RESEARCH ARTICLE

Incorporating Tumor Edge Information for Fine-Grained BI-RADS Classification of Breast Ultrasound Images

MENG XU[1], (Member, IEEE), JIANHUA HUANG[2], KUAN HUANG[1], (Member, IEEE), AND FEIFEI LIU[3,4]

[1]Department of Computer Science and Technology, Kean University, Union, NJ 07083, USA
[2]School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China
[3]Department of Ultrasound, Binzhou Medical University Hospital, Binzhou 256603, China
[4]Department of Ultrasound, Peking University People's Hospital, Beijing 100044, China

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

ABSTRACT Breast Ultrasound (BUS) imaging is an essential tool for the early detection of breast cancer. The Breast Imaging Reporting and Data System (BI-RADS) in BUS images helps standardize the interpretation and reporting process by categorizing breast tumors into multiple classes, which enables radiologists to make more accurate diagnoses and treatment plans. However, most existing classification methods distinguish only between benign and malignant categories. In addition, features extracted by classic convolutional neural networks tend to be insufficient when subdividing BUS images into fine-grained BI-RADS classes, as they typically do not consider prior knowledge in medical applications, such as foreground shape. To address the above problems, we propose a novel fine-grained BI-RADS classification approach that integrates tumor edges to provide more efficient discriminative features. Firstly, weakly supervised pseudo-label generation: we detect coarse tumor edge regions utilizing a pre-trained PiDiNet and two novel loss functions based on prior knowledge from our dataset. The detected tumor edges are subsequently used as pseudo-labels for the next step. Secondly, co-training a tumor edge detection network and a BI-RADS classification network: edge images generated by the edge detection network are used as weight masks to highlight tumor edge regions as discriminative parts for better classification results, especially for categories with high similarities. The proposed method is evaluated on a BUS image dataset of 1061 images with BI-RADS categories. Experimental results indicate that the proposed method significantly improves over the baseline model by 4.73% in terms of top-1 accuracy.

INDEX TERMS Weakly supervised learning, breast ultrasound, deep learning, BI-RADS classification, edge detection.

I. INTRODUCTION

Breast cancer is one of the most frequently diagnosed cancers among women [1]. Mammography and ultrasound are commonly used for the early detection of breast cancer, and magnetic resonance imaging has been reported to reveal additional information when necessary or for screening patients at high risk [2]. Breast ultrasound (BUS) imaging is a popular tool for the early detection of breast cancer due to its affordability and radiation-free characteristics [3], [4]. In addition, it reports the highest specificity based on breast density among the aforementioned diagnostic methods [1]. However, because of its imaging characteristics, such as low contrast, speckle-noise, and acoustic shadows, BUS images are difficult to interpret and therefore require experienced radiologists to make an accurate diagnosis [5],

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

TABLE 1. BI-RADS categories of breast ultrasound images [8], [9].

BI-RADS categories	Diagnosis
Category 0	Incomplete or inconclusive studies that require additional imaging or evaluation.
Category 1	Negative (no mass)
Category 2	Benign mass
Category 3	Probably benign mass (<2% probability of malignancy)
Category 4a	Low suspicion of being cancerous mass (2% - 10 % probability of malignancy)
Category 4b	Moderate suspicion of being cancerous mass (10% - 50% probability of malignancy)
Category 4c	High suspicion of being cancerous mass (50% - 95 % probability of malignancy)
Category 5	Highly suggestive of malignancy (>95% probability of malignancy)

[6]. Computer-Aided Diagnosis (CAD) systems can help the understanding of BUS images, offer a second opinion, and improve radiologists' diagnoses [7].

Radiologists commonly analyze BUS images according to the Breast Imaging Reporting and Data System (BI-RADS) [8] developed by the American College of Radiology. The BI-RADS system provides a set of standardized descriptors to classify the characteristics of breast tumors, such as shape, margin, echo pattern, and presence of calcifications. These descriptors help to standardize the interpretation of BUS images, which is helpful for the early detection and diagnosis of breast cancer. Table 1 introduces BI-RADS categories of BUS images.

Most existing CAD systems for BUS image classification [10], [11], [12], [13], [14] primarily aim at differentiating between benign and malignant tumors. This binary classification, however, only partially matches the nuanced assessments conducted by radiologists. Specifically, the benign category encompasses both BI-RADS 2 (benign tumors) and BI-RADS 3 (probably benign tumors) [15], distinctions critical for radiologists when determining the necessity and timing of further follow-ups or biopsy studies for patients. Yet, there needs to be more research focused on categorizing BUS images according to BI-RADS levels [9], [16], [17], [18].

In addition, automated BI-RADS classification of BUS images is a challenging task. It has more categories to classify compared to binary classification. With limited features in BUS images, the increased number of categories reduces inter-class variance and increases intra-class variance. Specifically, the binary BUS image classification task primarily focuses on distinctly different features between two classes (benign and malignant). However, the BI-RADS classification involves a finer categorization that reduces the inter-class variance, as the differences between adjacent BI-RADS categories are subtler than the binary classification. Additionally, BI-RADS classification increases the intra-class variance within each BI-RADS category as each category now encompasses more nuanced and diverse characteristics that are not differentiated in the binary model. For example, as shown in Figure 1, BI-RADS categories 2 and 3 tumors share some common characteristics, such as round shapes, clear boundaries, and growth in the mammary layer.

In binary classification, they are both classified into benign tumors in nearly all cases. In BI-RADS classification, more subtle and discriminative features are needed to distinguish them into BI-RADS categories 2 and 3. Similarly, tumors of BI-RADS categories 4 and 5 both have irregular shapes, unclear boundaries, and invasion of other layers. BI-RADS categories 4a, 4b, and 4c are particularly hard to classify as they have very close characteristics in shape, boundary, gray-level intensities, and sizes, making it difficult for human observers to distinguish between them. In binary classification, they are classified into either benign or malignant. But in BI-RADS classification, they need to be classified into four classes (BI-RADS categories 4a, 4b, 4c, and 5). The above observations are obtained by analyzing the BI-RADS categories diagnosed by radiologists and the biopsy results of all BUS images in our dataset. In summary, BI-RADS classification is more challenging than binary classification because a more precise method is needed to extract more subtle and discriminative features from BUS images.

To obtain more discriminative features, some classification methods indicate the effectiveness of utilizing prior medical knowledge, such as characteristics of tumor boundary and background region in BUS images [7], [19], [20]. However, most existing BI-RADS classification methods do not consider prior medical knowledge [16], [17], [18]. In addition, obtaining prior medical knowledge from BUS images often requires segmentation labels during the training phase, which are difficult to acquire due to the need to involve expert doctors and radiologists [20]. Huang et al. [9] propose to use a pre-trained region proposal network to get tumor marginal mask, central mask, and outer mask, which does not require segmentation labels. However, directly applying a pre-trained region proposal network to BUS images without fine-tuning does not guarantee generating accurate tumor boundaries. To the best of our knowledge, there is a notable gap in the literature regarding obtaining medical prior knowledge from BUS images using weakly supervised learning for the fine-grained BI-RADS classification of breast ultrasound images.

To address the aforementioned problems, we propose a novel framework for the fine-grained BI-RADS classification of BUS images. It incorporates prior medical knowledge in tumor edge regions, which contains crucial discriminative
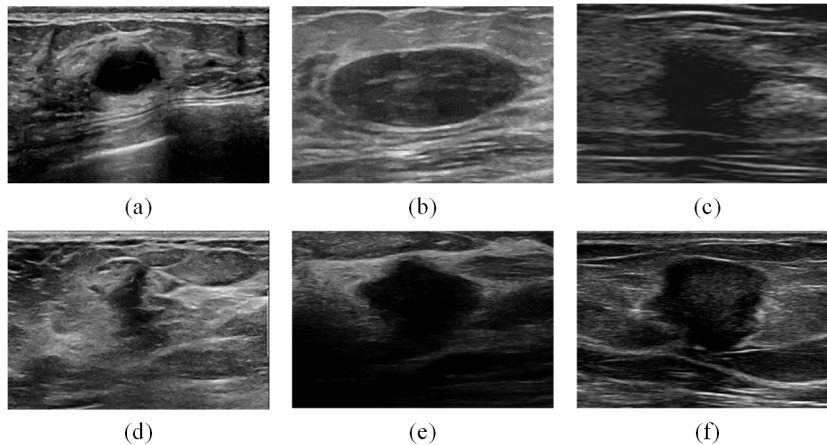
FIGURE 1. BUS images in different BI-RADS categories: (a) BI-RADS category 2, (b) BI-RADS category 3, (c) BI-RADS category 4a, (d) BI-RADS category 4b, (e) BI-RADS category 4c, (f) BI-RADS category 5.

features for tumor classification [8], [21], into a neural network framework to improve the overall BI-RADS classification results. Tumor edge regions are obtained in a weakly supervised manner without using edge segmentation labels. Our major contributions include:

- Proposing a novel framework for the BI-RADS classification of BUS images, which aligns with radiologists' assessment from a practical perspective [16], [17], [18]. The proposed method incorporates discriminative tumor edge information to boost the classification performance.
- Proposing a denoise loss function and an ellipse-fitting loss function for fine-tuning a pre-trained Pixel Difference Network (PiDiNet) [22] on the BUS image dataset in a weakly supervised manner without tumor edge segmentation labels. The proposed two loss functions greatly improve the tumor edge detection accuracy of the PiDiNet.
- Developing a novel co-training framework consisting of a tumor edge detection network and a BI-RADS classification network. The proposed co-training framework significantly improves the BI-RADS classification performance by employing edge images generated by the tumor edge detection network as weight masks to highlight tumor edge regions in BUS images.
- Conducting extensive experiments to demonstrate the superiority of the proposed method over state-of-the-art Fine-Grained Image Classification (FGIC) methods.

II. RELATED WORKS

A. BINARY BUS IMAGE CLASSIFICATION

The concept of using machine learning and digital image processing techniques to help doctors and radiologists make accurate breast cancer diagnoses can be traced back to the 1990s [23]. Binary BUS image classification between benign and malignant categories is crucial for providing diagnostic recommendations. Computer-aided diagnosis methods for

BUS image classification can be mainly divided into conventional machine learning and deep learning-based approaches. Conventional machine learning methods include linear discriminant analysis [24], Support Vector Machine (SVM) [25], and Gaussian Mixture Models (GMMs) [11]. They often comprise four steps: image pre-processing, tumor segmentation, feature extraction and selection, and tumor classification [26]. The classification result relies highly on the quality of tumor segmentation and manually selected features. The handcrafted features are sensitive to datasets. Furthermore, they require tumor segmentation labels for locating breast tumors. The performance of these methods is not reliably consistent when the dataset is changed.

With the development of deep Convolutional Neural Networks (CNNs) in image classification, more research is on deep learning-based BUS image classification, which can be categorized into three major directions: (1) network structure design, (2) incorporating prior knowledge via attention mechanisms or multi-task learning, and (3) combining deep learning and conventional machine learning methods. Network structure design methods, for example, Daoud et al. [27] employ a pre-trained CNN model to achieve accurate binary classification of BUS images. Xie et al. [28] propose a dual-sampling network structure combining the traditional convolutional and residual networks for binary BUS image classification. Deep CNNs outperform conventional machine learning methods and have higher robustness because they do not need manual feature extraction and have strong feature representation abilities. However, BUS images have distinctive imaging characteristics that can serve as valuable prior knowledge to increase classification accuracy, while most existing deep CNN methods do not incorporate them.

Auxiliary networks and attention mechanisms have incorporated prior knowledge to improve the classification results. For example, Xing et al. [29] utilize a spatial attention mechanism comprising instructive BI-RADS information into the network to enhance binary BUS image classification.

Liao et al. [30] propose a supervised block-based region segmentation algorithm to segment tumor regions and use a pre-trained VGG-19 for tumor classification. They incorporate strain elastography and BUS image features to improve classification accuracy. In addition, several preliminary studies have demonstrated that tumor, peritumoral (the tumor-adjacent area surrounding the tumor), and background regions in BUS images have a high correlation to BUS image classification results [20], [31]. For example, Xu et al. [7] develop a multi-task learning framework for simultaneous BUS image segmentation and classification, which uses tumor segmentation results as prior knowledge to improve classification. These deep learning-based methods exhibit high performance and robustness, but insufficient training samples limit their generalization ability.

Many methods combine conventional machine learning and deep learning to get higher performance with limited training samples. For example, Zhuang et al. [13] propose to combine characteristic image features (orientation, edge indistinctness, characteristics of posterior shadowing region, and shape complexity) with deep learning features extracted by VGG-16. Huang et al. [11] use ResNet-101 and VGG-16 to extract convolutional features and build two GMMs to classify benign and malignant tumors according to the extracted features. Experimental results of these methods show that the combination of conventional machine learning and deep learning outperformed each of them individually.

In summary, binary BUS image classification methods have been well studied. They design robust network architectures, incorporate prior knowledge via attention mechanisms or use multi-task learning, or combine conventional machine learning and deep learning to get better classification accuracy. However, fine-grained BI-RADS classification is more useful in practice, but research on it is insufficient.

B. FINE-GRAINED IMAGE CLASSIFICATION

Automated BI-RADS classification of BUS images is under the category of Fine-Grained Image Classification (FGIC), which refers to the task of categorizing images into detailed subcategories within a broader class [32], [33]. BI-RADS classification of BUS images presents a more significant challenge than binary classification due to the minor differences between classes and significant variations within the same class. Addressing this issue requires focusing on extracting discriminative patterns for different categories within BUS images, which can be achieved through two primary methods: 1) localizing discriminative patterns and 2) representing them in a way that highlights their differences.

In the first category, localization methods are integrated into classification networks to locate regions containing discriminative patterns. For example, Angelova and Zhu [34] first perform object region localization, followed by object rescaling and centering. They subsequently extract features from both original and object-centralized images and concatenate these features for fine-grained image classification.

Zhang et al. [35] develop a part-based R-CNN method to extract the discriminating parts on the whole image; then, the features from the entire image and discriminating parts are used together for FGIC. Wei et al. [36] use a novel Mask-CNN [37] model without the fully connected layers to locate the discriminative parts in images. Part detection scores are used to aggregate object- and part-level convolutional descriptors to improve the FGIC performance.

Detecting discriminative parts/objects and extracting features from these parts is the most straightforward way to improve FGIC. However, this approach is limited by the cost of obtaining annotations for these parts/objects. Weakly supervised object detection using image-level labels is a solution for the annotation problem. For example, Ge et al. [38] propose an FGIC method incorporating a weakly supervised image segmentation and object detection method. This method utilizes CAMs to initialize segmentation probability maps to be used in a conditional random field to extract higher-quality object instances. Liu et al. [39] propose a filtration and distillation learning model that matches predictions and proposal confidences to optimize region proposal for weakly supervised region detection. Santra et al. [40] develop an annotation-free part-based FGIC method. Discriminative parts are detected by unsupervised keypoint part proposal generation and bilinear pooling.

The second category method, discriminative pattern representation, often directly builds an end-to-end feature encoding module to highlight discriminative regions. For example, Wang et al. [41] proposes an FGIC method that enhances the mid-level learning capability of the classical CNN by introducing a bank of 1×1 convolutional filters. They use a destruction and construction learning model to break the order of regions in images and reconstruct the order to enhance the discriminative regions.

In summary, existing research indicates the significance of discriminative regions in improving the performance of FGIC. However, accurately locating these regions often necessitates pixel-level or key-point annotations. While end-to-end feature representation methods have lower annotation requirements, they may not always focus on the correct discriminative regions. For our specific task, BI-RADS classification of BUS images, we can leverage the prior knowledge that breast tumors generally have an approximately elliptical shape and that tumor edge contains crucial discriminative information for tumor classification. Considering that an exact tumor edge location is not necessarily required, we design a weakly supervised approach for tumor edge detection to improve the BI-RADS classification accuracy.

III. MATERIALS AND METHOD
A. DATASET
We collected a BUS image dataset for the BI-RADS classification with 1061 BUS images. These images were collected by various healthcare institutions, including Peking
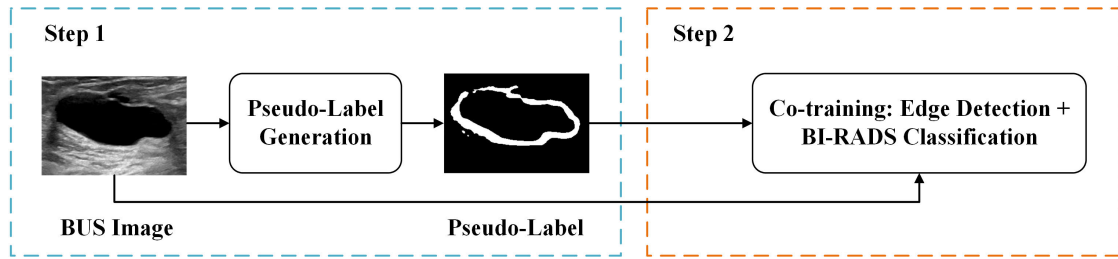
FIGURE 2. An overview of the proposed BUS image BI-RADS classification method.

University People's Hospital, Southeast University Zhongda Hospital, the First Affiliated Hospital of the Guangxi University of Chinese Medicine, and the First Affiliated Hospital of Zhengzhou University. The ethics committees of the four healthcare institutions have granted approval for this study. The dataset cannot be publicly shared without the explicit authorization of these institutions. Ultrasonic equipment used to collect these BUS images includes Aplio 500, GE Logic E9, Philips IU22, and Siemens S3000. The images initially have arbitrary sizes, approximately 600 × 400. Following the ACR BI-RADS® Atlas Fifth Edition, two experienced radiologists, who were unaware of the pathological results, assessed the suspicion of malignancy for each lesion independently and labeled the dataset for six BI-RADS categories. Both radiologists had over ten years of expertise in breast ultrasound diagnosis. The dataset also has actual diagnoses (benign or malignant) from the biopsy study. Six BI-RADS categories of the dataset include BI-RADS categories 2 (14 images), 3 (346 images), 4a (75 images), 4b (266 images), 4c (216 images), and 5 (144 images). The dataset has an imbalanced number of BUS images of each category. Therefore, several image augmentation techniques are used to alleviate this problem. Details are provided in section IV-C.

B. METHOD OVERVIEW

Figure 2 illustrates the overview of the proposed method for the BI-RADS classification of BUS images. The proposed method consists of two main steps: Firstly, a weakly supervised tumor edge detection method is employed to generate coarse tumor edges, which are subsequently used as pseudo labels to supervise the training of a co-training framework in the second step. Secondly, a co-training framework is implemented, integrating an edge detection network and an edge information-guided fine-grained image classification network. The edge detection network uses the pseudo labels generated in the first step and iteratively generates more accurate edge images during the co-training. These refined edge images then serve as weight masks, guiding the classification network to focus on the discriminative tumor edge regions. This way, the classification network learns a better feature representation for improved classification results.

C. TUMOR EDGE PSEUDO-LABEL GENERATION

In the first step, we detect tumor edges in BUS images with transfer learning from a pre-trained PiDiNet [22]. We employ PiDiNet as the tumor edge detector because of its good performance. It utilizes three types of pixel difference convolution (PDC), including central PDC, angular PDC, and radial PDC. They capture rich gradient information and improve edge detection accuracy by analyzing predefined local pixel differences. Specifically, PiDiNet calculates the differences between each pixel and its neighboring pixels in three different directions, using these differences in a convolutional manner to detect edges effectively.

The PiDiNet is originally a fully supervised learning method for edge detection. In our task, we use it in a weakly supervised manner because of the difficulty of acquiring tumor edge annotations in practical applications. Meanwhile, our specific task does not require precise tumor edge detection in each pixel, as the discriminative features exist within a small ring region encompassing the tumor edge. Thus, a coarse tumor edge detection is sufficient to provide the necessary discriminative information.

Specifically, we employ a PiDiNet pre-trained on BSDS500 (Berkeley Segmentation Data Set and Benchmark) dataset [42] and NYUDv2 (New York University Depth Dataset V2) [43]. Directly applying the PiDiNet pre-trained on natural images to BUS images for tumor edge detection in BUS images can lead to inaccurate results. To enhance the pre-trained PiDiNet's ability to predict breast tumor edges accurately, we propose fine-tuning it on the BUS image dataset with two novel loss functions: a denoise loss and an ellipse-fitting loss. By incorporating these two loss terms, PiDiNet achieves satisfactory accuracy in detecting tumor edges. The fine-tuned PiDiNet is then used to generate coarse tumor edge images as pseudo labels for supervising the training of the proposed co-training framework in section III-D. Figure 3 illustrates the fine-tuning process of PiDiNet and the generation of tumor edge pseudo-labels.

Let f represent the PiDiNet, and $I \in \mathbb{R}^{H \times W}$ represent an input grayscale BUS image, where H and W represent its height and width, respectively. The PiDiNet produces a probability map indicating the likelihood of edge regions in the input image I by:
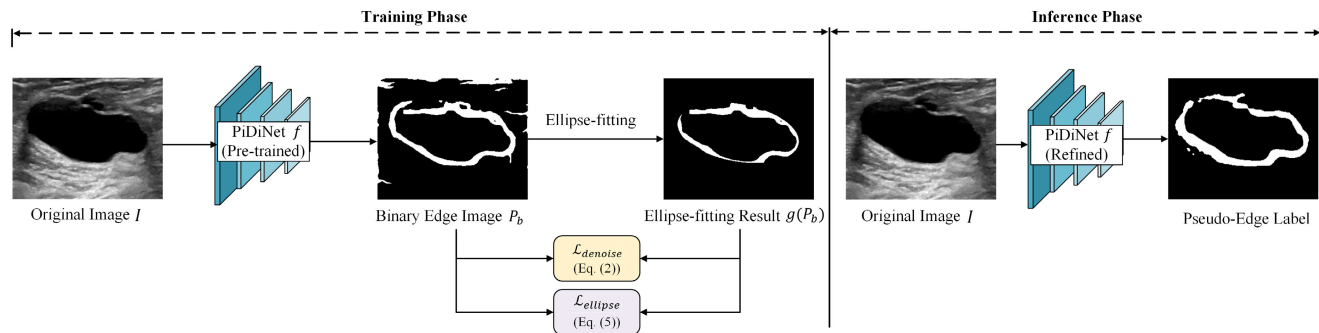
$$P = \sigma(f(I)) \tag{1}$$

FIGURE 3. Illustration of the fine-tuning process of PiDiNet.

where σ is a Sigmoid function and $P \in \mathbb{R}^{H \times W}$ denotes the probability map for edge. We design a denoise loss that calculates the Euclidean distance between all pixels marked as edge points and the current centroid of all edge pixels. We assume that the larger the distance, the less likely a pixel is considered an edge point. This loss aims to minimize the distance to reduce noise pixels and encourage the formation of clear edge patterns. The proposed denoise loss can be mathematically formulated as follows:

$$\mathcal{L}_{denoise} = \sum_{P(x,y)>\lambda} \|(x, y) - center\|_2 \tag{2}$$

where $P(x, y)$ represents the probability of a pixel at coordinates (x, y) in the image is an edge pixel, $center = (x_c, y_c)$ represents the coordinates of the centroid of all edge pixels, and $\lambda = 0.5$ represents the threshold used to convert the probability map into a binary edge image. Coordinates of the centroid are calculated by:

$$x_c = \frac{\sum\limits_{P(x,y)>\lambda} P(x, y) \times x}{|P(x, y) > \lambda|} \tag{3}$$

$$y_c = \frac{\sum\limits_{P(x,y)>\lambda} P(x, y) \times y}{|P(x, y) > \lambda|} \tag{4}$$

where $|P(x, y) > \lambda|$ represents the total number of pixels identified as edge pixels after applying the threshold λ.

Additionally, we observed that breast tumors generally have approximately elliptical shapes, and the majority of BUS images generally contain a single tumor. Based on these observations, we design an ellipse-fitting loss to guide the PiDiNet to learn elliptical edge patterns. Specifically, we first fit an ellipse on the binary edge image predicted by the PiDiNet. We then calculate a cross-entropy loss between the predicted edge probability image and the ellipse-fitted edge image. The ellipse-fitting loss is formulated as follows:

$$\mathcal{L}_{ellipse} = CrossEntropy(P, g(P_b)) \tag{5}$$

where P represents an edge probability map produced by PiDiNet (Eq. (1)), P_b represents the binary edge image after applying the threshold $\lambda = 0.5$, and $g(\cdot)$ represents the ellipse-fitting operation. After applying the ellipse-fitting operation, $g(P_b)$ is a binary image with a better edge

pattern. In this study, we adopt the least squares ellipse fitting algorithm for $g(\cdot)$, where its objective function can be formulated as follows:

$$\min \sum \left[\left(\frac{(x - x_c) \cdot \cos(\theta) + (y - y_c) \cdot \sin(\theta)}{a} \right)^2 + \left(\frac{(x - x_c) \cdot \sin(\theta) - (y - y_c) \cdot \cos(\theta)}{b} \right)^2 - 1 \right]^2 \tag{6}$$

where (x, y) represents the coordinates of an edge pixel in the binary edge image P_b; (x_c, y_c) represents the centroid coordinates of the fitted ellipse; a and b represents the semi-major and semi-minor axis length of the fitted ellipse, respectively; and θ represents the orientation angle of the fitted ellipse.

The initial centroid (x_c, y_c) is determined as the centroid of all edge pixels in a binary edge map. Then, data normalization is performed by subtracting the centroid from each edge pixel. Next, Principal Component Analysis (PCA) is applied to the set of edge pixels to calculate the eigenvectors and eigenvalues of the data distribution, which determines a, b, and θ. Specifically, the eigenvectors corresponding to the largest and smallest eigenvalues are selected as the major and minor axes, respectively. The lengths of the major and minor axes, a and b, are determined by scaling the square root of the eigenvalues and multiplying the results by 2. Last, the orientation angle θ is derived from the major axis eigenvector by calculating the angle between the major and horizontal axes. After obtaining the values of a, b, and θ for a predicted binary edge image, we estimate an ellipse on it by optimizing Eq. (6). Next, we apply a dilation operation to the obtained tumor edge using a square structuring element of size 5×5. This dilation operation ensures that the tumor edge has a sufficient width to provide enough discriminative information.

In summary, in the first step of the proposed framework, we employ a PiDiNet pre-trained on two natural image datasets and fine-tune it on the BUS image dataset using the proposed denoise loss and the ellipse-fitting loss for better tumor edge detection. The fine-tuning process is implemented in a weakly supervised manner without tumor edge segmentation labels. The loss function of the fine-tuning
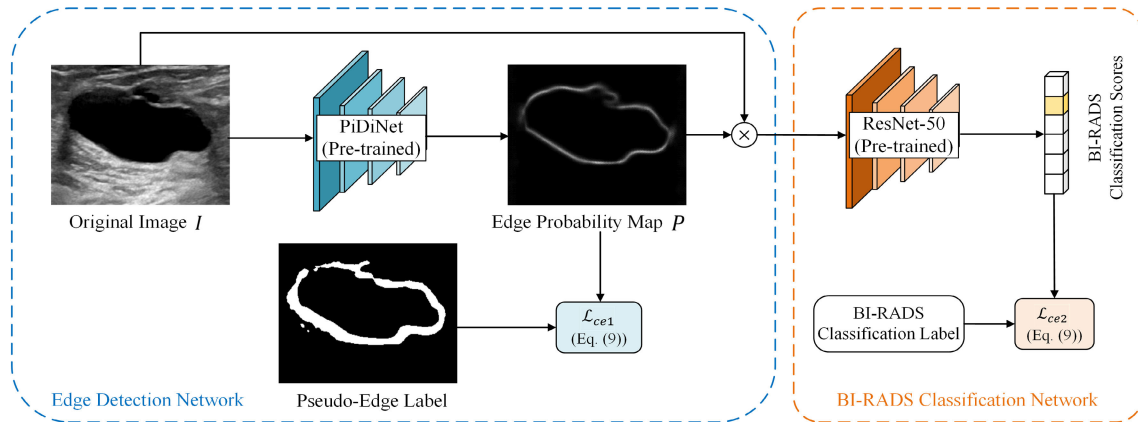
FIGURE 4. Illustration of the proposed co-training framework.

process is defined as follows:

$$\mathcal{L}_{edge} = \mathcal{L}_{denoise} + \mathcal{L}_{ellipse} \tag{7}$$

Last, the fine-tuned PiDiet is used to generate coarse edge images of all BUS images in the dataset. These coarse edge images are subsequently used as pseudo labels for the co-training framework in the next step.

D. CO-TRAINING FRAMEWORK

In the second step, we design a co-training framework for the BI-RADS classification of BUS images. It integrates an edge detection network (PiDiNet) and a BI-RADS classification network (ResNet-50 [44] pre-trained on ImageNet). In the proposed co-training framework, both networks are trained in a fully supervised way. Two networks enhance each other during training, resulting in a more accurate BI-RADS classification.

Figure 4 illustrates the proposed co-training framework, which is trained in a supervised manner. During the training process, PiDiNet uses the coarse edge images generated in section III-C as pseudo labels, and ResNet-50 uses BI-RADS annotations generated by experienced radiologists as labels. Firstly, an input BUS image is fed into PiDiNet to generate an edge probability map. The generated edge probability map, which serves as an edge weight mask, is multiplied with the input image to enhance the tumor edge area. The enhanced image is generated by Eq. (8):

$$I' = P \otimes I \tag{8}$$

where I is an input image, $\otimes$ represents the pixel-wise multiplication, and P is a probability map indicating the likelihood of edge regions in the input image. I' is the edge-enhanced image and is subsequently input to the classification network for BI-RADS classification. Secondly, the weighted input image I' is fed into ResNet-50 for BI-RADS classification.

The edge detection network (PiDiNet) has an edge detection loss, which is calculated by a cross-entropy loss ($\mathcal{L}_{ce1}$) between edge images produced by the PiDiNet in the co-training framework and the pseudo labels generated in Subsection III-C. The BI-RADS classification network

(ResNet-50) has a classification loss, which is a cross-entropy loss ($\mathcal{L}_{ce2}$) between the classification outputs and the BI-RADS classification labels. The total loss of the co-training framework is the sum of the edge detection loss and the classification loss, which can be formalized by:

$$\mathcal{L}_{total} = \mathcal{L}_{ce1} + \mathcal{L}_{ce2} \tag{9}$$

where $\mathcal{L}_{ce1}$ and $\mathcal{L}_{ce2}$ are loss functions of the edge detection network and BI-RADS classification network, respectively.

The quality of tumor edge pseudo-labels generated in section III-C is satisfactory, but there is still room for improvement. Therefore, in this section, we integrate PiDiNet in the co-training network, training it on the pseudo labels and refining its performance along with the classification network. The co-training framework employs tumor edges generated by PiDiNet as weights to enhance tumor edge areas in BUS images, which contain discriminative information crucial for BI-RADS classification. As a result, the classification network produces more accurate classification results, which, in turn, aids PiDiNet in generating more accurate edge information.

IV. EXPERIMENTAL SETUP AND RESULTS
In this section, we first present the experiment settings and performance evaluation metrics. We then describe the data augmentation techniques used in this study and their results. Finally, we present the experimental results of the proposed method and its comparison with the competing methods.

A. EXPERIMENT SETTINGS
The implementation of the proposed method is based on the public platform PyTorch 1.13.1. All experiments are conducted on a Ubuntu 20.04 system, AMD EPYC 7513 2.60 GHz CPU, and 8 NVIDIA GeForce RTX 3090 graphics cards with 24GB memory. In section III-C, a pre-trained model is loaded to initialize the network parameters of PiDiNet. Then, all 1061 images in the BUS dataset are used to fine-tune the network. The Stochastic gradient descent (SGD) optimizer is used with an initial learning rate of 1e-3 and a momentum of 0.9. The training epoch number is set to 40.

The batch size is set to 16. In section III-D, the dataset is randomly split into 852 training images and 209 test images. The SGD optimizer with an initial learning rate of 1e-3 and a momentum of 0.9 is used. The epoch number is set to 80. The patch size is set to 16. The learning rate is reduced to 0.1 times every 20 epochs. Parameter values are selected empirically. The complexity of the proposed method depends on ResNet and PiDiNet as it does not add additional trainable parameters.

B. EVALUATION METRICS

We evaluate the quality of pseudo labels generated in section III-C by the intersection over union (IoU) of the predicted edge images and the edge ground truth manually annotated by experienced radiologists, which is calculated by:

$$IoU = \frac{Prediction \cap GT}{Prediction \cup GT} \qquad (10)$$

where *Prediction* is an edge image generated by PiDiNet and *GT* is its corresponding ground truth image. We also provide some examples of the generated edge pseudo labels in Figure 5.

We evaluate the BI-RADS classification performance of the proposed co-training network in section III-D by Top-1 accuracy, which is calculated by:

$$\text{Top-1 Accuracy} = \frac{NCP}{N} \qquad (11)$$

where *NCP* represents the number of samples whose top-ranked predictions match the true class labels (number of correct predictions), and *N* represents the number of samples in the test set.

C. DATA AUGMENTATION

Medical image datasets are often limited in size because of the difficulty in data collection and annotation [45], [46]. It is hard to train a good deep neural network for BI-RADS classification of BUS images on a small-size dataset. To increase the training data size and improve the network's generalization ability, we adopt multiple data augmentation techniques, including horizontal flip, vertical flip, rotation (less than 90 degrees), center crop, and wavelet transformation [47]. Specifically, the rotation technique performs a random rotation of less than 90 degrees on a BUS image. The center crop technique crops a fixed-size region (256×256) from the central area of a BUS image. The wavelet transformation technique applies a wavelet transformation to a BUS image and generates a three-channel image. The original grayscale image is used as the first channel, the low-frequency information from the wavelet-transformed result is used as the second channel, and the high-frequency information from the wavelet-transformed result is used as the third channel.

To evaluate the performance of these augmentation techniques, we compare the results of using the baseline network

TABLE 2. Results of different data augmentation methods.

Method	Top-1 Accuracy
No augmentation	69.40%
Horizontal flip	69.62%
Vertical flip	67.80%
Rotation (less than 90 degrees)	70.51%
Center crop	70.70%
Wavelet Transformation	70.93%

(ResNet-50) for BI-RADS classification with each individual data augmentation technique applied to the BUS dataset. Grayscale input images of different sizes are normalized to 0 to 1 after each of the aforementioned data augmentation techniques. Results are summarized in Table 2. Specifically, we create modified copies of images in the BUS dataset by applying each individual augmentation technique to the original dataset. For example, we perform a horizontal flip on each image in the dataset and use the original images together with the augmented images to train a ResNet-50 for BI-RADS classification. Then, its Top-1 accuracy is listed in Table 2 as "Horizontal flip."

According to Table 2, all augmentation techniques, except for vertical flip, have a positive influence on the classification result. The classification accuracy drops when applying the vertical flip technique, possibly due to the specific characteristics of BUS images. In BUS images, there are five layers from top to bottom, including the pre-fat background area, fat layer, mammary layer, muscle layer, and retro-muscle layer, and breast tumors are generally located in the mammary layer [48], [49]. Therefore, changing the vertical orientation of BUS images may disrupt the network's learning of specific patterns from BUS images. In this study, we adopt all other augmentation techniques that increase the classification accuracy and integrate them into the proposed method, including horizontal flip, rotation (less than 90 degrees), center crop, and wavelet transformation. We perform these four techniques on each BUS image in the dataset to create four different augmented images. After data augmentation, we train the proposed method on the original images together with the augmented images.

D. EDGE DETECTION RESULTS

In this section, we compare the edge detection performance of two methods: the baseline network (a pre-trained PiDiNet on BSDS500 [42] and NYUDv2 [43]) and the proposed weakly supervised edge detection method (fine-tuning a pre-trained PiDiNet on the BUS image dataset using two proposed loss functions in Eq. (2) and Eq. (5).

Table 3 presents a comparison between the baseline method and the proposed method in terms of edge detection accuracy, as measured by IoU. The proposed method significantly improves the baseline method by 20.44%, which indicates the efficiency of the proposed two loss functions for fine-tuning the baseline method. As our task does not require
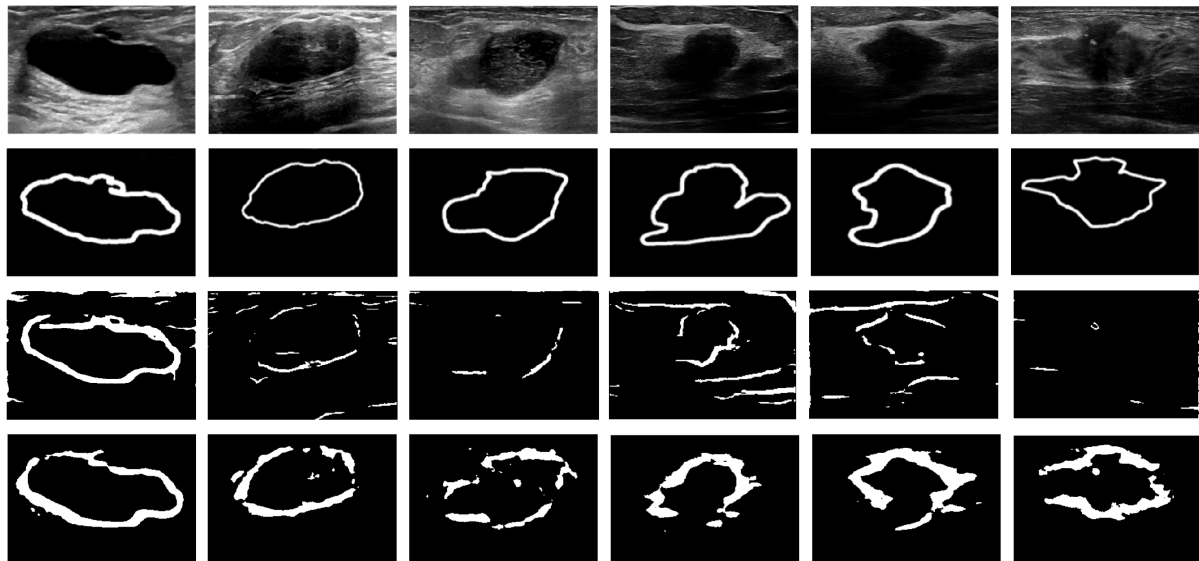
FIGURE 5. Edge detection results of six representative BUS images generated by two compared methods. First row: original BUS images. Second row: tumor edge labels generated by experienced radiologists. Third row: edge detection results generated by the baseline method (a pre-trained PiDiNet). Last row: edge detection results generated by the proposed method (fine-tuning a pre-trained PiDiNet on the BUS image dataset using the proposed two loss functions).

TABLE 3. Comparison of transfer learning-based tumor edge detection methods.

Method	IoU
Pre-trained PiDiNet	43.11%
The proposed method	**63.55**%

TABLE 4. Comparison of different baseline methods.

Methods	Accuracy
DenseNet-121	70.50%
Inception-v3	72.33%
VGG-19	71.83%
ResNet-50	71.51%
Pre-trained ResNet-50	**73.70%**

very accurate tumor edge detection in the first step, coarse tumor edges produced by the proposed method, which has an IoU of 63.55%, are sufficient to serve as pseudo labels for the co-training framework in the second step.

Additionally, Figure 5 displays six representative BUS images in the dataset, along with their corresponding edge annotations created by experienced radiologists, as well as the edge detection results obtained from both the baseline method and the proposed edge detection method. As can be seen from the figure, the baseline method is able to detect tumor edges that are in high contrast to the background (the first column) while failing to detect clear tumor edges for all other BUS images (second column to the last column). Additionally, it detects some background textures which do not align with our goal of identifying breast tumor edges. After the fine-tuning process using two proposed loss functions, the network learns better edge patterns and generates clearer and more accurate edge detection results (the last row).

E. BASELINE NETWORK COMPARISON

To choose a baseline network with the best performance for the BI-RADS classification task, we conduct experiments using a variety of well-known networks commonly used in image classification. These methods include DenseNet-121 [50], Inception-v3 [51], VGG-19 [52], ResNet-50 [44], and ResNet-50 pre-trained on the ImageNet dataset and then

fine-tuned on the BUS dataset. These networks are trained on the augmented dataset. A comparison of the results of different baseline networks is summarized in Table 4.

According to Table 4, the pre-trained ResNet-50 achieves the highest classification accuracy of 73.70% among all compared methods. Additionally, it shows superiority in both speed and accuracy compared to ResNet-50 without pre-training. Therefore, we adopt the pre-trained ResNet-50 as the baseline network of the proposed method for all the following experiments.

F. THREE FORMS OF EDGE ENHANCEMENT

In the co-training framework (section III-D), we use edge images generated by the PiDiNet as weight masks to enhance edge regions in the input BUS images for better classification accuracy. To demonstrate the effectiveness of the proposed edge-enhanced method, we compare the classification performance of the baseline method (a pre-trained ResNet-50) with and without edge enhancement and compare three forms of edge-enhanced methods.

The first edge-enhanced method is to concatenate an input BUS image and its corresponding binary edge image generated by the PiDiNet. Specifically, for a probability map P generated by the PiDiNet, we first convert it to a binary

TABLE 5. Results of three forms of edge-enhanced images and ablation study.

Methods	Accuracy
Baseline (pre-trained ResNet-50)	73.70%
Edge-enhanced method 1	74.24%
Edge-enhanced method 2	74.54%
Edge-enhanced method 3 (proposed)	**78.43%**

edge image P_b by applying a threshold $\lambda = 0.5$. Next, we perform a concatenation operation between the input BUS image $I \in \mathbb{R}^{H \times W}$ and P_b to enhance the edge region in I:

$$I' = Concatenate(I, P_b) \qquad (12)$$

where the output $I' \in \mathbb{R}^{H \times W \times 2}$.

The second edge-enhanced method is to pixel-wise multiply the binary edge image and the input BUS image:

$$I' = P_b \otimes I \qquad (13)$$

where the output I' only keeps pixels of the edge region, and other pixels are set to 0.

The third edge-enhanced method (the proposed method) is to pixel-wise multiply the probability map and the input BUS image, which is described in Eq. (8).

Results of the baseline method and three forms of edge-enhanced methods are shown in Table 5. All three forms of edge-enhanced methods improve the baseline method, which demonstrates the effectiveness of enhancing the edge region in BUS images for BI-RADS classification. Specifically, the first edge-enhanced method improves the baseline method by 0.54%, the second edge-enhanced method improves the baseline method by 0.84%, and the proposed edge-enhanced method improves the baseline method by 4.73%. The results indicate that focusing only on the edge region in BUS images is a little more effective than concatenation. Using an edge probability map as a weight mask proves to be the most effective and robust approach. The proposed method preserves more information from the input BUS images and quantitatively enhances or weakens each pixel within them.

G. COMPARATIVE EXPERIMENTS

To demonstrate the effectiveness of the proposed method, we also compare it with eight recent deep learning-based fine-grained image classification methods. These compared methods include: DFL-CNN [41], NTS-Net [53], DCL [54], LIO [55], IFBRC [17], PIM [56], HERBS [57], and IELT [58]. The first five methods use a ResNet-50 pre-trained on ImageNet as their backbone. PIM and HERBS use Swin Transformer [59], and IELT uses Vision Transformer [60] as the backbone, which are both pre-trained on ImageNet. All methods are fine-tuned on the BUS image dataset to achieve the best BI-RADS classification performance. Results of the comparative experiments are summarized in Table 6.

Among all the compared methods, the proposed method achieves the highest BI-RADS classification accuracy of 78.43%, which improves the second-best method (PIM) by

1.9%. The experiment results indicate that the proposed method has a decisive advantage in the BI-RADS classification of BUS images compared to other fine-grained image classification methods. Its good performance can be attributed to the following reasons: (1) Its classification network (ResNet-50) is pre-trained on a large natural image dataset (ImageNet) and then fine-tuned on the BUS dataset, allowing the proposed method to learn strong initial feature representations and then adapt them to specific characteristics of BUS images for a better BI-RADS classification. (2) The proposed method utilizes edge probability maps as weights to enhance the edge region in BUS images. Breast tumor edges contain discriminative features that are helpful for the BI-RADS classification. Enhancing the edge region guides the network focus more on the discriminative region in BUS images and, therefore, achieves better classification results. (3) Two tasks of the co-training framework promote each other during the training process, resulting in a more accurate BI-RADS classification.

V. DISCUSSIONS
A. ADVANTAGES AND POTENTIAL USEFULNESS
In this study, we propose a novel method incorporating tumor edge information to provide discriminative features for better BI-RADS classification performance of BUS images. In general, the advantages and potential usefulness of the proposed method can be summarized as follows:

First, the proposed method employs a pre-trained edge detection network (PiDiNet) on a large natural image dataset and fine-tunes it on the BUS image dataset in a weakly supervised manner. In this process, two novel loss functions are specifically designed to guide the PiDiNet to learn more apparent and elliptical tumor edge patterns. Although the edge detection results cannot reach the accuracy achieved by fully supervised methods, this approach offers a notable advantage: it does not require any pixel-level segmentation labels for tumor edges during the training, which possibly alleviates the difficulty of obtaining pixel-wise labels for medical images. Experimental results also indicate that breast tumor edge regions can provide important discriminative information for the BI-RADS classification of BUS images. This observation can be generalized to other BUS image classification methods. Moreover, this weakly supervised learning strategy, which uses appropriate loss functions to fine-tune a pre-trained network on a small-size medical image dataset, has the potential to be applied to other medical image tasks.

Second, this study investigates multiple factors that affect the performance of the fine-grained BI-RADS classification of BUS images: (1) The breast tumor edge region contains important discriminative features that are helpful for BI-RADS classification. This study also proposes an efficient way to enhance tumor edge regions by multiplying an edge probability map with the input BUS image. (2) This study compares five baseline methods for the BI-RADS

TABLE 6. Comparison with recent fine-grained image classification methods.

Methods	Backbone	Pre-trained	Accuracy
DFL-CNN [41]	ResNet-50	✓	73.80%
NTS-Net [53]	ResNet-50	✓	73.70%
DCL [54]	ResNet-50	✓	72.38%
LIO [55]	ResNet-50	✓	74.18%
IFBRC [17]	ResNet-50	✓	75.87%
PIM [56]	Swin Transformer	✓	76.53%
HERBS [57]	Swin Transformer	✓	72.30%
IELT [58]	Vision Transformer	✓	73.71%
The proposed method	ResNet-50	✓	**78.43%**

classification and finds that the pre-trained ResNet-50 achieves the highest classification accuracy. (3) This study proves the necessity of data augmentation and compares five data augmentation techniques. Our experiment results demonstrate that horizontal flipping, rotation (less than 90 degrees), center cropping, and wavelet transformation are helpful for improving the classification results. These findings and conclusions may provide valuable insights and potential directions for further investigations in this field.

B. LIMITATION AND FUTURE WORK

The proposed BUS image BI-RADS classification framework has some limits. First, the proposed edge detection method for pseudo-label generation in section III-C is trained in a weakly supervised manner without tumor edge segmentation labels. Its edge detection accuracy cannot reach fully supervised methods. For other tasks with enough training data and less difficulty in annotation generation, it may achieve better fine-grained image classification results when training a fully supervised method to generate edge probability maps as weights to enhance the edge region in the input images. Second, the proposed ellipse-fitting loss function is specifically designed for detecting objects with an elliptical shape, such as breast tumors. As a result, its generalization ability is limited and may not work on detecting other types of objects in images. The proposed denoise loss may have better generalization ability as it helps to reduce noise pixels in the detected edge regions without requiring a specific object shape. Third, due to the absence of the public BI-RADS BUS image dataset, the effectiveness of the proposed is only validated on a private dataset.

In the future, we will evaluate our proposed method on public datasets, if available. We will also explore more strategies to improve the generalization ability of the proposed method. Additionally, we plan to adopt self-learning methods [61] and class activation map [62] to further improve the weakly supervised edge detection method proposed in section III-C. Last, we will collect more BUS images for our dataset and ensure a balanced number of images in each BI-RADS category.

VI. CONCLUSION

In this study, we propose a novel and efficient framework for BUS image BI-RADS classification by incorporating tumor edge information as discriminative features. In the first step, we fine-tune a pre-trained PiDiNet on the BUS image dataset in a weakly-supervised manner. We design a denoise loss function and an ellipse-fitting loss function to guide the PiDiNet to learn better tumor edge patterns during the fine-tuning process. After fine-tuning, the PiDiNet produces coarse edge images for all images in the BUS dataset, which are subsequently used as pseudo-labels for supervising a co-training framework in the next step. In the second step, we propose a co-training framework that consists of an edge detection network (PiDiNet) and a classification network (ResNet-50). Edge probability maps generated by the edge detection network are used as weight masks to enhance the tumor edge region in the input BUS images. The edge-enhanced images are then fed into the classification network for classification. Two networks promote each other during training for a more accurate BI-RADS classification result. Extensive experiments demonstrate the effectiveness of the proposed method. It outperforms eight recent fine-grained image classification methods that use ResNet-50 or transformers as the backbone. Moreover, our findings and conclusions can provide valuable insights for further investigations in this field, such as the importance of tumor edge regions and the performance ranks of five data augmentation techniques for BUS image classification.

REFERENCES

[1] N. Aristokli, I. Polycarpou, S. C. Themistocleous, D. Sophocleous, and I. Mamais, "Comparison of the diagnostic performance of magnetic resonance imaging (MRI), ultrasound and mammography for detection of breast cancer based on tumor type, breast density and patient's history: A review," *Radiography*, vol. 28, no. 3, pp. 848–856, Aug. 2022.

[2] P. Jaglan, R. Dass, and M. Duhan, "Breast cancer detection techniques: Issues and challenges," *J. Inst. Eng. (India), Ser. B*, vol. 100, no. 4, pp. 379–386, Aug. 2019.

[3] O. Ginsburg, C.-H. Yip, A. Brooks, A. Cabanes, M. Caleffi, J. A. D. Yataco, B. Gyawali, V. McCormack, M. McLaughlin de Anderson, and R. Mehrotra, "Breast cancer early detection: A phased approach to implementation," *Cancer*, vol. 126, no. S10, pp. 2379–2393, May 2020.

[4] 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, no. 5, pp. 1–17, Dec. 2019.

[5] N. Karunanayake, W. Lohitvisate, and S. S. Makhanov, "Artificial life for segmentation of fusion ultrasound images of breast abnormalities," *Pattern Recognit.*, vol. 131, Nov. 2022, Art. no. 108838.

[6] Y. Luo, Q. Huang, and L. Liu, "Classification of tumor in one single ultrasound image via a novel multi-view learning strategy," *Pattern Recognit.*, vol. 143, Nov. 2023, Art. no. 109776.

[7] M. Xu, K. Huang, and X. Qi, "A regional-attentive multi-task learning framework for breast ultrasound image segmentation and classification," *IEEE Access*, vol. 11, pp. 5377–5392, 2023.

[8] C. J. D'Orsi, E. A. Sickles, E. B. Mendelson, and E. A. Morris, *ACR BI-RADS Atlas: Breast Imaging Reporting and Data System; Mammography, Ultrasound, Magnetic Resonance Imaging, Followup and Outcome Monitoring, Data Dictionary*. Reston, VA, USA: American College of Radiology, 2013.

[9] Y. Huang, L. Han, H. Dou, H. Luo, Z. Yuan, Q. Liu, J. Zhang, and G. Yin, "Two-stage CNNs for computerized BI-RADS categorization in breast ultrasound images," *Biomed. Eng. OnLine*, vol. 18, no. 1, pp. 1–18, Dec. 2019.

[10] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Deep learning approaches for data augmentation and classification of breast masses using ultrasound images," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 5, pp. 1–11, 2019.

[11] 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.

[12] W.-C. Shia and D.-R. Chen, "Classification of malignant tumors in breast ultrasound using a pretrained deep residual network model and support vector machine," *Computerized Med. Imag. Graph.*, vol. 87, Jan. 2021, Art. no. 101829.

[13] Z. Zhuang, Z. Yang, S. Zhuang, A. N. Joseph Raj, Y. Yuan, and R. Nersisson, "Multi-features-based automated breast tumor diagnosis using ultrasound image and support vector machine," *Comput. Intell. Neurosci.*, vol. 2021, pp. 1–12, May 2021.

[14] T. Xiao, L. Liu, K. Li, W. Qin, S. Yu, and Z. Li, "Comparison of transferred deep neural networks in ultrasonic breast masses discrimination," *BioMed Res. Int.*, vol. 2018, pp. 1–9, Jun. 2018.

[15] A. Rodríguez-Cristerna, W. Gómez-Flores, and W. C. de Albuquerque Pereira, "A computer-aided diagnosis system for breast ultrasound based on weighted BI-RADS classes," *Comput. Methods Programs Biomed.*, vol. 153, pp. 33–40, Jan. 2018.

[16] E. Kaplan, W. Y. Chan, S. Dogan, P. D. Barua, H. T. Bulut, T. Tuncer, M. Cizik, R.-S. Tan, and U. R. Acharya, "Automated BI-RADS classification of lesions using pyramid triple deep feature generator technique on breast ultrasound images," *Med. Eng. Phys.*, vol. 108, Oct. 2022, Art. no. 103895.

[17] Y. Xiao, K. Huang, S. Niu, and J. Huang, "Interpretable fine-grained BI-RADS classification of breast tumors," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2021, pp. 3617–3621.

[18] S. Boumaraf, X. Liu, C. Ferkous, and X. Ma, "A new computer-aided diagnosis system with modified genetic feature selection for BI-RADS classification of breast masses in mammograms," *BioMed Res. Int.*, vol. 2020, pp. 1–17, May 2020.

[19] 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.

[20] 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.

[21] W. K. Moon, C.-M. Lo, N. Cho, J. M. Chang, C.-S. Huang, J.-H. Chen, and R.-F. Chang, "Computer-aided diagnosis of breast masses using quantified BI-RADS findings," *Comput. Methods Programs Biomed.*, vol. 111, no. 1, pp. 84–92, Jul. 2013.

[22] Z. Su, W. Liu, Z. Yu, D. Hu, Q. Liao, Q. Tian, M. Pietikäinen, and L. Liu, "Pixel difference networks for efficient edge detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 5097–5107.

[23] M. L. Giger, Z. Huo, D. E. Wolverton, C. J. Vyborny, C. Moran, R. A. Schmidt, H. Al-Hallaq, R. M. Nishikawa, and K. Doi, "Computer-aided diagnosis of digital mammography and ultrasound images of breast mass lesions," in *Digital Mammography*. Springer, 1998, pp. 143–147.

[24] K. Horsch, M. L. Giger, L. A. Venta, and C. J. Vyborny, "Computerized diagnosis of breast lesions on ultrasound," *Med. Phys.*, vol. 29, no. 2, pp. 157–164, Jan. 2002.

[25] R.-F. Chang, W.-J. Wu, W. K. Moon, and D.-R. Chen, "Improvement in breast tumor discrimination by support vector machines and speckle-emphasis texture analysis," *Ultrasound Med. Biol.*, vol. 29, no. 5, pp. 679–686, May 2003.

[26] H. D. Cheng, J. Shan, W. Ju, Y. Guo, and L. Zhang, "Automated breast cancer detection and classification using ultrasound images: A survey," *Pattern Recognit.*, vol. 43, no. 1, pp. 299–317, Jan. 2010.

[27] M. I. Daoud, S. Abdel-Rahman, and R. Alazrai, "Breast ultrasound image classification using a pre-trained convolutional neural network," in *Proc. 15th Int. Conf. Signal-Image Technol. Internet-Based Syst. (SITIS)*, Nov. 2019, pp. 167–171.

[28] J. Xie, X. Song, W. Zhang, Q. Dong, Y. Wang, F. Li, and C. Wan, "A novel approach with dual-sampling convolutional neural network for ultrasound image classification of breast tumors," *Phys. Med. Biol.*, vol. 65, no. 24, Dec. 2020, Art. no. 245001.

[29] J. Xing, C. Chen, Q. Lu, X. Cai, A. Yu, Y. Xu, X. Xia, Y. Sun, J. Xiao, and L. Huang, "Using BI-RADS stratifications as auxiliary information for breast masses classification in ultrasound images," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 6, pp. 2058–2070, Jun. 2021.

[30] 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.

[31] 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.

[32] C. F. Flores, A. Gonzalez-Garcia, J. van de Weijer, and B. Raducanu, "Saliency for fine-grained object recognition in domains with scarce training data," *Pattern Recognit.*, vol. 94, pp. 62–73, Oct. 2019.

[33] X. Ke, Y. Cai, B. Chen, H. Liu, and W. Guo, "Granularity-aware distillation and structure modeling region proposal network for fine-grained image classification," *Pattern Recognit.*, vol. 137, May 2023, Art. no. 109305.

[34] A. Angelova and S. Zhu, "Efficient object detection and segmentation for fine-grained recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 811–818.

[35] N. Zhang, J. Donahue, R. Girshick, and T. Darrell, "Part-based R-CNNs for fine-grained category detection," in *Proc. Comput. Vis. ECCV 13th Eur. Conf.*, Zurich, Switzerland. Cham, Switzerland: Springer, 2014, pp. 834–849.

[36] X.-S. Wei, C.-W. Xie, J. Wu, and C. Shen, "Mask-CNN: Localizing parts and selecting descriptors for fine-grained bird species categorization," *Pattern Recognit.*, vol. 76, pp. 704–714, Apr. 2018.

[37] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

[38] W. Ge, X. Lin, and Y. Yu, "Weakly supervised complementary parts models for fine-grained image classification from the bottom up," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3029–3038.

[39] C. Liu, H. Xie, Z.-J. Zha, L. Ma, L. Yu, and Y. Zhang, "Filtration and distillation: Enhancing region attention for fine-grained visual categorization," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 7, 2020, pp. 11555–11562.

[40] B. Santra, A. K. Shaw, and D. P. Mukherjee, "Part-based annotation-free fine-grained classification of images of retail products," *Pattern Recognit.*, vol. 121, Jan. 2022, Art. no. 108257.

[41] Y. Wang, V. I. Morariu, and L. S. Davis, "Learning a discriminative filter bank within a CNN for fine-grained recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4148–4157.

[42] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2011.

[43] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from RGBD images," in *Proc. ECCV*, vol. 7576, 2012, pp. 746–760.

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

[45] M. Tardy and D. Mateus, "Looking for abnormalities in mammograms with self- and weakly supervised reconstruction," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2711–2722, Oct. 2021.

[46] M. Z. Liu, C. Swintelski, S. Sun, M. Siddique, E. Desperito, S. Jambawalikar, and R. Ha, "Weakly supervised deep learning approach to breast MRI assessment," *Academic Radiol.*, vol. 29, pp. S166–S172, Jan. 2022.

[47] K. Huang, H. D. Cheng, Y. Zhang, B. Zhang, P. Xing, and C. Ning, "Medical knowledge constrained semantic breast ultrasound image segmentation," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 1193–1198.

[48] 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.

[49] K. Huang, Y. Zhang, H.-D. Cheng, and P. Xing, "Trustworthy breast ultrasound image semantic segmentation based on fuzzy uncertainty reduction," *Healthcare*, vol. 10, no. 12, p. 2480, Dec. 2022.

[50] 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. 2261–2269.

[51] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

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

[53] Z. Yang, T. Luo, D. Wang, Z. Hu, J. Gao, and L. Wang, "Learning to navigate for fine-grained classification," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 420–435.

[54] Y. Chen, Y. Bai, W. Zhang, and T. Mei, "Destruction and construction learning for fine-grained image recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5152–5161.

[55] M. Zhou, Y. Bai, W. Zhang, T. Zhao, and T. Mei, "Look-into-object: Self-supervised structure modeling for object recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11771–11780.

[56] P.-Y. Chou, C.-H. Lin, and W.-C. Kao, "A novel plug-in module for fine-grained visual classification," 2022, *arXiv:2202.03822*.

[57] P.-Y. Chou, Y.-Y. Kao, and C.-H. Lin, "Fine-grained visual classification with high-temperature refinement and background suppression," 2023, *arXiv:2303.06442*.

[58] Q. Xu, J. Wang, B. Jiang, and B. Luo, "Fine-grained visual classification via internal ensemble learning transformer," *IEEE Trans. Multimedia*, vol. 25, pp. 9015–9028, 2023.

[59] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9992–10002.

[60] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, and S. Gelly, "An image is worth 16×16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Represent.*, 2020, pp. 1–12.

[61] J. Qin, J. Wu, X. Xiao, L. Li, and X. Wang, "Activation modulation and recalibration scheme for weakly supervised semantic segmentation," *Proc. AAAI Conf. Artif. Intell.*, vol. 36, no. 2, pp. 2117–2125, Jun. 2022.

[62] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2921–2929.

MENG XU (Member, IEEE) received the B.S. degree in management information systems from Tianjin University of Technology, Tianjin, China, in 2017, and the Ph.D. degree in computer science from Utah State University, Logan, UT, USA, in 2023.

Since 2023, she has been an Assistant Professor with the Department of Computer Science and Technology, Kean University, Union, NJ, USA. Her research interests include computer vision, deep learning, and medical image analysis.

Dr. Xu received several awards and honors, including the Presidential Doctoral Research Fellowship from Utah State University and the Outstanding Graduate Award from Tianjin University of Technology.

JIANHUA HUANG received the B.S. and M.S. degrees in pattern recognition and the Ph.D. degree in computer science from Harbin Institute of Technology, in 1989, 1992, and 2007, respectively.

In 1994, he was a Lecturer with the School of Computer Science and Technology, Harbin Institute of Technology. In 2000, he was promoted to an Associate Professor. In 2008, he was promoted to a Professor. Currently, he is mainly engaged in the research of pattern recognition, artificial intelligence, machine vision, medical image processing, and human motion intelligence.

Dr. Huang is a Committee Member of the Professional Committee of Science Popularization with the Chinese Association of Artificial Intelligence and a member of the Medical Artificial Intelligence Group, Chinese Society of Biomedical Engineering.
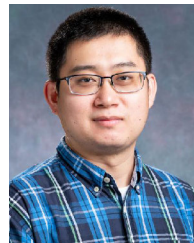
KUAN HUANG (Member, IEEE) received the B.Eng. degree in electrical engineering and automation from Harbin Institute of Technology, Harbin, China, in 2016, and the Ph.D. degree in computer science from Utah State University, Logan, UT, USA, in 2021.

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 received the Best Dissertation Award from Utah State University, in 2022.
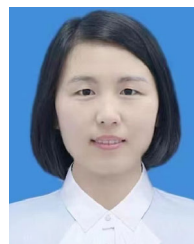
FEIFEI LIU received the Ph.D. degree from Peking University, Beijing, China, in 2022. She is currently a Doctoral and a Researcher with Binzhou Medical University Hospital, Shandong, China. Her research interests include musculoskeletal ultrasound, medical image processing, machine learning, and medical big data analysis.

• • •