IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 3 December 2022, accepted 2 January 2023, date of publication 13 January 2023, date of current version 19 January 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3236693

RESEARCH ARTICLE

A Regional-Attentive Multi-Task Learning Framework for Breast Ultrasound Image Segmentation and Classification

MENG XU[1], **(Graduate Student Member, IEEE), KUAN HUANG**[2], **(Member, IEEE), AND XIAOJUN QI**[1], **(Senior Member, IEEE)**

[1]Department of Computer Science, Utah State University, Logan, UT 84322, USA
[2]Department of Computer Science and Technology, Kean University, Union, NJ 07083, USA

Corresponding author: Kuan Huang (khuang@kean.edu)

This work was supported in part by the Graduate Research and Creative Opportunity (GRCO) Grant of Utah State University.

ABSTRACT Breast ultrasound (BUS) imaging is commonly used in the early detection of breast cancer as a portable, valuable, and widely available diagnosis tool. Automated BUS image classification and segmentation can assist radiologists in making accurate and fast decisions. Recent studies illustrate that tumor, peritumoral, and background regions of BUS images provide valuable information for BUS image segmentation or classification. However, few studies have investigated the influence of these three regions on multi-task learning. In this study, we propose an RMTL-Net to simultaneously segment tumor regions and classify tumors in BUS images into benign or malignant categories. To improve both segmentation and classification performance, we design a regional attention (RA) module that employs the predicted probability maps to automatically guide the classifier to learn important category-sensitive information in the tumor, peritumoral, and background regions and seamlessly fuse them to obtain a better feature representation. We conduct detailed ablation experiments of the proposed RA module and comparative experiments with four recent state-of-the-art peer multi-task learning methods, three single-task segmentation methods, and four single-task classification methods on two public BUS datasets. Experimental results show that the proposed RMTL-Net achieves the best overall segmentation and classification accuracy in terms of five segmentation metrics and six classification metrics.

INDEX TERMS Regional attention, multi-task learning, segmentation, classification, breast ultrasound.

I. INTRODUCTION

Breast cancer is a significant threat to women's health and is the most commonly diagnosed cancer and the leading cause of cancer mortality among women worldwide in 2020 [1]. Mortality rates are much higher in low- and middle-income countries than in high-income countries due to the delayed detection and treatment [2], [3]. Mammography and breast ultrasound (BUS) are two popular screening modalities for early breast cancer detection, which leads to appropriate treatment and increased survival rates. BUS has been commonly used in the early diagnosis of breast cancer in women of all ages, especially in low- and middle-income countries, because it is portable, widely available, low-cost, and highly sensitive [4], [5]. Computer-aided-diagnosis (CAD) systems are proposed to help radiologists interpret BUS images, make a more accurate diagnosis, and reduce their workload [6], [7]. In general, a CAD system for breast cancer detection includes automated segmentation and classification as two primary steps for further processing. Automated analysis of BUS images can help radiologists make efficient diagnoses of breast cancer. However, it is still challenging due to the lack of public training data and the high variability of tumors in shape, size, and location [8], [9].

BUS image segmentation methods can be classified into semi-automated [10], [11], [12] and fully automated

The associate editor coordinating the review of this manuscript and approving it for publication was Ravibabu Mulaveesala.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

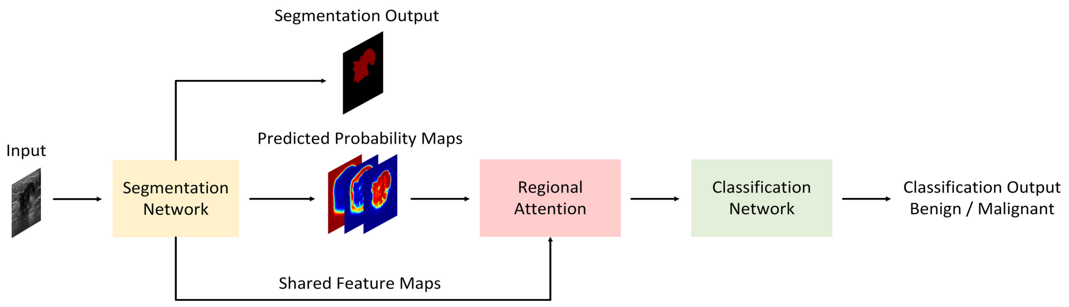


FIGURE 1. An overview of the proposed RMTL-Net.

methods [13], [14] based on human intervention. Fully automated BUS image segmentation is the trend in future BUS CAD systems since it is reproducible and suitable for large-scale tasks [15]. Fully automated deep learning-based methods, especially U-Net [16] based methods [9], [17], [18], have recently gained increased popularity. For example, Wang et al. [9] propose a fusion deep learning network to address issues of unclear boundaries and large variations in tumors in BUS images. It uses an encoder to capture the context information, a decoder to localize prediction, and a fusion to combine information from the encoder and the decoder. Amiri et al. [17] propose a two-stage U-Net architecture: one for tumor detection and one for tumor segmentation. They also prove that detection and its evaluation in the first stage improve segmentation results in the second stage.

Convolutional neural networks (CNNs) have recently achieved superior performance compared to traditional machine learning classification methods such as support vector machine [19], K-nearest neighbors [20], random forest [21], and Gaussian mixture models [22]. Among them, VGG [23], ResNet [24], and their variants are widely used for BUS image classification. Liao et al. [25] adopt a supervised block-based segmentation algorithm to separate tumor regions from BUS images and then use VGG-19 to classify segmented tumor regions as benign or malignant. Cui et al. [26] propose to use ResNet-34 as the backbone feature extractor and design a fused network to combine features of tumor, peritumoral, and combined-tumoral (combination of tumor and peritumoral) regions to achieve better classification results.

Multi-task learning (MTL) for simultaneous BUS segmentation and classification has recently been extensively studied in the computer vision community. Benign and malignant breast tumors have different characteristics [27], [28]. For example, benign tumors tend to be smooth, round, and well circumscribed whereas malignant tumors are typically rough and spiculated. In addition, malignant tumors tend to have spiculated margins and posterior acoustic shadows. Based on these observations, many MTL [29], [30], [31], [32] studies are proposed to join BUS image segmentation and classification tasks in one network to encourage feature sharing during training to improve both tasks. These MTL methods are mostly based on a U-Net structure (*i.e.*, an encoder-decoder network for segmentation) and some of them [30], [32] include attention mechanisms to achieve better classification performance. For example, Zhou et al. [29] propose an MTL framework with a light-weight multi-scale network to iteratively refine features to highlight tumor regions for better 3D BUS image classification. Chowdary et al. [31] propose an MTL framework with a dense branch to combine multi-scale features from different layers of the network for efficient classification of BUS images. Zhang et al. [30] propose an MTL framework with soft and hard attention mechanisms to guide the model to pay more attention to tumor regions to boost classification accuracy. Xu et al. [32] propose an MTL framework with a context-oriented self-attention (COSA) module to incorporate prior medical knowledge to guide the model to learn contextual relationships for better segmentation and classification performance.

Recently, several studies have demonstrated that tumor, peritumoral (the tumor-adjacent area surrounding the tumor), and background regions in BUS images help to improve the diagnosis accuracy of breast cancer in CAD methods [26], [33], [34], [35]. Lee et al. [34] use the mask R-CNN to extract tumor regions from BUS images and obtain peritumoral regions via a dilation operation. They then use a deep learning model to train tumor, peritumoral, and their combined-tumoral regions to predict axillary lymph node (ALN) metastasis status, which is important in guiding treatment in breast cancer. Sun et al. [33] build two models based on tumor, peritumoral, and combined-tumoral regions and compare their performance to show that peritumoral and combined-tumoral regions achieve significantly better performance in predicting ALN metastasis in BUS images for both models.

Tumor, peritumoral, and background regions of a BUS image have been further studied to provide important category-sensitive information to improve the aforementioned methods to achieve better segmentation or classification results. Specifically, the peritumoral region in BUS images was discussed in the BUS image classification task [26] and the ALN metastasis prediction task [33], [34] to further improve their accuracy. Cui et al. [26] use an encoder-decoder structure to obtain three tumoral regions

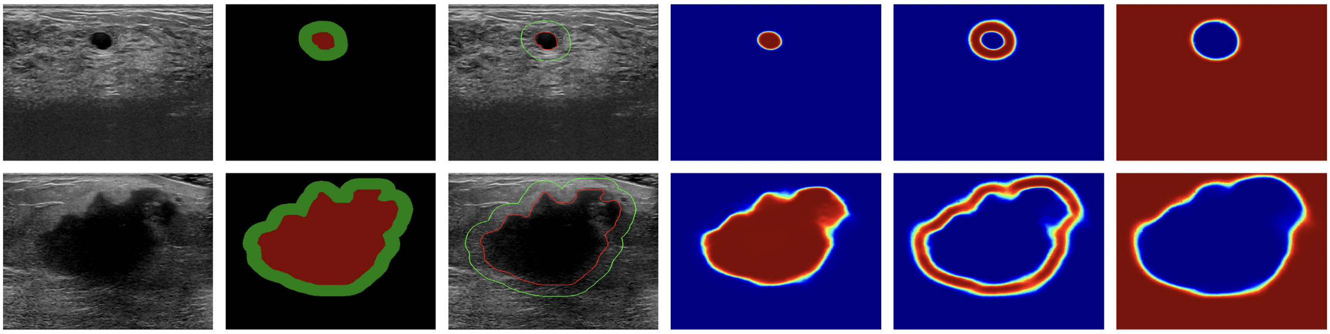


FIGURE 2. Illustration of two examples of BUS images, their ground truth and pseudo ground truth regions, and three probability maps generated by the proposed RMTL-Net. First column: Original BUS images with a benign tumor shown at the top row and a malignant tumor shown at the bottom row. Second column: Pseudo ground truth regions produced by the proposed pre-processing method, where the peritumoral region is shown in green and the background region is shown in black. The ground truth tumor region is shown in red. Third column: Three regions containing category-sensitive information overlaid on the original image, where the tumor region is within the red line, the peritumoral region is between green and red lines, and the background region is outside the green line. Fourth column: Probability map of the tumor region. Fifth column: Probability map of the peritumoral region. Sixth column: Probability map of the background region.

at different resolutions to extract tumor features (e.g., component, internal echo, and aspect ratio), peritumoral features (e.g., tumor boundary patterns), and background features (e.g., contextual relationship between the tumor and surrounding tissues). These features lead to higher computational costs but better classification results. Despite the success of the utilization of three tumoral regions, they have hardly been employed in simultaneous BUS image segmentation and classification. To the best of our knowledge, the research work of Xu et al. [32] is the pioneer in this direction. They employ three tumoral regions in a BUS image to improve the MTL performance. However, their extracted peritumoral region is small, which may not provide sufficient information for simultaneous BUS image segmentation and classification.

In this paper, we propose a regional attention (RA) module to learn corresponding category-sensitive features from three regions (e.g., tumor, peritumoral, and background regions) in BUS images and investigate their influence on MTL. We also apply the proposed RA module to a two-stage MTL framework to demonstrate its efficacy in BUS image segmentation and classification. The proposed regional-attentive multi-task learning framework (RMTL-Net) consists of an encoder-decoder network for segmentation and a light-weight network for classification. Both segmentation and classification share features extracted from the encoder. In addition, the RA module utilizes the predicted probability maps to guide the classification network to learn weighted region attentive features for more accurate classification. The overall framework of the proposed RMTL-Net is illustrated in Fig. 1. We conduct extensive experiments on two public BUS datasets that include 810 BUS images in total to evaluate the performance of RMTL-Net and its variants and compare RMTL-Net with several state-of-the-art single-task and multi-task methods. Experimental results show that RMTL-Net boosts the performance of both segmentation and classification tasks. Our main contributions are summarized as follows:

- We design a novel MTL framework, named RMTL-Net, for simultaneous tumor segmentation and classification in BUS images. The proposed RMTL-Net outperforms recent state-of-the-art segmentation and classification methods on two public BUS datasets.
- We propose a RA module to improve both segmentation and classification performance. It employs the predicted probability maps to automatically guide the classifier to learn important category-sensitive information in the tumor, peritumoral, and background regions.
- We conduct extensive experiments on two public BUS datasets. Experimental results prove its MTL efficacy in BUS image segmentation and classification and the importance of tumor, peritumoral, and background regions of BUS images.

II. MATERIALS AND METHODS

In this section, we first present the materials in terms of two datasets and the proposed pre-processing method to prepare the training images and their pseudo ground truth images. We then describe the proposed method in terms of its network architecture and the regional attention (RA) module.

A. MATERIALS

1) DATASETS

Two public BUS datasets used in this study are UDIAT [36] and BUSI [37]. Dataset UDIAT was collected by the UDIAT Diagnostic Centre of the Parc Taulı Corporation, Sabadell (Spain) using a Siemens ACUSON Sequoia C512 system 17L5 HD linear array transducer (8.5 MHz). It contains 163 BUS images with an average size of 760 × 570 pixels, where 110 images have benign tumors and 53 images have malignant tumors. These BUS images are obtained from different female patients and each BUS image presents one tumor. Ground truth is labeled by experienced radiologists. Dataset BUSI was collected by Baheya Centre for Early Detection and Treatment of Women's Cancer, Egypt using

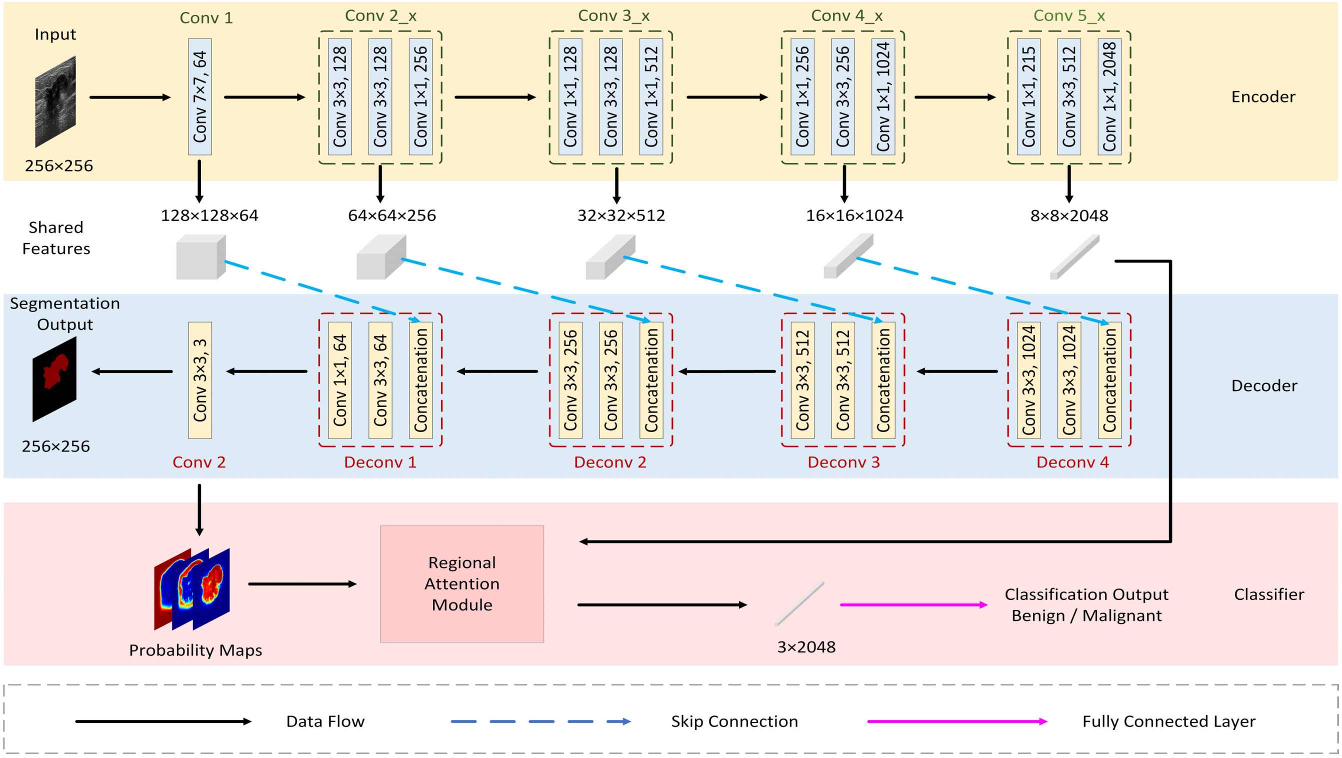


FIGURE 3. A detailed illustration of the proposed RMTL-Net.

LOGIQ E9 ultrasound and LOGIQ E9 Agile ultrasound system. It contains 780 BUS images with an average size of 500 × 500 pixels, where 437 images have benign tumors, 210 images have malignant tumors, and 133 images do not have any tumors. These BUS images are obtained from 600 female patients between the ages of 25 and 75 years old. We use 647 images with benign or malignant tumors in this dataset for binary classification in this study. Ground truth is labeled by radiologists from Baheya.

2) PRE-PROCESSING

In the proposed method, all images are resized to 256 × 256 by bilinear interpolation before being fed into RMTL-Net. Data augmentation techniques are carried out to augment images during the training process using four transformations: (i) rotation of an angle between -5 and 5 degrees at the image center, (ii) random flipping horizontally, vertically, or both, (iii) Gaussian blur, and (iv) Median blur. We perform these four transformations in the above order on each input BUS image to augment the training images during the training procedure.

Given a ground truth BUS image that contains the tumor contour, we generate two pseudo ground truth regions: peritumoral and background regions. First, we employ a Laplace edge detector on the ground truth image to find the contour of the tumor region. Second, we dilate the tumor region by 32 pixels and subtract the tumor region from the dilated result to obtain the peritumoral region. We choose 32 pixels in dilation to ensure the peritumoral region remains at the lowest resolution when a series of down-sampling operations take place in RMTL-Net. Third, we treat the remaining region as the background region. The first three columns in Fig. 2 present BUS example images, their ground truth tumor region labeled by radiologists and their pseudo ground truth peritumoral and background regions produced by the proposed pre-processing method, and three regions as shown on the original images. An image containing the ground truth tumor region, the pseudo ground truth peritumoral region, and the pseudo ground truth background region is further used during the training process to learn the boundaries delineating tumor, peritumoral, and background.

B. METHODS

The proposed RMTL-Net improves its peer MTL-COSA [32] from the following five aspects:

- Unlike MTL-COSA that generates a binary segmentation result, RMTL-Net generates a binary segmentation result and three probability maps for tumor, peritumoral, and background regions, respectively.
- Unlike MTL-COSA that uses the contour of segmented tumors to find binary segmentation masks for tumor, peritumoral, and background regions, RMTL-Net uses probability maps generated from the network to estimate tumor, peritumoral, and background regions in BUS images and feed them as estimated prior medical knowledge into the RA module to guide the classification task.

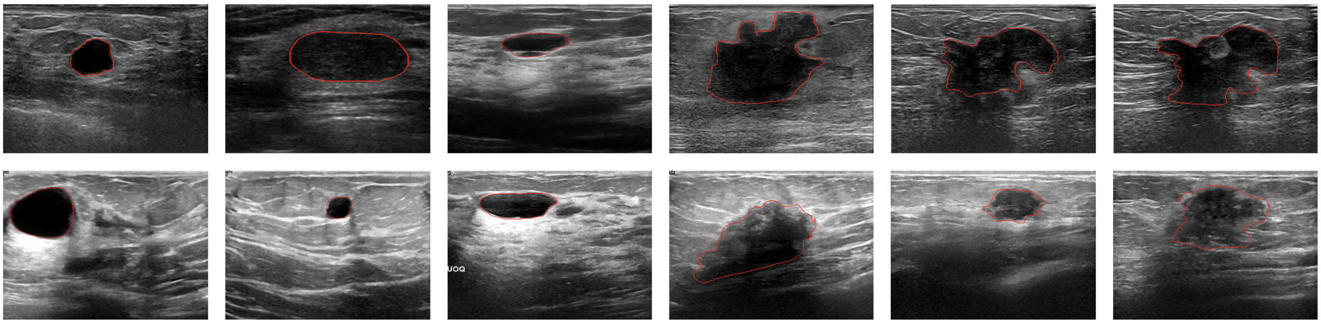


FIGURE 4. BUS images containing benign tumors (the first three columns) and malignant tumors (the last three columns). The first row shows images from dataset UDIAT and the second row shows images from dataset BUSI.

- Unlike MTL-COSA that extracts the peritumoral region by dilating the segmented tumor boundary, RMTL-Net is trained to generate respective probability maps for tumor, peritumoral, and background regions to gather more detailed categorical information than the binary masks extracted by MTL-COSA.
- Unlike MTL-COSA whose peritumoral region has a ring area of width of 5 pixels evenly covering the background and tumor areas, RMTL-Net extracts a bigger peritumoral region with a ring-like area of width of 32 pixels outside of the tumor to provide sufficient information at the lowest resolution to facilitate classification.
- Unlike MTL-COSA that uses self-attention to learn important classification features, RMTL-Net replaces it with the RA module to significantly reduce network parameters by 14.40% and reduce both training and testing times yet achieve better overall segmentation and classification performance.

1) NETWORK ARCHITECTURE

The detailed network architecture of the proposed RMTL-Net is illustrated in Fig. 3. RMTL-Net is a two-stage framework that consists of a segmentation stage and a classification stage. The segmentation stage utilizes a U-shape architecture consisting of an encoder, a decoder, and skip connections to extract multi-scale features and predict three respective probability maps for tumor, peritumoral, and background regions, as shown in the last three columns in Fig. 2. The classification stage uses shared features extracted from the encoder and three probability maps generated from the segmentation stage to produce classification results. Specifically, we use the peritumoral region to capture boundary characteristics, which are useful to differentiate benign and malignant tumors. We use the tumor region to capture the shape properties of tumors, which are useful for both tumor segmentation and classification. We use the background region to capture posterior acoustic shadowing, which is observed more for malignant lesions and less for benign tumors due to attenuation of the sonographic signal [27], [38]. Sharing features makes segmentation and classification promote each other during the training process. In addition, It addresses the problem of having insufficient training images for classification. Each pixel is a training sample in segmentation. Sharing features with the segmentation stage with sufficient training samples improves the overall accuracy and robustness of the classification stage.

We use ResNet-101 [24] as the backbone of the segmentation stage of RMTL-Net due to its great performance in BUS image segmentation and classification [18], [26]. The architecture of ResNet-101 remains the same. Specifically, the encoder utilizes one convolutional layer $Conv1$ together with four residual blocks ($Conv2\_x$ to $Conv5\_x$) to perform five down-sampling operations to extract multi-scale features from input images. Multi-scale features extracted by $Conv1$ to $Conv5\_x$ are of sizes $128 \times 128 \times 64$, $64 \times 64 \times 256$, $32 \times 32 \times 512$, $16 \times 16 \times 1024$, and $8 \times 8 \times 2048$, respectively. The decoder symmetrically utilizes four deconvolutional blocks ($Deconv4$ to $Deconv1$) and one convolutional layer ($Conv2$) followed by bilinear interpolation and softmax operations to perform up-sampling operations. Skip connections between the encoder and decoder combine feature maps in different scales to compensate for the loss of spatial information during down-sampling operations and to refine segmentation outcomes. As a result, multi-scale features are restored to the original input size and are further interpreted to predict three probability maps.

We use three probability maps generated from the segmentation stage of RMTL-Net and multi-scale high-level features shared by both segmentation and classification stages to produce classification results.

2) REGIONAL ATTENTION MODULE

Unlike classical image classification networks (e.g., VGG [23] and ResNet [24]), we add a regional attention (RA) model to further encourage information sharing. This RA model outputs a weighted feature vector of size 1×2048 that is passed to a fully connected layer to generate more accurate classification results.

We observe benign and malignant tumors exhibit different characteristics. For example, benign tumors tend to be smooth and round and malignant tumors are always rough with an aspect ratio of greater than 1 [27], [28]. Benign

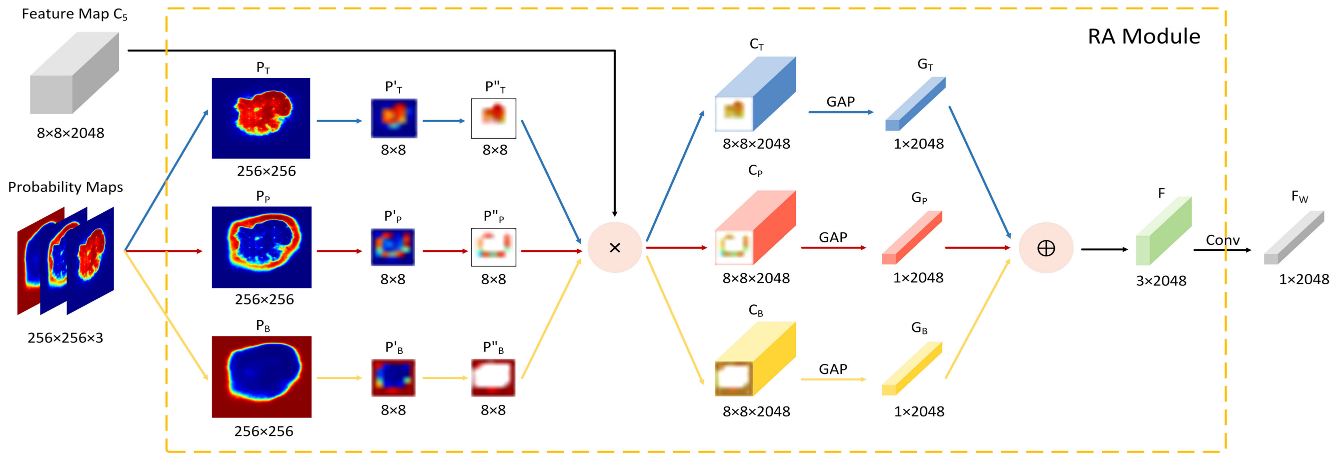


FIGURE 5. An overview of the proposed Regional Attention (RA) module.

tumors tend to have smooth, thin, and regular margins and malignant tumors tend to have spiculated, thick, and irregular margins. Benign tumors tend to have less posterior acoustic shadowing in the background region than malignant lesions. As a result, we propose to utilize tumor, peritumoral, and background regions to learn their inherently important characteristics including tumor features (e.g., component, internal echo, and aspect ratio), tumor boundary patterns (e.g., smoothness, shape, and contextual texture between tumor and surrounding tissues), and background features (posterior acoustic shadowing) [27], [35] to help with the joint segmentation and classification tasks. In addition, we propose to include a RA module in the classification stage of the RMTL-Net to encourage information sharing and output a weighted feature vector to facilitate classification. This RA module combines multi-scale high-level features with three probability maps generated from the segmentation stage to guide the learning of category-sensitive features from three regions, namely, tumor, peritumoral, and background regions. Category-sensitive features are represented as a weighted feature vector, which is passed to a fully connected layer to generate more accurate classification results. Fig. 4 shows six examples of BUS images from each of the two datasets that contain benign and malignant tumors, respectively. Tumor regions with high variability in shape, size, and location are delineated by red lines. When using these images as training images, we generate their pseudo ground truth peritumoral and background regions using the pre-processing method explained in Section II-A2. When using these images as testing images, RMTL-Net predicts their probability maps as shown in Fig. 2.

The structure diagram of the proposed RA module is shown in Fig. 5. The algorithmic view of the RA module is summarized below:

Input: C_5 (the feature map of size $8 \times 8 \times 2048$ extracted by $Conv5\_x$ of the encoder) and P (the probability map of size $256 \times 256 \times 3$ generated by the last convolutional layer $Conv2$ of the decoder).

Output: A weighted feature map F_W of size 1×2048.

1) Split P into three probability maps P_T, P_P, and P_B of size 256×256, where subscripts T, P, and B represent tumor, peritumoral, and background, respectively.
2) Employ the nearest neighbor method to resize P_T, P_P, and P_B to obtain coarse probability maps P'_T, P'_P, and P'_B of size 8×8.
3) Utilize a threshold of 0.5 to filter coarse probability maps P'_T, P'_P, and P'_B to obtain three noise-free probability maps P''_T, P''_P, and P''_B, respectively. Specifically, values greater than 0.5 in coarse probability maps are kept intact, and values smaller than or equal to 0.5 are set to 0:
$$P''_x = P'_x > 0.5 \,?\, P'_x : 0 \tag{1}$$
where subscript x can be replaced with T, P, or B.
4) Individually and elementwisely multiply P''_x with each channel of C_5 to generate multi-channel weighted regional feature maps C_x.
$$C_x = C_5 \cdot P''_x \tag{2}$$
5) Apply the global average pooling (GAP) on C_x to capture weights of each region in its corresponding G_x of size 1×2048:
$$G_x = GAP(C_x) \tag{3}$$
6) Concatenate G_T, G_P, and G_B to construct a new feature vector F of size 3×2048:
$$F = Concatenate(G_T, G_P, G_B) \tag{4}$$
7) Apply a 1×1 convolution filter to F to generate a weighted feature map F_W of size 1×2048.
$$F_W = f^{1\times1}(F) \tag{5}$$

It should be noted that all non-zero pixels in P''_T, P''_P, and P''_B have high likelihood values larger than 0.5, which indicate high strength of tumor, peritumoral, and background features, respectively. We choose 0.5 as the threshold because it classifies a pixel into one of the three classes. The multiplication of C_5 and P''_T, P''_P, and P''_B leads to multi-channel weighted tumor, peritumoral, and background features C_T, C_P, and C_B. The GAP operation further finds the features in each channel of C_T, C_P, and C_B to best represent three respective regions. The concatenation operation followed by the 1×1 convolution constructs a weighted sum of multi-view features from three parallel channels (*i.e.*, G_T, G_P, and G_B), which can be formulated as:

$$F_W = w_1 \cdot G_T + w_2 \cdot G_P + w_3 \cdot G_B \tag{6}$$

where w_1, w_2, and w_3 indicate the importance of tumor, peritumoral, and background regions, respectively. These weights are automatically learned during the training process. Finally, F_W is passed to a fully connected layer followed by a softmax activation function for automated tumor classification. F_W captures the importance of each region for better feature representation and therefore leads to better classification results than using a non-weighted feature map (*i.e.*, convolving C_5 with a feature vector of 1×2048). In summary, the proposed RA module follows the perspectives of radiologists to learn multi-view features from three regions in BUS images to achieve better segmentation and classification performance. Specifically, the tumor region helps to extract the basic features of breast tumors. The peritumoral region helps to capture tumor boundary patterns. The background region helps to collect contextual information.

3) LOSS FUNCTION

The overall loss of RMTL-Net is computed by the weighted sum of the loss of the segmentation task $\mathcal{L}_{seg}$ and the loss of the classification task $\mathcal{L}_{cls}$.

$$\mathcal{L} = \lambda \cdot \mathcal{L}_{seg} + (1 - \lambda) \cdot \mathcal{L}_{cls} \tag{7}$$

where λ and $1 - \lambda$ are contribution weights of losses from segmentation and classification tasks, respectively. Cross entropy is employed to compute both $\mathcal{L}_{seg}$ and $\mathcal{L}_{cls}$.

Let K denote the number of classes in a given task, N denote the number of images, and P denote the number of pixels in an image. In the segmentation task, there are 3 classes representing tumor, peritumoral, and background regions. In other words, $K = 3$. The pixel-wise cross entropy $\mathcal{L}_{seg}$ of the segmentation task is computed as follows:

$$\mathcal{L}_{seg} = -\frac{1}{P} \sum_{p}^{P} \sum_{k}^{K} y_{p,k} \cdot \log \hat{y}_{p,k} \tag{8}$$

where $y_{p,k}$ and $\hat{y}_{p,k}$ represent the true and predicted probability of pixel p belonging to class k, respectively. The true probability $y_{p,k}$ is either 0 or 1 since each pixel belongs to one of the three classes. The predicted probability $\hat{y}_{p,k}$ is in the range of [0, 1].

In the classification task, there are 2 classes representing benign and malignant tumors. In other words, $K = 2$. The image-wise cross-entropy $\mathcal{L}_{cls}$ of the classification task is computed as follows:

$$\mathcal{L}_{cls} = -\frac{1}{N} \sum_{n}^{N} \sum_{k}^{K} y_{n,k} \cdot \log \hat{y}_{n,k} \tag{9}$$

where $y_{n,k}$ and $\hat{y}_{n,k}$ represent the true and predicted category of image n belonging to class k, respectively. Both $y_{n,k}$ and $\hat{y}_{n,k}$ are either 0 or 1.

III. EXPERIMENTAL SETUP AND RESULTS

In this section, we first present the implementation details. We then describe the performance evaluation metrics followed by the competing methods. Finally, we present the experimental results of the proposed RMTL-Net method, its ablation study, and its comparison with the competing methods.

A. IMPLEMENTATION DETAILS

The implementation of the proposed method is based on the public platform PyTorch 1.4. All experiments are conducted on Ubuntu 18.04 system, Intel(R) Core(TM) CPU i5-11600K 3.9. All models are trained and tested on a GeForce RTX 3080 Ti graphics card with 12GB memory using the ADAM optimizer with momentum β_1 of 0.9, momentum β_2 of 0.99, a weight decay of 0.0001, and a learning rate initialized at 0.0001 and decayed at 10% after every 20 epochs. In the training procedure, the batch size is set as 16 and the number of training epochs is set as 100. Following the empirically optimal setup [24], we adopt batch normalization right after each convolution and before activation. To reduce overfitting, we adopt dropout with a probability of 0.5 in the fully connected layer of the classification network. The contribution weight of loss from the segmentation task (*i.e.*, λ) is empirically set to be 0.9. All competing methods, including ResNet, UResNet, MTL-Net, MTL-COSA, and RMTL-Net models, are pre-trained on ImageNet and fine-tuned with training images selected from datasets UDIAT and BUSI.

To evaluate the performance of different methods, we conduct five-fold cross-validation in all experiments, including multi-task learning, ablation, and comparative studies. Because the size of dataset UDIAT is small, there is 3% classification performance differences between multiple runs even if we use five-fold cross-validation to train and test on it. To increase the credibility of experimental results, we train all competing methods on two datasets together and test on two datasets separately. Specifically, for each dataset, we split the data into five groups, where each group keeps the same proportion of benign and malignant cases as in the original dataset. In each fold experiment, four groups of each dataset are combined and used as the training set, and the other group is used as the testing set. In this study, all experimental results are reported by averaging the five-fold cross-validation performance.

TABLE 1. A brief comparison of 11 SOTA methods and the proposed RMTL-Net.

Methods	Tasks			Feature Enhancement	
	Single Classification	Single Segmentation	Multi Task	Attention Mechanism	Skip Connection
VGG-16	✓				
DenseNet	✓				
ResNet-101	✓				
FCN		✓			
PSPNet		✓			
Deeplab v3+		✓			✓
U-ResNet		✓			✓
MTL-Net			✓		✓
MTL-COSA			✓	✓	✓
SHA-MTL			✓	✓	✓
Residual-U-Net			✓		✓
RMTL-Net (Proposed)			✓	✓	✓

B. PERFORMANCE EVALUATION

We employ commonly-used BUS segmentation metrics [9], [15], [17], [18], [30], [31] including sensitivity (SEN), specificity (SPE), accuracy (ACC), dice similarity coefficient (DSC), and intersection over the union of tumor (tumor IoU) to quantitatively evaluate the segmentation performance. Higher values of these metrics represent better segmentation performance. Specifically, SEN and SPE measure the ability of a model to correctly identify all tumor pixels and background pixels in BUS images, respectively; ACC reports the percent of correctly segmented tumor pixels in BUS images; both DSC and tumor IoU are positively correlated and measure the spatial overlap between the predicted segmentation result and ground truth. However, DSC tends to measure the average-case performance and tumor IoU tends to measure the worst-case performance. These metrics are calculated as follows:

$$SEN = \frac{TP}{TP + FN} \tag{10}$$

$$SPE = \frac{TN}{TN + FP} \tag{11}$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{13}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{14}$$

where TP represents true positives (*i.e.*, the number of true tumor pixels that are correctly predicted to be tumor pixels), FP represents false positives (*i.e.*, the number of true background pixels that are wrongly predicted to be tumor pixels), FN represents false negatives (*i.e.*, the number of true tumor pixels that are wrongly predicted to be background pixels), and TN represents true negatives (*i.e.*, the number of true background pixels that are correctly predicted to be background pixels). Since only two kinds of pixels (tumor and background) are involved in evaluating the segmentation performance, we consider all the pixels in the predicted background and peritumoral regions as background pixels and all the pixels in the predicted tumor region as tumor pixels.

We employ commonly-used BUS classification metrics [19], [22], [26], [30], [31] including SEN, SPE, ACC, precision (PRE), F1-score (F1), and area under receiver operating characteristic curve (AUC) to quantitatively evaluate the classification performance. Higher values of these metrics represent better classification performance. Specifically, SEN, SPE, and ACC are computed in the same manner as the segmentation metrics of the same names. However, TP, TN, FP, and FN are defined differently when evaluating classification. TP and TN respectively represent the number of BUS images that are correctly predicted as benign images (*i.e.*, a positive class) and malignant images (*i.e.*, a negative class). FP and FN respectively represent the number of BUS images that are incorrectly predicted as benign and malignant images. F1-score is the same as DSC. AUC is a summary of the receiver operating characteristic (ROC) curve, which shows the performance of a model at all classification thresholds. A higher AUC value represents better classification performance. PRE computes the ratio of correctly predicted positive samples to the total predicted positive samples. It is computed as follows:

$$PRE = \frac{TP}{TP + FP} \tag{15}$$

C. COMPETING METHODS

Table 1 briefly summarizes the task nature and enhanced features of the proposed RMTL-Net and 11 state-of-the-art (SOTA) methods. Specifically, we compare RMTL-Net with three recent single-task classification methods (e.g., VGG-16 [23], ResNet-101 [24], and DenseNet [39]), four recent single-task segmentation methods (e.g., FCN [40], PSPNet [41], Deeplab v3+ [42], and U-ResNet), and four recent MTL methods (e.g., MTL-Net, MTL-COSA [32], SHA-MTL [30], and Residual U-Net [31]). U-ResNet is a U-Net [16] with ResNet-101 as its backbone. MTL-Net passes features extracted by *Conv*5_*x* of U-ResNet into a GAP layer followed by a fully connected layer for classification. Table 1 shows that some of these compared methods employ feature enhancement strategies such as attention mechanism and skip connections to improve segmentation and classification performance.

D. RESULTS

1) MULTI-TASK LEARNING

All compared multi-task learning (MTL) methods including MTL-Net, MTL-COSA [32], SHA-MTL [30], Residual U-Net [31], and the proposed RMTL-Net compute their total loss as the weighted sum of both segmentation and classification losses. In other words, they use the hyperparameter λ in (7) to balance segmentation and classification performance during MTL. In this section, we evaluate the segmentation and classification performance of RMTL-Net under different λ values. We anticipate observing similar trends for the other compared multi-task methods since MTL-Net, MTL-COSA, and RMTL-Net use U-ResNet and others use a similar network as their backbones.

Fig. 6 compares the segmentation results of RMTL-Net under five λ values (e.g., 0.1, 0.3, 0.5, 0.7, and 0.9) on two datasets. We calculate all five segmentation metrics to evaluate the segmentation results on two datasets under five λ values. It is interesting to observe that SPE and ACC segmentation metrics yield similar values when using different λ values. Specifically, SPE oscillates between a range of 98.97% and 99.25% on dataset UDIAT and between a range of 97.75% and 98.02% on dataset BUSI. Similarly, ACC oscillates between a range of 98.20% and 98.79% on dataset UDIA and between a range of 94.96% and 96.28% on dataset BUSI. As a result, we remove SPE and ACC results in Fig. 6 to show values of segmentation metrics SEN, DSC, and IoU, where the narrow bar near the top of each bar indicates the standard deviation and the values above two selected narrow bars present the largest and smallest metric values obtained under five λ values in five-fold experiments. It demonstrates that SEN, DSC, and IoU values increase on both datasets when λ increases, except for $\lambda = 0.7$ on dataset UDIAT.

Fig. 7 compares the classification results of RMTL-Net under five λ values (e.g., 0.1, 0.3, 0.5, 0.7, and 0.9) on two datasets. We calculate all six classification metrics to evaluate the classification results on two datasets under five λ values. We re-scale AUC to the range of [0, 100] to ensure all classification values are in the same range for easy display and better understanding. Similar to Fig. 6, we use a narrow bar to indicate the standard deviation for each metric and present the largest and smallest metric values obtained under five λ values in five-fold experiments. It is clear that the overall classification performance of RMTL-Net tends to increase on both datasets when λ increases, except for the SEN values on both datasets.

RMTL-Net uses predicted probability maps to guide the classification task to learn better feature representations and achieve better classification results. As a result, accurate segmentation may lead to a better classifier. Fig. 6 and Fig. 7 confirm that both segmentation and classification accuracy tends to improve hand in hand when λ increases. Therefore, we set $\lambda = 0.9$ for RMTL-Net to ensure that more weights are given on the dominating task in the MTL framework. We also use the same setting for all MTL methods to ensure a fair comparison.

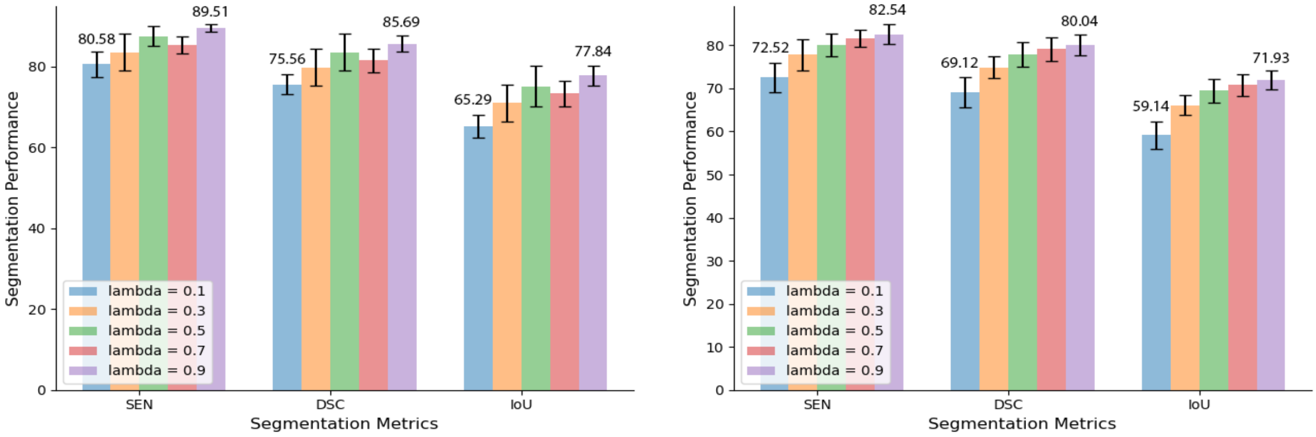
2) ABLATION STUDY OF RA MODULE

The regional attention (RA) module is a crucial component of RMTL-Net. It utilizes predicted probability maps to guide the classification network to learn multi-view features from tumor, peritumoral, and background regions in BUS images. To validate the effectiveness of the proposed RA module, we conduct a detailed ablation study by combining information from different region combinations. We list all variants of RMTL-Net below:

- **Variant 1 (MTL-Net):** None of the three regions are used.
- **Variant 2 (MTL-Net + P):** The peritumoral region is used.
- **Variant 3 (MTL-Net + T):** The tumor region is used.
- **Variant 4 (MTL-Net + B):** The background region is used.
- **Variant 5 (MTL-Net + T + P):** The tumor and peritumoral regions are used.
- **Variant 6 (MTL-Net + P + B):** The peritumoral and background regions are used.
- **Variant 7 (MTL-Net + T + B):** The tumor and background regions are used.
- **Variant 8 (proposed RMTL-Net):** The tumor, peritumoral, and background regions are used.

For Variant 1, the feature map extracted by *Conv5_x* of the encoder is directly passed to a GAP layer followed by a fully connected layer for classification. For Variants 2, 3, and 4, the weighted regional feature maps C_P, C_T, and C_B are respectively passed to a GAP layer to obtain a new feature vector G_P, G_T, and G_B of size 1×2048, which are then respectively passed to a fully connected layer for classification. For variants 5, 6, and 7, multi-channel weighted regional feature maps C_T and C_P, C_P and C_B, and C_T and C_B are respectively passed to a GAP layer and concatenated to obtain a new feature vector F of 2×2048. Their corresponding F is then filtered by a 1×1 convolution to get their associated weighted feature vector F_w of 1×2048. Lastly, their corresponding F_w is passed to a fully connected layer for classification.

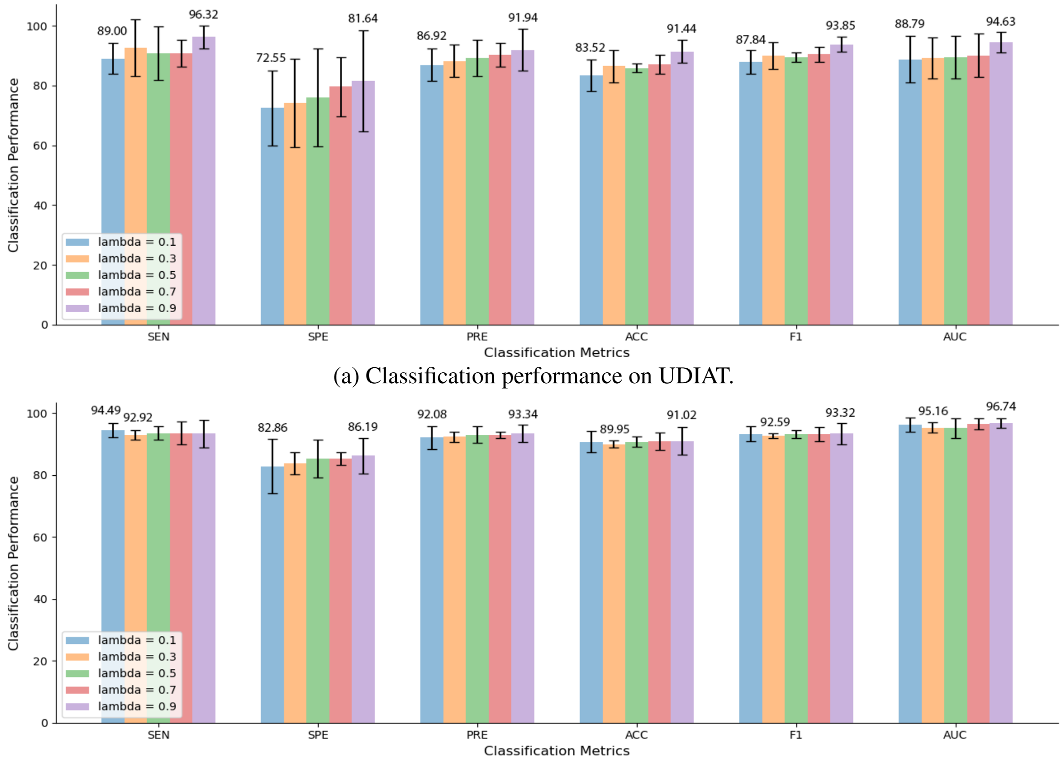
Tables 2 and 3 present the segmentation results of eight systems in the ablation study in terms of SEN, SPE, DSC, ACC, and Tumor IoU on datasets UDIAT and BUSI, respectively. Tables 4 and 5 present the classification results of eight systems in the ablation study in terms of SEN, SPE, PRE, ACC, F_1, and AUC on datasets UDIAT and BUSI, respectively. We observe the following from the results shown in these four tables: (1) Variant 1, which does not incorporate RA, achieves the worst overall segmentation performance when compared with the other seven variant systems. It achieves comparable overall classification performance as



(a) Segmentation performance on UDIAT.

(b) Segmentation performance on BUSI.

FIGURE 6. Segmentation results of RMTL-Net on two datasets under different λ values.



(a) Classification performance on UDIAT.

(b) Classification performance on BUSI.

FIGURE 7. Classification results of RMTL-Net on two datasets under different λ values.

the other seven variant systems. (2) Variant 8, which uses tumor, peritumoral, and background regions in the RA module, achieves the best overall segmentation and classification performance when compared with the other seven variant systems. (3) Comparing three variants that use a single region in the RA module, variant 4 involving the background region achieves the best overall performance. Variant 3 involving the tumor region achieves the second-best performance. Variant 2 involving peritumoral regions achieves the worst performance. (4) Comparing three variants that use two of the three regions in the RA module, variant 7 involving tumor and background regions achieves the best performance. Variant 5 involving tumor and peritumoral regions achieves the worst performance.

TABLE 2. Segmentation performance (Mean ± SD) of ablation study on dataset UDIAT.

Variants	SEN	SPE	DSC	ACC	Tumor IoU
MTL-Net	84.28 ± 5.05	99.25 ± 0.14	80.95 ± 5.00	98.65 ± 0.38	72.73 ± 5.49
MTL-Net + P	86.52 ± 1.06	99.20 ± 0.23	82.73 ± 2.52	98.68 ± 0.30	74.92 ± 2.86
MTL-Net + T	87.15 ± 2.74	99.23 ± 0.22	84.03 ± 2.91	98.71 ± 0.27	75.84 ± 3.24
MTL-Net + B	88.43 ± 2.99	99.09 ± 0.31	84.48 ± 3.23	98.65 ± 0.18	76.06 ± 3.57
MTL-Net + T + P	87.97 ± 3.07	99.22 ± 0.24	84.21 ± 4.62	98.69 ± 0.29	76.19 ± 5.17
MTL-Net + P + B	88.68 ± 2.29	99.17 ± 0.23	84.61 ± 2.84	98.68 ± 0.21	76.25 ± 3.16
MTL-Net + T + B	87.87 ± 3.76	99.24 ± 0.34	85.09 ± 2.33	98.72 ± 0.29	76.88 ± 2.51
RMTL-Net	**89.51 ± 0.91**	**99.25 ± 0.19**	**85.69 ± 2.00**	**98.79 ± 0.24**	**77.84 ± 2.45**

TABLE 3. Segmentation performance (Mean ± SD) of ablation study on dataset BUSI.

Variants	SEN	SPE	DSC	ACC	Tumor IoU
MTL-Net	78.91 ± 2.22	98.30 ± 0.25	77.76 ± 3.11	96.18 ± 0.15	69.33 ± 2.89
MTL-Net + P	81.02 ± 2.08	98.05 ± 0.58	79.46 ± 2.84	96.25 ± 0.44	71.31 ± 2.82
MTL-Net + T	81.63 ± 1.92	98.08 ± 0.51	79.55 ± 2.04	96.28 ± 0.27	71.31 ± 1.93
MTL-Net + B	81.84 ± 2.71	98.02 ± 0.43	79.53 ± 2.39	96.25 ± 0.21	71.44 ± 2.20
MTL-Net + T + P	81.30 ± 2.43	**98.08 ± 0.32**	79.64 ± 2.13	96.26 ± 0.20	71.47 ± 1.97
MTL-Net + P + B	81.98 ± 2.25	97.99 ± 0.44	79.79 ± 2.85	96.30 ± 0.24	71.72 ± 2.83
MTL-Net + T + B	81.84 ± 3.26	97.96 ± 0.40	79.91 ± 2.74	96.32 ± 0.30	71.82 ± 2.54
RMTL-Net	**82.54 ± 2.31**	98.00 ± 0.30	**80.04 ± 2.47**	**96.41 ± 0.27**	**71.93 ± 2.15**

TABLE 4. Classification performance (Mean ± SD) of ablation study on dataset UDIAT.

Variants	SEN	SPE	PRE	ACC	F_1	AUC
MTL-Net	89.96 ± 7.43	74.00 ± 13.63	87.99 ± 5.08	84.69 ± 2.85	88.65 ± 2.39	90.82 ± 6.46
MTL-Net + P	92.64 ± 4.12	68.91 ± 16.22	86.09 ± 6.62	84.73 ± 5.52	89.09 ± 3.79	88.26 ± 9.38
MTL-Net + T	91.73 ± 6.76	72.55 ± 17.96	87.66 ± 6.65	85.34 ± 4.82	89.34 ± 3.37	89.54 ± 8.75
MTL-Net + B	92.68 ± 6.87	70.36 ± 20.71	87.24 ± 7.79	85.30 ± 4.40	89.45 ± 2.81	89.90 ± 7.17
MTL-Net + T + P	93.55 ± 7.65	74.18 ± 07.20	88.10 ± 2.55	87.12 ± 3.94	90.56 ± 3.30	91.28 ± 4.65
MTL-Net + P + B	93.55 ± 6.15	77.82 ± 16.43	90.09 ± 6.65	88.33 ± 4.00	91.50 ± 2.84	91.87 ± 7.09
MTL-Net + T + B	94.50 ± 2.01	79.64 ± 11.81	90.62 ± 4.96	89.58 ± 3.41	92.43 ± 2.27	93.02 ± 6.50
RMTL-Net	**96.32 ± 3.82**	**81.64 ± 16.89**	**91.94 ± 6.97**	**91.44 ± 3.90**	**93.85 ± 2.58**	**94.63 ± 3.44**

TABLE 5. Classification performance (Mean ± SD) of ablation study on dataset BUSI.

Variants	SEN	SPE	PRE	ACC	F_1	AUC
MTL-Net	93.36 ± 2.37	84.67 ± 6.65	92.61 ± 3.18	90.18 ± 3.25	93.07 ± 2.41	96.20 ± 2.08
MTL-Net + P	92.21 ± 1.54	80.48 ± 6.16	90.79 ± 2.79	88.40 ± 2.94	91.49 ± 2.11	93.53 ± 2.49
MTL-Net + T	92.45 ± 0.67	82.38 ± 6.86	91.68 ± 2.99	89.17 ± 2.23	92.04 ± 1.52	94.31 ± 1.15
MTL-Net + B	91.30 ± 1.06	85.24 ± 3.91	92.81 ± 1.76	89.33 ± 1.42	92.04 ± 1.03	94.96 ± 1.56
MTL-Net + T + P	92.41 ± 4.98	83.81 ± 4.88	92.28 ± 2.13	89.63 ± 3.44	92.28 ± 2.67	95.87 ± 1.98
MTL-Net + P + B	91.28 ± 3.89	**88.10 ± 4.76**	94.13 ± 2.18	90.25 ± 3.01	92.64 ± 2.36	95.89 ± 1.63
MTL-Net + T + B	92.19 ± 3.51	88.10 ± 6.07	**94.21 ± 2.76**	90.87 ± 3.10	93.15 ± 2.33	95.92 ± 1.22
RMTL-Net	**93.34 ± 4.42**	86.19 ± 5.68	93.34 ± 2.80	**91.02 ± 4.42**	**93.32 ± 3.35**	**96.74 ± 1.48**

For most BUS images, we observe that the background region has the biggest size and the peritumoral region has the smallest size. As a result, we assume that the larger the region, the more information it can provide for both segmentation and classification tasks. The experimental results shown in Tables 2, 3, 4, and 5 seem to support this assumption. First, either background, tumor, or peritumoral region plays an important role in the segmentation task since variants 2, 3, and 4 outperform variant 1 without using the RA module in all segmentation metrics. Second, the background region of C_5 provides the most valuable information for both segmentation and classification tasks since variant 4 achieves the best performance among variants involving one region in the RA module. The tumor region of C_5 provides the second most valuable information followed by the peritumoral region. Third, variants involving two regions in the RA module outperform variants involving one region in the RA module since a combined larger region provides more information to facilitate the learning process. Fourth, the variant involving three regions in the RA module achieves the best performance. Fifth, the weighted feature vector F_W, which obtains valuable information from multiple regions, better represents BUS images than C_5 without using the RA module.

3) COMPARISON WITH COMPETING METHODS

We implement all compared methods except for SHA-MTL and Residual U-Net and conduct experiments using the same parameters to ensure a fair comparison. The authors of SHA-MTL and Residual U-Net did not provide sufficient

TABLE 6. Segmentation performance (Mean ± SD) of all compared methods on Dataset UDIAT.

Methods	SEN	SPE	DSC	ACC	Tumor IoU
FCN	78.78 ± 6.54	99.08 ± 0.24	76.90 ± 4.69	98.37 ± 0.27	66.49 ± 4.93
PSPNet	83.19 ± 4.60	**99.44 ± 0.16**	83.08 ± 4.58	98.74 ± 0.31	74.59 ± 5.30
Deeplabv3+	85.15 ± 3.54	99.34 ± 0.11	83.53 ± 4.28	98.77 ± 0.32	74.60 ± 4.88
UResNet	85.38 ± 3.84	99.20 ± 0.10	81.38 ± 4.86	98.57 ± 0.35	73.08 ± 5.28
MTL-Net	84.28 ± 5.05	99.25 ± 0.14	80.95 ± 5.00	98.65 ± 0.38	72.73 ± 5.49
MTL-COSA	86.97 ± 2.76	99.27 ± 0.25	84.07 ± 3.25	98.77 ± 0.26	76.05 ± 3.71
RMTL-Net	**89.51 ± 0.91**	99.25 ± 0.19	**85.69 ± 2.00**	**98.79 ± 0.24**	**77.84 ± 2.45**

TABLE 7. Segmentation performance (Mean ± SD) of all compared methods on Dataset BUSI.

Methods	SEN	SPE	DSC	ACC	Tumor IoU
FCN	78.40 ± 3.33	98.02 ± 0.20	76.87 ± 2.88	96.08 ± 0.12	67.05 ± 2.88
PSPNet	78.28 ± 2.36	**98.34 ± 0.28**	78.48 ± 2.73	96.31 ± 0.46	70.11 ± 2.51
Deeplabv3+	80.71 ± 2.40	98.15 ± 0.42	79.14 ± 2.84	96.37 ± 0.31	70.51 ± 3.04
UResNet	79.65 ± 2.16	98.05 ± 0.40	77.98 ± 2.91	96.06 ± 0.18	69.65 ± 2.92
MTL-Net	78.91 ± 2.22	98.30 ± 0.25	77.76 ± 3.11	96.18 ± 0.15	69.33 ± 2.89
MTL-COSA	79.31 ± 2.48	98.31 ± 0.11	78.90 ± 2.03	96.35 ± 0.16	70.65 ± 2.01
SHA-MTL	81.21 ± 4.83	97.36 ± 1.93	81.42 ± 1.53	95.56 ± 1.08	—
Residual-U-Net	**86.13**	—	**84.81**	88.08	—
RMTL-Net	82.54 ± 2.31	98.00 ± 0.30	80.04 ± 2.47	**96.41 ± 0.27**	**71.93 ± 2.15**

TABLE 8. Classification performance (Mean ± SD) of all compared methods on Dataset UDIAT.

Methods	SEN	SPE	PRE	ACC	F_1	AUC
VGG16	85.37 ± 4.86	63.09 ± 13.99	82.67 ± 5.49	77.99 ± 4.66	83.85 ± 3.28	86.24 ± 5.03
DenseNet	89.00 ± 2.42	66.73 ± 13.57	84.57 ± 5.62	81.62 ± 5.95	86.69 ± 4.02	86.93 ± 7.49
ResNet	91.69 ± 7.59	64.91 ± 14.65	84.52 ± 4.61	82.80 ± 1.93	87.65 ± 1.63	90.52 ± 5.08
MTL-Net	89.96 ± 7.43	74.00 ± 13.63	87.99 ± 5.08	84.69 ± 2.85	88.65 ± 2.39	90.82 ± 6.46
MTL-COSA	92.64 ± 7.63	75.82 ± 14.02	89.11 ± 5.41	87.08 ± 2.79	90.51 ± 2.29	93.61 ± 4.55
RMTL-Net	**96.32 ± 3.82**	**81.64 ± 16.89**	**91.94 ± 6.97**	**91.44 ± 3.90**	**93.85 ± 2.58**	**94.63 ± 3.44**

TABLE 9. Classification performance (Mean ± SD) of all compared methods on Dataset BUSI.

Methods	SEN	SPE	PRE	ACC	F_1	AUC
VGG16	93.80 ± 3.32	77.63 ± 3.61	89.71 ± 1.66	88.55 ± 2.86	91.69 ± 2.15	94.59 ± 2.67
DenseNet	94.26 ± 3.73	83.33 ± 5.32	92.23 ± 2.24	90.71 ± 2.42	93.18 ± 1.85	95.66 ± 2.123
ResNet	93.81 ± 2.41	80.95 ± 9.37	91.22 ± 3.93	89.63 ± 3.52	92.45 ± 2.49	95.74 ± 2.201
MTL-Net	93.36 ± 2.37	84.67 ± 6.65	92.61 ± 3.18	90.18 ± 3.25	93.07 ± 2.41	96.20 ± 2.08
MTL-COSA	93.57 ± 4.04	87.14 ± 3.19	93.81 ± 1.52	91.49 ± 3.02	93.66 ± 2.36	96.77 ± 1.57
SHA-MTL	96.13 ± 2.33	89.93 ± 5.59	—	94.12 ± 2.45	92.93 ± 3.31	96.28
Residual-U-Net	**98.79**	**94.65**	**98.12**	**97.86**	**98.45**	**99.99**
RMTL-Net	93.34 ± 4.42	86.19 ± 5.68	93.34 ± 2.80	91.02 ± 3.42	93.32 ± 3.35	96.74 ± 1.48

details on their methods and did not publish their code either. As a result, we directly use their reported segmentation and classification results on dataset BUSI in our comparison. We use the symbol of "—" to represent a missing result since they did not report their results on each metric. Both methods did not provide any results on dataset UDIAT. So they are not included when comparing segmentation and classification results on dataset UDIAT.

Table 6 summarizes the segmentation results of RMTL-Net and six methods in terms of five metrics on the dataset UDIAT. Among four single-task segmentation methods, Deeplabv3+ achieves the best overall segmentation performance with the highest values of SEN, DSC, ACC, and tumor IoU. PSPNet achieves the second-best overall segmentation performance, followed by UResNet and FCN. Among three MTL methods, the proposed RMTL-Net achieves the best segmentation performance in all metrics except for SPE. It improves the second-best method MTL-COSA by 2.54%, 1.62%, 0.02%, and 1.79% for SEN, DSC, ACC, and tumor IoU, respectively.

Table 7 summarizes the segmentation results of RMTL-Net and eight methods in terms of five metrics on the dataset BUSI. Single-task segmentation methods exhibit similar performance trends on dataset BUSI as on dataset UDIAT. The three MTL methods including MTL-Net, MTL-COSA, and RMTL-Net exhibit similar performance trends on dataset BUSI as on dataset UDIAT. The proposed RMTL-Net achieves the best overall segmentation performance and improves the second-best method MTL-COSA by 3.23%, 1.14%, 0.06%, and 1.28% for SEN, DSC, ACC, and tumor

IoU, respectively. Two MTL methods residual-U-Net and SHA-MTL seem to lack credibility since residual-U-Net did not report its standard deviation values for five runs on all evaluation metrics and SHA-MTL reported different values for two equivalent metrics DSC and F_1 without giving any explanation. In addition, residual-U-Net seems to have an overfitting issue since its AUC values of five runs are 0.98, 1, 0.99, 0.97, and 1. As a result, we do not include these two methods here for comparison and list their results in tables for completeness.

Table 8 summarizes the classification results of RMTL-Net and five methods in terms of six metrics on the dataset UDIAT. Among three single-task classification methods, ResNet achieves the best overall classification performance with the highest values of SEN, ACC, F_1, and AUC. DenseNet achieves the second-best overall classification performance, followed by VGG-16. Among three MTL methods, the proposed RMTL-Net achieves the best classification performance in all metrics. It improves the second-best method MTL-COSA by 3.68%, 5.82%, 2.83%, 4.36%, 3.34%, and 1.02% for SEN, SPE, PRE, ACC, F_1, and AUC, respectively.

Table 9 summarizes the classification results of RMTL-Net and seven methods in terms of six metrics on the dataset BUSI. Single-task classification methods exhibit similar performance trends on dataset BUSI as on dataset UDIAT. The three MTL methods including MTL-Net, MTL-COSA, and RMTL-Net exhibit similar performance trends on dataset BUSI as on dataset UDIAT. The proposed RMTL-Net achieves the second-best overall classification performance and MTL-COSA outperforms RMTL-Net by a little bit in all metrics. Due to the lack of credibility, residual U-Net and SHA-MTL are not included here for comparison and are listed in tables for completeness.

Tables 6, 7, 8, and 9 demonstrate that RMTL-Net achieves the best overall segmentation and classification results on both datasets. It incorporates the RA module to improve MTL-COSA by learning the importance of three predicted probability maps representing tumor, peritumoral, and background regions. MTL-COSA incorporates self-attention to improve MTL-Net by learning the importance of three regions constructed from the predicted binary segmentation mask. MTL-Net decreases the values of three segmentation metrics including SEN, DSC, and tumor IoU (*i.e.*, decreasing the segmentation performance) when compared with the best single-task segmentation method UResNet. This decrease in performance is caused by reduced segmentation weight, which was added to the classification task. Therefore, less weight is employed in training to reduce segmentation errors. However, incorporating attention to MTL-Net addresses this issue to achieve comparable or better segmentation results than UResNet and achieve comparable or better classification results than ResNet.

Table 10 lists the number of trainable parameters of all compared methods. It shows that MTL-Net increases trainable parameters of UResNet by 0.004% via adding a light-weight classification task. This simple addition utilizes segmentation results to guide the classification task, which leads to comparable segmentation results as single-task segmentation methods and better classification results than single-task classification methods. Table 10 also shows that both MTL-COSA and RMTL-Net increase the different amounts of trainable parameters in networks such as ResNet and UResNet by adding attention modules to learn important regions. RMTL-Net has a simpler attention mechanism than MTL-COSA and therefore leads to a reduction of 16.8% trainable parameters when compared with MTL-COSA. It also outperforms MTL-COSA in segmentation on both datasets and in classification on dataset UDIAT.

TABLE 10. Summary of the number of trainable parameters of all compared methods.

Methods	Number of Trainable Parameters
VGG-16	138,357,544
DenseNet	52,166,124
ResNet-101	44,677,034
FCN	134,270,278
PSPNet	70,295,620
Deeplabv3+	59,339,426
UResNet	93,500,842
MTL-Net	93,504,940
MTL-COSA	109,241,266
SHA-MTL	—
Residual U-Net	—
RMTL-Net	93,506,099

Fig. 8 presents the segmentation results of RMTL-Net and six compared methods on four representative BUS images: two in Dataset UDIAT as shown in the top two rows and two in Dataset BUSI as shown in the bottom two rows. The UDIAT BUS image on the first row contains a small tumor with an irregular boundary. All methods fail to predict a clear and accurate tumor boundary. FCN and MTL-Net completely fail to detect the tumor region. UResNet mistakenly segments a tumor-like region as a tumor. PSPNet, Deeplabv3+, and MTL-COSA segment a tumor partially overlapping with the ground truth. They achieve a tumor IoU value of 61.89%, 60.40%, and 69.23%, respectively. RMTL-Net yields a more accurate segmentation result with the highest IoU value of 76.65%. The UDIAT BUS image on the second row contains a small tumor. FCN, UResNet, and MTL-COSA segment a much bigger tumor region than the ground truth and yield a low tumor IoU value of 35.00%, 31.27%, and 40.98%, respectively. PSPNet, Deeplabv3+, and MTL-Net achieve better segmentation results with tumor IoU values of 59.42%, 51.01%, and 63.33%, respectively. RMTL-Net achieves the best segmentation result and the highest tumor IoU value of 81.60%. The BUSI BUS image on the third row contains a small tumor and a big tumor-like region. All methods except for RMTL-Net mistakenly segment the tumor-like region as tumor region and therefore yield low tumor IoU values less than 55.00%. RMTL-Net segments the correct tumor region and achieves large values close to 1 in almost all segmentation metrics (*i.e.*, 99.90% for SPE, 94.95% for DSC, 99.84% for ACC, and 90.39% for IoU). The BUSI BUS image on the last

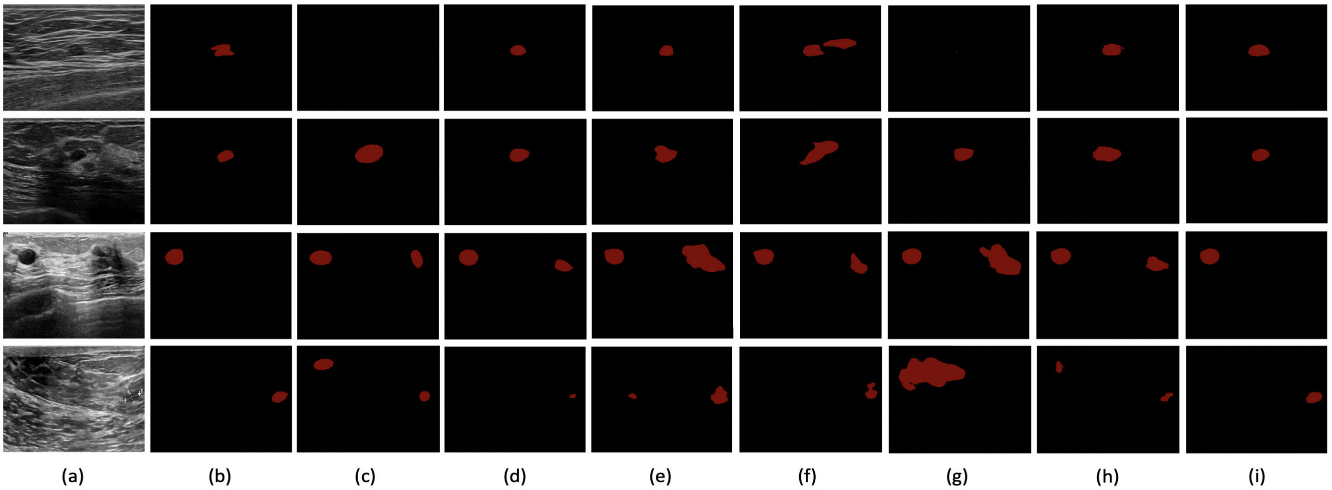


FIGURE 8. Illustration of segmentation results. (a) BUS images; (b) Ground truth; Segmentation results obtained by (c) FCN; (d) PSPNet; (e) Deeplabv3+; (f) UResNet; (g) MTL-Net; (h) MTL-COSA; (i) RMTL-Net.

row contains a small tumor with a blurry boundary. This small tumor locates on the right side towards the middle row. MTL-Net segments a completely wrong tumor region and obtains the lowest IoU value of 0.00%. FCN, Deeplabv3+, and MTL-COSA segment a partial tumor region and mistakenly segment another tumor-like region. Their tumor IoU values are 23.56%, 52.04%, and 32.67%, respectively. PSPNet and UResNet segment a partial tumor region with a low IoU value of 17.49%, and 32.58%, respectively. RMTL-Net segments the most accurate tumor region and achieves the largest values on all five segmentation metrics (94.98% for SEN, 99.91% for SPE, 92.37% for DSC, 99.86% for ACC, and 85.82% for IoU).

IV. DISCUSSIONS

A. ADVANTAGES AND POTENTIAL USEFULNESS

In this paper, we propose a novel MTL framework with a RA module for BUS image segmentation and classification. In general, advantages and potential usefulness of RMTL-Net can be summarized as follows:

First, RMTL-Net simultaneously performs segmentation and classification by utilizing predicted probability maps to guide the classification task to focus on regions of different importance. Single-task segmentation and classification methods have been well-studied in the BUS research community. However, simultaneous segmentation and classification is more practical and appealing than single segmentation and classification tasks, as it provides both tumor boundaries as well as tumor category. As a result, MTL in BUS image segmentation and classification is a promising direction that is worthy of more exploration. Our study clearly shows that adding a light-weight classification branch on most existing segmentation methods, at least U-Net-based ones (*e.g.*, UResNet), increases very few parameters but yields both good segmentation and classification results.

Second, RMTL-Net incorporates a three-region-based attention module (*i.e.*, RA module) to automatically assign appropriate weights to tumor, peritumoral, and background regions during the training procedure. The learned weights help to find regions of importance for better feature representations and therefore improve both the segmentation and classification performance of an MTL method. The RA module aligns well with doctors' clinical perspectives on the importance of tumor, peritumoral, and background regions. The proposed RA module can be easily applied to any existing MTL methods to incorporate prior medical knowledge into the attention model to improve the performance of multiple tasks.

B. LIMITATION AND FUTURE WORK

The proposed RMTL-Net has some limits. First, a preprocessing step is needed to generate pseudo ground truths of peritumoral and background regions, which are indispensable in the training procedure to help the network to learn and produce three regions in any test images. Second, more comparison between the proposed RA module and other traditional spatial or channel attention modules needs to be further conducted to prove the effectiveness of the RA module.

Due to the limited number of public BUS images, we do not have a separate testing set and use five-fold cross-validation to have every BUS image in the dataset validated and tested. As a result, more BUS images need to be collected to generate a large testing set to thoroughly test the generalization ability of RMLT-Net and other competing methods on new BUS images. These experiments are needed to prove the superiority of RMTL-Net without overfitting concerns.

In the future, we will test our RMTL-Net on larger nuclei segmentation and classification datasets and explore more strategies to improve its generalization ability. We will also compare the proposed RA module with more recent spatial and channel attention modules to not only validate its effectiveness but also find a new perspective to improve it.

V. CONCLUSION

In this study, we propose a regional-attentive multi-task learning framework (RMTL-Net) for simultaneous BUS image segmentation and classification. The proposed RMTL-Net adopts ResNet-101 as its backbone to extract features and utilizes a regional attention (RA) module to automatically learn weighted category-sensitive information from the tumor, peritumoral, and background regions in BUS images to more accurately represent each BUS image for better segmentation and classification performance. We conduct extensive five-fold cross-validation experiments on two public BUS datasets DIAT and BUSI. Extensive experiments show that RMTL-Net outperforms recent state-of-the-art single-task segmentation methods, single-task classification methods, and most MTL methods on two datasets.

Our proposed RMTL-Net sheds light on the new research direction toward multi-task learning (MTL) in general and simultaneous segmentation and classification for BUS images in particular. To this end, we can easily convert any existing segmentation network architecture to its counterpart MTL network architecture at a low cost by adding a classification branch to achieve comparable segmentation results and better classification results. Adding a RA module to incorporate prior medical knowledge regarding the importance of tumor, peritumoral, and background regions in BUS images can help to learn a better feature representation for better segmentation and classification results. The proposed RA module can be easily applied to any existing MTL methods and be easily modified based on different prior knowledge.

REFERENCES

[1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA, Cancer J. Clinicians*, vol. 71, no. 3, pp. 209–249, May 2021.

[2] M. M. Rivera-Franco and E. Leon-Rodriguez, "Delays in breast cancer detection and treatment in developing countries," *Breast Cancer, Basic Clin. Res.*, vol. 12, Jan. 2018, Art. no. 1178223417752677.

[3] C. A. Anyigba, G. A. Awandare, and L. Paemka, "Breast cancer in sub-Saharan Africa: The current state and uncertain future," *Exp. Biol. Med.*, vol. 246, no. 12, pp. 1377–1387, Jun. 2021.

[4] O. Ginsburg, C. H. Yip, A. Brooks, A. Cabanes, M. Caleffi, J. A. D. Yataco, B. Gyawali, V. McCormack, M. M. de Anderson, R. Mehrotra, and A. Mohar, "Breast cancer early detection: A phased approach to implementation," *Cancer*, vol. 126, pp. 2379–2393, May 2020.

[5] R. Sood, A. F. Rositch, D. Shakoor, E. Ambinder, K.-L. Pool, E. Pollack, D. J. Mollura, L. A. Mullen, and S. C. Harvey, "Ultrasound for breast cancer detection globally: A systematic review and meta-analysis," *J. Global Oncol.*, vol. 5, pp. 1–17, Dec. 2019.

[6] Q. Huang, F. Zhang, and X. Li, "Machine learning in ultrasound computer-aided diagnostic systems: A survey," *BioMed Res. Int.*, vol. 2018, pp. 1–10, Mar. 2018.

[7] S. Yang, X. Gao, L. Liu, R. Shu, J. Yan, G. Zhang, Y. Xiao, Y. Ju, N. Zhao, and H. Song, "Performance and reading time of automated breast U.S. with or without computer-aided detection," *Radiology*, vol. 292, no. 3, pp. 540–549, Sep. 2019.

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

[9] K. Wang, S. Liang, S. Zhong, Q. Feng, Z. Ning, and Y. Zhang, "Breast ultrasound image segmentation: A coarse-to-fine fusion convolutional neural network," *Med. Phys.*, vol. 48, no. 8, pp. 4262–4278, Aug. 2021.

[10] G. Pons, J. Martí, R. Martí, S. Ganau, and J. A. Noble, "Breast-lesion segmentation combining B-mode and elastography ultrasound," *Ultrason. Imag.*, vol. 38, no. 3, pp. 209–224, May 2016.

[11] Z. Zhou, W. Wu, S. Wu, P.-H. Tsui, C.-C. Lin, L. Zhang, and T. Wang, "Semi-automatic breast ultrasound image segmentation based on mean shift and graph cuts," *Ultrason. Imag.*, vol. 36, no. 4, pp. 256–276, Oct. 2014.

[12] H.-C. Kuo, M. L. Giger, I. Reiser, K. Drukker, J. M. Boone, K. K. Lindfors, K. Yang, A. Edwards, and C. A. Sennett, "Segmentation of breast masses on dedicated breast computed tomography and three-dimensional breast ultrasound images," *J. Med. Imag.*, vol. 1, no. 1, Apr. 2014, Art. no. 014501.

[13] Z. Hao, Q. Wang, X. Wang, J. B. Kim, Y. Hwang, B. H. Cho, P. Guo, and W. K. Lee, "Learning a structured graphical model with boosted top-down features for ultrasound image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2013, pp. 227–234.

[14] K. Huang, Y. Zhang, H. D. Cheng, P. Xing, and B. Zhang, "Semantic segmentation of breast ultrasound image with fuzzy deep learning network and breast anatomy constraints," *Neurocomputing*, vol. 450, pp. 319–335, Aug. 2021.

[15] M. Xian, Y. Zhang, H. D. Cheng, F. Xu, B. Zhang, and J. Ding, "Automatic breast ultrasound image segmentation: A survey," *Pattern Recognit.*, vol. 79, pp. 340–355, Jul. 2018.

[16] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.

[17] M. Amiri, R. Brooks, B. Behboodi, and H. Rivaz, "Two-stage ultrasound image segmentation using U-Net and test time augmentation," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 15, no. 6, pp. 981–988, Apr. 2020.

[18] M. Xu, K. Huang, Q. Chen, and X. Qi, "MSSA-Net: Multi-scale self-attention network for breast ultrasound image segmentation," in *Proc. IEEE 18th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2021, pp. 827–831.

[19] Y. Liu, L. Ren, X. Cao, and Y. Tong, "Breast tumors recognition based on edge feature extraction using support vector machine," *Biomed. Signal Process. Control*, vol. 58, Apr. 2020, Art. no. 101825.

[20] J. Ding, H. D. Cheng, M. Xian, Y. Zhang, and F. Xu, "Local-weighted Citation-kNN algorithm for breast ultrasound image classification," *Optik*, vol. 126, no. 24, pp. 5188–5193, 2015.

[21] N. Uniyal, H. Eskandari, P. Abolmaesumi, S. Sojoudi, P. Gordon, L. Warren, R. N. Rohling, S. E. Salcudean, and M. Moradi, "Ultrasound RF time series for classification of breast lesions," *IEEE Trans. Med. Imag.*, vol. 34, no. 2, pp. 652–661, Feb. 2015.

[22] K. Huang, M. Xu, and X. Qi, "NGMMs: Neutrosophic Gaussian mixture models for breast ultrasound image classification," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Nov. 2021, pp. 3943–3947.

[23] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[25] W.-X. Liao, P. He, J. Hao, X.-Y. Wang, R.-L. Yang, D. An, and L.-G. Cui, "Automatic identification of breast ultrasound image based on supervised block-based region segmentation algorithm and features combination migration deep learning model," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 4, pp. 984–993, Apr. 2020.

[26] W. Cui, Y. Peng, G. Yuan, W. Cao, Y. Cao, Z. Lu, X. Ni, Z. Yan, and J. Zheng, "FMRNet: A fused network of multiple tumoral regions for breast tumor classification with ultrasound images," *Med. Phys.*, vol. 49, no. 1, pp. 144–157, Jan. 2022.

[27] S. Gokhale, "Ultrasound characterization of breast masses," *Indian J. Radiol. Imag.*, vol. 19, no. 3, p. 242, 2009.

[28] W. Yang, S. Zhang, Y. Chen, W. Li, and Y. Chen, "Measuring shape complexity of breast lesions on ultrasound images," *Proc. SPIE*, vol. 6920, pp. 169–178, Mar. 2008.

[29] Y. Zhou, H. Chen, Y. Li, Q. Liu, X. Xu, S. Wang, P.-T. Yap, and D. Shen, "Multi-task learning for segmentation and classification of tumors in 3D automated breast ultrasound images," *Med. Image Anal.*, vol. 70, May 2021, Art. no. 101918.

[30] G. Zhang, K. Zhao, Y. Hong, X. Qiu, K. Zhang, and B. Wei, "SHA-MTL: Soft and hard attention multi-task learning for automated breast cancer ultrasound image segmentation and classification," *Int. J. Comput. Assist. Radiol. Surgery*, vol. 16, no. 10, pp. 1719–1725, Oct. 2021.

[31] J. Chowdary, P. Yogarajah, P. Chaurasia, and V. Guruviah, "A multi-task learning framework for automated segmentation and classification of breast tumors from ultrasound images," *Ultrason. Imag.*, vol. 44, no. 1, pp. 3–12, Jan. 2022.

[32] M. Xu, K. Huang, and X. Qi, "Multi-task learning with context-oriented self-attention for breast ultrasound image classification and segmentation," in *Proc. IEEE 19th Int. Symp. Biomed. Imag. (ISBI)*, Mar. 2022, pp. 1–5.

[33] Q. Sun, X. Lin, Y. Zhao, L. Li, K. Yan, D. Liang, D. Sun, and Z.-C. Li, "Deep learning vs. radiomics for predicting axillary lymph node metastasis of breast cancer using ultrasound images: Don't forget the peritumoral region," *Frontiers Oncol.*, vol. 10, p. 53, Jan. 2020.

[34] Y.-W. Lee, C.-S. Huang, C.-C. Shih, and R.-F. Chang, "Axillary lymph node metastasis status prediction of early-stage breast cancer using convolutional neural networks," *Comput. Biol. Med.*, vol. 130, Mar. 2021, Art. no. 104206.

[35] T. Liu, Q. Guo, C. Lian, X. Ren, S. Liang, J. Yu, L. Niu, W. Sun, and D. Shen, "Automated detection and classification of thyroid nodules in ultrasound images using clinical-knowledge-guided convolutional neural networks," *Med. Image Anal.*, vol. 58, Dec. 2019, Art. no. 101555.

[36] M. H. Yap, G. Pons, J. Marti, S. Ganau, M. Sentis, R. Zwiggelaar, A. K. Davison, and R. Marti, "Automated breast ultrasound lesions detection using convolutional neural networks," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 4, pp. 1218–1226, Jul. 2017.

[37] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data Brief*, vol. 28, Feb. 2020, Art. no. 104863.

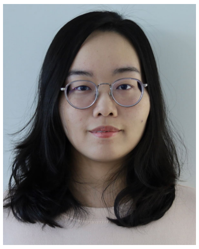
[38] K. Drukker, M. L. Giger, and E. B. Mendelson, "Computerized analysis of shadowing on breast ultrasound for improved lesion detection," *Med. Phys.*, vol. 30, no. 7, pp. 1833–1842, Jun. 2003.

[39] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

[40] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[41] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2881–2890.

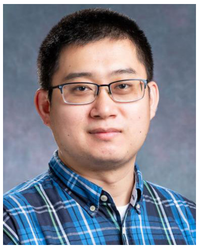
[42] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder–decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 801–818.



MENG XU (Graduate Student Member, IEEE) received the B.S. degree in management information systems from the Tianjin University of Technology, Tianjin, China, in 2017. She is currently pursuing the Ph.D. degree in computer science with Utah State University, Logan, UT, USA.

From 2018 to 2022, she was a Graduate Student Instructor with the Department of Computer Science, Utah State University. Her research interests include computer vision, deep learning, and medical image analysis.

Ms. Xu's awards and honors include the Presidential Doctoral Research Fellowship at Utah State University and the Outstanding Graduate Award at the Tianjin University of Technology.



KUAN HUANG (Member, IEEE) received the B.Eng. degree in electrical engineering and automation from the Harbin Institute of Technology, Harbin, China, in 2016, and the Ph.D. degree in computer science from Utah State University, Logan, UT, USA, in 2021.

From 2020 to 2021, he was a Postdoctoral Researcher with the Lester and Sue Smith Breast Center, Baylor College of Medicine, Houston, TX, USA. Since 2022, he has been an Assistant Professor with the Department of Computer Science and Technology, Kean University, Union, NJ, USA. His research interests include computer vision, deep learning, pattern recognition, medical image analysis, and fuzzy logic.

Dr. Huang was a recipient of the Best Dissertation Award at Utah State University, in 2022.

XIAOJUN QI (Senior Member, IEEE) received the B.S. degree in computer science from Donghua University, in 1993, the M.S. degree in computer science from the Shenyang Institute of Automation, Chinese Academy of Sciences, in 1996, and the Ph.D. degree in computer science from Louisiana State University, in 2001.

In 2002, she joined the Department of Computer Science, Utah State University, as a Tenure-Track Assistant Professor. In 2008, she received tenure and was promoted to an Associate Professor. In 2015, she was promoted to a Professor. She has expertise in artificial intelligence focusing on machine learning and computer vision. She has established and sustained a nationally recognized research program to solve challenging research problems. She has acquired competitive funding from reputable agencies and has published more than 110 high-quality manuscripts. She has worked as a PI on more than 20 research projects and has supervised 83 students. She has also gained her professional leadership experience by serving on technical program committees, NSF panels, and as a reviewer.