Computers in Biology and Medicine 173 (2024) 108319

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Multi-task learning for segmentation and classification of breast tumors from ultrasound images

Qiqi He [a,b], Qiuju Yang [a,*], Hang Su [a], Yixuan Wang [a]

[a] School of Physics and Information Technology, Shaanxi Normal University, Xi'an, China
[b] School of Life Science and Technology, Xidian University, Xi'an, China

ARTICLE INFO

Keywords:
Multi-task learning
Breast tumor
Ultrasound images
Segmentation
Classification

ABSTRACT

Segmentation and classification of breast tumors are critical components of breast ultrasound (BUS) computer-aided diagnosis (CAD), which significantly improves the diagnostic accuracy of breast cancer. However, the characteristics of tumor regions in BUS images, such as non-uniform intensity distributions, ambiguous or missing boundaries, and varying tumor shapes and sizes, pose significant challenges to automated segmentation and classification solutions. Many previous studies have proposed multi-task learning methods to jointly tackle tumor segmentation and classification by sharing the features extracted by the encoder. Unfortunately, this often introduces redundant or misleading information, which hinders effective feature exploitation and adversely affects performance. To address this issue, we present ACSNet, a novel multi-task learning network designed to optimize tumor segmentation and classification in BUS images. The segmentation network incorporates a novel gate unit to allow optimal transfer of valuable contextual information from the encoder to the decoder. In addition, we develop the Deformable Spatial Attention Module (DSAModule) to improve segmentation accuracy by overcoming the limitations of conventional convolution in dealing with morphological variations of tumors. In the classification branch, multi-scale feature extraction and channel attention mechanisms are integrated to discriminate between benign and malignant breast tumors. Experiments on two publicly available BUS datasets demonstrate that ACSNet not only outperforms mainstream multi-task learning methods for both breast tumor segmentation and classification tasks, but also achieves state-of-the-art results for BUS tumor segmentation. Code and models are available at https://github.com/qqhe-frank/BUS-segmentation-and-classification.git.

1. Introduction

Female breast cancer has surpassed lung cancer as the most commonly diagnosed cancer worldwide, and ranked fifth as the leading cause of cancer deaths worldwide [1]. Ultrasound imaging has become an important tool for breast cancer diagnosis due to its versatility, safety, and sensitivity. However, evaluating a whole breast ultrasound (BUS) examination slice by slice is highly time-consuming, even for experienced radiologists [2], and there is a high inter-observer variation rate. Therefore, an effective computer-aided diagnosis (CAD) system is essential in assisting physicians with the diagnosis and treatment of breast cancer. Segmentation and classification of BUS tumors are highly relevant tasks and are two fundamental objectives of CAD systems, because both share generic image features, such as shape and boundary features [3]. Nevertheless, designing this system for BUS images is challenging due to the problems such as large variations in tumor size and shape, irregular and blurred tumor boundaries, and low signal-to-noise ratios in ultrasound images, as shown in Fig. 1.

Convolutional Neural Networks (CNNs) have achieved remarkable success in medical image analysis due to their powerful capability for automatic feature extraction [4–10]. Specifically, in medical image segmentation tasks, UNet [11] is preferred for its ability to reconstruct high-resolution segmentation images by integrating multi-level information. In BUS image segmentation, Almajalid et al. [12] enhanced the performance of UNet in the precision of breast lesion segmentation by implementing strategies such as contrast enhancement and speckle noise reduction. Ning et al. [13] pointed out that a simple modeling framework is difficult to obtain ideal segmentation results on BUS images due to the complexity of the echo patterns of ultrasound images and the interference from surrounding tissues. Additionally, Zhao et al. [14] pointed out issues with the encoder-decoder model architecture, emphasizing that while high-quality segmentation relies on useful

* Corresponding author.
E-mail address: yangqiuju@snnu.edu.cn (Q. Yang).

https://doi.org/10.1016/j.compbiomed.2024.108319
Received 11 July 2023; Received in revised form 3 March 2024; Accepted 12 March 2024
Available online 18 March 2024

encoder features, direct transfer through skip connections could introduce misleading contextual information, resulting in underutilization of valuable features and potential introduction of irrelevant information leading to pixel classification errors. Moreover, the feature extraction capacity at a single level is limited, causing the global contextual information captured by the encoder at deeper levels to be progressively lost during the up-sampling process in the decoder [15].

Similarly, the progress in deep learning has significantly advanced tumor classification in BUS imaging. CNNs adopt a hierarchical structure of multiple convolutional and down-sampling layers to capture deep contextual features, an architecture capable of recognizing patterns ranging from simple edges to complex high-order arrangements. This allows the model to learn advanced semantic features in images and effectively identify diverse visual content. Notable studies in BUS tumor classification include Daoud et al.'s [16] investigation of deep feature extraction and transfer learning techniques to enhance the classification accuracy. In addition, Saad et al. [17] introduced the BreastUS Transformer, a Transformer model that incorporates self-attention mechanisms specifically designed for BUS image classification.

Both the breast tumor segmentation and classification tasks rely heavily on the accurate identification and detailed interpretation of common morphological features of tumors, including their shape, size, texture, and edge characteristics [3]. Proper delineation of tumor boundaries for segmentation and determination of their benign or malignant nature for classification are closely related goals [18]. Accurate tumor segmentation can provide key morphological features that are essential for accurate classification, while a sound understanding of the classification will aid in more accurate tumor segmentation. This interdependence and information sharing between tasks lead to higher accuracy and efficiency in breast cancer diagnosis.

However, single-task CNNs designed for BUS imaging may face limitations in scalability and dependence on substantial amounts of annotated data when tackling specific segmentation or classification tasks. In contrast, multi-task learning (MTL), a well-established learning paradigm in machine learning, has demonstrated its ability to improve the performance of individual tasks by sharing information between related tasks [19]. The inductive bias introduced by MTL serves as a regularization mechanism, effectively preventing overfitting and promoting the acquisition of more robust feature representations. In addition, MTL proves beneficial in mitigating the challenges associated with small datasets by facilitating knowledge transfer between different tasks and improving the overall generalizability of the model [20].

For example, Shareef et al. [20] showed that multi-task learning outperformed single-task learning methods in the context of breast ultrasound classification. Recent research on MTL [2,21,22] advocates the integration of segmentation and classification tasks in BUS image analysis within a unified network framework. This approach promotes the sharing of breast tumor-specific features, effectively improving the performance of both tasks and emerging as a promising avenue for further investigation. However, most existing multi-task learning architectures rely on an encoder-decoder approach and often overlook the impact of redundant or misleading information within the encoder on the results. Such oversight can hinder the selective transfer of valuable contextual information from the encoder to the decoder [14]. Additionally, tumor boundaries in BUS images are often poorly defined and there is significant variability in tumor size and shape. This variability requires a network that can adapt to tumors of different sizes.

In this paper, we present ACSNet, an end-to-end multi-task learning network tailored for the simultaneous learning of BUS tumor segmentation and classification tasks. The architecture consists of an encoder-decoder network dedicated to segmentation and a robust multi-scale feature aggregation network for benign and malignant classification. To refine the information flow during skip connections, gate unit modules for segmentation are designed to control the feature extraction process of the encoder. Furthermore, a deformable spatial attention module (DSAModule) is introduced to capture the complex morphological variations of breast tumors. In the classification task, a channel attention mechanism within the multi-scale feature extraction network highlights salient features while mitigating the influence of extraneous information, thereby improving the accuracy of the network in distinguishing between benign and malignant breast tumors. In addition, to suppress noise and further improve the multi-task learning performance, LossNet, a generic no-training loss network proposed by Zhao et al. [23], is used to supervise the feature mapping extracted from the bottom to the top layer, providing crucial structural detail supervision to the feature layers. We evaluate the effectiveness of the proposed method through extensive experiments on two publicly available datasets. Our main contributions can be summarized as follows:

(1) We present ACSNet, a novel multi-task learning model for simultaneous segmentation and classification of BUS tumors, which effectively addresses issues related to tumor size and shape variability and irregular boundaries in BUS images.
(2) For tumor segmentation, we are developing the DSAModule and gate units. The DSAModule effectively adapts to morphological changes, while the gate unit optimizes the information flow between the encoder and decoder, thereby enhancing the accuracy of segmentation.
(3) For classification, ACSNet integrates a channel attention mechanism and a multi-scale feature fusion method, achieving excellent performance in breast cancer classification.

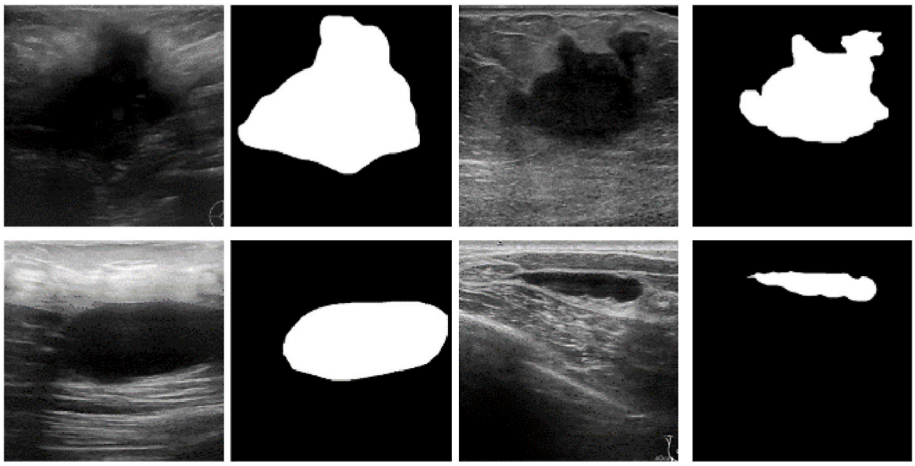

Fig. 1. Heavily shadowed areas resembling lesions in ultrasound images.

(4) Our experiments on two BUS tumor datasets show that ACSNet outperforms current mainstream multi-task learning methods for both segmentation and classification tasks. Furthermore, ACSNet achieves state-of-the-art results for BUS tumor segmentation.

The rest of the paper is organized as follows. Section 2 provides an overview of breast ultrasound image segmentation, classification, and multi-task learning for both tasks. Section 3 describes the proposed ACSNet in detail. Section 4 introduces the datasets used in our research and provides experimental details. The experimental results are presented in Section 5. Further discussions are given in Section 6, and Section 7 concludes the study.

2. Related work

2.1. Breast ultrasound image segmentation

BUS tumor segmentation methods can be broadly categorized into traditional methods and deep learning-based approaches. Traditional methods include region-based methods [24], deformable models [25,26], graph-based methods [27,28], and learning-based methods [29,30]. These methods rely heavily on hand-designed features and texture analysis to detect and segment breast lesion regions in ultrasound images, which could lead to misidentification of breast lesions in complex environments [31].

In recent years, deep learning-based approaches have shown considerable progress in BUS medical image processing. Shareef et al. [32] introduced row-column convolutional kernels that adapt to the anatomical structure of the breast and fused contextual information at different scales to segment small breast tumors. Hu et al. [33] proposed to combine an expanded convolutional network with a phase-based active contour model for automatic breast lesion region segmentation. Yap et al. [34] comparatively studied patched-based LeNet [35], UNet [11], and migration learning methods for breast ultrasound lesion detection. Zhu et al. [36] developed a second-order sub-region pooling network for breast lesion segmentation by using second-order statistics for multiple feature sub-regions. Vakanski et al. [37] proposed to integrate visual saliency into a deep learning model for breast tumor segmentation in ultrasound images. He et al. [38] designed a cross CNN-Transformer network, named HCTNet, for breast ultrasound lesion segmentation.

2.2. Breast ultrasound image classification

For BUS image classification, deep learning-based methods have shown superior feature extraction capabilities when compared to traditional machine learning methods. Qi et al. [39] proposed a deep CNN with multi-scale kernels and skip connections for diagnosing breast ultrasonography images. Byra et al. [40] developed a Selective Kernel (SK) UNet, which utilized an attention mechanism to adjust the receptive field of the network and fused the feature maps extracted by dilation and traditional convolution. Qian et al. [41] proposed a multi-pathway deep learning architecture for automated breast cancer risk prediction, using multimodal and multi-view BUS images to mimic routine clinical working processes and improve clinical applicability. Cui et al. [42] constructed a CNN that adaptively fuses information from multiple tumor regions for breast tumor classification in the testing process using images without segmentation masks. Zhuang et al. [43] utilized transfer learning [44] methods with the original BUS images, as well as enhanced and bilaterally filtered images, to improve breast lesion classification. Mo et al. [45] proposed a HoVer-Trans model that associates Transformers with CNNs for breast cancer diagnosis in BUS images.

2.3. Multi-task learning for breast ultrasound image segmentation and classification

Many multi-task learning (MTL) methods have been used to improve the results of both tumor classification and segmentation in BUS images. Zhou et al. [2] developed a multi-task learning network for 3D automatic BUS images, utilizing a multi-scale feature connectivity network for tumor classification and an iterative feature refinement training strategy to focus on tumor regions. Xu et al. [46] proposed a region-focused MTL framework (RMTL-Net) for simultaneous BUS image segmentation and classification, using a region-attention (RA) module to automatically learn weighted category-sensitive information in tumor, peri-tumor, and background regions. Wang et al. [47] proposed a multi-feature-guided CNN architecture for enhancement, segmentation, and classification of bone surfaces in ultrasound images. Zhang et al. [21] proposed the use of an Attention Gate (AG) module to integrate segmentation and classification tasks and improve the accuracy by effectively using lesion region information. Multi-task learning has also shown advantages in other medical image analysis tasks. For example, Chen et al. [48] proposed a fully automated framework for gadolinium-enhanced magnetic resonance image (GE-MRI) left atrial segmentation based on deep learning, Qu et al. [49] proposed a new multi-task learning method for cell nucleus segmentation and classification within a unified framework.

3. Methodology

Fig. 2 shows the proposed multi-task learning network, ACSNet, which integrates the BUS image segmentation and classification into a unified end-to-end model. ACSNet takes a BUS image as input and produces two outputs: a tumor segmentation map and an image-level classification probability.

More specifically, the architecture of ACSNet consists of two main components: a segmentation network and a classification branch. The core of the segmentation network is a U-shaped encoder-decoder structure that includes stages 1–5 for the encoder and stages 6–9 for the decoder. The final output stage of the decoder, stage 9, effectively achieves the segmentation of tumor lesions in ultrasound images. At the same time, the classification branch utilizes features from stages 4, 5, and 6 to classify breast tumors in these images as either benign or malignant.

To optimize the flow of information, gate units have been introduced at each skip connection between the encoder and decoder for the segmentation task. These gate units are designed to extract more robust features from the lesion areas during segmentation, thereby making more efficient use of the encoder features and generating more accurate predictive segmentation maps. In addition, the DSAModule is developed to enhance the spatial feature information of the lesions at different stages of the decoder, taking into account the variation in tumor size and the interference caused by background noise.

For the classification task, multi-scale features are extracted by merging feature maps from different scales, and a channel attention mechanism [50] is employed to enhance the learning capability of the network for feature representation. By jointly handling the segmentation and classification tasks and utilizing common features, MTL enhances the model's ability to analyze and process breast tumors. Furthermore, this approach improves data efficiency and reduces the dependence on large, task-specific datasets.

3.1. Multi-task learning network

The proposed ACSNet utilizes UNet [11] as the backbone architecture for multi-task learning due to its excellent performance in medical image segmentation. UNet consists of encoding paths, decoding paths, and skip connections between them. Skipping connections are essential mechanisms for disseminating spatial information, bridging semantic

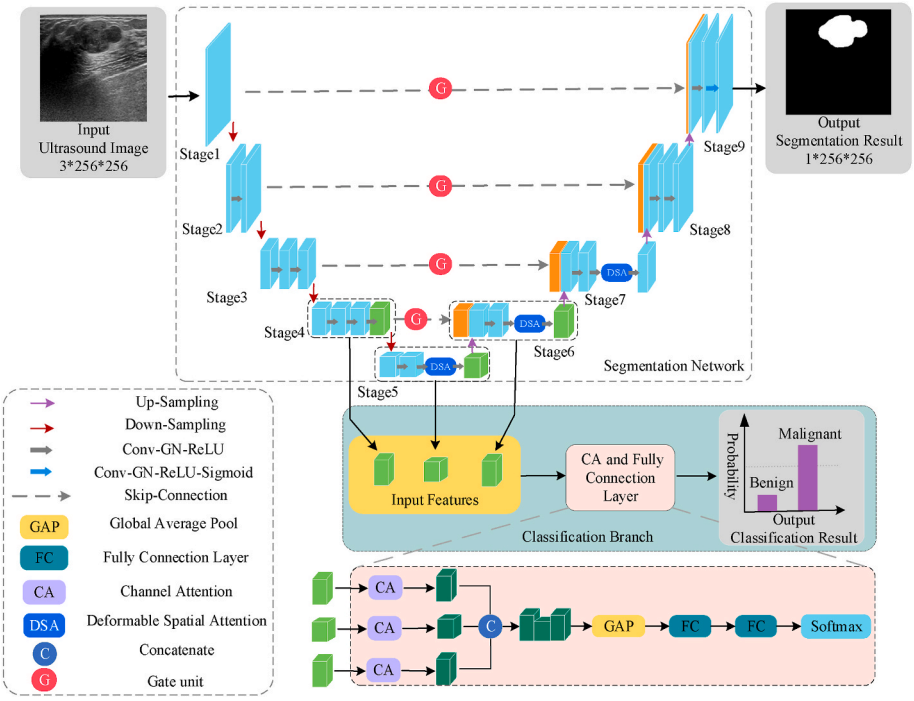

Fig. 2. Overview of the proposed multi-task learning network. The architecture consists of a segmentation network and a classification branch. The segmentation network includes gate units to regulate the features extracted from different encoder levels, while the DSAModules enhance the network's adaptability to tumor deformations. The input to the classification branch consists of tumor features at different scales, which are aggregated using the channel attention mechanism.

gaps, and refining segmentation results. These connections greatly enhance the model's ability to capture and utilize global and local information, making UNet a powerful architecture for medical image segmentation tasks. Given UNet's excellent performance in segmentation tasks, strong feature fusion capability, adaptability to different shapes, effectiveness on limited data, and previous success in related works [22,40], we make it the backbone of ACSNet. As shown in the segmentation network in Fig. 2, UNet contains convolutional layers at each stage. The encoding path employs four down-sampling operations to extract high-level semantic features, while the decoding path uses four upsampling operations to restore the feature maps to the original input size; in addition, the feature maps obtained from the encoder are connected to the decoder using skip connections to propagate spatial information and refine the segmentation results. Each convolution operation is followed by group normalization (GN) and ReLU.

Image classification networks mostly use the high-level feature maps of CNNs such as VGG [51] and ResNet [52]. Inspired by this, we add a classification branch at the bottom of UNet. First, the feature maps from stages 4, 5, and 6 of UNet are passed into the classification network. Then, the shared feature maps from these stages are fused using the channel attention (CA) mechanism [50] to explicitly model interdependencies between channels and adaptively recalibrate the feature responses of the channels. Finally, the fused features are input into two fully connected (FC) layers and a Softmax layer to predict the benign and malignant categories of the input images.

3.2. Gate control mechanism

To overcome the problem of redundant or misleading information in the encoder during multi-task learning, several solutions have been proposed by researchers. Among them, the application of gating mechanisms plays a crucial role, not only by improving the quality of feature extraction, but also by enhancing the adaptability of the models. For example, the Gated Dual-branch Network (GateNet) [14] was designed to optimize the transfer of contextual information from the encoder to the decoder. Inspired by this, we have integrated gate units at each skip connection in ACSNet, allowing for precise control and optimization of the feature information flowing from the encoder. In contrast to the approach of Zhao et al. [14], which evaluates the importance of different stages by combining encoder and decoder features, our gate unit employs a distinctive combination of Global Average Pooling (GAP) and Global Max Pooling (GMP). This approach enhances the selectivity of the feature control process, facilitating effective identification and retention of key features. Consequently, our method significantly improves the overall multi-task learning performance.

Fig. 3 shows the designed gate unit. The gate unit allows the information selection of the input features and improves the effectiveness of the output features. The gate unit takes the convolutional feature map $X \in \mathbb{R}^{C\times H\times W}$ as input, where H and W are the spatial height and width, and C is the number of channels. With the convolution operation, the input X is first divided into two feature maps F_1 and F_2, (F_1, $F_2 \in R^{1\times H\times W}$), which allows the gate unit to pay attention to effective information from different representation subspaces, and then F_1 and F_2 are normalized to the interval [0,1] using the Sigmoid function to obtain $F_1^{'}$ and $F_2^{'}$. A pair of gate values G^1 and G^2 are computed by using the global average pooling and global maximum pooling operations on $F_1^{'}$ and $F_2^{'}$. Finally, G^1 and G^2 are used to select the contextual information from the input features respectively, and to sum and fuse the information to obtain the output feature $Y \in \mathbb{R}^{C\times H\times W}$. In this way, the gate unit can selectively send the features from the encoder to the decoder during the skip connection process. The whole calculation process for the gate unit can be expressed as follows:

$$G^1 = GAP\left(\underbrace{Sigmoid(Conv(X))}_{F_1^{'}}\right), \tag{1}$$

$$G^2 = GMP\left(\underbrace{Sigmoid(Conv(X))}_{F_2^{'}}\right), \tag{2}$$

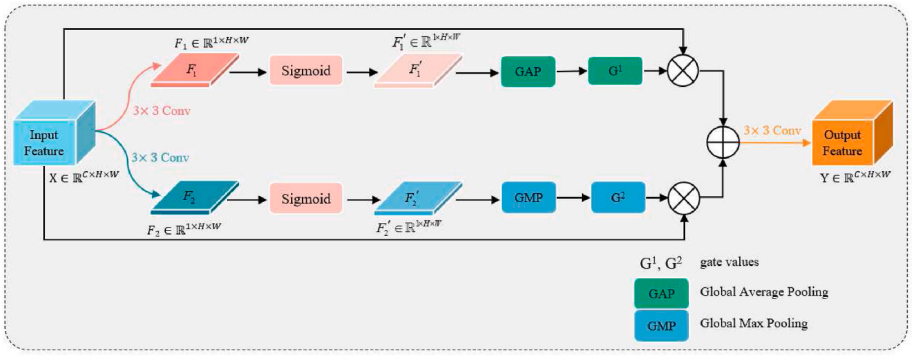

Fig. 3. Overview of the proposed gate unit. The input feature map is directed to two branches with spatial attention mechanisms. These branches utilize GAP and GMP, respectively, to extract relevant feature details. After summing the feature maps from both branches, a refined output is generated via a 3 × 3 Conv.

$$Y = Conv\left(X \times G^1 + X \times G^2\right), \tag{3}$$

where *GAP* denotes global average pooling operation and *GMP* denotes global maximum pooling operation.

3.3. Deformable spatial attention module

In the proposed ACSNet, we further improve the segmentation performance of the model by addressing the problem of under- and over-segmentation due to the variation in lesion size and different morphological characteristics of the tumor. We design a deformable spatial attention module (DSAModule) that overcomes the limitations of conventional convolution by applying deformable convolution [53] to the significant changes of the lesion region accordingly. Specifically, while traditional convolutional operations in image processing rely on a fixed and regular grid, deformable convolution introduces learnable offsets that allow the convolutional kernel sampling points to be dynamically adjusted based on image features. This design provides greater flexibility to adapt to local variations in the image. We use deformable convolution to address the limitations of traditional convolution when dealing with the diverse morphologies of tumors. In addition, we integrate a spatial attention mechanism into the DSAModule that merges max-pooling and average-pooling operations. This integration effectively extracts and consolidates critical information from the feature maps, resulting in the generation of the final feature map. This inclusion of the spatial attention mechanism significantly improves the model's ability to identify and focus on critical regions within the images.

Fig. 4 shows the proposed DSAModule. DSAModule takes the feature map $X \in \mathbb{R}^{C\times H\times W}$ of the current stage of the encoder as input, and first obtain $X^{'} \in \mathbb{R}^{C\times H\times W}$ by 3×3 deformable convolution. To further enhance the spatial details and improve the segmentation performance, we add a spatial attention module after the deformable convolution. In the spatial attention module, we use average pooling and maximum pooling operations for $X^{'}$ along the channel dimension C to obtain information about the spatial context of the object and the important information in the features, respectively [54]. Then the 3×3 convolutional layers and Sigmoid function are used to generate the corresponding spatial attention map $S_{1,2} \in \mathbb{R}^{1\times H\times W}$ respectively, and the attention-weighted feature maps $A_{1,2} \in \mathbb{R}^{C\times H\times W}$ are obtained by multiplying $S_{1,2} \in \mathbb{R}^{1\times H\times W}$ with the deformable convolved features $X^{'}$, respectively. Finally, A_1 and A_2 are concatenated and the lesion features are again enhanced by a deformable convolution to obtain the final output $Y \in \mathbb{R}^{C\times H\times W}$. The whole DSAModule process can be expressed as follows:

$$X^{'} = DConv(X), \tag{4}$$

$$S_1 = Sigmoid(Conv(Avg_{axis=channel}(X^{'}))), \tag{5}$$

$$S_2 = Sigmoid(Conv(Max_{axis=channel}(X^{'}))), \tag{6}$$

$$Y = DConv\left(Cat\left(\underbrace{X^{'} \times S_1}_{A_1}, \underbrace{X^{'} \times S_2}_{A_2}\right)\right), \tag{7}$$

where *DConv* is the deformable convolution.

3.4. Multi-task loss function

In the classification task, we use the binary cross-entropy [55] loss function to train the model to classify the whole image instead of each pixel, which can be expressed as:

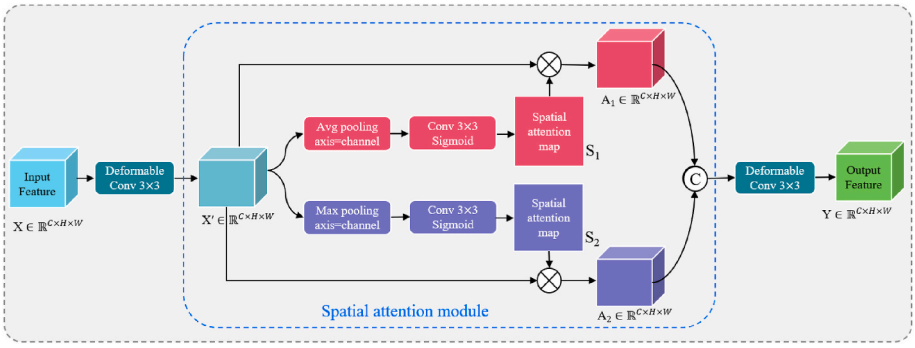

Fig. 4. Overview of the proposed DSAModule. The input feature maps first pass through a deformable convolution to preliminarily extract the features of the tumor regions. These features are then fed into a spatial attention module to refine relevant feature details, and finally processed through another 3 × 3 deformable convolution to obtain enhanced output features.

$$L_{cls}(p^k, g^k) = -\frac{1}{N}\sum_{k=1}^{N}\left(g^k \log p^k + (1-g^k)\log(1-p^k)\right), \tag{8}$$

where g^k is the ground truth class of sample k and p^k is the predicted classification probability from the proposed network for sample k, respectively.

In the segmentation task, we use a segmentation loss based on the Dice coefficient to address the issue of class imbalance between the foreground (object) and background in the image, which can cause segmentation bias. The Dice coefficient measures the similarity between two sets of data, and is commonly used as an evaluation metric in segmentation tasks. The segmentation loss is defined as:

$$L_{dice}(P_{seg}, Y_{seg}) = 1 - \frac{2P_{seg}Y_{seg}+1}{P_{seg}+Y_{seg}+1}, \tag{9}$$

where L_{dice} is the dice segmentation loss, P_{seg} and Y_{seg} denote the predicted segmentation map from the proposed network and the ground truth respectively.

Additionally, inspired by the polyp segmentation described in Zhao et al. [23], we further employed their LossNet to refine the segmentation results, ensuring comprehensive supervision of both details and structures at the feature level. To achieve this, we extract the predicted and ground truth multi-scale features using an ImageNet pre-trained classification network, namely VGG-16 [51]. Then, the feature difference is computed as a loss $L_{LossNet}$:

$$L_{LossNet} = l^1_{LossNet} + l^2_{LossNet} + l^3_{LossNet} + l^4_{LossNet}, \tag{10}$$

$$l^i_{LossNet} = \left\|F^i_P - F^i_G\right\|_2, i = 1,2,3,4 \tag{11}$$

where F^i_P and F^i_G separately represent the i-th level feature maps extracted from the prediction and ground truth, and $l^i_{LossNet}$ is calculated as their Euclidean distance (L2-Loss). Fig. 5 illustrates the calculation of LossNet.

Thus, the total segmentation loss function can be expressed as:

$$L_{seg}(L_{dice}, L_{LossNet}) = L_{dice} + \gamma \times L_{LossNet}, \tag{12}$$

where $\gamma \in [0,1]$ is the $L_{LossNet}$ weight in the total segmentation loss function.

In our approach, the classification loss L_{cls} and the segmentation loss L_{seg} are linearly combined into a multi-task loss by a hyperparameter λ. The multi-task loss is defined as:

$$L_{total} = \lambda L_{cls} + (1-\lambda)L_{seg}, \tag{13}$$

where L_{total} is denoted as multi-task loss and $\lambda \in [0,1]$ is the classification task weight. In our experiment, segmentation and classification tasks are weighted differently, with more weights given to the more difficult segmentation tasks.

4. Data and experiments

4.1. Dataset

We evaluated the effectiveness of the proposed ACSNet using two publicly available breast ultrasound image datasets. The first dataset, BUSI [56], was collected at Baheya Women's Cancer Early Detection and Treatment Hospital in Cairo, Egypt in 2018, utilizing the LOGIQ E9 ultrasound and LOGIQ E9 Agile ultrasound systems. The BUSI dataset consists of 780 breast cancer ultrasound images from 600 female patients aged 25–75 years with an average image size of 500 × 500 pixels. It includes 437 benign cases, 210 malignant masses, and 133 normal cases. The selection of the BUSI dataset was motivated by its inherent challenges, including variations in tumor size, irregular boundaries, and significant morphological differences.

The second dataset, BUS [34], was collected by the UDIAT Diagnostic Centre of Parc Taulí in Sabadell, using a Siemens ACUSON Sequoia C512 system and a 17L5HD linear array transducer (8.5 MHz) to acquire ultrasound images. The BUS dataset consists of 163 breast cancer ultrasound images from different women, with an average image size of 760 × 570 pixels, including 53 images of malignant cases and 110 images of benign cases. Since tumors tend to be smaller in the early stages of clinical detection, accurate segmentation or classification becomes a challenge. Therefore, the BUS dataset deliberately focuses on smaller tumors, which pose a significant challenge due to their limited size and the possibility of superimposed benign or malignant features, complicating the segmentation and classification process.

As our multi-task learning aims to segment and classify benign and malignant lesions in breast ultrasound images, we removed normal cases without breast lesions from the BUSI dataset. In addition, a five-fold cross-validation method was used to evaluate the performance of different methods on the two aforementioned datasets.

4.2. Experiment details

All experiments were conducted on an Ubuntu 16.04 system equipped with an Intel(R) Xeon(R) CPU E5-2680 V3. All models were implemented using the deep learning toolbox PyTorch [57] and were trained and tested on an NVIDIA GeForce GTX 1080Ti with 11,264 MB of memory. Using the Adam optimizer with an initialized learning rate

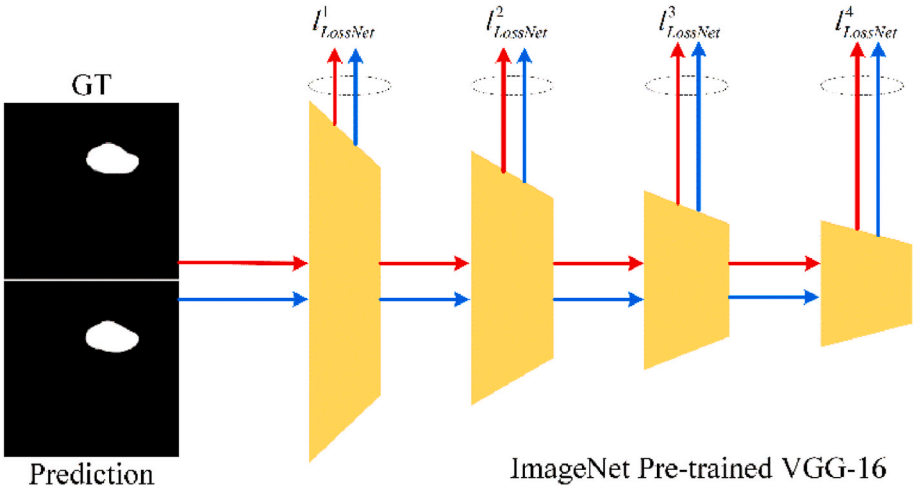

Fig. 5. Illustration of the calculation of LossNet.

of 1e-4, a momentum β_1 of 0.9, a momentum β_2 of 0.99, and a weight decay of 0.0001. We employed LambdaLR to adjust the learning rate during training. The hyperparameter γ was set to 0.2 and λ was set to 0.3. For the BUSI dataset, we set the GPU batch size to 16 and trained the model for 100 epochs, whereas for the BUS dataset, we set the GPU batch size to 4 and trained the model for 90 epochs. All images were resized to 256 × 256. We employed standard data augmentation techniques including random rotation, horizontal flip, and vertical flip. Rotations were applied randomly within a range of 10–30° range to introduce variability in tumor orientation. Horizontal and vertical flips were used to simulate different tumor appearances from different orientations and sides of the image. These well-established augmentation methods are commonly used in deep learning-based medical image analysis. By artificially expanding the variety and size of the dataset, they play a key role in enhancing the generalization ability of the model, a critical aspect for robust performance in various clinical scenarios.

4.3. Evaluation metrics

To quantitatively evaluate the effectiveness of tumor segmentation, we utilized four commonly used evaluation metrics: Dice similarity coefficient (Dice) [58], Jaccard index (Jaccard) [31], 95% asymmetric Hausdorff distance (95HD) [59], and average surface distance (ASD) [60]. These indices were calculated as follows:

$$Dice(A,B) = \frac{2 \times |A \cap B|}{|A| + |B|}, \tag{14}$$

$$Jaccard(A,B) = \frac{|A \cap B|}{|A \cup B|}, \tag{15}$$

$$95HD(A,B) = \max\left\{ \sup_{x \subset A} \inf_{y \subset B} \|x - y\|, \sup_{y \subset B} \inf_{x \subset A} \|x - y\| \right\}, \tag{16}$$

$$ASD(A,B) = \frac{\sum_{x \in A} \min_{y \in B} d(x,y)}{|A|}, \tag{17}$$

where A and B are the segmentation results and the ground truth, and x and y are the pixels in A and B, respectively. Dice and Jaccard are sensitive to tumor size and 95HD sensitive to tumor shape.

For the tumor classification task, we used recall (REC), precision (PRE), and accuracy (ACC) for the quantitative evaluation:

$$ACC(\%) = \frac{TP + TN}{TP + TN + FN + FP} \times 100, \tag{18}$$

$$REC(\%) = \frac{TP}{TP + FN} \times 100, \tag{19}$$

$$PRE(\%) = \frac{TP}{TP + FP} \times 100, \tag{20}$$

where TP, FP, TN, FN are the number of true positive, false positive, true negative and false negative, respectively. See Ref. [58] for details of these classification metrics.

5. Results

5.1. Ablation study

In this section, we conducted ablation experiments to evaluate the effectiveness of the principal components of our network, including CA, gate unit, DSAModule, and multi-scale feature fusion. We performed these experiments on the BUSI dataset since it contains the most challenging data with significant lesion morphological variation and the largest number of samples in both datasets.

Table 1 presents the comparison results of our method with different combinations of components. Specifically, the single-task models ClsNet and UNet are used as baseline models for classification and segmentation, respectively. ClsNet uses the encoding path of UNet followed by a GAP layer and two FC layers for classification. In the multi-task models, the baseline is denoted as CUNet. CUNet uses UNet as the framework for the segmentation task and adds a single-scale classification branch to the UNet model, which uses the features of stage 5 of UNet as the classification features. We then add the CA module to the single-scale classification branch to obtain C_AUNet. Based on CUNet, by using the gate unit to learn the skip connections, we obtain CU_GNet. Based on CUNet, we add DSAModule to enhance the spatial location information of lesions, resulting in CU_ANet, and use both gate unit and DSAModule to obtain CU_{GA}Net. In addition, we use a multi-scale classification branch for multi-task learning instead of a single-scale classification branch (denoted as C_{MS}UNet) and add the proposed CA module to the classification branch for multi-task learning (denoted as C_{AMS}UNet). Finally, we integrate all proposed components to obtain the final model ACSNet.

Table 1
Networks constructed in ablation study. Cls: classification; Seg: segmentation.

Model	Task	CA	Gate unit	DSAModule	Multi-scale
UNet	Seg	×	×	×	×
ClsNet	Cls	×	×	×	×
CUNet	Seg + Cls	×	×	×	×
CU_GNet	Seg + Cls	×	✓	×	×
C_AUNet	Seg + Cls	✓	×	×	×
CU_ANet	Seg + Cls	×	×	✓	×
CU_{GA}Net	Seg + Cls	×	✓	✓	×
C_{MS}UNet	Seg + Cls	×	×	×	✓
C_{AMS}UNet	Seg + Cls	✓	×	×	✓
ACSNet	Seg + Cls	✓	✓	✓	✓

The quantitative results of the ablation experiments are presented in Table 2. These results show that the integration of all principal components in our model leads to the best performance compared to all other configurations. Specifically, for the segmentation tasks, the proposed ACSNet achieved remarkable results, with the highest Dice score of 84.90%, the highest Jaccard index of 78.62%, the lowest 95HD of 13.04 mm, and the second-lowest ASD of 3.45 mm among all configurations. ACSNet also performed outstandingly in the classification tasks, achieving the highest ACC of 94.44%, the highest PRE of 94.61% and a remarkably high REC of 93.86%. These results highlight the critical importance of each component, including the channel attention mechanism, gate unit, DSAModule, and multi-scale feature fusion, in improving the accuracy of both tumor segmentation and classification in breast ultrasound images.

Furthermore, a comparison between CU_GNet and CUNet shows that the gate unit optimizes the features extracted from the encoder to a certain extent. This optimization helps to avoid interference from redundant or misleading information, thereby enhancing the performance of the model in breast tumor segmentation and classification. Additionally, when comparing CU_{GA}Net with CU_GNet, it can be seen that the DSAModule effectively improves the model's ability to detect deformations in breast tumors. The arrows accompanying each indicator indicate the desired trend: an upward arrow (↑) indicates a preference for higher values of the indicator, while a downward arrow (↓) indicates a desire for lower values to achieve optimal performance.

When comparing the results of CUNet and UNet, it is clear that CUNet shows superior performance in the segmentation task. This improvement clearly confirms the benefit of integrating classification information into segmentation, and underscores the advantages of multi-task learning over conventional single-task approaches. Furthermore, the inclusion of a gate unit in the skip connections of the CUNet model, as seen in CU_GNet, leads to improved results across all metrics for both segmentation and classification. This enhancement signifies the effectiveness of the gate unit in selectively filtering the features extracted by the encoder, ensuring that only the most relevant features are passed to the decoder. Additionally, the significant improvements

Table 2
Quantitative results on BUSI for all networks constructed in ablation study. The optimal results are shown in bold.

Model	Segmentation (Mean ± Std)				Classification (Mean ± Std)		
	Dice (%) ↑	Jaccard (%) ↑	95HD ↓	ASD ↓	ACC (%) ↑	PRE (%) ↑	REC (%) ↑
UNet	83.04 ± 1.64	76.34 ± 2.04	15.95 ± 2.29	5.14 ± 0.67	–	–	–
ClsNet	–	–	–	–	92.38 ± 2.59	92.54 ± 2.29	91.38 ± 2.30
CUNet	83.56 ± 1.72	76.36 ± 2.19	15.07 ± 2.47	4.01 ± 1.13	92.54 ± 2.87	92.92 ± 2.39	92.20 ± 2.04
CU_GNet	84.26 ± 1.14	77.58 ± 1.36	14.04 ± 1.15	3.75 ± 0.61	93.02 ± 2.21	93.20 ± 2.13	91.70 ± 1.62
C_AUNet	83.44 ± 1.28	76.98 ± 1.17	14.33 ± 1.39	3.95 ± 0.93	93.81 ± 1.46	93.89 ± 1.40	92.89 ± 1.25
CU_ANet	84.48 ± 1.26	77.78 ± 0.94	14.03 ± 1.75	4.22 ± 1.33	93.33 ± 2.91	93.61 ± 2.43	92.82 ± 2.12
$CU_{GA}Net$	84.76 ± 0.83	78.20 ± 0.94	13.12 ± 1.08	**3.33 ± 0.68**	93.17 ± 2.91	93.35 ± 2.68	92.03 ± 2.74
$C_{MS}UNet$	83.76 ± 1.90	77.12 ± 1.99	14.84 ± 2.27	4.26 ± 1.29	94.16 ± 2.77	94.35 ± 2.11	93.76 ± 1.67
$C_{AMS}UNet$	84.02 ± 1.18	77.12 ± 1.29	14.63 ± 1.68	4.45 ± 0.95	94.29 ± 2.10	94.46 ± 2.07	**94.07 ± 1.48**
ACSNet	**84.90 ± 1.69**	**78.62 ± 1.75**	**13.04 ± 2.58**	3.45 ± 1.59	**94.44 ± 2.07**	**94.61 ± 1.56**	93.86 ± 2.33

observed in CU_ANet and C_AUNet, as compared to CUNet, are due to their respective attentional mechanisms, which contribute to performance enhancements by focusing on critical features, thereby refining the overall model accuracy and effectiveness. Furthermore, $C_{MS}UNet$, unlike CUNet, employs multi-scale feature extraction for classification, resulting in superior performance over both ClsNet and CUNet in classification tasks. This result highlights the effectiveness of multi-scale feature extraction and provides a compelling argument for its superiority over single-scale approaches. $C_{MS}UNet$'s ability to capture a wider range of feature details allows for more accurate classification, further validating the model's robustness and versatility.

Fig. 6 visually compares the segmentation results of various methods in the ablation experiment. This paper designs a gate unit and a DSA-Module for tumor segmentation. From Fig. 6, we can conclude that the segmentation performance of CU_GNet has greatly improved compared to CUNet, because its gate unit can select and control the feature information in the encoder; Also, the segmentation accuracy of CU_ANet is superior to CUNet, and over and under-segmentation problems are significantly reduced, indicating that the developed DSAModule can learn the location information of lesions in the feature map. $CU_{GA}Net$ combines the strengths of the gate unit and DSAModule, further improving the segmentation performance of CU_GNet and CU_ANet, resulting in more accurate lesion location and improved boundaries. In addition, $CU_{GA}Net$ achieves significantly better segmentation results than C_AUNet, $C_{MS}UNet$, and $C_{AMS}UNet$, which do not contain the gate unit and DSAModule. All the results demonstrate that the effectiveness of the proposed gate unit and DSAModule in breast ultrasound image segmentation.

5.2. Comparison with the state-of-the-arts

Table 3 compares the proposed ACSNet with ten advanced methods, including six recent single-task segmentation methods (FCN [61], UNet [11], DeeplabV3+ [62], FPN [63], Unet++ [64], and TransUnet [65]) and four state-of-the-art MTL methods (CUNet, Wang et al. [47], Chen et al. [48], and Qu et al. [49]) on the BUSI dataset. CUNet is the model discussed in ablation study. For a fair comparison, we used the same backbone (UNet) as the proposed method in all compared methods, and applied five-fold cross-validation to all models.

Table 4 reports the mean and standard deviation values of the

Table 3
Briefly comparing the proposed ACSNet with 10 advanced methods.

Methods	Backbone	Tasks		Feature Enhancement
		Single Segmentation	Multi-Task	Attention Mechanism
UNet [11]	ResNet-18	✓		
FCN [61]	ResNet-18	✓		
DeepLabV3+ [62]	ResNet-18	✓		
FPN [63]	ResNet-18	✓		
Unet++ [64]	ResNet-18	✓		
TransUnet [65]	R50+ViT-B_16	✓		
CUNet	ResNet-18		✓	
Wang et al. [47]	ResNet-18		✓	✓
Qu et al. [49]	ResNet-18		✓	✓
Chen et al. [48]	ResNet-18		✓	✓
ACSNet	ResNet-18		✓	✓

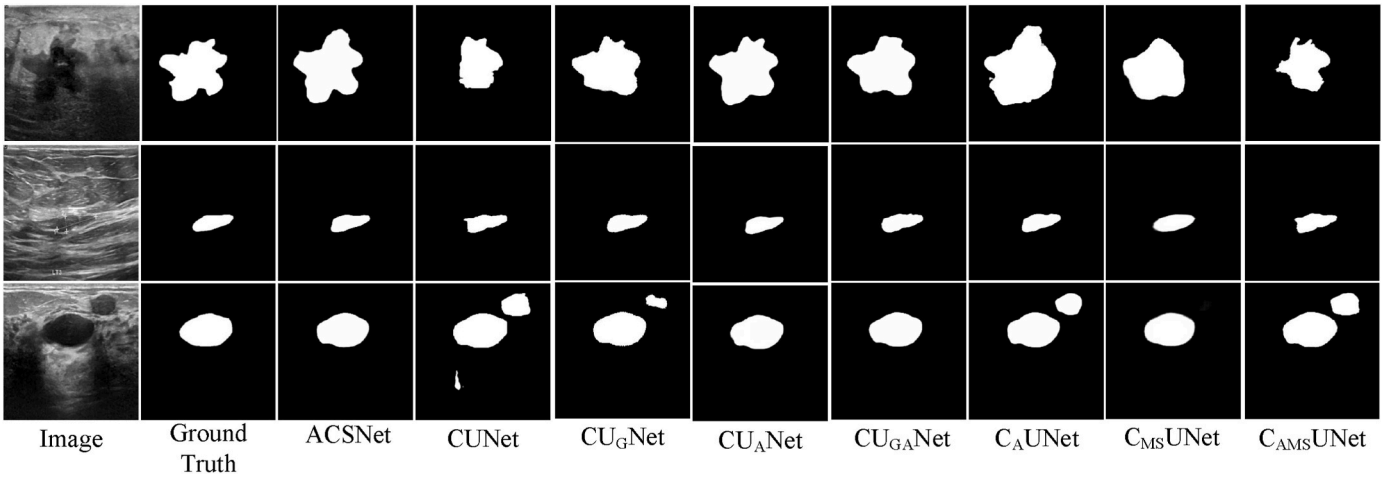

Fig. 6. Visual segmentation results of ablation study.

Table 4
Classification performance (Mean ± SD) of all compared methods on BUSI.

Methods	ACC (%) ↑	PRE (%) ↑	REC (%) ↑
CUNet	92.54 ± 2.87	92.92 ± 2.39	92.20 ± 2.04
Wang et al. [47]	92.86 ± 2.84	92.93 ± 2.93	92.13 ± 3.61
Qu et al. [49]	94.13 ± 1.98	94.51 ± 1.48	93.58 ± 1.36
Chen et al. [48]	92.22 ± 1.91	92.32 ± 1.88	92.25 ± 1.89
ACSNet	**94.44 ± 2.07**	**94.61 ± 1.56**	**93.86 ± 2.33**

classification metrics (ACC, PRE, and REC) among ACSNet and all the competitors on BUSI. Among the five MTL methods, the proposed ACSNet achieves the best performance in all metrics. Low-level features are usually used for classification because they mainly capture shape and boundary information, while high-level features extract semantic features of the targets. Capturing the features of small objects is challenging when the network depth increases with more convolution and down-sampling operations. To address this issue, ACSNet combines and fuses the multi-scale features for the classification task. Specifically, ACSNet inputs the feature maps from stage 4 to stage 6 of the segmentation network to the classification branch, which enables the multi-scale features to contain more useful information through channel attention CA, thus achieving the best classification performance on the BUSI.

Table 5 summarizes the segmentation results of ACSNet and ten compared methods on Dice, Jaccard, 95HD, and ASD on the BUSI, including six single-task segmentation methods (i.e., UNet, FCN, FPN, DeepLabV3+, UNet++ and TransUnet) and five MTL methods (i.e., CUNet, Wang et al. [47], Chen et al. [48], Qu et al. [49], and ACSNet). Among the six single-task segmentation methods, TransUnet obtained the highest Jaccard and Dice values, the lowest 95HD and ASD values, and the best overall segmentation performance on BUSI, because it combines the advantages of Transformer and UNet, with Transformer encoding the tokenized image patches from the feature maps of CNNs as the input sequence for extracting the global semantics, while the decoder upsamples the encoded features and then combines them with the high-resolution CNN feature maps for precise localization. DeepLabV3+ obtained the second-best overall segmentation performance, followed by FPN. Among the five MTL methods, the proposed ACSNet achieved the best segmentation performance on all four metrics, with the lowest 95HD and ASD of 1.65 mm and 0.66 mm, respectively, while Dice and Jaccard were 1.08% and 1.50% higher than the sub-optimal method (i.e., TransUnet), respectively.

Table 5
Segmentation performance of all compared methods on BUSI.

Methods	Dice (%) ↑	Jaccard (%) ↑	95HD ↓	ASD ↓	FLOPs (G)	Param (M)
UNet [11]	83.04 ± 1.64	76.34 ± 2.04	15.95 ± 2.29	4.14 ± 0.67	25.69	29.38
FCN [61]	81.46 ± 1.50	73.90 ± 1.66	16.14 ± 1.69	4.50 ± 0.36	9.54	22.35
FPN [63]	82.68 ± 0.69	75.88 ± 0.72	15.98 ± 1.27	4.26 ± 0.68	14.41	28.91
DeepLabV3+ [62]	83.34 ± 0.96	76.30 ± 0.88	14.87 ± 1.18	4.26 ± 0.64	15.08	27.78
UNet++ [64]	82.80 ± 0.87	76.04 ± 1.03	14.83 ± 0.67	4.18 ± 1.26	35.20	16.80
TransUnet [65]	83.82 ± 0.89	77.12 ± 0.84	14.69 ± 1.68	4.11 ± 1.09	32.23	93.23
CUNet	83.56 ± 1.72	76.36 ± 2.19	15.22 ± 1.59	4.01 ± 1.13	25.69	29.12
Wang et al. [47]	82.76 ± 1.62	76.30 ± 2.03	14.39 ± 2.79	4.45 ± 1.76	37.86	29.38
Chen et al. [48]	82.38 ± 0.82	75.48 ± 0.47	16.51 ± 0.67	5.20 ± 1.52	27.90	34.72
Qu et al. [49]	82.28 ± 0.99	75.48 ± 1.03	17.71 ± 2.12	5.41 ± 1.45	5.82	14.55
ACSNet	**84.90 ± 1.69**	**78.62 ± 1.75**	**13.04 ± 2.58**	**3.45 ± 1.59**	30.96	42.99

In addition, Table 5 also provides a clear comparison of floating-point operations (FLOPs) and the number of parameters (in millions) for each model. ACSNet has a computational complexity of 30.96 G FLOPs and 42.99 M parameters. Although ACSNet's parameters are the largest among these models, its model complexity is low, and in terms of results, ACSNet achieves optimal results on the task of breast tumor segmentation and classification from ultrasound images, which we believe strikes a favorable balance between performance and efficiency.

The segmentation results of ACSNet and all competitors on BUSI are given in Fig. 7. Among the single-task segmentation methods (Fig. 7(d)–(i)), TransUnet has the best overall segmentation performance due to the advantage of using the Transformer to obtain global information, while the CNN models (Fig. 7(d)–(h)) tend to ignore breast lesion details or misclassify lesion regions as non-breast lesions in the predicted segmentation maps. The multi-task models shown in Fig. 7(j)–(m) have poor results for tumor boundary segmentation and suffer from more serious over- or under-segmentation problems, which are due to the fact that the tumor morphology is highly variable and the features extracted by the encoder are not fully suitable for the segmentation task. The proposed ACSNet achieved the best segmentation performance among the five MTL methods and provides more accurate segmentation of the breast lesion region, with fewer over- or under-segmentation errors.

Table 6 summarizes the classification results of ACSNet and four compared methods on BUS. It is clear that ACSNet has the highest accuracy, precision, and recall and improves over the second-best method (i.e., Qu et al. [49]) by 0.63%, 0.13%, and 3.4%, respectively.

Table 7 reports the mean and standard deviation values of the four segmentation metrics among our ACSNet and ten compared methods on the dataset BUS. Among the single-task segmentation methods (UNet, FCN, FPN, DeepLabV3+, UNet++, and TransUnet), TransUnet has the best overall segmentation performance, with the highest values on Dice and Jaccard and the lowest values on 95HD. Among the five MTL methods (CUNet, Wang et al. [47], Chen et al. [48], Qu et al. [49], and ACSNet), ACSNet achieved the best segmentation performance on all four metrics, with increased Dice and Jaccard values over the second-best method (TransUnet) by 1.26% and 1.66%, respectively, while reducing 95HD and ASD by 0.63 mm and 0.69 mm, respectively.

Fig. 8 shows the visual segmentation results of different segmentation methods on the dataset BUS. The proposed ACSNet effectively mitigated the influence of tumor size, surrounding tissue, and shadow regions, providing segmentation results closer to the ground truth and fewer missed and false detections. The comprehensive evaluation results and visual effects demonstrate the effectiveness of the proposed ACSNet for breast ultrasound image segmentation and classification.

6. Discussion

BUS tumor segmentation and classification face significant challenges, such as variations in tumor size and shape, irregular and blurred tumor boundaries, and low signal-to-noise ratio of ultrasound images [2]. To address these challenges, we propose ACSNet, a novel multi-task learning model specifically designed for tumor segmentation and classification in BUS images. ACSNet starts by extracting basic patterns and features at shallow levels and gradually progresses to complex, high-level semantic feature extraction. As the model's receptive field expands, it not only performs more accurate tumor feature analysis, but

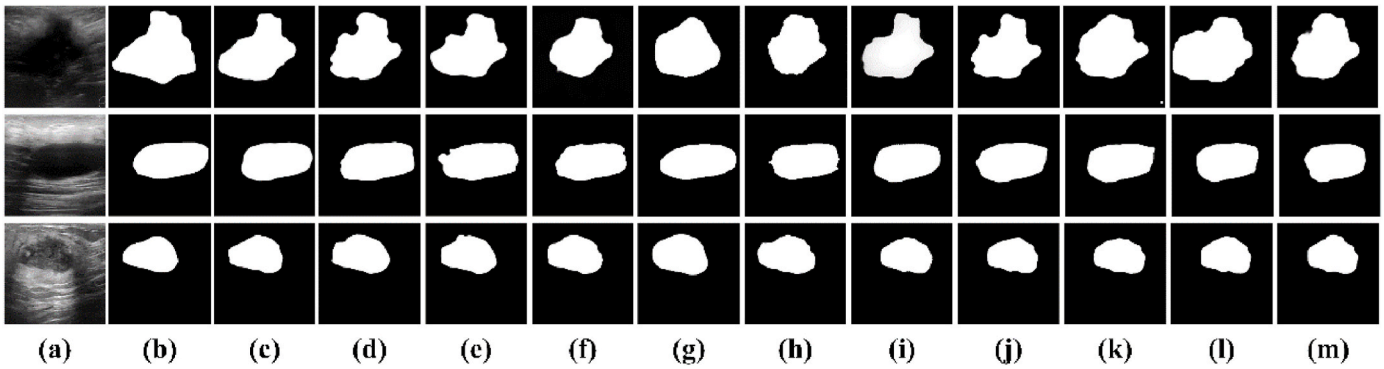

Fig. 7. Visualization of segmentation results of different methods on the dataset BUSI, (a)–(m) from left to right are (a) input image, (b) ground truth, and (c) the segmentation results of ACSNet (Ours), (d) UNet, (e) Unet++, (f) FPN, (g) FCN, (h) DeeplabV3+, (i) TransUnet, (j) CUNet, (k) Wang et al. [47], (l) Chen et al. [48], and (m) Qu et al. [49] respectively.

Table 6
Classification performance (Mean ± SD) of all compared methods on BUS.

Methods	ACC (%) ↑	PRE (%) ↑	REC (%) ↑
CUNet	87.08 ± 3.70	87.00 ± 3.75	83.91 ± 4.54
Chen et al. [48]	84.02 ± 3.69	84.20 ± 3.56	80.38 ± 4.52
Qu et al. [49]	88.33 ± 4.16	89.16 ± 2.89	84.12 ± 6.99
Wang et al. [47]	85.89 ± 2.42	86.68 ± 1.97	83.22 ± 3.27
ACSNet	**88.96 ± 1.49**	**89.29 ± 1.46**	**87.52 ± 2.66**

Table 7
Segmentation performance (Mean ± SD) of all compared methods on BUS.

Methods	Dice (%) ↑	Jaccard (%) ↑	95HD ↓	ASD ↓
UNet [11]	87.36 ± 1.99	81.28 ± 2.41	8.77 ± 1.12	2.56 ± 0.90
FCN [61]	84.36 ± 3.31	76.62 ± 4.21	8.57 ± 4.30	2.31 ± 2.21
FPN [63]	83.00 ± 3.73	75.52 ± 3.36	14.28 ± 1.87	4.97 ± 1.61
DeepLabV3+ [62]	86.78 ± 3.70	80.34 ± 4.58	9.81 ± 5.06	3.11 ± 2.24
Unet++ [64]	87.37 ± 2.29	81.30 ± 2.89	9.05 ± 1.49	2.19 ± 0.72
TransUnet [65]	87.86 ± 2.51	81.38 ± 3.28	8.39 ± 1.39	2.74 ± 1.21
CUNet	87.26 ± 3.70	80.90 ± 3.74	10.95 ± 4.33	3.35 ± 2.36
Chen et al. [48]	86.32 ± 2.19	79.98 ± 1.87	9.79 ± 2.48	3.11 ± 1.96
Qu et al. [49]	83.06 ± 4.05	75.56 ± 3.68	13.51 ± 4.29	3.50 ± 1.74
Wang et al. [47]	85.76 ± 4.54	79.40 ± 4.91	9.98 ± 4.37	3.35 ± 2.46
ACSNet	**89.12 ± 2.31**	**83.04 ± 2.60**	**7.76 ± 1.73**	**2.05 ± 1.38**

also effectively ignores background information in non-tumor regions, utilizing contextual information to enhance accuracy.

We have systematically evaluated the effectiveness of ACSNet's components and parameters through ablation studies. Experimental results show that ACSNet achieves the best performance in breast cancer classification experiments, highlighting the critical role of the integration of the channel attention mechanism and the multi-scale feature fusion network in achieving superior classification results. ACSNet also outperforms other methods in the segmentation task. The gate units optimize the encoder's feature information, enhancing the model's ability to accurately identify and exploit the contextual information. At the same time, the DSAModule adapts to the morphological variability of tumors, further improving the accuracy of the model in segmenting breast tumors.

For the multitask learning of breast tumor segmentation and classification from ultrasound images, the deep learning architecture is of paramount importance. Additionally, we utilize data augmentation techniques such as random rotation and flipping to mitigate the limitations of small datasets. By implementing weight regularization, we prevent the model from overfitting, ultimately improving its generalization ability and overall performance. Through these strategies, ACSNet effectively identifies and handles common patterns and artifacts in images, ensuring accurate analysis and recognition of tumor features, thereby significantly enhancing the overall segmentation and classification performance and effectively exploiting contextual information.

The analysis of the experimental results on the segmentation task leads to several conclusions. U-shaped structures with skip connections effectively merge low-level features from the encoding stage with high-level features from the decoding stage, leading to commendable segmentation results. Furthermore, multi-task learning shows improved

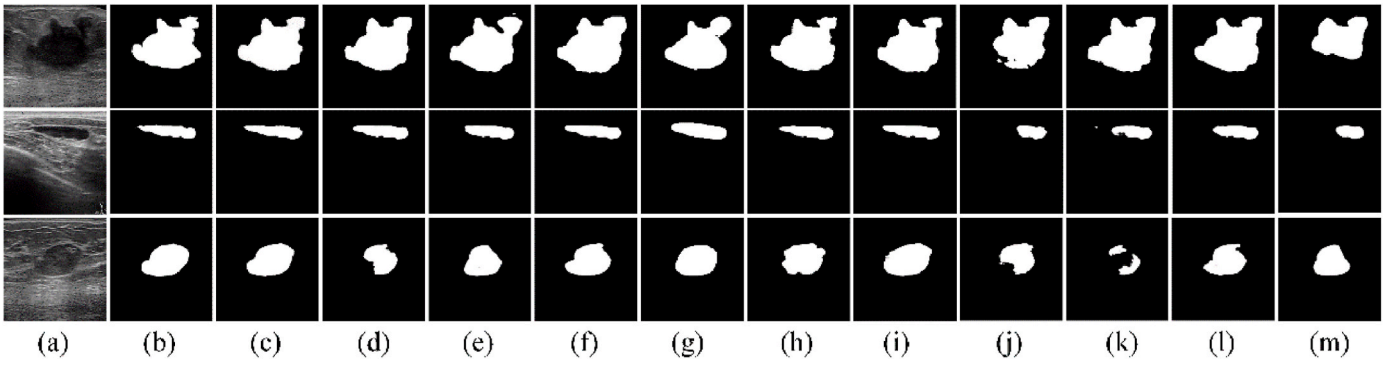

Fig. 8. Visualization of segmentation results of different methods on BUS, (a)–(m) from left to right are (a) input image, (b) ground truth, and (c) the segmentation results of ACSNet (Ours), (d) UNet, (e) Unet++, (f) FPN, (g) FCN, (h) DeeplabV3+, (i) TransUnet, (j) CUNet, (k) Qu et al. [49], (l) Chen et al. [48], and (m) Wang et al. [47] respectively.

segmentation performance compared to single-task segmentation methods. Visual results from breast lesion segmentation experiments highlight the effectiveness of our proposed method, showing improved performance compared to TransUnet. The gate unit and the spatial attention mechanism play critical roles in overcoming challenges such as irregular boundaries and variable morphologies, contributing to the observed improvement. In the segmentation task, ACSNet excels on the Dice and Jaccard metrics, demonstrating its ability to accurately identify and delineate tumor regions. The robust performance observed highlights the effectiveness of our advanced feature extraction and segmentation techniques in improving the accuracy of breast cancer diagnosis.

In classification tasks, our model demonstrates high accuracy and precision, both of which are critical for clinical diagnosis. This suggests that ACSNet not only accurately segments tumor regions, but also provides important information about the nature of the tumors, thereby helping to formulate treatment strategies. Despite its state-of-the-art performance, ACSNet has some limitations. Firstly, there may be a classification bias due to the imbalance between benign and malignant tumor samples. Secondly, the manual adjustment of weights in the multi-task loss function requires further research into more intelligent methods. Thirdly, segmentation challenges include over- and under-segmentation of complex BUS images, which requires the development of an effective boundary detection module. Future research directions include the development of a unified large-model approach integrated with a CAD system to accelerate the diagnosis of tumor diseases and to address clinical needs in ultrasound imaging tasks.

7. Conclusion

In this paper, we have presented ACSNet, a multi-task learning network designed for simultaneous segmentation and classification of BUS images. Our network incorporates a multi-scale feature connectivity network for classification, and integrates a channel attention mechanism to enhance classification features. For segmentation, we designed the DSAModule, which uses deformable convolutions to implement a spatial attention mechanism that captures features representing tumor morphological changes. In addition, our gate unit module enables the selective transfer of valuable contextual information from the encoder to the decoder, thereby reducing the impact of redundant or misleading information on the results. We evaluated the effectiveness of ACSNet using two publicly available breast ultrasound image datasets. The experimental results showed that the proposed ACSNet not only outperformed mainstream multi-task learning methods for both breast tumor segmentation and classification tasks, but also achieved state-of-the-art results for BUS tumor segmentation.

CRediT authorship contribution statement

Qiqi He: Writing – original draft, Methodology, Conceptualization. **Qiuju Yang:** Writing – review & editing, Supervision, Resources, Funding acquisition. **Hang Su:** Validation, Supervision, Investigation. **Yixuan Wang:** Visualization, Validation, Investigation.

Declaration of competing interest

All 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.

Acknowledgments

This work was supported by the Natural Science Basic Research Plan in Shaanxi Province of China under Grant 2023-JC-YB-228, and the Open Fund of State Key Laboratory of Loess and Quaternary Geology under Grant SKLLQGZR2201.

References

[1] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, Ca - Cancer J. Clin. 71 (2021) 209–249.
[2] Y. Zhou, H. Chen, Y. Li, Q. Liu, X. Xu, S. Wang, P.-T. Yap, D. Shen, Multi-task learning for segmentation and classification of tumors in 3D automated breast ultrasound images, Med. Image Anal. 70 (2021) 101918.
[3] K. Yang, A. Suzuki, J. Ye, H. Nosato, A. Izumori, H. Sakanashi, Multi-task learning with consistent prediction for efficient breast ultrasound tumor detection, in: 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2022, pp. 3201–3208.
[4] Y. Chen, X.-H. Yang, Z. Wei, A.A. Heidari, N. Zheng, Z. Li, H. Chen, H. Hu, Q. Zhou, Q. Guan, Generative adversarial networks in medical image augmentation: a review, Comput. Biol. Med. 144 (2022) 105382.
[5] M. Rostami, M. Oussalah, K. Berahmand, V. Farrahi, Community detection algorithms in healthcare applications: a systematic review, IEEE Access 11 (2023) 30247–30272.
[6] J. Zhou, Z. Wu, Z. Jiang, K. Huang, K. Guo, S. Zhao, Background selection schema on deep learning-based classification of dermatological disease, Comput. Biol. Med. 149 (2022) 105966.
[7] J. Wang, D. Wu, Y. Gao, X. Wang, X. Li, G. Xu, W. Dong, Integral real-time locomotion mode recognition based on GA-CNN for lower limb exoskeleton, JBE 19 (2022) 1359–1373.
[8] Y. Wang, T. Bai, T. Li, L. Huang, Osteoporotic vertebral fracture classification in X-rays based on a multi-modal semantic consistency network, JBE 19 (2022) 1816–1829.
[9] C.-f. Chen, Z.-j. Du, L. He, Y.-j. Shi, J.-q. Wang, W. Dong, A novel gait pattern recognition method based on LSTM-CNN for lower limb exoskeleton, JBE 18 (2021) 1059–1072.
[10] Q. Guan, Y. Chen, Z. Wei, A.A. Heidari, H. Hu, X.-H. Yang, J. Zheng, Q. Zhou, H. Chen, F. Chen, Medical image augmentation for lesion detection using a texture-constrained multichannel progressive GAN, Comput. Biol. Med. 145 (2022) 105444.
[11] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: N. Navab, J. Hornegger, W.M. Wells, A.F. Frangi (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Springer International Publishing, Cham, 2015, pp. 234–241.
[12] R. Almajalid, J. Shan, Y. Du, M. Zhang, Development of a deep-learning-based method for breast ultrasound image segmentation, in: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 2018, pp. 1103–1108.
[13] Z. Ning, S. Zhong, Q. Feng, W. Chen, Y. Zhang, SMU-net: saliency-guided morphology-aware U-net for breast lesion segmentation in ultrasound image, IEEE Trans. Med. Imag. 41 (2022) 476–490.
[14] X. Zhao, Y. Pang, L. Zhang, H. Lu, L. Zhang, Suppress and balance: a simple gated network for salient object detection, in: A. Vedaldi, H. Bischof, T. Brox, J.-M. Frahm (Eds.), Computer Vision – ECCV 2020, Springer International Publishing, Cham, 2020, pp. 35–51.
[15] S. Feng, H. Zhao, F. Shi, X. Cheng, M. Wang, Y. Ma, D. Xiang, W. Zhu, X. Chen, CPFNet: context pyramid fusion network for medical image segmentation, IEEE Trans. Med. Imag. 39 (2020) 3008–3018.
[16] M.I. Daoud, S. Abdel-Rahman, R. Alazrai, Breast ultrasound image classification using a pre-trained convolutional neural network, in: 2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), 2019, pp. 167–171.
[17] M. Saad, M. Ullah, H. Afridi, F.A. Cheikh, M. Sajjad, BreastUS: vision transformer for breast cancer classification using breast ultrasound images, in: 2022 16th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), 2022, pp. 246–253.
[18] W. Yang, S. Zhang, Y. Chen, W. Li, Y. Chen, Measuring shape complexity of breast lesions on ultrasound images, in: Medical Imaging 2008: Ultrasonic Imaging and Signal Processing, SPIE, 2008, pp. 169–178.
[19] Y. Zhang, Q. Yang, An overview of multi-task learning, Natl. Sci. Rev. 5 (2018) 30–43.
[20] B. Shareef, M. Xian, A. Vakanski, H. Wang, Breast ultrasound tumor classification using a hybrid multitask CNN-transformer network, in: H. Greenspan, A. Madabhushi, P. Mousavi, S. Salcudean, J. Duncan, T. Syeda-Mahmood, R. Taylor (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2023, Springer Nature Switzerland, Cham, 2023, pp. 344–353.
[21] G. Zhang, K. Zhao, Y. Hong, X. Qiu, K. Zhang, 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. Surg. 16 (2021) 1719–1725.
[22] M. Xu, K. Huang, X. Qi, Multi-task learning with context-oriented self-attention for breast ultrasound image classification and segmentation, in: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), 2022, pp. 1–5.
[23] X. Zhao, L. Zhang, H. Lu, Automatic Polyp Segmentation via Multi-Scale Subtraction Network, Springer International Publishing, 2021, pp. 120–130.
[24] J. Shan, Y. Wang, H.D. Cheng, Completely automatic segmentation for breast ultrasound using multiple-domain features. 2010 IEEE International Conference on Image Processing, 2010, pp. 1713–1716.
[25] J. Shan, H.D. Cheng, W. Yuxuan, A novel automatic seed point selection algorithm for breast ultrasound images. 2008 19th International Conference on Pattern Recognition, 2008, pp. 1–4.

[26] A. Madabhushi, D.N. Metaxas, Combining low-, high-level and empirical domain knowledge for automated segmentation of ultrasonic breast lesions, IEEE Trans. Med. Imag. 22 (2003) 155–169.
[27] C.-M. Chen, H.H.-S. Lu, Y.-S. Huang, Cell-based dual snake model: a new approach to extracting highly winding boundaries in the ultrasound images, Ultrasound Med. Biol. 28 (2002) 1061–1073.
[28] M. Xian, Y. Zhang, H.D. Cheng, Fully automatic segmentation of breast ultrasound images based on breast characteristics in space and frequency domains, Pattern Recogn. 48 (2015) 485–497.
[29] X. Guofang, M. Brady, J.A. Noble, Z. Yongyue, Segmentation of ultrasound B-mode images with intensity inhomogeneity correction, IEEE Trans. Med. Imag. 21 (2002) 48–57.
[30] B. Liu, H.D. Cheng, J. Huang, J. Tian, X. Tang, J. Liu, Fully automatic and segmentation-robust classification of breast tumors based on local texture analysis of ultrasound images, Pattern Recogn. 43 (2010) 280–298.
[31] C. Xue, L. Zhu, H. Fu, X. Hu, X. Li, H. Zhang, P.-A. Heng, Global guidance network for breast lesion segmentation in ultrasound images, Med. Image Anal. 70 (2021) 101989.
[32] B. Shareef, A. Vakanski, P.E. Freer, M. Xian, ESTAN: enhanced small tumor-aware network for breast ultrasound image segmentation, Healthcare 10 (2022) 2262.
[33] Y. Hu, Y. Guo, Y. Wang, J. Yu, J. Li, S. Zhou, C. Chang, Automatic tumor segmentation in breast ultrasound images using a dilated fully convolutional network combined with an active contour model, Med. Phys. 46 (2019) 215–228.
[34] M.H. Yap, G. Pons, J. Marti, S. Ganau, M. Sentis, R. Zwiggelaar, A.K. Davison, R. Marti, Automated breast ultrasound lesions detection using convolutional neural networks, IEEE Journal of Biomedical and Health Informatics 22 (2018) 1218–1226.
[35] Y. LeCun, B.E. Boser, J.S. Denker, D. Henderson, R.E. Howard, W.E. Hubbard, L. D. Jackel, Handwritten Digit Recognition with a Back-Propagation Network, NIPS, 1989.
[36] L. Zhu, R. Chen, H. Fu, C. Xie, L. Wang, L. Wan, P.-A. Heng, A Second-Order Subregion Pooling Network for Breast Lesion Segmentation in Ultrasound, Springer International Publishing, 2020, pp. 160–170.
[37] A. Vakanski, M. Xian, P.E. Freer, Attention-enriched deep learning model for breast tumor segmentation in ultrasound images, Ultrasound Med. Biol. 46 (2020) 2819–2833.
[38] Q. He, Q. Yang, M. Xie, HCTNet: a hybrid CNN-transformer network for breast ultrasound image segmentation, Comput. Biol. Med. 155 (2023) 106629.
[39] X. Qi, L. Zhang, Y. Chen, Y. Pi, Y. Chen, Q. Lv, Z. Yi, Automated diagnosis of breast ultrasonography images using deep neural networks, Med. Image Anal. 52 (2019) 185–198.
[40] M. Byra, P. Jarosik, A. Szubert, M. Galperin, H. Ojeda-Fournier, L. Olson, M. O'Boyle, C. Comstock, M. Andre, Breast mass segmentation in ultrasound with selective kernel U-Net convolutional neural network, Biomed. Signal Process Control 61 (2020) 102027.
[41] X. Qian, J. Pei, H. Zheng, X. Xie, L. Yan, H. Zhang, C. Han, X. Gao, H. Zhang, W. Zheng, Q. Sun, L. Lu, K.K. Shung, Prospective assessment of breast cancer risk from multimodal multiview ultrasound images via clinically applicable deep learning, Nat. Biomed. Eng. 5 (2021) 522–532.
[42] W. Cui, Y. Peng, G. Yuan, W. Cao, Y. Cao, Z. Lu, X. Ni, Z. Yan, J. Zheng, FMRNet: a fused network of multiple tumoral regions for breast tumor classification with ultrasound images, Med. Phys. 49 (2022) 144–157.
[43] Z. Zhuang, Y. Kang, A.N. Joseph Raj, Y. Yuan, W. Ding, S. Qiu, Breast ultrasound lesion classification based on image decomposition and transfer learning, Med. Phys. 47 (2020) 6257–6269.
[44] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (2010) 1345–1359.
[45] Y. Mo, C. Han, Y. Liu, M. Liu, Z. Shi, J. Lin, B. Zhao, C. Huang, B. Qiu, Y. Cui, L. Wu, X. Pan, Z. Xu, X. Huang, Z. Li, Z. Liu, Y. Wang, C. Liang, HoVer-trans: anatomy-aware HoVer-transformer for ROI-free breast cancer diagnosis in ultrasound images, IEEE Trans. Med. Imag. (2023) 1, 1.
[46] M. Xu, K. Huang, X. Qi, A regional-attentive multi-task learning framework for breast ultrasound image segmentation and classification, IEEE Access 11 (2023) 5377–5392.
[47] P. Wang, V.M. Patel, I. Hacihaliloglu, Simultaneous Segmentation and Classification of Bone Surfaces from Ultrasound Using a Multi-Feature Guided CNN, Springer International Publishing, 2018, pp. 134–142.
[48] C. Chen, W. Bai, D. Rueckert, Multi-task Learning for Left Atrial Segmentation on GE-MRI, Springer International Publishing, 2019, pp. 292–301.
[49] H. Qu, G. Riedlinger, P. Wu, Q. Huang, J. Yi, S. De, D. Metaxas, Joint Segmentation and Fine-Grained Classification of Nuclei in Histopathology Images, IEEE.
[50] J. Hu, L. Shen, G. Sun, Squeeze-and-Excitation networks, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.
[51] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, 2014 arXiv preprint arXiv:1409.1556.
[52] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.
[53] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, Y. Wei, Deformable convolutional networks, IEEE.
[54] F. Yuan, L. Zhang, X. Xia, Q. Huang, X. Li, A gated recurrent network with dual classification assistance for smoke semantic segmentation, IEEE Trans. Image Process. 30 (2021) 4409–4422.
[55] P.-T. De Boer, D.P. Kroese, S. Mannor, R.Y. Rubinstein, A tutorial on the cross-entropy method, Ann. Oper. Res. 134 (2005) 19–67.
[56] W. Al-Dhabyani, M. Gomaa, H. Khaled, A. Fahmy, Dataset of breast ultrasound images, Data Brief 28 (2020) 104863.
[57] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, PyTorch: an imperative style, high-performance deep learning library, in: Proceedings of the 33rd International Conference on Neural Information Processing Systems, Curran Associates Inc., 2019. Article 721.
[58] B. Chen, Y. Liu, Z. Zhang, G. Lu, A.W.K. Kong, TransAttUnet: multi-level attention-guided U-net with transformer for medical image segmentation, IEEE Transactions on Emerging Topics in Computational Intelligence (2023) 1–14.
[59] H.Y. Zhou, J. Guo, Y. Zhang, X. Han, L. Yu, L. Wang, Y. Yu, nnFormer: volumetric medical image segmentation via a 3D transformer, IEEE Trans. Image Process. 32 (2023) 4036–4045.
[60] B.N. Li, C.K. Chui, S. Chang, S.H. Ong, A new unified level set method for semi-automatic liver tumor segmentation on contrast-enhanced CT images, Expert Syst. Appl. 39 (2012) 9661–9668.
[61] E. Shelhamer, J. Long, T. Darrell, Fully convolutional networks for semantic segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2017) 640–651.
[62] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation, Springer International Publishing, 2018, pp. 833–851.
[63] T.Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 936–944.
[64] Z. Zhou, M.M. Rahman Siddiquee, N. Tajbakhsh, J. Liang, UNet++: A Nested U-Net Architecture for Medical Image Segmentation, Springer International Publishing, 2018, pp. 3–11.
[65] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, Alan, Y. Zhou, arXiv Preprint Server, in: TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation, 2021.