

PH-Net: Semi-Supervised Breast Lesion Segmentation via Patch-wise Hardness

Siyao Jiang¹, Huisi Wu^{1*}, Junyang Chen¹, Qin Zhang¹, Jing Qin²

¹ College of Computer Science and Software Engineering, Shenzhen University

² Centre for Smart Health, School of Nursing, The Hong Kong Polytechnic University

2200271019@email.szu.edu.cn, {hswu, junyangchen, qinzhang}@szu.edu.cn, harry.qin@polyu.edu.hk

Abstract

We present a novel semi-supervised framework for breast ultrasound (BUS) image segmentation, which is a very challenging task owing to (1) large scale and shape variations of breast lesions and (2) extremely ambiguous boundaries caused by massive speckle noise and artifacts in BUS images. While existing models achieved certain progress in this task, we believe the main bottleneck nowadays for further improvement is that we still cannot deal with hard cases well. Our framework aims to break through this bottleneck, which includes two innovative components: an adaptive patch augmentation scheme and a hard-patch contrastive learning module. We first identify hard patches by computing the average entropy of each patch and then shield hard patches to prevent them from being cropped out while performing random patch cutmix. Such a scheme is able to prevent hard regions from being inadequately trained under strong augmentation. We further develop a new hard-patch contrastive learning algorithm to direct model attention to hard regions by applying extra contrast to pixels in hard patches, further improving segmentation performance on hard cases. We demonstrate the superiority of our framework to state-of-the-art approaches on two famous BUS datasets, achieving better performance under different labeling conditions. The code is available at <https://github.com/jjjsyyy/PH-Net>.

1. Introduction

Breast cancer is the most common cancer among females worldwide. According to newly released statistics [27], one in eight newly diagnosed cancer patients develops breast cancer. Due to its unclear etiology and unpredictable variability, early screening of breast cancer plays a vital role in early intervention and hence significantly reduces mortality [9]. Currently, ultrasound imaging is the most widely used tool for breast cancer screening, owing to its non-

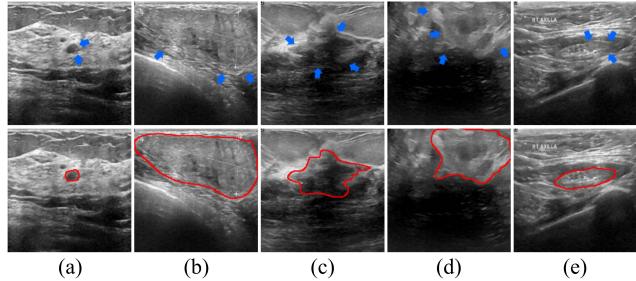


Figure 1. Typical challenges in segmentation of BUS images: (a)-(c) large variations in scale and shape of lesions, (d)-(e) lesions with low contrast and ambiguous boundaries. Blue arrows and red lines indicate blurring boundaries and ground truth, respectively.

invasiveness, cost-effectiveness, and real-time nature [26]. However, manual screening of breast ultrasound (BUS) images is labor-intensive, time-consuming, and error-prone. In addition, it heavily depends on physicians' expertise and hence there usually exists large intra- and inter-observer variability. To the end, automated screening tools are highly demanded in clinical practice to improve diagnosis accuracy and efficacy.

One of the essential tasks in developing such an automated screening system is to automatically segment lesions from BUS images. However, segmentation of BUS images presents a highly challenging task, as shown in Fig. 1. First, lesions exhibit highly ambiguous boundaries due to the low contrast of BUS images. Second, tumors have substantial variability in scale and shape. Third, due to heavy speckle noise and artifacts in BUS images, some lesions are exceptionally challenging to recognize. In this regard, a lot of effort has been dedicated to tackling this challenging task.

In recent years, with the rapid developments of deep learning techniques, fully supervised methods [30, 34, 37, 46] have achieved remarkable performance in BUS image segmentation. However, these fully supervised methods heavily rely on extensive annotated data, while acquiring large-scale pixel-level annotations for segmentation remains time-consuming and labor-intensive. To mitigate this

*Corresponding Author

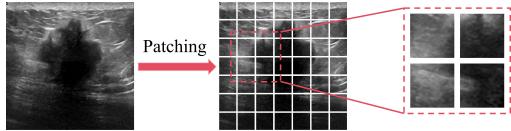


Figure 2. Detailed visualization of the patches. Small patches can capture lesions, background and border information.

limitation, many researchers have shifted their attention to semi-supervised learning, which aims at achieving satisfactory performance by utilizing limited labeled images alongside massive unlabeled images.

Some semi-supervised models have been proposed to address the challenges of BUS image segmentation with limited training data. Attention-based methods [6, 45] attempted to capture features from multiple receptive fields through attention modules to more robustly determine the location and size of lesions. However, it is difficult for them to tackle ambiguous boundaries with limited pixel-level annotations. Later, generative adversarial network (GAN)-based methods [14, 41] endeavored to reduce supervisory signals by improving both the capability of the discriminator and the quality of the predicated map produced by the generator. However, extensive speckle noise and artifacts significantly affect the judgment of the discriminator, and thus impact the stability of adversarial training as well as the quality of segmentation. Different from the above methods, RA-UGMT [11] proposes a residual-attention-based uncertainty-guided mean teacher framework to estimate the certainty of all the outputs and then select the optimal output. Nevertheless, this method only processes reliable samples, which results in insufficient training of difficult samples, making the model not perform well on hard cases.

In this paper, we propose a novel semi-supervised BUS image segmentation framework via patch-wise hardness, we call it *PH-Net*. The proposed PH-Net is constructed based on classical teacher-student architecture, and includes two innovative components: an adaptive patch augmentation scheme and a hard-patch contrastive learning module. Inspired by [29], in the teacher network, we divide the input image into multiple small patches, as shown in Fig. 2, and then compute the average entropy of each patch as a hardness score. After that, we shield hard patches to prevent them from being cropped out while performing random patch cutmix on the remaining patches, which ensures adequate model training on difficult regions. Moreover, to further enhance model learning on these hard regions, we incorporate a projector into the student branch. We acquire reliable samples via rigorous sampling for additional contrastive learning, improving intra-class compactness and inter-class discriminability within the hard patches. We extensively evaluate our proposed method on two well-known public datasets, UDIAT [40] and BUSI [1], demonstrating

the superiority of our proposed model to state-of-the-art approaches. Our contributions can be summarized as follows:

- We propose a novel semi-supervised breast lesion segmentation framework based on patch-wise hardness, which aims at preventing hard regions from being inadequately trained under strong augmentation.
- We develop a new hard-patch contrastive learning algorithm to direct model attention to challenging regions by applying extra contrast to pixels in hard patches, further improving segmentation performance on hard cases.
- We demonstrate the superiority of our proposed method to state-of-the-art approaches on two famous BUS datasets, achieving better performance under different labeling conditions.

2. Related work

2.1. Breast ultrasound image segmentation

In recent years, convolutional neural networks (CNNs) based BUS image segmentation has made remarkable advancements owing to rapid developments in deep learning. Methods including RDAU-Net [46], PPU-Net [42] and SK-U-Net [5], modifying on the classical U-Net [25] convolutional neural network, have achieved promising results in the early stage. GG-Net [37] utilizes multi-level integrated feature maps as guiding information and incorporates additional boundary detection to enhance segmentation quality. MCRNet [22] performs multi-level contextual refinement within the encoder-decoder architecture to achieve fully automatic semantic segmentation. Although achieving impressive segmentation results, all of these methods rely on fully supervised training strategies and demand a large amount of labeled data. However, acquiring precisely annotated BUS images is a time-consuming and labor-intensive process and requires the expertise of medical professionals. Consequently, semi-supervised learning offers a more promising alternative that requires limited labeled data to achieve outstanding performance, considerably reducing the model's dependence on extensive labeled data.

2.2. Semi-supervised segmentation

Based on existing research, semi-supervised segmentation methods can be divided into three categories: (1) Pseudo-labelling based methods [18, 24, 38, 47], which generate pseudo-labels for unlabeled images and subsequently re-training the model. (2) Consistency regularisation based methods [8, 12, 21, 39, 43, 44], which perturb the unlabeled images and constrain the outputs to be consistent to train a model resistant to these perturbations. (3) Contrastive learning based methods [2, 13, 31, 32], which minimize the feature space distances between positive pairs and maximize distances for negative pairs. In particular, UniMatch [39] introduces a dual-stream perturbation technique to utilize a

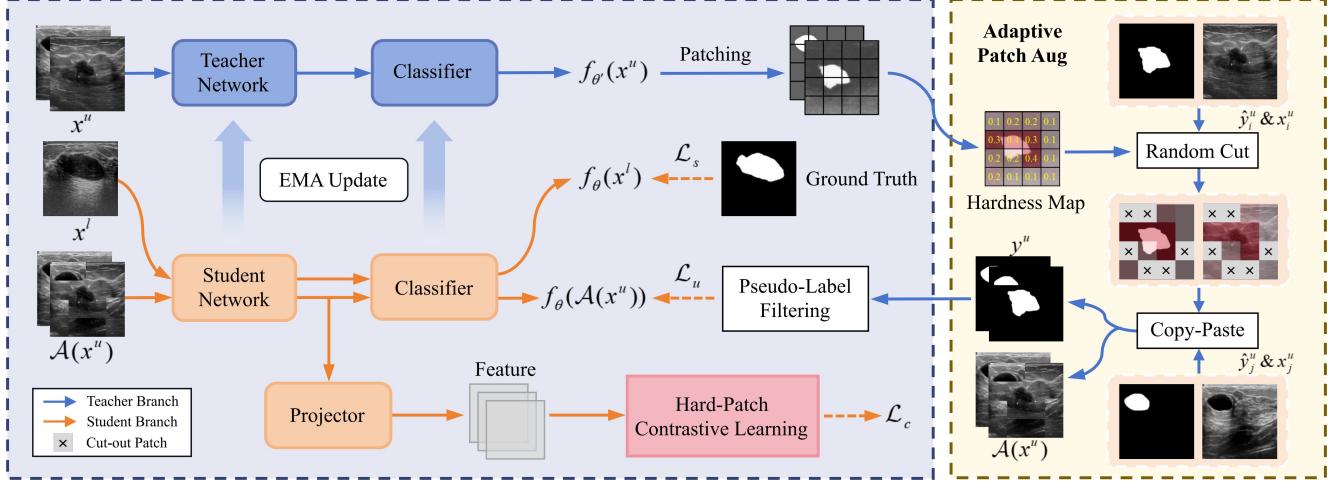


Figure 3. Framework of our proposed method. In the teacher-student framework, we employ the teacher model to generate pseudo-labels for unlabeled images, which together with the original images are subjected to adaptive patch augmentation based on the hardness maps at the patch level. In this process, we shield the patches with high hardness, randomly cut out the remaining patches, and paste them with another image correspondingly. The augmented images, together with the labeled images, are then fed into the student model for supervised training. A projector is additionally installed and the output features are involved in hard-patch contrastive learning for further training. \hat{y}^u is obtained from $f_{\theta'}(x^u)$ via $\text{argmax}(\cdot)$. The dashed line indicates the loss calculation.

wider perturbation space. However, BUS images containing substantial noise are not amenable to extensive perturbations. U²PL [32] separates reliable and unreliable pixels based on predicted entropy, using the latter as negative samples for contrastive learning. Though effectively utilizes unreliable pixels, it is inapplicable to medical image few-class segmentation tasks. Moreover, AugSeg [44] performs data augmentation by randomly and adaptively mixing labeled and unlabeled samples to stabilize training on unlabeled data. iMAS [43] performs model-adaptive data perturbation and weighting on each unlabeled instance to enhance model generalization during training. However, these methods perform adaptive perturbation at the image or instance level, overlooking the inconsistent learning difficulty of the model for pixels in different regions of the image. We argue that pixels with varying difficulty in images should be handled distinctly, with difficult regions receiving more training from the model compared to easy regions.

2.3. Semi-supervised medical segmentation

Many researchers have made a lot of efforts in semi-supervised medical segmentation in recent years and proposed many excellent methods [3, 4, 19, 23, 33, 36]. Among them, BCP [3] encourages unlabeled data to learn comprehensive public semantics from labeled data both inwardly and outwardly by bidirectionally copy-pasting labeled and unlabeled data. MCF [33] proposes a new mutual correction framework to rectify cognitive biases in the model. PatchCL [4] provides additional guidance to facilitate contrastive learning through pseudo-labeling, using average

patch entropy to provide guidance on positive and negative sampling. However, these methods do not consider intra- and inter-image variability in learning difficulty. Furthermore, PDF-UNet [16] designed for BUS images, combines data expansion, probability map generator, and U-shaped pyramid-dilated fusion to expand the model’s receptive field for accurate tumor detection at different scales. Although partially addresses the multi-scale issue in BUS images, this method does not attach importance to challenges like indistinct lesion boundaries and artifacts. Unlike the above methods, our method performs adaptive perturbation and contrastive learning by calculating the hardness scores of patches to evaluate challenging regions such as edges.

3. Method

3.1. Motivation

Considering the ambiguous and low-contrast characteristics of BUS images, in combination with the limited scale of publicly available datasets, we conduct our study with semi-supervised segmentation methods via patch-wise hardness. Our motivation is to train a model focusing on inherently challenging regions. Due to the considerable variation in learning difficulty of different regions in BUS images, we divide images into multiple small patches. As depicted in Fig. 2, small patches can effectively capture lesion and background details. Patches encompassing ambiguous areas typically constitute hard ones. Our method incorporates an adaptive patch augmentation (APA) module to minimize the exposure of challenging patches during the application

of cutmix augmentation, allowing these regions to be adequately trained. In addition, we provide additional contrast using the hard-patch contrastive learning (HPC) module to further train these challenging patches.

3.2. Overview

Our semi-supervised segmentation method PH-Net operates on a limited labeled training set D_l and a large unlabeled training set D_u . The overall framework of PH-Net is illustrated in Fig. 3. Based on the common teacher-student semi-supervised framework, we incorporate a projector into the student network. During the training process, gradient propagation is only performed within the student network, while the teacher weights θ' get updated via the exponential moving average (EMA) of the student weights θ :

$$\theta' \leftarrow \alpha\theta' + (1 - \alpha)\theta \quad (1)$$

where α is a momentum parameter, set to 0.999 followed [28]. In the first few epochs of training, we only feed labeled images to pre-train the model, establishing basic segmentation capabilities. For each subsequent training step, labeled images $\mathcal{B}_l = \{(x_i^l, y_i^l)\}$ and unlabeled images $\mathcal{B}_u = \{x_i^u\}$ are concurrently sampled. Supervised training on labeled images employs the standard cross-entropy (CE) loss:

$$\mathcal{L}_s = \frac{1}{|\mathcal{B}_l|} \sum_{i=1}^{|\mathcal{B}_l|} \ell_{ce}(f_\theta(x_i^l), y_i^l) \quad (2)$$

where $|\cdot|$ denotes set length, $f_\theta(x_i^l)$ represents the student network's prediction for the i -th labeled image x_i^l , and y_i^l is the ground truth. For unsupervised learning on unlabeled images, we perform consistency regularization by adaptive patch augmentation and further train through hard-patch contrastive learning, which will be described in detail in Sec. 3.3 and Sec. 3.4 respectively. The total loss is calculated as follows:

$$\mathcal{L} = \mathcal{L}_s + \lambda_u \mathcal{L}_u + \lambda_c \mathcal{L}_c \quad (3)$$

where λ_u and λ_c are the weights of unsupervised loss \mathcal{L}_u and contrastive loss \mathcal{L}_c respectively.

3.3. Adaptive patch augmentation

Traditional cutmix [12] augmentation method has shown its effectiveness in semi-supervised segmentation tasks, with a proliferation of existing cutmix-based improvement methods [17, 43, 44]. However, these methods all ignore the discrepancy in learning difficulty between challenging and simple regions in the image. To mitigate this, we propose adaptive patch augmentation (APA) to identify and shield hard regions. Specifically, we calculate the hardness score of each patch and shield high-hardness patches, while performing random patch cutmix on the remaining patches.

Hardness metrics. As shown in Fig. 3, our image augmentation is performed in the teacher branch. By default, x^u has already undergone weak augmentation (flipping, scaling, and cropping). The probabilistic prediction $f_{\theta'}(x^u)$ is divided into $N \times N$ patches of size $h \times w$. The hardness score of the patch is calculated as follows:

$$r_m = -\frac{1}{\log C} \frac{1}{h \times w} \sum_{k=1}^{h \times w} \sum_{c=0}^{C-1} p_{m,k}(c) \log(p_{m,k}(c)) \quad (4)$$

where $p_{m,k} \in R^C$ represents the activation probability distribution on pixel k in the m -th patch. C represents the total number of categories, and $p_{m,k}(c)$ represents the probability value for class c . Obviously, the hardness score is patch-specific and reflects the model's confidence of pixels within the patch. It increases with greater prediction ambiguity within a patch and decreases as predictions become more certain.

Patch cutmix augmentation. Further training of the model relies on patches with higher hardness scores, as these patches cannot be predicted with high confidence. To prevent patches with high hardness scores from being cut out during cutmix augmentation, we shield them first. The patches are sorted by descending hardness score, and the shielded patch number is set dynamically as follows:

$$k = \beta \frac{t}{T} \cdot N^2 \quad (5)$$

where t and T represent the current training epoch and the total epoch respectively, β is a hyperparameter used to control the patch shielding ratio. While applying cutmix augmentation on the image in patch units, we shield the top k patches (i.e., keeping them unmasked) and randomly cut the remaining $(N^2 - k)$ patches. We then fill with the image arranged in the following position to create the augmented image. Besides, the pseudo-labels are modified accordingly. The adaptive patch cutmix can be expressed as:

$$\mathcal{A}(x_i^u) \leftarrow M \odot x_i^u + (1 - M) \odot x_{i+1}^u \quad (6)$$

$$y_i^u \leftarrow M \odot \hat{y}_i^u + (1 - M) \odot \hat{y}_{i+1}^u \quad (7)$$

where x_i^u and x_{i+1}^u represent the i -th and $(i + 1)$ -th unlabeled images in a batch, \hat{y}_i^u and \hat{y}_{i+1}^u denote their pseudo-labels respectively. M is the randomly selected patch mask used for cutmix.

Unsupervised loss calculation. For consistency, the output prediction of the augmented images should be consistent with the original images. However, considering that unreliable pseudo-labels are not suitable for supervision, we input the augmented image $\mathcal{A}(x_i^u)$ into the teacher model to filter the pseudo-labels during the calculation of the unsupervised

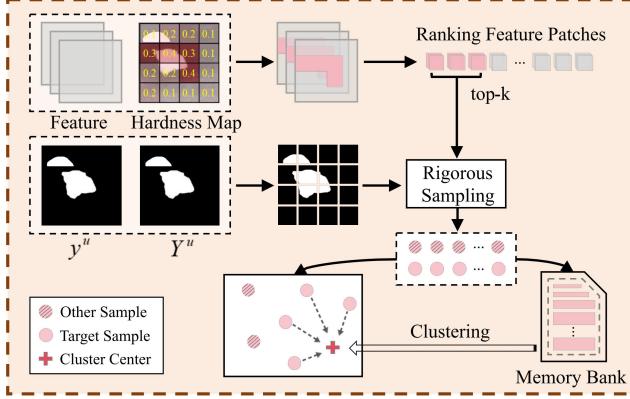


Figure 4. Conceptual explanation of hard-patch contrastive learning (HPC). y^u and Y^u represent the pseudo-labels produced twice by the teacher model. The dashed line indicates that the samples converge towards the center. Features generate multiple feature patches based on the patch-level hardness map, the top- k are taken for rigorous sampling, forming two groups: target samples and other samples. The memory bank stores the target sample in each iteration, and sample centers are generated through clustering.

loss:

$$\mathcal{L}_u = \frac{1}{|\mathcal{B}_u|} \sum_{i=1}^{|\mathcal{B}_u|} \sum_{j=1}^{H \times W} M_{ij}^f \cdot \ell_{ce}(f_\theta(\mathcal{A}(x_{ij}^u)), y_{ij}^u) \quad (8)$$

where

$$M_{ij}^f = \mathbb{1}[\max(f_{\theta'}(\mathcal{A}(x_{ij}^u))) \geq \gamma] \quad (9)$$

where H and W represent the height and width of the prediction respectively, γ is a pre-set confidence threshold to filter unreliable pixels. $f_\theta(\mathcal{A}(x_{ij}^u))$ and $f_{\theta'}(\mathcal{A}(x_{ij}^u))$ are the output predictions for the $\mathcal{A}(x_i^u)$ at the j -th pixel from the student network and the teacher network, respectively.

3.4. Hard-patch contrastive learning

Previous semi-supervised segmentation methods [2, 31, 32] have applied contrastive learning by sampling all image features or filtering unreliable ones. However, they have not emphasized the model's learning of challenging pixels or regions. Recent work [32] explores the utilization of unreliable pixels, but is inapplicable to two-class segmentation tasks. Therefore, we propose a hard-patch contrastive learning (HPC) module, conducting additional contrastive learning on hard patches identified by our adaptive patch augmentation. Specifically, we select the top k hardest patches, rigorously sample features, and categorize them into target and other (non-target) classes for contrastive learning.

Rigorous sampling. As the pseudo-labels generated directly by the model cannot guarantee their accuracy, to obtain more reliable samples, we use $f_{\theta'}(\mathcal{A}(x_{ij}^u))$ to rigorously filter the pseudo-labels:

$$M_{ij}^{inter} = \mathbb{1}\{y_{ij}^u = Y_{ij}^u\} \quad (10)$$

where

$$Y_{ij}^u = \operatorname{argmax}(f_{\theta'}(\mathcal{A}(x_{ij}^u))) \quad (11)$$

Under the rigorous supervision of dual pseudo-labels, the pixel is considered reliable only when its outputs remain consistent. Notably, we only sample anchor features from the hardest k patches. We categorize the sampled features into two sets as follows:

$$\mathcal{Z}_c = \{z_{ij} | M_{ij}^{inter} = 1, y_{ij}^u = c, z_{ij} \in \mathcal{Z}_k\} \quad (12)$$

$$\mathcal{Z}_c^- = \{z_{ij} | M_{ij}^{inter} = 1, y_{ij}^u \neq c, z_{ij} \in \mathcal{Z}_k\} \quad (13)$$

where z represents the output feature from the projector in the student model, \mathcal{Z}_k is the feature set of patches with hardness values in the top k . Specifically, we divide the reliable features in \mathcal{Z}_k into two feature sets \mathcal{Z}_c and \mathcal{Z}_c^- based on y_{ij}^u . Fig. 4 provides a conceptual explanation.

To prevent limited anchors leading to an overly localized and unstable sample center, we store the anchor feature set \mathcal{Z}_c by using a memory bank. The calculation of positive samples can be expressed as:

$$z_c^+ = \frac{1}{|\mathcal{R}_c|} \sum_{z \in \mathcal{R}_c} z \quad (14)$$

where \mathcal{R}_c represents the feature set of class c stored in the memory bank. Note that when the memory bank is saturated, we remove old features to leave enough space to store the latest features.

Contrastive loss calculation. For each anchor feature, we assign one positive sample and multiple negative samples, aiming to encourage anchor features to converge towards the same-class cluster center z^+ while diverging from features of different classes. The contrastive loss is calculated as follows:

$$\mathcal{L}_c = -\frac{1}{C \times |\mathcal{Z}_c|} \sum_{c=0}^{C-1} \sum_{z_{ci} \in \mathcal{Z}_c} \log \left[\frac{e^{(\langle z_{ci}, z_c^+ \rangle) / \tau)}}{e^{(\langle z_{ci}, z_c^+ \rangle) / \tau)} + \sum_{z_{-} \in \mathcal{Z}_c^-} e^{(\langle z_{ci}, z_{-} \rangle) / \tau)} \right] \quad (15)$$

where $\langle \cdot, \cdot \rangle$ is the cosine similarity between the two features, τ is a temperature coefficient, set to 0.5 followed [20].

4. Experiments

4.1. Experimental setup

Datasets. We validate our method on two public breast ultrasound image datasets:

- **UDIAT dataset** [40] collected from the UDIAT Diagnostic Center of the Parc Taulí Corporation, Sabadell, Spain. The dataset contains 163 breast ultrasound images, comprising 110 benign and 53 malignant cases.

Method	Backbone	UDIAT						BUSI					
		1/2		1/4		1/8		1/2		1/4		1/8	
		Dice±std(%)	P-value										
SupOnly	ResNet-50	84.62±0.56	0.019	83.45±0.47	0.022	81.10±0.85	0.024	74.36±0.89	0.021	73.69±1.02	0.031	70.99±1.21	0.030
CPS [8]	ResNet-50	84.87±0.84	0.043	83.76±0.64	0.024	81.29±0.40	0.028	74.83±0.91	0.037	74.02±1.31	0.028	71.31±1.34	0.040
PS-MT [21]	ResNet-50	85.45±1.18	0.035	84.20±0.48	0.041	82.21±0.87	0.030	75.76±0.54	0.036	74.21±1.05	0.033	71.88±1.27	0.045
U ² PL [32]	ResNet-50	86.42±0.46	0.037	84.97±0.83	0.028	83.04±0.67	0.039	76.34±0.24	0.036	75.73±0.87	0.040	72.49±0.46	0.042
iMAS [43]	ResNet-50	86.53±0.84	0.027	85.31±0.61	0.043	83.48±1.20	0.041	76.24±1.35	0.038	75.81±0.79	0.029	72.74±1.17	0.036
AugSeg [44]	ResNet-50	86.98±0.56	0.036	85.47±0.94	0.034	84.01±0.73	0.037	76.93±0.67	0.029	76.16±1.13	0.023	72.85±1.06	0.036
BCP [3]	U-Net	86.28±0.88	0.038	84.93±0.97	0.041	83.52±0.75	0.037	76.47±1.06	0.042	75.35±0.84	0.038	72.56±1.10	0.030
PDF-UNet [16]	U-Net	85.70±0.41	0.029	84.32±0.45	0.035	82.65±0.83	0.033	75.51±0.73	0.040	74.47±0.96	0.044	72.05±0.42	0.034
RA-UGMT [11]	U-Net	86.54±0.60	0.039	85.16±0.38	0.027	83.71±0.52	0.025	76.57±0.94	0.026	75.79±0.57	0.039	72.41±0.76	0.045
Ours	U-Net	86.83±0.32	0.027	85.40±0.49	0.034	83.91±0.62	0.027	77.05±0.37	0.030	76.26±0.55	0.028	72.79±0.53	0.033
	ResNet-50	87.76±0.47	-	86.23±0.50	-	84.99±0.33	-	78.19±0.28	-	76.84±0.52	-	73.61±0.48	-

Table 1. Quantitative comparison using different state-of-the-art methods on the UDIAT and BUSI datasets.

- **BUSI dataset** [1] gathered from 600 females aged 25 to 75 at Baheya Hospital in Cairo, Egypt, with a total of 780 images (133 normal, 437 benign and 210 malignant). We perform the experiment using only 647 abnormal images. Both datasets are randomly divided into training, validation and test sets at a ratio of 8:1:1. The training sets are further partitioned into labeled and unlabeled subsets according to the 1/2, 1/4, and 1/8 partitioning protocols.

Evaluation metrics. We use ResNet-50 [15] and U-Net [25] as the backbone and evaluate the segmentation performance using dice coefficients (Dice) and intersection over union (IoU). We also perform standard deviation (std) and Wilcoxon signed-rank tests on Dice. All evaluations are based on the output of the teacher network.

Implementation details. Following most of the previous work [21, 32, 35, 44], we use ResNet [15] pre-trained on ImageNet [10] as the backbone and DeepLabv3+ [7] as the decoder. Both our classifier and projector consist of two Conv-BN-ReLU blocks, with the difference being the output feature dimensions of 2 and 256, respectively. We adopt identical training configurations for both datasets. The student model is trained using a stochastic gradient descent (SGD) optimizer with 0.9 momentum and 0.0001 weight decay. The initial learning rate is set to 0.003, which is decayed by a polynomial strategy: $lr = lr_{init} \cdot (1 - \frac{iter}{iter_{total}})^{0.9}$. The image resolution is uniformly 500×500 , and training lasted for 200 epochs (first 10 for pre-training) with a batch size of 8. In a batch, labeled and unlabeled images count equally and both introduce weak data augmentation, including random horizontal flipping, random scaling, and random cropping. The strong data augmentation of cutmix is only applied to the unlabeled images. For hyperparameters, both λ_u and λ_c are set to 1 in Eq. (3).

4.2. Comparison with state-of-the-art methods

We compare our method with eight recent semi-supervised segmentation methods, where iMAS [43], AugSeg [44] and BCP [3] are primarily image augmentation-based methods,

PDF-UNet [16] and RA-UGMT [11] are methods for BUS images. To ensure a fair comparison, our method experiments with both ResNet-50 and U-Net backbones, and all methods are under the same experimental setup.

As shown in Tab. 1, we evaluate our method on two datasets using Dice metrics. Under all partitioning protocols, our method shows a statistical improvement on Dice at the 5% level (all P-values are less than 0.05). Notably, under the 1/2, 1/4, and 1/8 partition protocols of the UDIAT dataset, our method (with ResNet-50 as the backbone) outperforms the baseline (SupOnly) by +3.14%, +2.78%, and +3.89% in Dice, respectively. Particularly, in the 1/8 partition (with only 15 labeled images), our method outperforms the previous state-of-the-art method, AugSeg [44], by +0.98% in Dice. Meanwhile, on the BUSI dataset, our method (with U-Net as the backbone) achieves superior Dice performance compared to the latest semi-supervised BUS image segmentation method, RA-UGMT [11], by margins of +0.48%, +0.47%, and +0.38% under the 1/2, 1/4, and 1/8 partition protocols, respectively.

Furthermore, for a more intuitive demonstration of our method’s superiority, we present the output predictions of several breast ultrasound images across all compared methods in Fig. 5. Regardless of the target’s scale or boundary clarity, our method consistently produces superior segmentations, especially with ResNet-50 as the backbone. Additionally, in Fig. 6, we provide an additional evaluation of our method against three recent image-augmented methods (iMAS [43], AugSeg [44], BCP [3]) in terms of prediction accuracy on unlabeled image sets. Obviously, our method consistently provides more accurate segmentation for unlabeled images under all partitions in both the UDIAT and BUSI datasets, which illustrates the ability of our PH-Net to make the model better trained for unlabeled images.

4.3. Ablation studies

To validate the effectiveness of our method, we perform extensive ablation experiments on the 1/4 partition of the

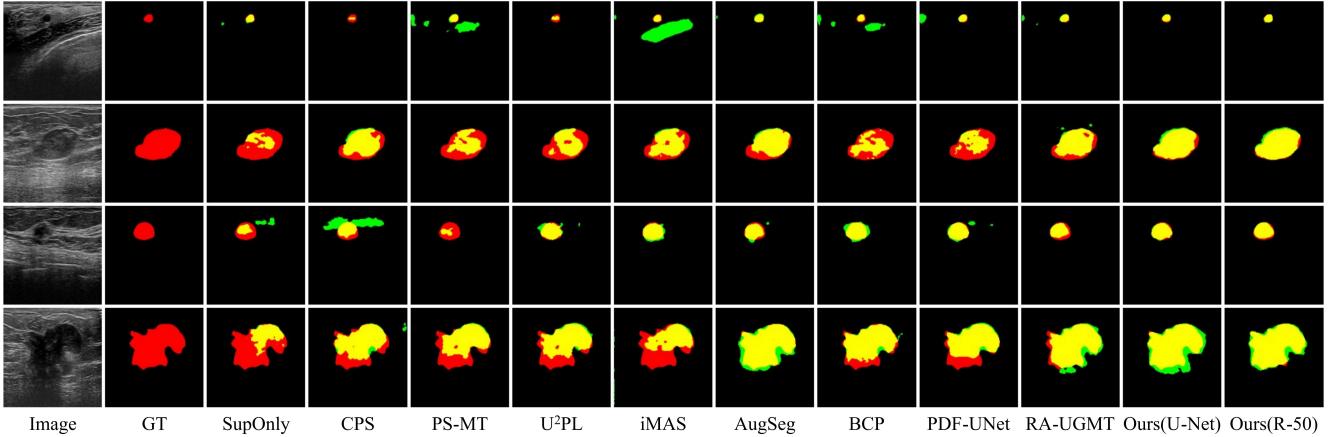


Figure 5. Visual comparison of different state-of-the-art methods on the UDIAT and BUSI datasets. All models are trained in a 1/4 partition protocol. Red, green and yellow regions represent ground truth, prediction and overlapping regions, respectively.

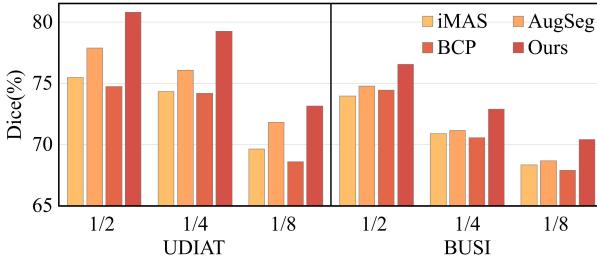


Figure 6. Comparison of prediction for unlabeled images using different image-augmented methods under all partition protocols.

UDIAT dataset using ResNet-50 as the backbone.

Effectiveness of components. We perform a stepwise ablation study on each component of PH-Net in Tab. 2. The model trained with only supervised learning (SupOnly) serves as our baseline, achieving a Dice of 83.45%. Adding adaptive patch augmentation (APA) improves the baseline by +1.89% in Dice, validating the APA module can provide promising information for the model’s consistency regularization. Further incorporating hard-patch contrastive learning (HPC) without a memory bank (MB) increases the baseline by +2.48% in Dice. The full model achieves 86.23% in Dice and 75.80% in IoU, which fully demonstrates the effectiveness of our proposed PH-Net for semi-supervised breast lesion segmentation.

To better visualize the advantages of each component, Fig. 7 illustrates visualizations of our model’s output predictions. We represent the prediction probability as the heat map. It is clear that with the progressive application of the enhancement components, the predicted lesion boundaries become increasingly fine-grained, indicating increasingly accurate predictions for challenging regions. It illustrates the effectiveness of our method for processing ambiguous regions in breast ultrasound images.

Method	APA	HPC	MB	Dice(%)	IoU(%)
SupOnly				83.45	71.59
I	✓			85.34	74.43
II	✓	✓		85.93	75.33
III	✓	✓	✓	86.23	75.80

Table 2. Ablation study of different components on UDIAT dataset under 1/4 partition protocol. APA: Adaptive Patch Augmentation. HPC: Hard-Patch Contrastive Learning. MB: Memory Bank.

Impact of different augmentation settings. As shown in Tab. 3a, we compare the performance effects of our hard patch cutmix augmentation with normal cutmix augmentation, and our method is significantly superior. While cutmix improves the weakly augmented baseline by +0.60% in Dice, our method achieves a remarkable improvement of +2.38% in Dice. In addition, different shielding strategies also impact performance. Random patch shielding without considering hardness (rand-patch) improves +0.46% in Dice over the baseline, and choosing easy patches for shielding (easy-patch) leads to +0.32% in Dice. Demonstrating that our hard-patch augmentation effectively alleviates the issue of inadequate training in challenging areas induced by standard cutmix.

Ablation of hyper-parameters. Tab. 3b illustrates the performance impact of the patch shielding parameter β in Eq. (5). The patch shielding ratio is affected by β and increases from 0 to β with the training epochs. We find that the best performance is achieved at $\beta = 30\%$. A larger β indicates that more challenging patches in the images are shielded, preventing them from being randomly cropped and allowing them to undergo additional contrastive learning. While a smaller β results in the undesired possibility of difficult patches being cropped out at the later training

Method	Mode	Dice(%)	IoU(%)
w	-	83.85	72.19
	-	84.45	73.08
w+patch cutmix	rand	84.31	72.87
	easy	84.17	72.66
	hard	86.23	75.80

(a) Different augmentation methods.

β	Dice(%)	IoU(%)
0	84.31	72.87
10%	84.69	73.44
20%	85.35	74.44
30%	86.23	75.80
40%	85.96	75.37

(b) Patch shielding parameter β .

γ	Dice(%)	IoU(%)
0.70	85.03	73.96
0.80	85.61	74.84
0.85	86.07	75.55
0.90	86.23	75.80
0.95	86.11	75.61

(c) Confidence threshold γ .

Table 3. A set of ablation studies on UDIAT dataset under 1/4 partition protocol.

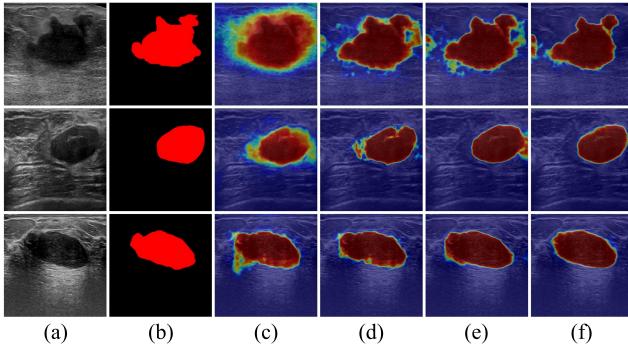


Figure 7. Visual comparison of component ablation studies. (a) Input image. (b) Ground truth. (c) SupOnly. (d)-(f) Method I-III in Tab. 2. Predicted probability is represented by the heat map, green indicates low confidence while red and blue denote high ones.

stages. Tab. 3c shows the ablation study of the confidence threshold γ for filtering pseudo-labels in Eq. (9). We set $\gamma = 0.9$. Using an appropriate threshold can effectively prevent the model from learning incorrect pseudo-labels. A smaller γ may cause erroneous predictions to disturb loss calculations, while too high a threshold makes it difficult for the model to acquire learning information. In addition, as shown in Fig. 8, we ablate different patch sizes. Our method achieves the best performance on both datasets with a patch size of 25×25 , substantiating that oversized or undersized patches insufficiently capture local information.

4.4. Limitations

Although our method achieved state-of-the-art performance, some limitations still exist. As shown in Fig. 9, our method struggles to yield well-segmented results when the lesions with extremely ambiguous contours or minimal distinction from the background.

5. Conclusion

In this paper, we propose a novel semi-supervised breast lesion segmentation framework via patch-wise hardness (PH-Net). Considering the challenging boundaries in the BUSI image, we evaluate region-wise hardness from a patch per-

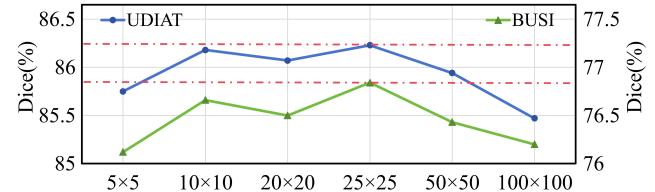


Figure 8. Ablation study of different patch sizes $h \times w$ on UDIAT and BUSI datasets under 1/4 partition protocol.

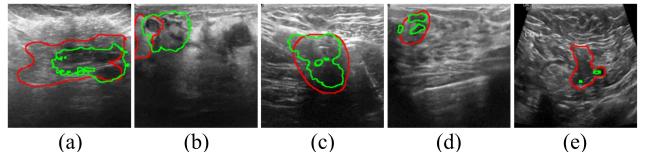


Figure 9. Failure cases. Red and green contours represent ground truth and prediction, respectively.

spective. We employ adaptive patch augmentation (APA) to ensure that challenging regions receive sufficient training without excessive perturbation during strong augmentation. Furthermore, to emphasize the model’s focus on challenging regions, we introduce hard-patch contrastive learning (HPC) to provide additional training, enhancing intra-class compactness and inter-class discriminability of pixels within hard patches. Extensive experiments on two public datasets demonstrate the superiority of our method over the state-of-the-art semi-supervised segmentation methods.

Acknowledgments

This work was supported partly by National Natural Science Foundation of China (Nos. 62273241 and 62206179), Natural Science Foundation of Guangdong Province, China (Nos. 2024A1515011946 and 2022A1515010129), the University Stability Support Program of Shenzhen (20220811121315001), and the General Research Fund from Hong Kong Research Grants Council (project no. 15218521).

References

- [1] Walid Al-Dhabyani, Mohammed Gomaa, Hussien Khaled, and Aly Fahmy. Dataset of breast ultrasound images. *Data in brief*, 28:104863, 2020. 2, 6
- [2] Iñigo Alonso, Alberto Sabater, David Ferstl, Luis Montesano, and Ana C Murillo. Semi-supervised semantic segmentation with pixel-level contrastive learning from a class-wise memory bank. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8219–8228, 2021. 2, 5
- [3] Yunhao Bai, Duowen Chen, Qingli Li, Wei Shen, and Yan Wang. Bidirectional copy-paste for semi-supervised medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11514–11524, 2023. 3, 6
- [4] Hritam Basak and Zhaozheng Yin. Pseudo-label guided contrastive learning for semi-supervised medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19786–19797, 2023. 3
- [5] Michal Byra, Piotr Jarosik, Aleksandra Szubert, Michael Galperin, Haydee Ojeda-Fournier, Linda Olson, Mary O’Boyle, Christopher Comstock, and Michael Andre. Breast mass segmentation in ultrasound with selective kernel u-net convolutional neural network. *Biomedical Signal Processing and Control*, 61:102027, 2020. 2
- [6] Gongping Chen, Lei Li, Yu Dai, Jianxun Zhang, and Moi Hoon Yap. Aau-net: An adaptive attention u-net for breast lesions segmentation in ultrasound images. *IEEE Transactions on Medical Imaging*, 42(5):1289–1300, 2023. 2
- [7] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 801–818, 2018. 6
- [8] Xiaokang Chen, Yuhui Yuan, Gang Zeng, and Jingdong Wang. Semi-supervised semantic segmentation with cross pseudo supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2613–2622, 2021. 2, 6
- [9] Heng-Da Cheng, Juan Shan, Wen Ju, Yanhui Guo, and Ling Zhang. Automated breast cancer detection and classification using ultrasound images: A survey. *Pattern recognition*, 43(1):299–317, 2010. 1
- [10] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255. Ieee, 2009. 6
- [11] Muhammad Umar Farooq, Zahid Ullah, and Jeonghwan Gwak. Residual attention based uncertainty-guided mean teacher model for semi-supervised breast masses segmentation in 2d ultrasonography. *Computerized Medical Imaging and Graphics*, 104:102173, 2023. 2, 6
- [12] Geoff French, Samuli Laine, Timo Aila, Michal Mackiewicz, and Graham Finlayson. Semi-supervised semantic segmentation needs strong, varied perturbations. *arXiv preprint arXiv:1906.01916*, 2019. 2, 4
- [13] Ran Gu, Jingyang Zhang, Guotai Wang, Wenhui Lei, Tao Song, Xiaofan Zhang, Kang Li, and Shaoting Zhang. Contrastive semi-supervised learning for domain adaptive segmentation across similar anatomical structures. *IEEE Transactions on Medical Imaging*, 42(1):245–256, 2022. 2
- [14] Luyi Han, Yunzhi Huang, Haoran Dou, Shuai Wang, Sahar Ahamad, Honghao Luo, Qi Liu, Jingfan Fan, and Jiang Zhang. Semi-supervised segmentation of lesion from breast ultrasound images with attentional generative adversarial network. *Computer methods and programs in biomedicine*, 189:105275, 2020. 2
- [15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016. 6
- [16] Ahmed Iqbal and Muhammad Sharif. Unet: A semi-supervised method for segmentation of breast tumor images using a u-shaped pyramid-dilated network. *Expert Systems with Applications*, 221:119718, 2023. 3, 6
- [17] Jang-Hyun Kim, Wonho Choo, and Hyun Oh Song. Puzzle mix: Exploiting saliency and local statistics for optimal mixup. In *International Conference on Machine Learning*, pages 5275–5285. PMLR, 2020. 4
- [18] Dong-Hyun Lee et al. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on challenges in representation learning, ICML*, page 896. Atlanta, 2013. 2
- [19] Tao Lei, Dong Zhang, Xiaogang Du, Xuan Wang, Yong Wan, and Asoke K Nandi. Semi-supervised medical image segmentation using adversarial consistency learning and dynamic convolution network. *IEEE Transactions on Medical Imaging*, 2022. 3
- [20] Shikun Liu, Shuaifeng Zhi, Edward Johns, and Andrew J Davison. Bootstrapping semantic segmentation with regional contrast. *arXiv preprint arXiv:2104.04465*, 2021. 5
- [21] Yuyuan Liu, Yu Tian, Yuanhong Chen, Fengbei Liu, Vasileios Belagiannis, and Gustavo Carneiro. Perturbed and strict mean teachers for semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4258–4267, 2022. 2, 6
- [22] Meng Lou, Jie Meng, Yunliang Qi, Xiaorong Li, and Yide Ma. Mrcnet: Multi-level context refinement network for semantic segmentation in breast ultrasound imaging. *Neurocomputing*, 470:154–169, 2022. 2
- [23] Huayu Mai, Rui Sun, Tianzhu Zhang, Zhiwei Xiong, and Feng Wu. Dualrel: Semi-supervised mitochondria segmentation from a prototype perspective. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19617–19626, 2023. 3
- [24] Mamshad Nayeem Rizve, Kevin Duarte, Yogesh S Rawat, and Mubarak Shah. In defense of pseudo-labeling: An uncertainty-aware pseudo-label selection framework for semi-supervised learning. *arXiv preprint arXiv:2101.06329*, 2021. 2

- [25] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III* 18, pages 234–241. Springer, 2015. [2](#), [6](#)
- [26] Rupali Sood, Anne F Rositch, Delaram Shakoor, Emily Ambinder, Kara-Lee Pool, Erica Pollack, Daniel J Mollura, Lisa A Mullen, and Susan C Harvey. Ultrasound for breast cancer detection globally: a systematic review and meta-analysis. *Journal of global oncology*, 5:1–17, 2019. [1](#)
- [27] Hyuna Sung, Jacques Ferlay, Rebecca L Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 71(3):209–249, 2021. [1](#)
- [28] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems*, 30, 2017. [4](#)
- [29] Haochen Wang, Kaiyou Song, Junsong Fan, Yuxi Wang, Jin Xie, and Zhaoxiang Zhang. Hard patches mining for masked image modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10375–10385, 2023. [2](#)
- [30] Ke Wang, Shujun Liang, and Yu Zhang. Residual feedback network for breast lesion segmentation in ultrasound image. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I* 24, pages 471–481. Springer, 2021. [1](#)
- [31] Xiaoyang Wang, Bingfeng Zhang, Limin Yu, and Jimin Xiao. Hunting sparsity: Density-guided contrastive learning for semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3114–3123, 2023. [2](#), [5](#)
- [32] Yuchao Wang, Haochen Wang, Yujun Shen, Jingjing Fei, Wei Li, Guoqiang Jin, Liwei Wu, Rui Zhao, and Xinyi Le. Semi-supervised semantic segmentation using unreliable pseudo-labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4248–4257, 2022. [2](#), [3](#), [5](#), [6](#)
- [33] Yongchao Wang, Bin Xiao, Xiuli Bi, Weisheng Li, and Xinbo Gao. Mcf: Mutual correction framework for semi-supervised medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15651–15660, 2023. [3](#)
- [34] Huisi Wu, Xiaoting Huang, Xinrong Guo, Zhenkun Wen, and Jing Qin. Cross-image dependency modelling for breast ultrasound segmentation. *IEEE Transactions on Medical Imaging*, 2023. [1](#)
- [35] Huisi Wu, Wende Xie, Jingyin Lin, and Xinrong Guo. Aclnet: semi-supervised polyp segmentation via affinity contrastive learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 2812–2820, 2023. [6](#)
- [36] Xiaozheng Xie, Jianwei Niu, Xuefeng Liu, Qingfeng Li, Yong Wang, and Shaojie Tang. Dk-consistency: A domain knowledge guided consistency regularization method for semi-supervised breast cancer diagnosis. In *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 3435–3442. IEEE, 2021. [3](#)
- [37] Cheng Xue, Lei Zhu, Huazhu Fu, Xiaowei Hu, Xiaomeng Li, Hai Zhang, and Pheng-Ann Heng. Global guidance network for breast lesion segmentation in ultrasound images. *Medical image analysis*, 70:101989, 2021. [1](#), [2](#)
- [38] Lihe Yang, Wei Zhuo, Lei Qi, Yinghuan Shi, and Yang Gao. St++: Make self-training work better for semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4268–4277, 2022. [2](#)
- [39] Lihe Yang, Lei Qi, Litong Feng, Wayne Zhang, and Yinghuan Shi. Revisiting weak-to-strong consistency in semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7236–7246, 2023. [2](#)
- [40] Moi Hoon Yap, Gerard Pons, Joan Martí, Sergi Ganau, Melcior Sentis, Reyer Zwiggelaar, Adrian K Davison, and Robert Martí. Automated breast ultrasound lesions detection using convolutional neural networks. *IEEE journal of biomedical and health informatics*, 22(4):1218–1226, 2017. [2](#), [5](#)
- [41] Donghai Zhai, Bijie Hu, Xun Gong, Haipeng Zou, and Jun Luo. Ass-gan: Asymmetric semi-supervised gan for breast ultrasound image segmentation. *Neurocomputing*, 493:204–216, 2022. [2](#)
- [42] Yijun Zhao, Dashun Que, Jiaqi Tan, Yang Xiao, and Yanyan Yu. Automated breast lesion segmentation from ultrasound images based on ppu-net. In *2019 International Conference on Medical Imaging Physics and Engineering (ICMIP)*, pages 1–4, 2019. [2](#)
- [43] Zhen Zhao, Sifan Long, Jimin Pi, Jingdong Wang, and Luping Zhou. Instance-specific and model-adaptive supervision for semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23705–23714, 2023. [2](#), [3](#), [4](#), [6](#)
- [44] Zhen Zhao, Lihe Yang, Sifan Long, Jimin Pi, Luping Zhou, and Jingdong Wang. Augmentation matters: A simple-yet-effective approach to semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11350–11359, 2023. [2](#), [3](#), [4](#), [6](#)
- [45] Yue Zhou, Houjin Chen, Yanfeng Li, Xuyang Cao, Shu Wang, and Dinggang Shen. Cross-model attention-guided tumor segmentation for 3d automated breast ultrasound (abus) images. *IEEE Journal of Biomedical and Health Informatics*, 26(1):301–311, 2022. [2](#)
- [46] Zhemin Zhuang, Nan Li, Alex Noel Joseph Raj, Vijayalakshmi GV Mahesh, and Shunmin Qiu. An rdaunet model for lesion segmentation in breast ultrasound images. *PloS one*, 14(8):e0221535, 2019. [1](#), [2](#)
- [47] Barret Zoph, Golnaz Ghiasi, Tsung-Yi Lin, Yin Cui, Hanxiao Liu, Ekin Dogus Cubuk, and Quoc Le. Rethinking pre-training and self-training. *Advances in neural information processing systems*, 33:3833–3845, 2020. [2](#)