepancies between our results (yellow) and the ground truth segmentations (red). Segmentation DSC of tumors for (a) and (b) are 0.404 and 0.415, respectively.

References

Al-Dhabyani, W., Gomaa, M., Khaled, H., Aly, F., 2019. Deep learning approaches for data augmentation and classification of breast masses using ultrasound images. Int. J. Adv. Comput. Sci. Appl. 10, 681–627. doi:10.14569/IJACSA.2019.0100579.

Bian, C., Lee, R., Chou, Y.H., Cheng, J.Z., 2017. Boundary regularized convolutional neural network for layer parsing of breast anatomy in automated whole breast ultrasound, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 259–266. doi:10.1007/978-3-319-66179-7_30.

Bleicher, R.J., Ruth, K., Sigurdson, E.R., Beck, J.R., Ross, E., Wong, Y.N., Patel, S.A., Boraas, M., Chang, E.I., Topham, N.S., Egleston, B.L., 2016. Time to surgery and breast cancer survival in the united states. JAMA oncology 2, 330–339. doi:10.1001/jamaoncol.2015.4508.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A., 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians 68, 394–424. doi:10.3322/caac.21492.

Brem, R.F., Lenihan, M.J., Lieberman, J., Torrente, J., 2015. Screening breast ultrasound: past, present, and future. American Journal of Roentgenology 204, 234–240. doi:10.2214/AJR.13.12072.

Cao, X., Chen, H., Li, Y., Peng, Y., Wang, S., Cheng, L., 2020. Uncertainty aware temporal-ensembling model for semi-supervised abus mass segmentation. IEEE Transactions on Medical Imaging , 1–1doi:10.1109/TMI.2020.3029161.

Chakravarty, A., Sivaswamy, J., 2018. A deep learning based joint segmentation and classification framework for glaucoma assesment in retinal color fundus images. ArXiv .

Chen, C., Bai, W., Rueckert, D., 2018. Multi-task learning for left atrial segmentation on ge-mri, in: International workshop on statistical atlases and computational models of the heart, Springer. pp. 292–301. doi:10.1007/978-3-030-12029-0_32.

Chen, C.M., Chou, Y.H., Chen, C.S., Cheng, J.Z., Ou, Y.F., Yeh, F.C., Chen, K.W., 2005. Cell-competition algorithm: a new segmentation algorithm for multiple objects with irregular boundaries in ultrasound images. Ultrasound in medicine & biology 31, 1647–1664. doi:10.1016/j.ultrasmedbio.2005.09.011.

Cheng, J.Z., Chou, Y.H., Huang, C.S., Chang, Y.C., Tiu, C.M., Chen, K.W., Chen, C.M., 2010a. Computer-aided us diagnosis of breast lesions by using cell-based contour grouping. Radiology 255, 746–754. doi:10.1148/radiol.09090001.

Cheng, J.Z., Chou, Y.H., Huang, C.S., Chang, Y.C., Tiu, C.M., Yeh, F.C., Chen, K.W., Tsou, C.H., Chen, C.M., 2010b. Accomp: augmented cell competition algorithm for breast lesion demarcation in sonography. Medical physics 37, 6240–6252. doi:10.1118/1.3512799.

Chough, D.M., Berg, W.A., Bandos, A.I., Rathfon, G.Y., Hakim, C.M., Lu, A.H., Gizienski, T.A., Ganott, M.A., Gur, D., 2020. A prospective study of automated breast ultrasound screening of women with dense breasts in a digital breast tomosynthesis-based practice. Journal of Breast Imaging 2, 125–133. doi:10.1093/jbi/wbaa006.

Flores, W.G., de Albuquerque Pereira, W.C., Infantosi, A.F.C., 2015. Improving classification performance of breast lesions on ultrasonography. Pattern Recognition 48, 1125–1136. doi:10.1016/j.patcog.2014.06.006.

Galińska, M., Ogiegło, W., Wijata, A., Juszczyk, J., Czajkowska, J., 2017. Breast cancer segmentation method in ultrasound images, in: Conference on Innovations in Biomedical Engineering, Springer. pp. 23–31. doi:10.1007/978-3-319-70063-2_3.

Giger, M.L., Inciardi, M.F., Edwards, A., Papaioannou, J., Drukker, K., Jiang, Y., Brem, R., Brown, J.B., 2016. Automated breast ultrasound in breast cancer screening of women with dense breasts: Reader study of mammography-negative and mammography-positive cancers. American Journal of Roentgenology 206, 1341–1350. doi:10.2214/AJR.15.15367.

Gomez, W., Leija, L., Alvarenga, A., Infantosi, A., Pereira, W., 2010. Computerized lesion segmentation of breast ultrasound based on marker-controlled watershed transformation. Medical physics 37, 82–95. doi:10.1118/1.3265959.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.

Jesorsky, O., Kirchberg, K.J., Frischholz, R.W., 2001. Robust face detection using the hausdorff distance, in: International conference on audio-and video-based biometric person authentication, Springer. pp. 90–95. doi:10.1007/3-540-45344-X_14.

Kendall, A., Gal, Y., Cipolla, R., 2018. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7482–7491.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 .

Kong, X., Tan, T., Bao, L., Wang, G., 2018. Classification of breast mass in 3d ultrasound images with annotations based on convolutional neural networks. Chinese Journal of Biomedical Engineering 37, 414–422. doi:10.3969/j.issn.0258-8021.2018.04.004.

Kozegar, E., Soryani, M., Behnam, H., Salamati, M., Tan, T., 2017. Mass segmentation in automated 3-d breast ultrasound using adaptive region growing and supervised edge-based deformable model. IEEE transactions on medical imaging 37, 918–928. doi:10.1109/TMI.2017.2787685.

Le, E., Wang, Y., Huang, Y., Hickman, S., Gilbert, F., 2019. Artificial intelligence in breast imaging. Clinical radiology 74, 357–366. doi:10.1016/j.crad.2019.02.006.

Lee, C.Y., Chang, T.F., Chou, Y.H., Yang, K.C., 2020. Fully automated lesion segmentation and visualization in automated whole breast ultrasound (abus) images. Quantitative Imaging in Medicine and Surgery 10, 568. doi:10.21037/qims.2020.01.12.

Lei, B., Huang, S., Li, H., Li, R., Bian, C., Chou, Y.H., Qin, J., Zhou, P., Gong, X., Cheng, J.Z., 2020. Self-co-attention neural network for anatomy segmentation in whole breast ultrasound. Medical Image Analysis , 101753doi:10.1016/j.media.2020.101753.

Lei, B., Huang, S., Li, R., Bian, C., Li, H., Chou, Y.H., Cheng, J.Z., 2018. Segmentation of breast anatomy for automated whole breast ultrasound images with boundary regularized convolutional encoder–decoder network. Neurocomputing 321, 178–186. doi:10.1016/j.neucom.2018.09.043.

Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017a. Feature pyramid networks for object detection, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2117–2125.

Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017b. Focal loss for dense object detection, in: Proceedings of the IEEE international conference on computer vision, pp. 2980–2988.

Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. Medical Image Analysis 42, 60–88. doi:10.1016/j.media.2017.07.005.

Liu, W., Rabinovich, A., Berg, A.C., 2015. ParseNet: Looking wider to see better. arXiv:1506.04579 [cs] .

Milletari, F., Navab, N., Ahmadi, S.A., 2016. V-net: Fully convolutional neural networks for volumetric medical image segmentation, in: 2016 fourth international conference on 3D vision (3DV), IEEE. pp. 565–571. doi:10.1109/3DV.2016.79.

Moon, W.K., Lo, C.M., Chang, J.M., Huang, C.S., Chen, J.H., Chang, R.F., 2012. Computer-aided classification of breast masses using speckle features of automated breast ultrasound images. Medical physics 39, 6465–6473. doi:10.1118/1.4754801.

Moon, W.K., Shen, Y.W., Huang, C.S., Chiang, L.R., Chang, R.F., 2011.

Computer-aided diagnosis for the classification of breast masses in automated whole breast ultrasound images. Ultrasound in medicine & biology 37, 539–548. doi:10.1016/j.ultrasmedbio.2011.01.006.

Nair, T., Precup, D., Arnold, D.L., Arbel, T., 2020. Exploring uncertainty measures in deep networks for multiple sclerosis lesion detection and segmentation. Medical image analysis 59, 101557. doi:10.1016/j.media.2019.101557.

Nalepa, J., Marcinkiewicz, M., Kawulok, M., 2019. Data augmentation for brain-tumor segmentation: A review. Frontiers in Computational Neuroscience 13. doi:10.3389/fncom.2019.00083.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. PyTorch: An imperative style, high-performance deep learning library, in: Advances in Neural Information Processing Systems 32. Curran Associates, Inc., pp. 8026–8037.

Qu, H., Riedlinger, G., Wu, P., Huang, Q., Yi, J., De, S., Metaxas, D., 2019. Joint segmentation and fine-grained classification of nuclei in histopathology images, in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), IEEE. pp. 900–904. doi:10.1109/ISBI.2019.8759457.

Schroeder, W.J., Lorensen, B., Martin, K., 2004. The visualization toolkit: an object-oriented approach to 3D graphics. Kitware.

Shen, W., Zhou, M., Yang, F., Yang, C., Tian, J., 2015. Multi-scale convolutional neural networks for lung nodule classification, in: International Conference on Information Processing in Medical Imaging, Springer. pp. 588–599. doi:10.1007/978-3-319-19992-4_46.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 .

Singh, V.K., Rashwan, H.A., Abdel-Nasser, M., Sarker, M., Kamal, M., Akram, F., Pandey, N., Romani, S., Puig, D., 2019. An efficient solution for breast tumor segmentation and classification in ultrasound images using deep adversarial learning. arXiv preprint arXiv:1907.00887 .

Skaane, P., Bandos, A.I., Niklason, L.T., Sebuødegård, S., Østerås, B.H., Gullien, R., Gur, D., Hofvind, S., 2019. Digital mammography versus digital mammography plus tomosynthesis in breast cancer screening: The oslo tomosynthesis screening trial. Radiology 291, 23–30. doi:10.1148/radiol.2019182394.

Song, L., Lin, J.P., Wang, Z.J., Wang, H., 2020. An end-to-end multi-task deep learning framework for skin lesion analysis. IEEE journal of biomedical and health informatics 39, 1876–1886. doi:10.1109/JBHI.2020.2973614.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818–2826.

Tu, Z., Bai, X., 2009. Auto-context and its application to high-level vision tasks and 3d brain image segmentation. IEEE transactions on pattern analysis and machine intelligence 32, 1744–1757. doi:10.1109/TPAMI.2009.186.

Uniyal, N., Eskandari, H., Abolmaesumi, P., Sojoudi, S., Gordon, P., Warren, L., Rohling, R.N., Salcudean, S.E., Moradi, M., 2014. Ultrasound rf time series for classification of breast lesions. IEEE transactions on medical imaging 34, 652–661. doi:10.1109/TMI.2014.2365030.

Vourtsis, A., 2019. Three-dimensional automated breast ultrasound: Technical aspects and first results. Diagnostic and Interventional Imaging 100, 579–592. doi:10.1016/j.diii.2019.03.012.

Wang, N., Bian, C., Wang, Y., Xu, M., Qin, C., Yang, X., Wang, T., Li, A., Shen, D., Ni, D., 2018a. Densely deep supervised networks with threshold loss for cancer detection in automated breast ultrasound, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 641–648. doi:10.1007/978-3-030-00937-3_73.

Wang, P., Patel, V.M., Hacihaliloglu, I., 2018b. Simultaneous segmentation and classification of bone surfaces from ultrasound using a multi-feature guided cnn, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 134–142. doi:10.1007/978-3-030-00937-3_16.

Wang, Y., Choi, E.J., Choi, Y., Zhang, H., Jin, G.Y., Ko, S.B., 2020. Breast cancer classification in automated breast ultrasound using multiview convolutional neural network with transfer learning. Ultrasound in Medicine & Biology 46, 1119–1132. doi:10.1016/j.ultrasmedbio.2020.01.001.

Wu, L., Xin, Y., Li, S., Wang, T., Heng, P.A., Ni, D., 2017. Cascaded fully convolutional networks for automatic prenatal ultrasound image segmentation, in: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), pp. 663–666. doi:10.1109/ISBI.2017.7950607.

Xie, X., Shi, F., Niu, J., Tang, X., 2018. Breast ultrasound image classification and segmentation using convolutional neural networks, in: Pacific Rim Conference on Multimedia, Springer. pp. 200–211. doi:10.1007/978-3-030-00764-5_19.

Xing, J., Li, Z., Wang, B., Yu, B., Zanjani, F.G., Zheng, A., Duits, R., Tan, T., 2019. Automated segmentation of lesions in ultrasound using semi-pixel-wise cycle generative adversarial nets. arXiv preprint arXiv:1905.01902 .

Yang, W., Zhang, S., Chen, Y., Li, W., Chen, Y., 2008. Measuring shape complexity of breast lesions on ultrasound images, in: Medical Imaging 2008: Ultrasonic Imaging and Signal Processing, International Society for Optics and Photonics. p. 69200J.

Zanotel, M., Bednarova, I., Londero, V., Linda, A., Lorenzon, M., Girometti, R., Zuiani, C., 2018. Automated breast ultrasound: basic principles and emerging clinical applications. La Radiologia Medica 123, 1–12. doi:10.1007/s11547-017-0805-z.

Zeiler, M.D., Fergus, R., 2014. Visualizing and understanding convolutional networks, in: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (Eds.), Computer Vision – ECCV 2014, Springer International Publishing. pp. 818–833. doi:10.1007/978-3-319-10590-1_53.

Zeimarani, B., Costa, M.G.F., Nurani, N.Z., Bianco, S.R., Pereira, W.C.D.A., Costa Filho, C.F.F., 2020. Breast lesion classification in ultrasound images using deep convolutional neural network. IEEE Access 8, 133349–133359. doi:10.1109/ACCESS.2020.3010863.

Zhang, R., Shen, J., Wei, F., Li, X., Sangaiah, A.K., 2017. Medical image classification based on multi-scale non-negative sparse coding. Artificial intelligence in medicine 83, 44–51. doi:10.1016/j.artmed.2017.05.006.

Yue Zhou: Conceptualization, Methodology, Software, Writing- Original draft preparation, **Houjin Chen**: Writing- Reviewing and Editing, Supervision, **Yanfeng Li**: Writing- Reviewing and Editing, **Qin Liu**: Methodology, Software, Formal analysis, **Xuanang Xu:** Methodology, Software, Formal analysis, **Shu Wang**: Data curation, **Pew-Thian Yap**: Writing- Reviewing and Editing, **Dinggang Shen**: Supervision, Methodology.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: