IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 4 February 2024, accepted 14 March 2024, date of publication 18 March 2024, date of current version 25 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3379135

RESEARCH ARTICLE

A Tumor Segmentation Method Based on Mean-Teacher Reusing Pseudo-Labels

CHENGYU JIANG, SHANGKUN LIU, JINGYU WANG, SHICHENG GUO, AND WEIMIN ZHENG
Shandong University of Science and Technology, Qingdao, Shandong 266590, China

Corresponding author: Weimin Zheng (zhengwm901@126.com)

This work was supported by the Natural Science Foundation of Shandong Province under Grant ZR2022LZH014.

ABSTRACT Breast tumor is a common female physiological disease, and the malignant tumor is one of the main fatal diseases of women. Accurate examination and assessment of tumor shape can facilitate subsequent treatment and improve the cure rate. With the development of deep learning, automatic detection systems are designed to assist doctors in diagnosis. However, the blurry edges, poor visual quality, and irregular shapes of breast tumors pose significant challenges to design a highly efficient detection system. In addition, the lack of publicly available labeled data is a major obstacle in developing highly accurate and robust deep learning models for breast tumor detection. To overcome the aforementioned issues, we propose SRU-PMT+, a pseudo-label reusing Mean-Teacher architecture based on squeeze-and-excitation residual (SE-Res) attention. We utilize the proposed segmentation network, SRU-Net++, to generate pseudo-labels for unlabeled data, and guide the learning of the student model using the generated pseudo-labels and groundtruth, improving the accuracy and robustness of the model. Our proposed semi-supervised method has been rigorously evaluated on the available labeled dataset, i.e., Breast Ultrasound Images (BUSI) dataset. Results show that our proposed method outperforms current segmentation methods and has good performance. Importantly, our strategy of reusing pseudo-labels improves the performance of breast tumor segmentation.

INDEX TERMS Semi-supervised learning, mean-teacher, ultrasound images, tumor segmentation.

I. INTRODUCTION

Cancer has always been a major public health issue worldwide. In recent years, the incidence and mortality rates of malignant tumors have gradually been increasing [1]. Early detection of breast cancer can significantly improve the cure rate of patients. The study shows that patients with early stage cancer have a 98% probability of surviving 5 years, compared with only 22% for patients with advanced cancer [2]. Ultrasound images have become an important means of diagnosis and treatment for doctors because of their painless, real-time, and prompt characteristics. Traditional ultrasonic tumor detection is a time-consuming and laborious method [3]. In addition, factors such as blurred edges, large motion artifacts, and irregular shapes of breast tumor ultrasound images increase the difficulty of tumor edge segmentation, as shown in Figure 1.

With the development of technology, artificial intelligence (AI) has been widely applied in various fields [4], [5], [6], [7], [8]. Deep learning is widely used in the field of image segmentation [9]. Convolutional Neural Network (CNN) has successfully achieved the extraction of image feature representation, and generates pixel prediction through fully connected layer to achieve segmentation [10]. However, the limitation of this architecture lies in the subsequent fully connected layer, which results in high space complexity and degrades overall performance. In 2015, Long et al. proposed the Fully Convolutional Network (FCN) [11] as a way to overcome the limitations of CNN. The FCN uses a fully convolutional encoder-decoder architecture to capture semantic representation. It adopts upsampling operations in the decoder to increase the spatial dimension of fine-grained segmentation results. This provides pixel-level predictions for achieving image segmentation. Ronneberger et al. are inspired by the FCN architecture to develop the classical medical image segmentation network model U-Net [12], which is suitable for

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

various applications of medical image analysis and is capable of handling a variety of image types. In recent years, many network models have evolved based on U-Net as infrastructure, such as: 3D U-Net [13], Attentation U-Net [14], U-Net++ [15], U-Net3+ [16], Graph U-Net [17], PDF U-Net [18], SGU-Net [19] and etc. These network models are demonstrated promising results in medical image segmentation tasks.

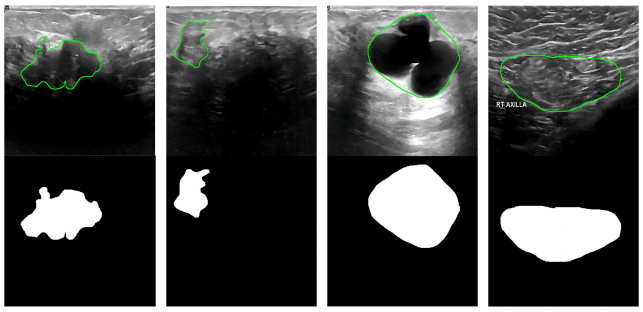

FIGURE 1. Examples of different disadvantages in breast tumor ultrasound images, (a) and (b) are ultrasound images of malignant tumors with irregular borders and shapes, (c) is ultrasound images of benign tumors with uneven shapes, and (d) is large motion artifact images ultrasound image of a benign tumor. Tumor boundaries are indicated in green. The images in the second row is the groundtruth of tumors.

Semi-supervised semantic segmentation (SSS) is a feasible solution to the scarcity of labeled data and can eliminate the need for manual expert labeling. Semi-supervised learning can improve segmentation performance by capturing information from unlabeled data, surpassing the performance of using only a small set of annotated examples. Consistency regularization is currently the dominant SSS method, which effectively incorporates training on unlabeled data into standard supervised learning. To improve segmentation performance, SSS relies on retaining labeled data or model perturbations to create prediction divergence on the same input. This allows the model to use the partly unknown real information of unlabeled data by maintaining consistency of predicted or intermediate features under different perturbations. Recently, some unsupervised prototype segmentation methods employ a feature matching operation to generate pseudo-labels in semi-supervised segmentation tasks, and then add consistency constraints between model predictions and prototype predictions to improve model performance.

Mean-Teacher based methods use a consistency loss to supervise unlabeled data with augmentation. DTC [20] is a dual-task consistency architecture for all kinds of data. ContrasMask [21] uses contrastive learning to extract features from labeled and unlabeled data. These methods extract features from data using different learning paradigms and guide model learning solely based on the similarity between labeled data and labels. However, labeled data may not fully cover the distribution and information differences of the entire dataset, which causes the model to lose some image features from the unlabeled data during training and fail to learn the full image features.

To solve the aforementioned issues, we propose a novel semi-supervised medical image segmentation method. To tackle the problem of poor image quality and inadequate feature extraction of breast tumor images, we propose a new segmentation network called SRU-Net++ that improves feature extraction for fuzzy edges. To address the issue of limited publicly available datasets, we propose a pseudo-label reuse strategy and a new semi-supervised training framework called SRU-PMT+. This enhances the model's learning of unlabeled data, which is necessary for improving the model's performance due to the need for a large amount of data in deep learning. Our proposed method significantly enhances the network's performance in tumor segmentation. The main contributions of this paper are as follows:

- Drawing inspiration from U-Net++ and combining it with the SE-Res attention mechanism, we propose a new segmentation network called SRU-Net++. This network achieves excellent accuracy in tumor edge segmentation, demonstrating a high level of segmentation precision.
- We propose a pseudo-label reuse strategy that generates pseudo-labels for unlabeled data using a network trained on a limited amount of labeled data, which is then utilized in semi-supervised training. By training the limited labeled data with SRU-Net++, it gains the ability to learn and generate pseudo-labels for the remaining unlabeled data. These pseudo-labels have well-defined edges and provide effective guidance for network feature extraction.
- We incorporate the pseudo-labels into the Mean-Teacher architecture, where the pseudo-labels and groundtruths guide the Student model's learning, and consistency is calculated between the Student and Teacher models. We refer to this architecture as SRU-MT+. This method significantly improves the model's segmentation accuracy and robustness.

II. RELATED WORK

A. MEDICAL IMAGE SEGMENTATION

Medical image segmentation plays a crucial role in a wide range of applications, as it involves dividing an entire image into a set of regions [22]. This process utilizes various image features, such as brightness, color, texture, shape, size, and location, to partition the image into multiple non-overlapping regions [23]. The resulting segmentation of these regions can provide clinicians with detailed and comprehensive image information, which can better support medical diagnosis and treatment decisions.

B. U-NET AND ITS VARIANTS

The encoder-decoder architecture has gained widespread adoption in medical image segmentation, and U-Net [12] is recognized as the most prevalent and classic model in this domain. The network has been enhanced by multiple scholars in recent years. For its improvement, it is also diverse, resetting the skip connection, introducing residual convolution blocks, introducing attention mechanisms, etc [24].

To enhance the segmentation performance of medical images, it incorporates skip connection fusion features into the encoder-decoder architecture of FCN [11]. With the wide application of the attention mechanism in the field of deep learning, Attention U-Net [14] was proposed in 2018. To improve the segmentation accuracy, this network introduces the attention mechanism into the U-Net and takes soft attention in the skip connection. However, this enhancement also leads to increased computational overhead and a reduction in the calculation rate.

U-Net++ [15] is a new segmentation method that utilizes nested and dense skip connections. It is capable of providing the decoder with accurate semantics and coarse grading information by passing information from all previous layers to the current layer. This network utilizes direct connections to tackle the vanishing gradient problem, enhance image features, and achieve accurate segmentation results for blurry medical images [25]. Nonetheless, the dense skip connections of U-Net++ substantially increase the number of parameters, which can lead to computational challenges. While this network indirectly merges features from various receptive fields, it only integrates information from the next layer and overlooks information from previous layers. Consequently, the decoder may lack the necessary level of detail, which can affect its ability to accurately segment images. To improve image segmentation performance, several U-Net architectures that incorporate transformer [26] have been proposed [27], [28], [29], such as TransU-Net [30] and TransFuse [31].

These architectures leverage the self-attention mechanism of the transformer to capture long-range dependencies and contextual information. These architectures have demonstrated good performance in various applications. However, a relatively large amount of high-quality labeled data is required for training, which is unrealistic for medical images. This kind of data set requires expert annotation to provide reliable support, which greatly increases the cost of manual annotation.

C. SEMI-SUPERVISED MEDICAL IMAGE SEGMENTATION

Semi-supervised learning is to improve the performance of the model by training unlabeled data together with labeled data. A simple description is given a training set $\mathbf{D}$, where the training set $\mathbf{D}$ is divided into labeled data sets $\mathbf{D_{label}} = \{x_i^{label}, y_i\}_{i=1}^{n}$ and the unlabeled dataset $\mathbf{D_{unlabel}} = \{x_i^{unlabel}\}_{i=1}^{m}$, where n and m are the data volume of the two data sets, y_i is the groundtruth of the labeled data, $\mathbf{D_{label}}$ is a smaller subset of $\mathbf{D}$. Semi-supervised segmentation is to combine $\mathbf{D_{label}}$ and $\mathbf{D_{unlabel}}$ to construct a data-efficient deep learning model, so that its performance is similar to the optimal model trained on the full-label data set. At present, many methods have been developed for semi-supervised medical image segmentation tasks, mainly divided into three strategies [3]: pseudo-label [32], unsupervised regularization [33], [34], [35] and knowledge prior [36]. Among them, semi-supervised learning using pseudo-labels is the most widely used method. Liu et al. [37] evaluate the pseudo-labels output by the model through the trust module, set a threshold to select a high confidence level to improve the quality of the pseudo-labels. Li et al. [38] proposed a method for building state-of-the-art self-ensemble strategies using exponential moving average to reduce the noise and instability of pseudo-labels. Although the pseudo-label strategy is widely used in the field of semi-supervised learning, this strategy is mainly constrained by the quality of the pseudo-label. When a model overfits on a small amount of labeled data during training, it is unable to correct errors, resulting in excessive amplification of errors during predictions. Poor quality pseudo-labels cannot provide effective consistency guidance during training, and may even cause the model to learn incorrect information, resulting in decreased segmentation performance. Therefore, high-quality pseudo-labels are crucial for semi-supervised segmentation task.

D. MEAN-TEACHER ARCHITECTURE

The Mean-Teacher architecture is a semi-supervised learning algorithm proposed in 2018 [39]. Compared to the Temporal Ensembling algorithm [40], this method addresses the issue of high computational cost. The architecture primarily focuses on the model's weights and divides it into two components: the Teacher and the Student. In the Mean-Teacher architecture, the Teacher model generates learning objectives for the Student model, which is trained based on these objectives and inherits the weights of the Teacher model. The final model is obtained through a weighted average of temporal memory. Additionally, the algorithm introduces perturbations to the input data and improves the model's robustness through consistency regularization.

In recent years, many methods for automatic segmentation of medical images have been developed using the Mean-Teacher architecture. In 2021, Wang et al. [41] introduced feature uncertainty and segmentation methods into the Mean-Teacher model. Furthermore, research teams such as Qual [42], Li et al. [43], and Bortsova et al. [44] have conducted studies on improving the robustness and segmentation performance of the Mean-Teacher model by addressing disturbances such as rotation, elastic deformation, and model regularization techniques. These methods mainly focus on changes in data and the combination of different loss functions, without considering the guiding role of pseudo-labels. In the original Mean-Teacher architecture, pseudo-labels are primarily used to guide consistency between the Student and Teacher models, which can lead to long training epochs and limited segmentation accuracy. To address this issue, we propose a Mean-Teacher architecture with pseudo-label reusing, which we call SRU-PMT+.

III. METHOD

In this section, we will provide a detailed introduction to the proposed segmentation network SRU-Net++, and the upgraded Mean-Teacher architecture SRU-PMT+.

A. SRU-NET++ NETWORK ARCHITECTURE

Due to the large noise and blurred boundaries in breast tumor ultrasound images, it is challenging for neural networks to extract features effectively. To address this issue, we propose a novel segmentation network called SRU-Net++, which is constructed based on the U-Net++ architecture. SRU-Net++ adopts the dense skip connections and multi-depth integration of U-Net++, addressing the issues of high parameter and computational complexity. The main idea is to use residual channel attention modules and multi-channel attention modules to improve the feature extraction capability of the network. The specific architecture of SRU-Net++ as shown in the Figure 2. We combine residual connections and attention mechanisms to form the block and use SE-Res blocks to replace the convolutional operations in U-Net++. Additionally, we also add multi-channel attention modules in the skip connection part to refine the feature information. The deconvolution operation in the decoder is replaced with bilinear interpolation upsampling to ensure the quality of the feature information and the connection between information. To mitigate the issue of a high number of network parameters in U-Net++, we have integrated the Primary feature conservation (PFC) module proposed by Xu et al. [45] in 2023. The PFC module consists of depthwise separable convolutional kernels, which reduce computational cost and parameter count while maximizing the extraction of low-level semantic features. The architecture of PFC module is illustrated in Figure 2. This module reduces the amount of parameters and calculations of the network model and ensures the accuracy of calculations. We use SE-Res blocks to suppress noise in the image and obtain better image features. Through residual connections, the module can better propagate and utilize the original features, reducing information loss during the computation process and mitigating the impact of noise on feature extraction. The adoption of channel attention allows for selective enhancement or attenuation of feature responses in different channels, reducing the extraction of irrelevant information and improving the capability of extracting useful features. It exhibits superior noise suppression ability compared to regular convolution operations.

We resize an image with a size of 256 × 256 pixels as input, and obtain the mask through iterative training of the SRU-Net++ network, and generate pseudo-labels of unlabeled data according to the weights obtained from the training of the labeled data for the training of the Mean-Teacher framework. The specific process is in section III-C for details. The training process of the data in the segmentation network is specifically described as the following operations. First, the calculation of the data in the PFC module is:

$$\tilde{X} = \sigma_{\text{pwise}}\left(\sigma_{\text{conv}}(X) + \sigma_{\text{dwise}}\left(\sigma_{\text{conv}}(X)\right)\right) \tag{1}$$

where $X \in R^{H\times W\times C}$, σ_{conv} is 3 × 3 convolution operation, σ_{dwise} is 7 × 7 deep convolution operation, σ_{pwise} is 1 × 1 point convolution operation. After transforming the data in the PFC module, it enters the encoder section, whose specific calculation can be described as follows:

$$x_{i+1,0} = \vartheta\left(\sigma_{mp}\left(x_{i,0}\right)\right) \tag{2}$$

where $i \in \{1, 2, 3\}$, ϑ is residual channel attention module, σ_{mp} is max pooling operation. We compress the features of the previous layer using maximum pooling to increase the perceptual field, achieve nonlinearity and invariance. This enhances the feature extraction of the previous layer by the next layer. Eq (3) describes the calculation process of the skip connection and decoder part:

$$x_{i,j} = \vartheta\left(x_{i,j-1}, \delta\left(x_{i+1,j-1}\right)\right) \tag{3}$$

where $i \in \{0, 1, 2, 3, 4\}, j \in \{1, \cdots, 4-i\}$, δ is the bilinear upsampling operation, and finally the data computed by the decoder is passed through the 1 × 1 convolution operation and the sigmoid activation function to obtain the Mask map of segmentation. After a certain number of rounds of training on labeled data, the weights obtained from the training are given to the network model to generate pseudo-labels of unlabeled data for subsequent segmentation.

B. SQUEEZE-AND-EXCITATION RESIDUAL ATTENTION MODULE

1) RESIDUAL BLOCK

In the field of medical image analysis, deep learning is widely applied, with CNN being the most commonly used, and numerous enhancement methods have been proposed to improve the performance. Increasing the depth of a network can improve its learning ability to some extent, but this operation can cause the problem of vanishing gradients, which slows down convergence and is not conducive to improving segmentation accuracy. To solve this issue, there are generally two approaches: adjusting the optimization method or adjusting the model structure to make it easier to optimize. The author of ResNet [46] adopted the second approach, introducing residual connections into the construction of basic blocks, which changes the way the model is solved:

$$H(x) = F(x) + x \tag{4}$$

where $H(x)$ is the expected potential mapping, $F(x)$ is the fitting function, and Eq (4) is the solution method of residual block. The module architecture is shown in Figure 3(a).

2) SE BLOCK

In addition to the superposition of network layers, the attention mechanism is also a methods to improve the performance of CNN. CNN is composed of a series of convolutional layers, nonlinear layers, and down-sampling layers. The convolutional kernel usually aggregates spatial and feature information within a local receptive field to obtain global information, but it may not capture fine details. Therefore, SE block [47] comprises of two components, namely squeezing and excitation. It leverages global information to selectively amplify relevant feature

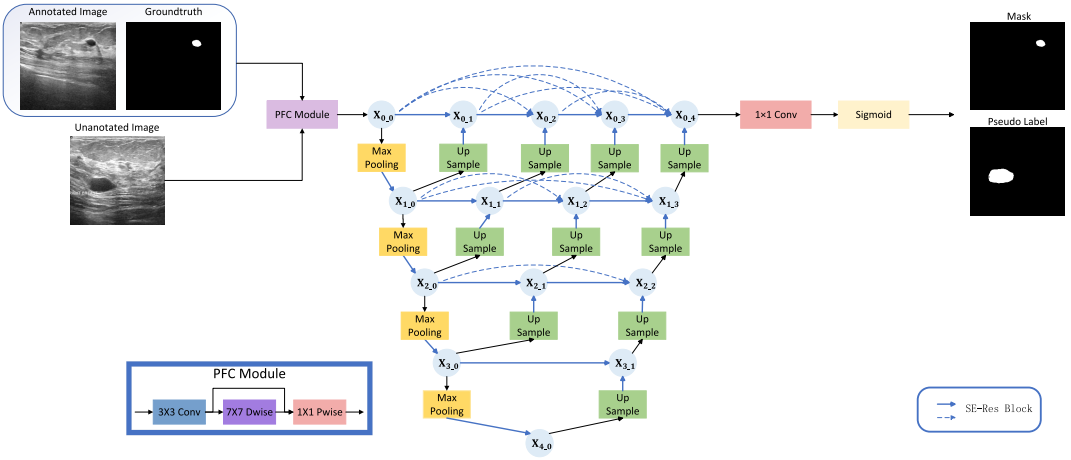

FIGURE 2. Schematic illustration of SRU-Net++ network and PFC module. The SRU-Net++ network adopts a globally symmetric dense skip-connection structure and primarily utilizes SE-Res blocks, bilinear upsampling operations, and max pooling operations. We train the SRU-Net++ using labeled data and then use it to generate pseudo-labels for unlabeled data, which are used for subsequent processing.

channels and attenuate irrelevant ones, thereby achieving adaptive feature recalibration. On the basis of the original module, we lightweighted it by replacing convolutional layers with linear layers, reducing the computational and parameter complexity of the module. We also replaced the max pooling operation with adaptive average pooling to ensure that the module extracts global information from the image. The module is shown in Figure (3b), and the calculation is as in Eq (5):

$$\tilde{X} = F_{se}(X, \theta)X = \sigma\left(W_2 \cdot \delta\left(W_1 \cdot GAP(X)\right)\right) \cdot X \quad (5)$$

where GAP is adaptive average pooling, W_1 and W_2 are linear layers, δ is ReLU layers, and σ is sigmoid functions.

3) SE-RES BLOCK

Combining the above two methods, we abandons the convolution operation and adopt the residual channel attention mechanism to enhance the model's learning of image details and improve the segmentation accuracy of the model. The module is shown in Figure (3c). In this module, the Res block and SE block are combined, and the downsampling operation and channel attention module are added to the skip connection part, so that different semantic information can be divided into different levels of importance when extracting image features, avoiding excessive. It can effectively learn useless information and supplement the detailed information of the segmentation area, and also increase the number of shallow network layers and avoid the problem of gradient disappearance.

C. SRU-PMT+ ARCHITECTURE

Our research proposes an improved segmentation network called SRU-PMT+ to address the issues of high noise and blurry boundaries in breast tumor ultrasound images. In the training process of the neural network model, obtaining high-quality data and a sufficient quantity of data is crucial

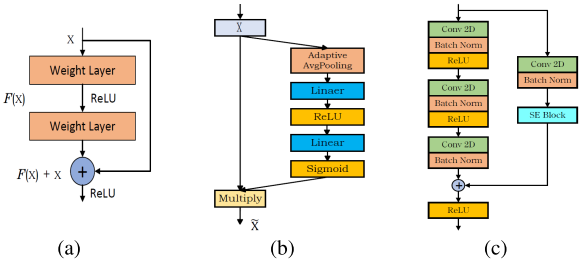

FIGURE 3. (a) is the residual block architecture, (b) is the SE block architecture, (c) is the SE-Res block architecture.

for achieving good results. However, the availability of annotated data for breast tumor ultrasound images is limited. To overcome this issue, we introduce semi-supervised learning and leverage a combination of limited labeled data and a large amount of unlabeled data to enhance the model's capabilities. SRU-PMT+ is an enhancement of the original Mean-Teacher architecture. In the original Mean-Teacher framework, pseudo-labels are only used to calculate the consistency loss based on uncertainty between the Student and Teacher models. We train SRU-Net++ on labeled data to provide it with learning abilities and then generate pseudo-labels for the unlabeled data. These generated pseudo-labels are used to compute the supervised loss function for the Student model. By using pseudo-labels, the Student model can better learn the feature information from unlabeled data, guided by both the Teacher model and its own trained model. This improves the segmentation accuracy for edge details.

To further improve the segmentation performance of SRU-Net++, we employ it as both the Student and Teacher models, and we preprocess the input data before training the network architecture. We apply three data augmentation techniques, including horizontal flipping, translation scaling rotation, and coarse dropout, to enhance the model's learning

capabilities. Additionally, we retain the parameters of the Student model from the Mean-Teacher framework and update the weights of the Teacher model using a weighted average method based on temporal memory. SRU-PMT+ combines the segmentation capabilities of SRU-Net++, semi-supervised learning, and an improved Mean-Teacher architecture. The framework is shown in Figure 4. Our approach also incorporates data augmentation operations and a weight updating strategy to further enhance the model's performance and robustness. The update of the parameters for the Teacher model in SRU-PMT+ can be represented by the Eq (6). the translation of the overall training process as described in Algorithm 1.

$$\theta'_t = \alpha\theta'_{t-1} + (1-\alpha)\theta_t \tag{6}$$

where θ'_t is the parameters of the Teacher model in the t-th epoch, θ_t is the parameters of the Student model in the t-th epoch, α is the hyperparameter for smoothing coefficient. It is used to control the weight update speed of the teacher model. Setting its value to 0.99 is to maintain the relative stability of the teacher model and avoid drastic changes in weights during the training process. Setting this hyperparameter ensures that the teacher model generates stable predictions, which helps improve the model's generalization ability and robustness.

Algorithm 1: SRU-PMT+ Algorithm

Input: Labeled data set D_{labeled}, unlabeled data set $D_{\text{unlabeled}}$, network architecture SRU-Net++, number of iterations T, learning rate lr, pseudo-label generation threshold threshold, weight update factor α.

Output: Trained SRU-Net++ model.

1: Initialize the parameters θ_s and θ_t for the Student and Teacher models of SRU-Net++.
2: **for** $t = 1$ to T **do**
3: Perform forward and backward propagation using the labeled data set D_{labeled}, calculate the supervised loss for the Student model, and update the parameters θ_s of the Student model.
4: Perform forward propagation using the unlabeled data set $D_{\text{unlabeled}}$ and generate pseudo-labels.
5: Perform forward and backward propagation using the pseudo-labels and the labeled data set D_{labeled}, calculate the consistency loss for the Student model, and update the parameters θ_s of the Student model.
6: Update the parameters θ_t of the Teacher model using the weight update factor α, according to the equation: $\theta_t = \alpha \cdot \theta_{t-1} + (1-\alpha) \cdot \theta_s$.
7: **end for**
8: Return the trained SRU-Net++ model.

The semi-supervised segmentation framework we propose changes the inherent guidance of using pseudo-labels for uncertainty consistency, improves the utilization value of pseudo-labels, and makes unlabeled data play a greater role in the entire training process.

D. LOSS FUNCTION

The output of SRU-Net++ is the result of two classifications (tumor and background), which is the process of classifying each pixel, thus converting the breast tumor segmentation problem into a binary classification problem. In the SRU-PMT+ framework, our ultimate goal is to find the optimal solution to the combined loss function. The combined function is shown in Eq (7):

$$Loss = L_{\text{sup}} + \lambda L_{\text{con}} \tag{7}$$

where L_{sup} is supervised loss, L_{con} is consistency loss, λ is a constant,the calculation process is detailed in Section IV-E. For the breast tumor segmentation problem, we choose the Dice function as the supervision loss function, and the binary cross-entropy (BCE) function as the consistency loss function, and the specific calculations are shown in Eq (8) and Eq (9):

$$L_{\text{sup}} = 1 - \frac{2\sum_{i=1}^{N} y_i \hat{y}_i}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \widehat{y_i}} \tag{8}$$

$$L_{\text{con}} = -\frac{1}{N}\sum_{i=1}^{N} y_i \log\left(p\left(y_i\right)\right) + \left(1-y_i\right)\log\left(1-p\left(y_i\right)\right) \tag{9}$$

where y_i is the label value (0 or 1) of pixel i, y_i is the predicted value of the pixel, N represents the total number of pixels, and $p(y_i)$ is the probability that the output belongs to y_i.

IV. EXPERIMENTS AND RESULTS

In this section, we will introduce the dataset used in the experiment, the preprocessing operations on the data, as well as the evaluation indicators and related environments used in the experiment.

A. DATASET

We evaluate our proposed architecture using the open-source dataset BUSI [48]. The BUSI dataset is widely used in the field of breast ultrasound image processing and contains 600 cases of breast ultrasound of women aged 25-75. The dataset consists of 780 images with a size of 500 × 500 pixels, including 437 benign cases, 210 malignant cases, and 133 normal cases. To more carefully evaluate the performance of our vector for malignant and benign tumor segmentation, we train and test on the BUSI dataset for benign and malignant tumors, respectively. Due to the limited number of images in both datasets, we adopted a data augmentation technique inspired by Salazar et al. to increase the data volume and enable model to learn a wider range of features [49], [50]. Specifically, we augmented the benign dataset to 1050 images and expanded the malignant dataset to 850 images. In the experiment, we use 25% of the training set as the labeled data set, and the remaining data sets are unlabeled data sets. The specific allocation is shown in Table 1.

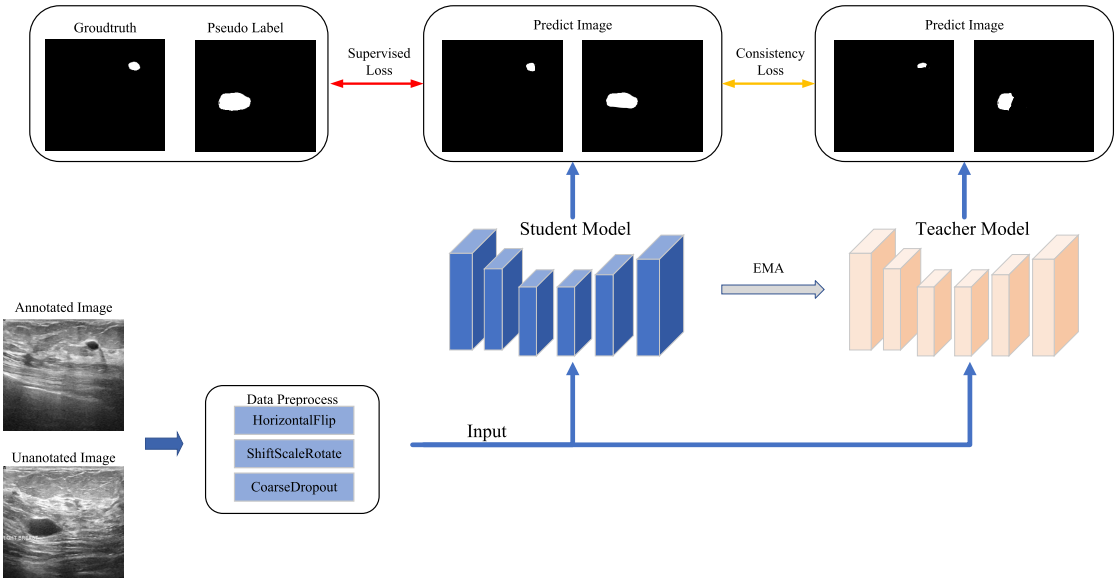

FIGURE 4. Schematic illustration of SRU-PMT+ architecture and the segmentation process of tumor ultrasound images. We use SRU-Net++ as the Student and Teacher models, where mixed labeled and unlabeled data undergo noise processing and are input into both models. Unlike Mean-Teacher, we introduce pseudo-labels to guide the learning of the Student model, enhancing the model's segmentation accuracy on unlabeled data.

TABLE 1. Illustration of BUSI dataset used in this work. The 25% of training set is labeled data, and the rest is unlabeled data.

Dataset	Train		Test	Total
	Labeled	Unlabeled		
Benign	250	750	50	1050
Malignant	200	600	50	850

B. DATA PREPROCESSING

The breast tumor segmentation task requires precise localization and segmentation of tumor regions, which are usually complex and contain many edges and texture features. Therefore, before training the model, we used three methods of data preprocessing including horizontal flipping, shift-scale-rotation, and coarse dropout to improve the accuracy and efficiency of the model.

Horizontal Flip can help the model better adapt to different directions of tumor morphology. In breast tumor segmentation tasks, tumors are usually not perfect circles or ellipses, but present various shapes. Using horizontal flipping can expand the dataset and increase the model's ability to recognize different tumor shapes. Shift-scale-rotation can help the model better adapt to tumors of different sizes and rotation angles. In breast tumor segmentation tasks, the size and shape of tumors may vary. Using shift-scale-rotation can increase the diversity of data and enable the model to better adapt to tumors of different sizes and rotation angles. Coarse dropout can help the model better recognize and classify tumor regions with missing parts. In breast tumor segmentation tasks, some tumor regions may be missing due to pathological reasons or image quality issues. Using coarse dropout can randomly delete some pixels, enabling the model to better recognize and classify tumor regions with missing parts. Through these three preprocessing operations, we improved the diversity and quality of the data, and increased the accuracy and efficiency of the model. The specific results are shown in the Figure 5.

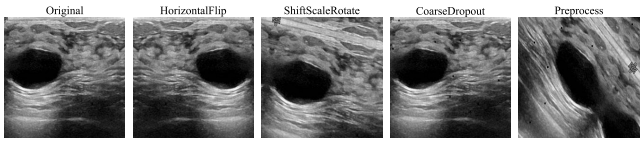

FIGURE 5. Illustration of data processing. From left to right are images after data preprocessing.

C. EXPERIMENT CONFIGURATIONS AND EVALUATION METRICS

All experiments are performed in parallel on a single RTX 3090 GPU using the deep learning pytorch framework. We use the proposed SRU-Net++ network as the backbone network, and use this network to generate pseudo-labels for unlabeled data. In the network training, both Student model and Teacher model are set to SRU-Net++, and the EMA method is used to update the parameters between two models, and the upper limit of the ema coefficient is set to 0.99. Before the image is input into SRU-MT+, aussian noise and salt and pepper noise are added to the input of the two networks to improve the robustness of the model.

In order to quantitatively compare the segmentation results of different methods on breast tumors, we used several

TABLE 2. Quantitative comparison of our proposed SRU-PMT+ on BUSI benign tumor dataset, with the existing supervised breast tumor segmentation models.

Tumor Type	Method	DSC (%)	Jaccard (%)	Accuracy (%)	Precision (%)	Recall (%)
benign	U-Net [12]	66.52±3.09	56.45±2.89	95.99	74.76	66.18
	U-Net++ [51]	68.82±3.52	61.35±3.35	96.56	72.58	69.71
	U-Net3+ [16]	65.12±3.37	56.09±3.16	96.22	73.60	67.57
	ResU-Net [52]	57.16±2.81	45.04±2.58	94.90	57.12	68.43
	ResU-Net++ [15]	53.10± 3.52	43.66± 3.17	95.42	62.39	59.10
	AttentionU-Net [14]	58.24±3.75	50.25±3.48	96.24	66.58	60.05
	GhostU-Net [53]	67.85±1.09	57.83±1.06	96.51	76.71	65.96
	DSCU-Net [54]	57.19±1.30	47.91±1.15	95.86	68.89	55.37
	DCSAU-Net [45]	71.37±1.13	62.80±1.07	96.53	73.86	74.69
	SRU-PMT+	**74.56±1.16**	**68.14±1.04**	**97.03**	**77.63**	**73.39**
malignant	U-Net [12]	58.74±2.09	44.43±1.95	87.22	56.38	73.45
	U-Net++ [51]	58.70±1.53	44.73±1.11	88.55	58.24	71.06
	U-Net3+ [16]	59.65±1.66	45.77±1.26	88.55	60.88	70.57
	ResU-Net [52]	54.04±1.79	39.09±1.10	85.61	51.85	69.22
	ResU-Net++ [15]	58.04±2.72	45.54±2.47	88.62	58.79	69.53
	AttentionU-Net [14]	59.92±2.00	45.46±1.90	89.87	65.60	63.03
	GhostU-Net [53]	63.31±1.44	49.76±1.02	89.95	69.70	68.03
	DSCU-Net [54]	57.49±2.47	44.13±2.21	88.51	63.24	62.53
	DCSAU-Net [45]	63.37±2.29	50.07±2.25	89.71	64.61	69.45
	SRU-PMT+	**64.31±2.30**	**51.38±2.26**	**89.17**	**60.96**	**74.95**

evaluation indicators such as Dice similarity coeffieient(DSC), Jaccard, Accuracy, Precision, and Recall. These evaluation metrics are defined as follows:

$$DSC = \frac{2 \times \mathrm{TP}}{2 \times \mathrm{TP} + \mathrm{FN} + \mathrm{FP}} \tag{10}$$

$$Jaccard = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN}} \tag{11}$$

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

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{13}$$

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{14}$$

D. MODEL PERFORMANCE ANALYSIS

We compared the results of the proposed SRU-PMT+ framework with existing state-of-the-art models on the benign and malignant tumor datasets in BUSI. The quantitative results with state-of-the-art models are shown in Table 2. On the benign tumor dataset, our method demonstrates significant advantages over other models in DSC, Jaccard, Accuracy, and Precision metrics. On the malignant tumor dataset, our method outperforms other models in DSC, Jaccard, and Recall metrics. Based on the results presented in Table 2, it is evident that our proposed SRU-PMT+ method exhibits superior performance compared to the leading DCSAU-Net model across various evaluation metrics on the benign tumor dataset. Specifically, the SRU-PMT+ method demonstrates improvements of 3.19% in DSC, indicating a more accurate capture of tumor regions. Moreover, the Jaccard index shows a notable enhancement of 5.34%, implying a higher degree of overlap in the segmented results. Additionally, the SRU-PMT+ method achieves a 0.5% improvement in Accuracy and a 3.17% improvement in Precision, signifying better overall segmentation accuracy and precision.

Similarly, as depicted in Table 2, the SRU-PMT+ method outperforms the best-performing DCSA-UNet on the malignant tumor dataset. Specifically, it achieves improvements of 0.94% in DSC, 1.31% in Jaccard index, and 5.5% in Recall. These results indicate that the SRU-PMT+ architecture excels in accurately delineating malignant tumor regions, achieving a higher degree of overlap with ground truth annotations, and capturing a greater number of true positive instances. Overall, these experimental results provide strong evidence that our proposed SRU-PMT+ architecture possesses robust segmentation capabilities, outperforming the state-of-the-art models on both benign and malignant tumor datasets.

To verify the effectiveness of our proposed semi-supervised segmentation method, SRU-PMT+, we compared it with widely applied semi-supervised segmentation methods, as shown in Table 3. In the comparison, the models of the contrastive semi-supervised segmentation methods were all set as U-Net. From the table, it can be observed that SRU-PMT+ outperforms other methods in all metrics, indicating that our proposed method has excellent segmentation performance.

In medical image segmentation tasks, ensuring both accurate lesion segmentation and result balance is crucial. Result balance refers to the consistency of segmentation results across different samples, avoiding the phenomenon of excellent performance only on specific types of images. To evaluate the balance of these segmentation methods, the standard deviation of Dice coefficient and Jaccard index is also calculated.

As the Table 2 and Table 3 shown. In the benign dataset, our method has a standard deviation of 1.16% for the Dice coefficient and 1.04% for the Jaccard index. The

TABLE 3. Quantitative comparison of our proposed SRU-PMT+ on BUSI dataset, with the existing semi-supervised segmentation methods.

Tumor type	Method	Jaccard (%)	DSC(%)	Accuracy (%)	Precision (%)	Recall (%)
benign	Mean-Teacher [39]	69.12±1.88	58.09±1.96	96.54	76.90	66.79
	CCT [42]	70.87±1.37	60.91±1.62	95.11	76.85	68.33
	URPC [55]	71.68±1.99	67.46±1.89	96.70	77.05	70.72
	SRU-PMT+	**74.56±1.16**	**68.14±1.04**	**97.03**	**77.63**	**73.39**
malignant	Mean-Teache [39]r	59.92±3.57	49.36±2.45	88.47	59.62	73.59
	CCT [42]	61.22±3.21	50.19±2.36	89.77	60.87	72.43
	URPC [55]	60.39±2.82	50.05±2.77	89.18	60.20	71.01
	SRU-PMT+	**64.31±2.30**	**51.38±2.39**	**89.17**	**60.96**	**74.95**

worst Dice coefficient values of our method are higher than the best Dice coefficient values of the compared fully supervised segmentation network and semi-supervised segmentation method. The range of Jaccard index values is also exceptionally high compared to the compared methods. This indicates that our method not only has good segmentation performance but also has generality.In the malignant dataset, our method has a standard deviation of 2.30% for the Dice coefficient and 2.26% for the Jaccard index, which is lower than the compared semi-supervised segmentation method. Although the standard deviation values are higher than some fully supervised methods, the worst segmentation results generated by our method still have higher values for these two metrics compared to those methods. This indicates that our method exhibits good generality and segmentation accuracy. Therefore, our method demonstrates both good segmentation performance and generality in both benign and malignant datasets.

In addition to comparing the results with these fully supervised segmentation networks, we also compared the SRU-Net++ network with these models in terms of parameter count, computational complexity, and model size. As shown in Table 4, our proposed SRU-Net++ network has a relatively small model size, low computational complexity, and fewer parameters. This indicates that by using the SE-Res module instead of convolutional kernels in the U-Net++ model, we have reduced the complexity of the network and also decreased the utilization of computational resources. This demonstrates that the SRU-Net++ network achieves a lightweight and low-computation model while still achieving high segmentation performance and maintaining good accuracy.

Visualization of the segmentation results is shown in Figure 6 and Figure7. The visualization results show that due to the high noise and blurry boundaries of tumor ultrasound images, most models exhibit missing edges and over-segmentation. Our method can segment most of the tumor edges on benign tumors. Although there are still some missing edges and over-segmentation on malignant tumors, most of the tumors are segmented without segmenting normal breast tissue or missing tumor edges.

E. ABLATION STUDY

In order to validate the segmentation performance of our network, we explored SRU-Net++ with different depths, as shown in Figure 8. Table 5 demonstrates that the depth of

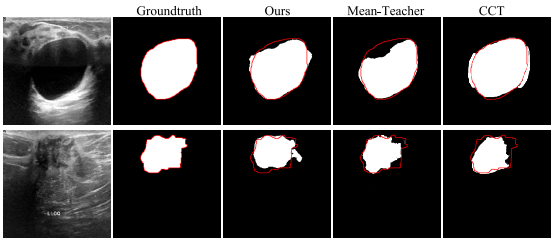

FIGURE 6. Comparison of tumor segmentation results with some semi-supervised segmentation methods.

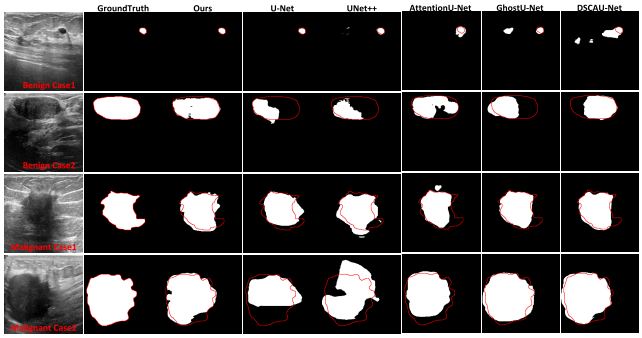

FIGURE 7. Comparison of tumor segmentation results with some supervised segmentation methods. Tumor boundaries are indicated in red.

TABLE 4. Quantitative comparison of our proposed SRU-PMT+ on computational complexity and parameter, with the existing supervised breast tumor segmentation models.

Method	Model Size(M)	FLOPs(G)	Params(M)
U-Net [12]	30.02	14.10	7.85
U-Net++ [51]	35.05	34.9	9.16
U-Net3+ [16]	95.94	280.14	25.12
ResU-Net [52]	49.84	80.98	13.04
ResU-Net++ [15]	55.39	70.99	14.48
AttentionU-Net [14]	133.19	66.63	34.88
GhostU-Net [53]	64.78	31.87	16.94
DSCU-Net [54]	22.95	14.14	5.99
DCSAU-Net [45]	10.20	6.92	2.60
SRU-Net++	**17.20**	**22.58**	**4.41**

SRU-Net++ is optimal. In the benign dataset, SRU-Net++ performed the best in DSC, Jaccard, Accuracy, and Recall metrics. In the malignant dataset, it performed the best in all metrics. We also compared SRU-MPT+ with Mean-Teacher, and the results are shown in Table 6. The results reveals that the SRU-PMT+ architecture exhibits enhanced segmentation

performance, compared to the Mean-Teacher architecture, on both the benign and malignant tumor datasets. In the benign tumor segmentation task, it improved by 4.52%, 6.41%, 0.16%, and 2.36% on the five metrics. In the malignant tumor segmentation task, it improved by 6.06%, 5.97%, 0.41%, and 10.86% on DSC, Jaccard, Accuracy, and Recall metrics, respectively.

In addition, we explored the coefficient lambda in Eq (7), and the results are shown in Table 7. It is evident that our method achieves optimal results when the loss function coefficient is chosen as 0.9.

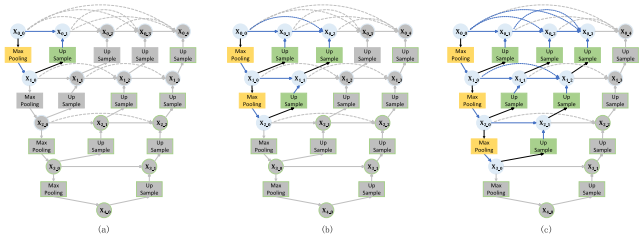

FIGURE 8. The architectures of SRU-Net++ with different depths, (a) is SRU-Net++(L^1), (b) is SRU-Net++(L^2), (c) is SRU-Net++(L^3).

Medical image segmentation is a binary classification problem that aims to classify foreground and background pixels in an image, such as distinguishing tumor regions in breast ultrasound images. However, in cases where the tumor contours are small and the distribution of foreground and background pixels is imbalanced, performance evaluation often relies on PR (Precision-Recall) curves and ROC (Receiver Operating Characteristic) curves.

TABLE 5. Performance of SRU-Net++ with different depths on BUSI dataset.

Tumor type	Architecture	DSC (%)	Jaccard (%)	Accuracy (%)	Precision (%)	Recall (%)
benign	SRU-Net++(L^1)	50.91	39.36	94.37	62.43	54.25
	SRU-Net++(L^2)	56.28	45.49	95.18	65.53	59.78
	SRU-Net++(L^3)	51.96	37.50	85.34	52.04	64.32
	SRU-Net++	**68.52**	**58.59**	**96.55**	**71.37**	**71.84**
malignant	SRU-Net++(L^1)	54.29	39.60	83.62	50.26	74.12
	SRU-Net++(L^2)	52.51	37.98	85.48	50.39	64.75
	SRU-Net++(L^3)	60.38	50.8	95.49	71.46	63.4
	SRU-Net++	**62.36**	**48.51**	**88.63**	**58.55**	**75.41**

TABLE 6. Quantitative comparison of our proposed SRU-PMT+ on BUSI dataset, with the original mean-teacher architecture models.

Tumor type	Method	DSC (%)	Jaccard (%)	Accuracy (%)	Precision (%)	Recall (%)
benign	MT + SRU-Net++	70.04	61.73	96.87	73.07	71.03
	SRU-PMT+	**74.56**	**68.14**	**97.03**	**77.63**	**73.39**
malignant	MT + SRU-Net++	58.25	45.41	88.76	63.14	64.09
	SRU-PMT+	**64.31**	**51.38**	**89.17**	**60.96**	**74.95**

To better validate the improvement in model performance achieved by our proposed pseudo-label reuse semi-supervised learning mechanism, we also employed these two metrics to evaluate the classification performance of SRU-PMT+ in terms of foreground and background pixels. As shown in Figure 9, SRU-PMT with pseudo-label reuse achieved

TABLE 7. Segmentation performance of SRU-PMT+ with different λ coefficients on the BUSI dataset.

Tumor type	λ	Jaccard (%)	DSC(%)	Accuracy (%)	Precision (%)	Recall (%)
benign	0.5	65.94	72.08	96.9	73.8	72.43
	0.6	65.08	71.41	96.74	75.23	70.05
	0.7	66.35	72.45	96.94	74.4	72.16
	0.8	64.51	70.85	96.6	76.09	68.86
	0.9	**68.14**	**74.56**	**97.03**	**77.63**	**73.39**
	1.0	64.34	71.62	96.76	72.95	73.51
malignant	0.5	51.55	63.70	90.48	69.48	65.24
	0.6	51.22	64.11	90.48	66.54	68.69
	0.7	51.82	64.43	90.55	68.55	67.50
	0.8	48.85	61.05	90.09	63.53	64.21
	0.9	**51.38**	**64.31**	**89.17**	**60.96**	**74.95**
	1.0	49.75	62.91	90.34	67.20	66.18

an average precision that was 4.47% higher than that of the standalone SRU-Net++ model in the benign tumor scenario and 4.43% higher in the malignant tumor scenario. As shown in Figure 10, SRU-PMT+ outperformed SRU-Net++ in terms of AUC (Area Under the Curve) in both the benign and malignant tumor scenarios. The experimental results indicate that our proposed pseudo-label reuse semi-supervised learning mechanism enhances the performance of the network model when faced with significantly imbalanced pixel class distributions. This ensures the accuracy of lesion pixel classification by the network model.

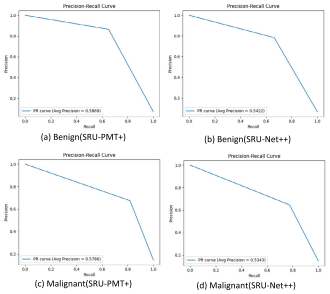

FIGURE 9. The PR curves of the SRU-PMT+ method and a single SRU-Net++ network without employing a semi-supervised learning mechanism.

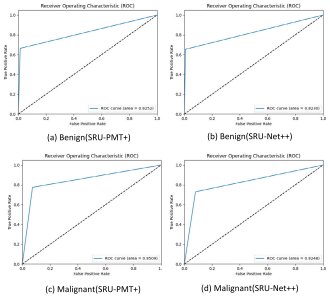

FIGURE 10. The ROC curves of the SRU-PMT+ method and a single SRU-Net++ network without employing a semi-supervised learning mechanism.

V. CONCLUSION

This study proposes a segmentation network based on residual attention mechanism, called SRU-Net++. We use SRU-Net++ to generate pseudo-labels for unlabeled tumor

images and incorporates them into the Mean-Teacher architecture to enable the reuse of pseudo-labels. We also introduce a novel semi-supervised segmentation architecture, SRU-MPT+, which achieves excellent segmentation accuracy in breast tumor segmentation tasks and provides strong auxiliary support for medical diagnosis by doctors. The introduction of this architecture offers improved image segmentation results for doctors and has the potential to enhance the accuracy and efficiency of breast tumor diagnosis.

REFERENCES

[1] M. Xu, T. Zhang, and D. Zhang, "MedRDF: A robust and retrain-less diagnostic framework for medical pretrained models against adversarial attack," *IEEE Trans. Med. Imag.*, vol. 41, no. 8, pp. 2130–2143, Aug. 2022.

[2] R. L. Siegel, K. D. Miller, N. S. Wagle, and A. Jemal, "Cancer statistics, 2023," *CA, Cancer J. Clinicians*, vol. 73, no. 1, pp. 17–48, Jan. 2023.

[3] X. Yang, Z. Song, I. King, and Z. Xu, "A survey on deep semi-supervised learning," *IEEE Trans. Knowl. Data Eng.*, vol. 109, no. 2, pp. 373–440, Nov. 2020.

[4] X. Wang, J.-S. Pan, and S.-C. Chu, "A parallel multi-verse optimizer for application in multilevel image segmentation," *IEEE Access*, vol. 8, pp. 32018–32030, 2020.

[5] J.-L.-Z. J.-L. Zhou, S.-C.-C. J.-L. Zhou, A.-Q.-T. S.-C. Chu, Y.-J.-P. A.-Q. Tian, and J.-S.-P. Y.-J. Peng, "Intelligent neural network with parallel salp swarm algorithm for power load prediction," *J. Internet Technol.*, vol. 23, no. 4, pp. 643–657, Jul. 2022.

[6] P. Hu, J.-S. Pan, S.-C. Chu, and C. Sun, "Multi-surrogate assisted binary particle swarm optimization algorithm and its application for feature selection," *Appl. Soft Comput.*, vol. 121, May 2022, Art. no. 108736.

[7] J.-S. Pan, P. Hu, V. Snášel, and S.-C. Chu, "A survey on binary metaheuristic algorithms and their engineering applications," *Artif. Intell. Rev.*, vol. 56, no. 7, pp. 6101–6167, Jul. 2023.

[8] N. Liu, S. Liu, Q.-W. Chai, and W.-M. Zheng, "A method for analyzing Stackelberg attack–defense game model in 5G by tCPSO," *Expert Syst. Appl.*, vol. 228, Oct. 2023, Art. no. 120386.

[9] S.-C. Chu, Z. Zhuang, J.-S. Pan, A. W. Mohamed, and C.-C. Hu, "Enhanced SparseEA for large-scale multi-objective feature selection problems," *Complex Intell. Syst.*, vol. 10, no. 1, pp. 485–507, Feb. 2024.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 1–12.

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

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351, Munich, Germany. Cham, Switzerland: Springer, 2015, pp. 234–241.

[13] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, Cham, Switzerland: Springer, 2016, pp. 424–432.

[14] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*.

[15] D. Jha, P. H. Smedsrud, M. A. Riegler, D. Johansen, T. D. Lange, P. Halvorsen, and H. D. Johansen, "ResUNet++: An advanced architecture for medical image segmentation," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2019, pp. 225–2255.

[16] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu, "UNet 3+: A full-scale connected UNet for medical image segmentation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 1055–1059.

[17] H. Gao and S. Ji, "Graph U-Nets," in *Proc. 36th Int. Conf. Mach. Learn.*, 2019, pp. 2083–2092.

[18] A. Iqbal and M. Sharif, "PDF-UNet: A semi-supervised method for segmentation of breast tumor images using a U-shaped pyramid-dilated network," *Expert Syst. Appl.*, vol. 221, Jul. 2023, Art. no. 119718.

[19] T. Lei, R. Sun, X. Du, H. Fu, C. Zhang, and A. K. Nandi, "SGU-Net: Shape-guided ultralight network for abdominal image segmentation," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 3, pp. 1431–1442, Mar. 2023.

[20] X. Lai, Z. Tian, L. Jiang, S. Liu, H. Zhao, L. Wang, and J. Jia, "Semi-supervised semantic segmentation with directional context-aware consistency," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 1205–1214.

[21] X. Wang, K. Zhao, R. Zhang, S. Ding, Y. Wang, and W. Shen, "ContrastMask: Contrastive learning to segment every thing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 11594–11603.

[22] S. Liu, Y. Li, Q.-W. Chai, and W. Zheng, "Region-scalable fitting-assisted medical image segmentation with noisy labels," *Expert Syst. Appl.*, vol. 238, Mar. 2024, Art. no. 121926.

[23] S.-C. Guo, S.-K. Liu, J.-Y. Wang, W.-M. Zheng, and C.-Y. Jiang, "CLIP-driven prototype network for few-shot semantic segmentation," *Entropy*, vol. 25, no. 9, p. 1353, Sep. 2023.

[24] W. Zheng, L. Zhou, Q. Chai, J. Xu, and S. Liu, "Fully automatic analysis of muscle B-mode ultrasound images based on the deep residual shrinkage U-Net," *Electronics*, vol. 11, no. 7, p. 1093, Mar. 2022.

[25] L. Zhou, S. Liu, and W. Zheng, "Automatic analysis of transverse musculoskeletal ultrasound images based on the multi-task learning model," *Entropy*, vol. 25, no. 4, p. 662, Apr. 2023.

[26] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–8.

[27] Y. Gao, M. Zhou, and D. N. Metaxas, "UTNet: A hybrid transformer architecture for medical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2021, pp. 61–71.

[28] Y. Ji, "Multi-compound transformer for accurate biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2021, pp. 326–336.

[29] A. Tragakis, C. Kaul, R. Murray-Smith, and D. Husmeier, "The fully convolutional transformer for medical image segmentation," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2023, pp. 3649–3658.

[30] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "TransUNet: Transformers make strong encoders for medical image segmentation," 2021, *arXiv:2102.04306*.

[31] Y. Zhang, H. Liu, and Q. Hu, "TransFuse: Fusing transformers and CNNs for medical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*, 2021, pp. 14–24.

[32] X. Yuan, K. Chen, J. Zhang, W. Zhang, N. Yu, and Y. Zhang, "Pseudo label-guided model inversion attack via conditional generative adversarial network," 2023, *arXiv:2302.09814*.

[33] C. Ouyang, C. Biffi, C. Chen, T. Kart, H. Qiu, and D. Rueckert, "Self-supervision with superpixels: Training few-shot medical image segmentation without annotation," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 762–780.

[34] W. Feng, L. Wang, L. Ju, X. Zhao, X. Wang, X. Shi, and Z. Ge, "Unsupervised domain adaptive fundus image segmentation with category-level regularization," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Cham, Switzerland: Springer, 2022, pp. 497–506.

[35] L. Yang, L. Qi, L. Feng, W. Zhang, and Y. Shi, "Revisiting weak-to-strong consistency in semi-supervised semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 7236–7246.

[36] X. Chen, Y. Yuan, G. Zeng, and J. Wang, "Semi-supervised semantic segmentation with cross pseudo supervision," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2613–2622.

[37] Y. Liu, Y. Tian, Y. Chen, F. Liu, V. Belagiannis, and G. Carneiro, "Perturbed and strict mean teachers for semi-supervised semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 4248–4257.

[38] X. Li, L. Yu, H. Chen, C.-W. Fu, L. Xing, and P.-A. Heng, "Transformation-consistent self-ensembling model for semisupervised medical image segmentation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 2, pp. 523–534, Feb. 2021.

[39] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–21.

[40] S. Laine and T. Aila, "Temporal ensembling for semi-supervised learning," 2016, *arXiv:1610.02242*.

[41] K. Wang, "Tripled-uncertainty guided mean teacher model for semi-supervised medical image segmentation," in *Proc. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, 2021, pp. 450–460.
[42] Y. Ouali, C. Hudelot, and M. Tami, "Semi-supervised semantic segmentation with cross-consistency training," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 12671–12681.
[43] X. Li, L. Yu, H. Chen, C.-W. Fu, and P.-A. Heng, "Semi-supervised skin lesion segmentation via transformation consistent self-ensembling model," 2018, *arXiv:1808.03887*.
[44] G. Bortsova, F. Dubost, L. Hogeweg, I. Katramados, and M. de Bruijne, "Semi-supervised medical image segmentation via learning consistency under transformations," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Shenzhen, China: Springer, 2019, pp. 810–818.
[45] Q. Xu, Z. Ma, N. He, and W. Duan, "DCSAU-Net: A deeper and more compact split-attention U-Net for medical image segmentation," *Comput. Biol. Med.*, vol. 154, Mar. 2023, Art. no. 106626.
[46] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[47] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
[48] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data Brief*, vol. 28, Feb. 2020, Art. no. 104863.
[49] A. Salazar, L. Vergara, and E. Vidal, "A proxy learning curve for the Bayes classifier," *Pattern Recognit.*, vol. 136, Apr. 2023, Art. no. 109240.
[50] A. Salazar, L. Vergara, and G. Safont, "Generative adversarial networks and Markov random fields for oversampling very small training sets," *Expert Syst. Appl.*, vol. 163, Jan. 2021, Art. no. 113819.
[51] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Proc. Int. Workshop Deep Learn. Med. Image Anal.*, 2018, pp. 3–11.
[52] X. Xiao, S. Lian, Z. Luo, and S. Li, "Weighted res-UNet for high-quality retina vessel segmentation," in *Proc. 9th Int. Conf. Inf. Technol. Med. Educ. (ITME)*, Oct. 2018, pp. 327–331.
[53] I. A. Kazerouni, G. Dooly, and D. Toal, "Ghost-UNet: An asymmetric encoder–decoder architecture for semantic segmentation from scratch," *IEEE Access*, vol. 9, pp. 97457–97465, 2021.
[54] L. Xu, D. Ming, T. Du, Y. Chen, D. Dong, and C. Zhou, "Delineation of cultivated land parcels based on deep convolutional networks and geographical thematic scene division of remotely sensed images," *Comput. Electron. Agricult.*, vol. 192, Jan. 2022, Art. no. 106611.
[55] X. Luo, G. Wang, W. Liao, J. Chen, T. Song, Y. Chen, S. Zhang, D. N. Metaxas, and S. Zhang, "Semi-supervised medical image segmentation via uncertainty rectified pyramid consistency," *Med. Image Anal.*, vol. 80, Aug. 2022, Art. no. 102517.

CHENGYU JIANG is currently pursuing the degree with the College of Computer Science and Technology, Shandong University of Science and Technology, Qingdao, Shandong, China. His research interests include semi-supervised medical image segmentation and noisy image processing.

SHANGKUN LIU received the B.S. degree from Shandong University of Science and Technology, Qingdao, China, in 2019, where she is currently pursuing the Ph.D. degree. Her recent research interests include image processing and deep learning.

JINGYU WANG is currently pursuing the degree with the College of Computer Science and Technology, Shandong University of Science and Technology, Qingdao, Shandong, China. His research interest includes few-shot segmentation.

SHICHENG GUO is currently pursuing the degree with the College of Computer Science and Technology, Shandong University of Science and Technology, Qingdao, Shandong, China. His research interests include multimodal segmentation and few-shot segmentation.

WEIMIN ZHENG received the Ph.D. degree from Harbin Institute of Technology, Heilongjiang, China, in 2001. He is currently a Full Professor with the College of Computer Science and Technology, Shandong University of Science and Technology, Qingdao, Shandong, China. His current research interests include artificial intelligence, network security, and blockchain.