Weakly Supervised Breast Ultrasound Image Segmentation Based on Image Selection

Tzu-Han Lin, Daehan Kwak, and Kuan Huang*

***Abstract*—Automatic segmentation in Breast Ultrasound (BUS) imaging is vital to BUS computer-aided diagnostic systems. Fully supervised learning approaches can attain high accuracy, yet they depend on pixel-level annotations that are challenging to obtain. As an alternative, weakly supervised learning methods offer a way to lessen the dependency on extensive annotation requirements. Existing weakly supervised learning methods are typically trained on the entire dataset, but not all samples are effective in training a robust image segmentation model. To overcome this challenge, we have developed a new weakly supervised learning approach for BUS image segmentation. Our framework includes three key contributions: 1) A novel image selection method using Class Activation Maps is proposed to identify high-quality candidates for generating pseudo-segmentation labels; 2) The 'Segment Anything' is utilized for pseudo-label generation; 3) A segmentation model is trained using a Mean Teacher method, incorporating both pseudo-labeled and non-labeled images. The proposed framework is evaluated on a public BUS image dataset and achieves an Intersection over Union score that is 82.9% of what is attained by fully supervised methods.**

***Index Terms*—breast ultrasound imaging, weakly supervised learning, semi-supervised learning, class activation map.**

I. Introduction

Breast cancer is the most commonly diagnosed cancer among U.S. women (excluding nonmelanoma skin cancers) and the second leading cause of cancer death among women overall [1]. Breast ultrasound (BUS) imaging is a commonly used tool for early breast cancer diagnosis because it is affordable, radiation-free, non-invasive, fast-imaging, and has high sensitivity and accuracy [2]. Accurate segmentation of tumors from BUS images is essential for radiotherapy planning and clinical diagnosis. Traditional tumor segmentation from BUS images has been a manual process, requiring highly experienced medical professionals. This method is time-consuming and labor-intensive. Recently, significant advancements in deep learning for image segmentation [3], [4], along with the achievements of U-Net [5] in medical image segmentation, have led to the development of deep learning-based methods. These include the HCNet (a convolutional neural network and transformer-based method) [6], CMU-Net (a variant of U-Net) [7], and RMTL-Net (a MultiTask BUS image segmentation and classification network) [8].

Fully supervised BUS image segmentation methods require pixel-level annotations, which are hard to acquire because of the need for professional radiologists and doctors. To reduce labeling costs, researchers have proposed weakly supervised image segmentation methods. Typical forms of weak supervision for segmentation tasks include the use of bounding boxes [9], scribbles [10], points [11], multi-instance learning [12] and image-level labels [13]. Among these, image-level labels are the simplest to acquire and thus have become widely adopted in the field. In image-level weakly supervised image segmentation, classification networks are employed to generate Class Activation Maps (CAMs) [14] as intermediaries, subsequently thresholded into pseudo-segmentation labels. These labels are then used to train a segmentation network in a fully supervised manner. However, CAMs from classification networks typically highlight only the most discriminative regions, often covering just a portion of the object. To address this, several studies [15], [16] focus on methods to broaden CAM coverage to the entire object. Others [13], [17] concentrate on improving pseudo-label derivation from CAMs for more effective segmentation. Weakly supervised image segmentation is also proposed in BUS images. In [18], a weakly supervised method for segmenting BUS images is introduced. CAMs approximate the breast tumor's initial location, and a level-set algorithm [19] is utilized to produce the final segmentation result. In [20], a two-step method for segmenting BUS images is proposed. The initial step involves a semi-supervised approach for segmenting breast anatomy. The second step combines CAMs and a level-set method to segment breast tumors.

Although refining CAMs can improve the quality of pseudo-segmentation labels and thereby enhance the training of the final segmentation model, and while weakly supervised image segmentation has been previously studied in the context of BUS images, we observed that pseudo-segmentation labels generated from CAMs in BUS images are not consistently reliable. We categorize the images into three distinct groups based on their CAMs, as illustrated in Fig. 1.

- **CAMs encompassing the entire tumor (Figs. 1(a) and (b)):** The CAMs encompass the entire tumor, providing a distinct boundary that separates the tumor from the surrounding tissues. This clear delineation achieved by the CAMs is advantageous for generating pseudo-segmentation labels of high quality, as seen in the fourth row in Figs. 1(a) and (b).
- **CAMs highlighting the tumor boundary (Figs. 1(c) and (d)):** The CAMs mainly focus on the tumor boundaries. By meticulously setting the threshold, these boundary-emphasizing CAMs prove to be valuable in

*Corresponding author: Kuan Huang khuang@kean.edu

Tzu-Han Lin, Daehan Kwak, and Kuan Huang are with the Department of Computer Science and Technology, Kean University, Union, NJ, 07083 {lintzuh, dkwak, khuang}@kean.edu

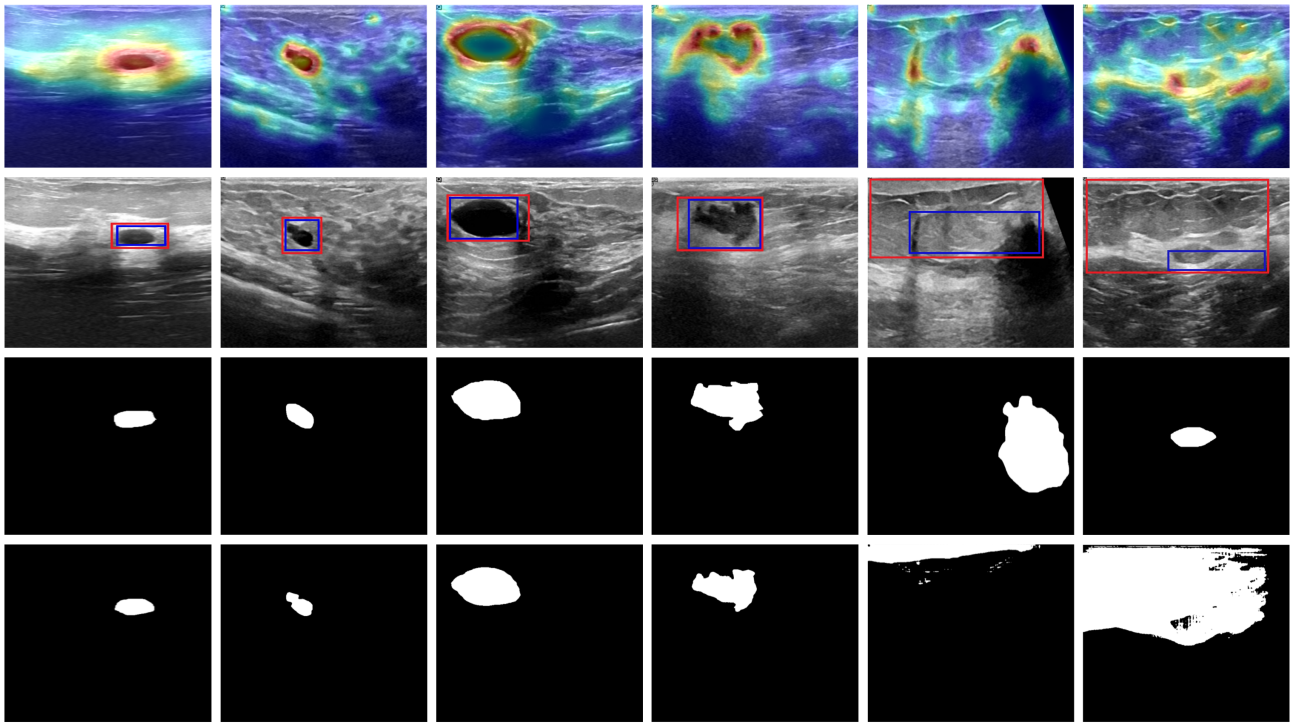

Fig. 1. Three distinct categories of images, their corresponding CAMs, real segmentation labels, and pseudo-segmentation labels: The first row displays the CAMs produced by the Activation Modulation and Recalibration (AMR) model [16]. The second row shows bounding boxes created using two thresholds: red boxes result from a higher threshold, while blue boxes derive from a lower threshold. The third row illustrates the real segmentation labels. Finally, the fourth row depicts pseudo-segmentation labels generated by the 'Segment Anything' model.

creating effective pseudo-segmentation labels, as seen in the fourth row in Figs. 1(c) and (d).

- **CAMs not accentuating significant tumor areas (Figs.1(e) and (f)):** The CAMs do not emphasize regions critical to the tumor, resulting in an inability to produce influential pseudo-segmentation labels, as seen in the fourth row of Figs. 1(e) and (f).

Our observations above show that despite using the most advanced techniques to improve CAMs, certain images are ineffective in generating reliable pseudo-segmentation labels in BUS images. Utilizing the entire training set for generating pseudo-segmentation labels can lead to suboptimal results in the trained segmentation model, particularly with images similar to Figs. 1(e) and (f). To address this challenge, we propose an automated approach to identify high-quality candidates for pseudo-label generation, aiming to enhance the performance of weakly supervised image segmentation in BUS images. The proposed method consists of four steps. First, we utilize image-level labels (tumor/normal) to train an Activation Modulation and Recalibration (AMR) model [16], which generates CAMs. Second, we propose a novel threshold-based technique to select candidate images likely to yield high-quality pseudo labels. Third, we use the 'Segmentation Anything' model [21] to create pseudo-segmentation labels for the selected images. Fourth, our dataset will consist of a mix of images: some with pseudo-segmentation labels and others without. To effectively train the segmentation model using both pseudo-labeled and unlabeled images, we adopt the Mean Teacher method [22]. This strategy enhances the learning process through the use of high-quality pseudo-segmentation labels and unlabeled images. Our contributions are summarized as follows:

- We propose an innovative method for weakly supervised breast tumor segmentation in BUS images, employing solely image-level labels.
- Instead of utilizing the entire dataset for pseudo-segmentation label generation, we propose a novel image selection approach that efficiently selects the most suitable candidates for creating optimal pseudo labels and leverages the Mean Teacher method to train a more effective segmentation model.
- Extensive experiments demonstrate that the proposed technique surpasses state-of-the-art weakly supervised image segmentation methods on a public BUS image dataset [23].

II. Method

The proposed methodology includes the following steps: 1) The complete dataset is randomly divided into a training set (N) and a test set (K) at ratios of 80% and 20%, respectively; 2) A classification model is trained on the training dataset (N), with benign and malignant classified as class 1 and normal as class 0. CAMs are generated for all images in N; 3) Images

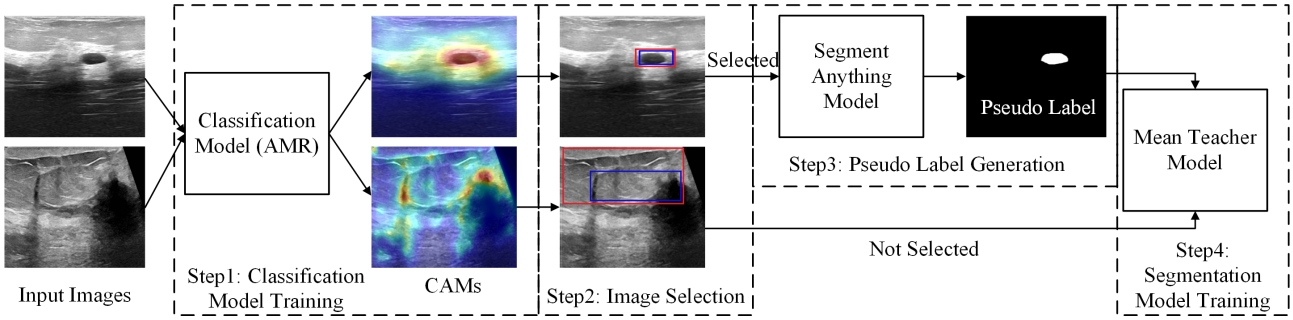

Fig. 2. An overview of the proposed method is as follows. Step 1: Train a classification model and produce CAMs. Step 2: Identify high-quality samples for the generation of pseudo-segmentation labels using the proposed image selection method. Step 3: Create pseudo-segmentation labels for the selected samples using the 'Segment Anything' technique. Step 4: Utilize the Mean Teacher model within a semi-supervised learning framework to train a segmentation model, leveraging both pseudo-labeled and unlabeled samples.

classified as normal are removed from both N and K to form the training and test datasets for segmentation, denoted as V and X, respectively. We only focus exclusively on those images that contain tumors; 4) The proposed image selection method is applied to identify candidate images from V that are most likely to produce high-quality pseudo-segmentation labels. These selected images are placed in subset L, with the rest going into subset U, where $V = L + U$; 5) Pseudo-segmentation labels are created for images in L using the 'Segment Anything' [21]; 6) A U-Net model is trained using the Mean Teacher method on both subsets L and U, and tested on X. A detailed illustration of our approach is shown in Fig. 2. Table I indicates the number of images in each set. More details about the dataset are given in Subsection III-A.

TABLE I
THE NUMBER OF IMAGES IN DIFFERENT SUBSETS

Dataset	Benign	Malignant	Normal	Total
Entire	437	210	133	780
N	357	161	106	624
K	80	49	27	156
V	357	161	0	518
X	80	49	0	129
L	112	52	0	164
U	245	109	0	354

A. Classification Model Training and CAMs Generation

Our class activation maps are generated using the AMR model [16], consisting of two branches: the spotlight and compensation branches. The spotlight branch is based on the original ResNet50 architecture and focuses on the most discriminative regions of the target object. Meanwhile, the compensation branch employs ResNet50 but incorporates Attention Modulation Modules (AMMs) to recalibrate the spotlighted CAMs, refining the CAMs. This dual-branch approach extends the CAMs' coverage to include the entire tumor. The AMR has two classification losses and an additional loss related to the CAMs from both branches. For training the AMR module, we utilize all images $I \in \mathbb{R}^{W \times H}$ and their respective classification labels (tumor/normal), where W and H are the width and height of the images. The AMR module is then used to generate CAMs $CAM \in \mathbb{R}^{W \times H}$ in the tumor class for all the images.

B. Selection of Candidates for Generating Pseudo Labels

Our selection method is based on the regions of interest (ROIs) derived from the CAM in Subsection II-A. The CAM is first normalized to 0 to 255. Then, the normalized CAM and a percentage $\alpha\%$ are input to Algorithm 1 to select an intensity T. The pixels in the normalized CAM between [T, 250] are selected and their coordinates are defined as $S = (x_1, y_1), (x_2, y_2), ..., (x_P, y_P)$. The coordinates in set S are used to form the ROIs. The coordinates for the upper left corner of the ROI are $(x_{leftmost}, y_{topmost})$, determined by finding the minimum values of x and y in S. Similarly, the coordinates for the lower right corner, $(x_{rightmost}, y_{bottommost})$, are the maximum values of x and y in the set S.

Algorithm 1 Pseudo Label Generation Candidate Selection

Input: Normalized $CAM \in \mathbb{R}^{W \times H}$ in the tumor class, percentage $\alpha\%$.
1: Calculate histogram $H(i)$ for $i = 0$ to 255.
2: Calculate cumulative sum $C(i)$: $C(i) = \sum_{j=i}^{255} H(j)$, for $i = 255$ to 0.
3: Calculate total number of pixels P_{total} in CAM.
4: Find the minimum intensity T: $argmin_T(C(T) <= \alpha\% \times P_{total})$.
Output: Intensity T.

For each image in V, we choose two different percentage values, $\alpha_1\%$ and $\alpha_2\%$, where $\alpha_2\%$ is marginally smaller than $\alpha_1\%$. They are used to extract two different ROIs on images, as shown in the second row of Fig. 1 (red boxes and blue boxes). We then compute the Intersection over Union (IoU) for the two ROIs. If the IoU is greater than 0.7, we identify the image as a promising candidate and add this image to a subset L. This approach is based on the idea that a high IoU indicates the CAMs are stable enough to produce two similar ROIs, consequently implying the corresponding images are high quality and have less noise, making them more suitable for generating pseudo labels. In our approach, we have chosen $\alpha_1\%$ and $\alpha_2\%$ values of 2.5% and 1%, respectively. Selected images are then used to generate pseudo labels.

C. Pseudo Label Generation and Model Training

After identifying the promising candidates in subset L and their broader ROIs defined by $\alpha_1\%$ as detailed in Subsection II-B, we utilize the 'Segment Anything' model [21] to create pseudo-segmentation labels (as illustrated in the fourth row of Fig. 1). The original images and their ROIs are input into the 'Segment Anything' method, facilitating the generation of pseudo-segmentation labels.

After generating pseudo-segmentation labels for images in L, we adopt the Mean Teacher method [22] to train a segmentation model on the labeled subset L and unlabeled set U. Specifically, the Mean Teacher strategy has student and teacher networks that are two of the same U-Nets. Images in L and U are passed to the student and teacher networks. The outputs from the student network of images in L are used to compute $\mathcal{L}_{CE}$ and $\mathcal{L}_{DC}$ for the segmentation task. Meanwhile, a consistency loss $\mathcal{L}_{MSE}$ is computed on the outputs of student and teacher networks using both images in L and U. The weights of the student model are updated with gradient descent. The weights of the teacher network are updated as the Exponential Moving Average (EMA) weights of the student network through the training procedure. The loss function of the mean teacher method is as follows:

$$\mathcal{L}_{total} = \frac{1}{2}(\mathcal{L}_{CE} + \mathcal{L}_{DC}) + \beta\mathcal{L}_{MSE} \tag{1}$$

where $\mathcal{L}_{CE}$ and $\mathcal{L}_{DC}$ are the cross-entropy loss and Dice's loss; $\mathcal{L}_{MSE}$ is mean square error loss; β represents the weight of consistency loss that is calculated following [22].

III. Experiment

A. Dataset and Implementation

A public dataset [23] is employed in this research, comprising 780 BUS images, including 210 images of malignant tumors, 437 of benign tumors, and 133 normal images. We partition each category—malignant, benign, and normal—into two subsets: 80% for training and 20% for testing. The number of images utilized in various phases and the number of images selected by our proposed image selection algorithm are detailed in Section II and Table I. In the AMR training phase, we categorize both the malignant and benign images as Category 1 and the normal images as Category 0 for a binary classification task (tumor/normal). When training the segmentation model, we only use images with tumors (malignant or benign) to generate pseudo labels because there is no need to segment tumors for non-tumor images. The segmentation model is built upon a U-Net architecture with a ResNet50 as the backbone and trained using the Mean Teacher method. During the Mean Teacher method training, we form a subset L to produce pseudo-segmentation labels from the training benign and malignant BUS images. The model is evaluated on the test benign and malignant images to ensure a fair comparison. For training the AMR and segmentation model, we set 60 epochs, with a batch size set at 32. We employ the Adam optimizer, configured with a learning rate of 0.001 and the β_1 and β_2 of 0.9 and 0.999, to facilitate steady convergence of the model. We employ the IoU metric to assess performance, comparing the segmentation outcomes with the real ground truths.

To augment our training data and enhance model robustness, we incorporate various transformations. These include random horizontal and vertical flips, as well as random rotations.

TABLE II
Ablation Study Results

Methods	Tumor IoU	Background IoU	mIoU
GT-Whole	**67.9%**	**94.8%**	**81.3%**
GT-Subset	56.3%	93.9%	75.1%
GT-MeanTeacher	64.0%	94.1%	79.0%
Pseudo-Whole	48.5%	91.7%	70.1%
Pseudo-Subset	35.2%	91.1%	63.1%
Pseudo-MeanTeacher	**56.3%**	**91.3%**	**73.8%**

B. Ablation Study

In this section, we conduct an ablation study to validate the effectiveness of pseudo labels for training segmentation models. Additionally, we aim to substantiate our hypothesis that generating pseudo-segmentation labels from high-quality candidates and employing the Mean Teacher semi-supervised method for training the segmentation model yields superior results compared to using the entire dataset with a pseudo-segmentation model. This study focuses solely on benign and malignant BUS images. We select a subset L of training images based on our selection algorithm described in Subsection II-B from the entire benign and malignant training set V. The images not included in L are categorized under U. Our study involves six experiments: 1) training a U-Net with the complete training set V, using all images with their real ground truths (GT-Whole), 2) training a U-Net with the subset L, using images in L with their real ground truths (GT-Subset), 3) training a U-Net using the Mean Teacher method with both L and U, but only L has the real ground truths (GT-MeanTeacher), 4) training a U-Net with the complete training set V, using all images with their pseudo ground truths (Pseudo-Whole), 5) training a U-Net with the subset L, using images in L with their pseudo ground truths (Pseudo-Subset), 6) training a U-Net using the Mean Teacher method with both L and U, but only L has the pseudo ground truths (Pseudo-MeanTeacher). All six models are tested on test set X.

The analysis of the ablation study, as shown in Table II, indicates that the fully supervised learning approach using the entire training set achieves the best results on real ground truths. Comparing the first and second rows as well as the fourth and fifth rows of the table, it is evident that training the U-Net with subset L, whether using pseudo or real ground truths, leads to a decline in performance. However, implementing the Mean Teacher method, which incorporates unlabeled data, significantly enhances performance, particularly when using pseudo-ground truths. The Pseudo-MeanTeacher approach shows a notable improvement of 7.8% in tumor IoU, reaching 56.3%, and a 3.7% increase in mean IoU.

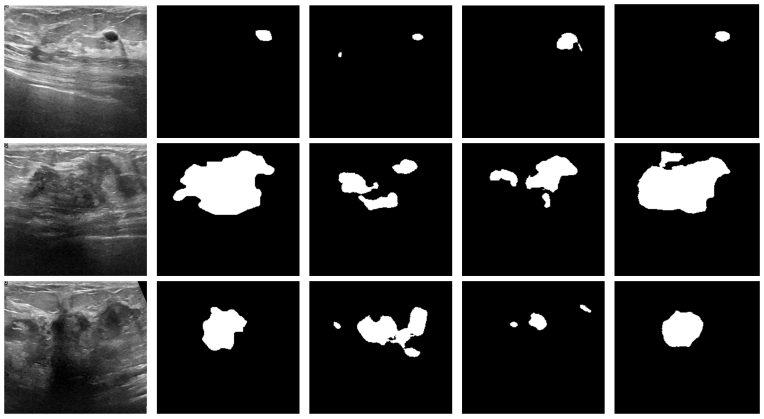

Fig. 3. Semantic segmentation results: (a) Original BUS images, (b) Real segmentation ground truths, (c) Pseudo-Subset; (d) Pseudo-Whole, (e) Pseudo-MeanTeacher.

This improvement is attributed to pseudo-labels generated from unselected images being less accurate, and using these for training can be detrimental. Thus, it is more beneficial to employ these images in a semi-supervised manner. The visualized segmentation results in Fig. 3 corroborate these findings. They demonstrate that utilizing the selected subset L to generate pseudo-labels and training the segmentation model using the Mean Teacher method leads to improved outcomes, as seen in Fig. 3(e), compared to using only subset L (Fig. 3(c)) or the entire dataset (Fig. 3(d)).

C. Comparison with State-of-the-art Methods

In our research, we performed a comparative analysis (as shown in Table III) against various alternative methods that use different classification models to generate CAMs, including employing VGG16 and ResNet50 networks to train tumor/normal classification models and generate CAMs, as cited in [14]. Additionally, we examined the use of the OAA method [15] to refine CAMs generated from ResNet50. These methods employ Dense Conditional Random Fields (CRF) [24] to generate pseudo-segmentation labels from CAMs. Dense CRF is a widely utilized technique for enhancing segmentation output, considering individual pixel features and spatial interactions among pixels. Subsequently, a U-Net model is explicitly trained to segment BUS images using the pseudo-labels. Our study also includes comparisons with a weakly supervised medical image segmentation method, Swin MIL [12], and two state-of-the-art natural image weakly supervised methods, AMR [16] and ReCAM [25].

The results in Table III demonstrate that our method achieves the highest tumor IoU and mean IoU, with increases of 19.1% and 9.5%, respectively, compared to the second-best performing method. Additionally, our method attains the second-highest background IoU, reaching 91.3% IoU. A significant drawback of previous methods is their reliance on using the entire dataset to generate pseudo-labels. However, not all images produce high-quality pseudo-labels, which diminishes the overall performance of the models. In contrast, our method focuses on selectively identifying promising candidates for pseudo-label generation and effectively harnesses unlabeled data to enhance model performance. This approach effectively addresses the specific challenges of BUS image segmentation and efficiently uses the available data, resulting in improved performance and more dependable outcomes.

TABLE III
QUANTITATIVE COMPARISONS TO OTHER METHODS

Methods	Tumor IoU	Background IoU	mIoU
VGG16-CAM [14]	24.1%	89.1%	56.6%
ResNet50-CAM [14]	37.2%	**91.4%**	64.3%
OAA-CAM [15]	30.4%	85.3%	57.8%
Swin MIL [12]	25.4%	89.3%	57.6%
ReCAM [25]	31.7%	88.1%	59.9%
AMR [16]	35.8%	87.5%	61.7%
Pseudo-MeanTeacher (Ours)	**56.3%**	91.3%	**73.8%**

IV. CONCLUSION

Due to the high cost and resource requirements for medical image annotation, it is essential to design weakly supervised segmentation methods that perform similarly to fully supervised methods. In this study, we propose a novel weakly supervised BUS image segmentation framework that employs an efficient selection method to find high-quality candidates for generating pseudo-segmentation labels. The proposed method utilizes the Mean Teacher method to train the segmentation model on both pseudo-labeled and unlabeled images. The primary benefit of the proposed method is decreasing the costs and labor associated with comprehensive annotation while upholding high-performance standards. This benefit is particularly crucial in medical imaging, where the precision and dependability of models are critical. The proposed method is evaluated on a public BUS image dataset. Extensive experiments demonstrate its effectiveness. It outperforms six recent weakly supervised segmentation methods and achieves state-of-the-art performance.

ACKNOWLEDGMENT

This work was partially supported by the Office of Research and Sponsored Programs, Kean University and the CAHSI-Google Institutional Research Program 2023-2024.

REFERENCES

[1] Angela N Giaquinto, Hyuna Sung, Kimberly D Miller, Joan L Kramer, Lisa A Newman, Adair Minihan, Ahmedin Jemal, and Rebecca L Siegel, "Breast cancer statistics, 2022," *CA: a cancer journal for clinicians*, vol. 72, no. 6, pp. 524–541, 2022.

[2] Qinghua Huang, Yaozhong Luo, and Qiangzhi Zhang, "Breast ultrasound image segmentation: a survey," *International journal of computer assisted radiology and surgery*, vol. 12, pp. 493–507, 2017.

[3] Jonathan Long, Evan Shelhamer, and Trevor Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[4] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[5] 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*. Springer, 2015, pp. 234–241.

[6] Qiqi He, Qiuju Yang, and Minghao Xie, "Hctnet: A hybrid cnn-transformer network for breast ultrasound image segmentation," *Computers in Biology and Medicine*, vol. 155, pp. 106629, 2023.

[7] Fenghe Tang, Lingtao Wang, Chunping Ning, Min Xian, and Jianrui Ding, "Cmu-net: a strong convmixer-based medical ultrasound image segmentation network," in *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2023, pp. 1–5.

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

[9] Jifeng Dai, Kaiming He, and Jian Sun, "Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1635–1643.

[10] Di Lin, Jifeng Dai, Jiaya Jia, Kaiming He, and Jian Sun, "Scribblesup: Scribble-supervised convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3159–3167.

[11] Amy Bearman, Olga Russakovsky, Vittorio Ferrari, and Li Fei-Fei, "What's the point: Semantic segmentation with point supervision," in *European conference on computer vision*. Springer, 2016, pp. 549–565.

[12] Ziniu Qian, Kailu Li, Maode Lai, Eric I-Chao Chang, Bingzheng Wei, Yubo Fan, and Yan Xu, "Transformer based multiple instance learning for weakly supervised histopathology image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2022, pp. 160–170.

[13] Jiwoon Ahn, Sunghyun Cho, and Suha Kwak, "Weakly supervised learning of instance segmentation with inter-pixel relations," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 2209–2218.

[14] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba, "Learning deep features for discriminative localization," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2921–2929.

[15] Peng-Tao Jiang, Ling-Hao Han, Qibin Hou, Ming-Ming Cheng, and Yunchao Wei, "Online attention accumulation for weakly supervised semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 10, pp. 7062–7077, 2021.

[16] Jie Qin, Jie Wu, Xuefeng Xiao, Lujun Li, and Xingang Wang, "Activation modulation and recalibration scheme for weakly supervised semantic segmentation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022, pp. 2117–2125.

[17] Jiwoon Ahn and Suha Kwak, "Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4981–4990.

[18] Yongshuai Li, Yuan Liu, Zhili Wang, and Jianwen Luo, "Weakly-supervised deep learning for breast tumor segmentation in ultrasound images," in *2021 IEEE International Ultrasonics Symposium (IUS)*. IEEE, 2021, pp. 1–4.

[19] Tony F Chan and Luminita A Vese, "Active contours without edges," *IEEE Transactions on image processing*, vol. 10, no. 2, pp. 266–277, 2001.

[20] Yongshuai Li, Yuan Liu, Lijie Huang, Zhili Wang, and Jianwen Luo, "Deep weakly-supervised breast tumor segmentation in ultrasound images with explicit anatomical constraints," *Medical image analysis*, vol. 76, pp. 102315, 2022.

[21] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al., "Segment anything," *arXiv preprint arXiv:2304.02643*, 2023.

[22] 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*, vol. 30, 2017.

[23] Walid Al-Dhabyani, Mohammed Gomaa, Hussien Khaled, and Aly Fahmy, "Dataset of breast ultrasound images," *Data in brief*, vol. 28, pp. 104863, 2020.

[24] Philipp Krähenbühl and Vladlen Koltun, "Efficient inference in fully connected crfs with gaussian edge potentials," *Advances in neural information processing systems*, vol. 24, 2011.

[25] Zhaozheng Chen, Tan Wang, Xiongwei Wu, Xian-Sheng Hua, Hanwang Zhang, and Qianru Sun, "Class re-activation maps for weakly-supervised semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 969–978.